Quanta and Qualia

Adrian Kent
Centre for Quantum Information and Foundations, DAMTP,
Centre for Mathematical Sciences, University of Cambridge,
Wilberforce Road, Cambridge, CB3 0WA, U.K. and
Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON N2L 2Y5, Canada.[*]
(Dated: August 2016)

I sketch a line of thought about consciousness and physics that gives some motivation for the hypothesis that conscious observers deviate – perhaps only very subtly and slightly – from quantum dynamics. Although it is hard to know just how much credence to give this line of thought, it does motivate a stronger and more comprehensive programme of quantum experiments involving quantum observers.

INTRODUCTION

I take the following stances on consciousness. They are all highly debatable and most of them have been criticised and defended by many. (See e.g. Ref. [1–3] and references therein for some expositions and reviews.) My goal here is not to make new cases or new defences for these stances, but simply to sketch them and then try to say something about their implications for quantum theory and experimental tests involving conscious observers.

1. Consciousness – the collection of perceptions, sensations, thoughts, emotions, thoughts about perceptions, and so on, that we experience – is a natural phenomenon. We say something about the world when we say that we are conscious, just as we do when we say that the Earth is roughly round and that solid objects tend to fall towards it. Saying that an individual's brain runs algorithms that include models of the individual, or that their body tends to respond in a relatively predictable way to stimuli, also saying something about the world – but it does not say or logically imply anything about their consciousness, including its existence.

2. The goal of physics is to give compressed descriptions of natural phenomena. Physical laws reduce a very large set of data to a much smaller set. For example, Newton's laws of gravity and of motion not only characterise and quantify how and when solid objects fall towards the Earth, but give us a unified description that includes the large-scale behaviour of liquids and gases, the motion of celestial and terrestrial bodies, and laws governing tides and atmospheres. So it is a reasonable ambition for physics to look for a compressed, lawlike description of consciousness.

3. The only certain examples we have, our brains and nervous systems, suggest that consciousness is intimately bound up with the properties of matter. So a reasonable ansatz, or starting point, for a lawlike description of consciousness would be a relatively compressed set of rules from which we can infer that when a physical system is in state S its consciousness is in state $C(S)$. We certainly want to allow $C(S)$ to be empty, since we don't want to assume that every physical system is conscious. We should also allow for the possibility of a physical system having more than one separate consciousness, since a human family or a city appear to be examples, and perhaps even a single human or animal brain can be. So really we should say "its consciousness is in state $C(S)$, or its consciousnesses are in states $C_i(S)$ for a list i in some index set $I(S)$". For brevity we leave this implicit below.

4. By the admittedly high standards of successful laws of physics, we don't have any remotely satisfactory lawlike description of consciousness. We can say the waking human brain is generally conscious, and that specific types of consciousness – visual imagery, or smell, or formulating speech – are associated with activities in various regions of the brain (generally many such regions for any given activity). These seem like raw observational data which any theory should aim to explain. Perhaps, more charitably, they could also be seen as steps towards high level laws in a high level description, which should eventually be superseded by more fundamental laws. In physical terms they seem roughly comparable to the observations that stones fall to the ground, ducks go up in the air when they flap their wings hard and clouds tend to float around in the sky. That is, they are generally true, though slightly vague, statements about quite complex physical systems. The history of physics encourages us to try to describe the underlying phenomena better – more completely, more simply, more precisely – by formulating simple and precise mathematical laws governing a smaller range of more elementary objects or quantities.

5. The sort of law suggested so far is consistent with consciousness as an epiphenomenon. Suppose that the laws of physics are complete, or complete enough to describe physics in many regimes, including the behaviour of matter on Earth. Or at least that they are completable, in the sense that there is an as yet undiscovered unified theory T of the sort physicists conventionally imagine. That is, one that includes quantum theory and gravity, and fully describes the dynamics of matter, fields and spacetime, perhaps also including a theory of initial conditions and/or other constraints – but which makes no reference to consciousness. And suppose, just to simplify the language of the discussion, that T allows a sort of effective reductionism in many contexts, so that with appropriate modelling, which in principle can be justified from the fundamental principles of T, we can describe physical systems S interacting with their environment $E(S)$, modelled in a way derived from the laws of physics encapsulated in T. In shorthand, we say such systems S follow the laws of physics given by T. Now, some of these physical systems S – human brains, for example – have associated non-empty consciousnesses $C(S)$. But by (conventional) hypothesis, S follows the laws of physics given by T whatever the form of its consciousness $C(S)$. We don't need to know anything about $C(S)$ to predict the physical behaviour of S, or any of its physical properties (other than those of $C(S)$ itself). Indeed, we don't even need to be aware of the phenomenon of consciousness in order to predict the physical behaviour of S. On the view so far, a complete understanding the physics of S involves understanding T, deriving the predictions that T makes for S, and then adding, as an extra interesting detail, that S has a particular (maybe empty) form of consciousness $C(S)$. This detail is generally time-dependent, our experiences tell us: as the physical state of S changes over time, the consciousness $C(S)$ generally also changes.[20]

6. However, if consciousness is an epiphenomenon, and its epiphenomenal association with the material world is described by simple laws, then it is very hard to understand how and why we evolved to have rich consciousnesses that contain a great deal of data highly relevant to our survival. Darwinian evolution takes place in the material physical world. If consciousness hitches a free ride on that world, then there is no particular need *a priori* for evolutionarily successful creatures to be conscious. [4] Even if they are, there is no need for their consciousnesses to contain data relevant to survival. We could equally well be agilely escaping a tiger while conscious of nothing, or aware only of the fermion numbers of our patellae, or any other physical variables associated with our material selves. On the epiphenomenal view of consciousness, the laws of physics encoded in T are all that is relevant to our body and brain functions during the escape; they are also all that is relevant to describing the evolution of those body and brain functions over aeons that include successful and unsuccessful encounters with tigers by earlier generations. All that we need to explain evolutionary is that our direct ancestors tended to be over-represented in the successful encounters (et cetera).

7. It is also very hard to understand how, if consciousness is purely an epiphenomenon, we can talk about the contents of our conscious minds, listen to ourselves doing so, and feel that we accurately represented ourselves.

8. As if these were not already devastating enough problems for the epiphenomenal view of consciousness, there is more. [4] Our conscious sensations also seem to include highly pleasant and unpleasant sensations, pleasures and pains. By and large, the pleasures seem associated with evolutionarily advantageous activities (food, drink, friendship, bonding, raising of status, sex, ...), and the pains with disadvantageous ones (raging thirst, injury, lowering of status, rejection, ...). Yet, on an epiphenomenal view, there seems no possibility of an evolutionary explanation for these correlations. Evolution of our material selves explains that the laws of physics encoded in T caused our brains and bodies to tend to seek out evolutionary advantageous activities and avoid disadvantageous ones. It does not then matter whether our epiphenomenal consciousnesses find the former pleasurable and the latter painful, or vice versa.

9. So, consciousness is not an epiphenomenon.[21]

BACKWARD OR ONWARD?

At this point, one really needs to pause and take a breath, because the terrain is not going to become easier if one presses further. Following the logic of the argument so far, there should be a physical theory of consciousness, but it should not be an epiphenomenal theory. But are there any other coherent options? And even if there might possibly be, how could they do any better in explaining the puzzles of the evolution of consciousness? Even if one is willing to dream up equations somehow trying to characterize a dynamical interaction between conscious states and familiar material physical states, would they not necessarily work equally well if we relabelled painful states as pleasurable and vice versa?

When an argument runs into such difficulties, one should question one's premises. Perhaps the whole line of thinking about consciousness we have set out is just misguided? Perhaps one of the other standard lines of thought is more promising after all? Well, perhaps. A review is beyond my scope here. But I'm not convinced: it seems to me they also end up either falsely[22] denying any possibility of scientific progress on the hard problem or creating insoluble puzzles of their own.

If every line of thought runs into deep problems, one should arguably pursue the one that offers most hope of bringing new data. I will now argue that the one I have set out does at least suggest the possibility of experimental progress on the problem of consciousness, and with that, the possibility of saying at least something more about how consciousness evolved. I don't see how to take it far enough to sketch any plausible conjecture about a satisfactory solution to the problem of pain and pleasure. Still, even a small chance of experimental progress is worth pursuing, especially given the huge implications. And if there is experimental progress, perhaps it will bring conceptual and theoretical progress, in the new light of which these puzzles might seem less daunting.

People must, I imagine, once have thought it pointless to ask why stones always fall, birds sometimes fly, and clouds generally float in the sky. Those were just part of the definition of stones, birds, and clouds. It must have seemed useless to such people to speculate that we might be able to understand all this falling, flying and floating better if stones, birds, clouds and everything else in the natural world turned out to be made up of smaller constituents. After all, even if they were, it must have seemed that we would just be left asking essentially the same question: why stone-constituents do fall (at least when assembled into stones) whereas bird-constituents sometimes fly (at least when assembled into birds), and so on. In a sense, on this last point, they were right. Even now, we *do* still ask why the laws of general relativity and quantum theory hold and not others. But even if the essence of the question is in some sense still the same, its form has changed as our understanding developed, from an obstinate gatekeeper seemingly preventing progress to a faithfully helpful guide along the long path to modern physics.

So, let us continue.

QUANTA AND QUALIA

Quantum theory and the brain

The hard problem of consciousness was a problem when we believed the world was described by classical physics. It may still be a problem if and when quantum theory and general relativity are superseded. There is no compelling reason of principle to believe that quantum theory is the right theory in which to try to formulate a theory of consciousness, or that the problems of quantum theory must have anything to do with the problem of consciousness.

That said, physics is where it is. Quantum theory is our best current fundamental theory. It works extremely successfully in describing microscopic physics and some aspects of macroscopic physics. It also has problems, of which in my view the sharpest is the long-standing problem of finding some description of objective reality consistent with quantum theory.[23]

So, let us start by supposing that quantum theory applies pretty well to systems like human brains. However, let us keep an open mind on whether it captures absolutely everything that physics can say about them – since this has certainly not been well tested – and see where this takes us.

Qualia

According to one popular line of thought (see e.g. [1, 5, 6]) our consciousnesses can be thought of as composed of very large numbers of individual sensation-components, or qualia. The analogy here is with (what was once) the atomic hypothesis: that matter, in all its rich variety, can be understood as composed of various types of elementary objects, atoms, in various proportions and combinations. Modern chemistry eventually led to the classification of the elements, and hence the elementary atoms, and to the postulated understanding of (macroscopic terrestrial) matter as combinations of atoms. Similarly, one might think, visual perceptions can maybe be understood as some combination of a finite number of colour and relationship qualia, emotions as combinations of finitely many elementary emotional qualia, and so on.

There is absolutely no evidence in favour of this qualia-as-atoms-of-consciousness model. If consciousness is indeed something that can be modelled in any scientifically familiar way, it could be as a field, or a manifold. It could also, of course, be that there is some mathematical model that looks nothing like anything we have encountered in physics so far. Still, if we are going to speculate about the relationship of consciousness to the rest of physics at all, we

need some language, and the qualia model gives a useful way of thinking about how a connection might be made. So we will use it, while emphasizing that our tentative conclusions are meant to apply more generally. The same fundamental questions arise whether consciousness is built from atomic qualia or described by some other quantities. And, importantly, the same conclusions follow.

Qualia from quantum states: a cartoon

Whatever our consciousnesses are, they are certainly not identical to the physical states of our brains. Even on a classical description, the vast majority of the information carried by a brain's physical state is not carried by the associated consciousness. Most microscopic details of the brain's structure and operation seem to have nothing directly to do with our consciousness. Many important neural processes seem to be entirely unconscious. So, if there is a physical theory of consciousness, it seems to involve a great deal of data selection. Perhaps we can model this as a *data selection principle*: some rule that maps the large amount of information in the physical state to the much smaller amount of information contained in the consciousness.

I am not sure that anyone currently has any compellingly plausible idea as to how this might work in any detail.[24] Certainly I don't. So let me instead give a cartoon: not an idea to be taken seriously, but an illustration of the sort of thing that would count as a data selection principle. Suppose that nature has fixed a cubic lattice with a certain scale L, where L is larger than a small molecule and maybe not much larger than a neuron.[25] Take the local density matrix describing the quantum state of matter within each volume L^3 cube. Update these density matrices at each time interval L/c, supposing that nature has also fixed a one-dimensional lattice in time. Suppose there is some local rule according to which a quale Q_j is associated with the cube C at discrete time point T provided that the configuration of local density matrices for nearby and recent cubes (within distance NL and within past time NL/c, for some number $N > 1$) satisfies some property $P_j(\rho_1, \ldots, \rho_M)$. Here $M \approx 8N^4$ is the number of nearby and recent cubes, $j \in J$ is an index over the possible types of quale (which we might perhaps take to be finite), and the properties P_j are sets of mathematical constraints (which to simplify the cartoon we might take to be exclusive, so that each cube is associated with at most one quale). If none of the constraints P_j hold, then there is no quale associated with the given cube at the given time. The consciousness $C(S)$ associated at any given time with a system S to which these rules are applied is the collection of all the qualia defined at that time.

Within our cartoon, this rule is meant to be fundamental, not tailored to the specifics of human brains. It is supposed to give us a general algorithm for identifying $C(S)$ for any system S. So we really should extend the cartoon to give some cartoon-level story about how we can tell whether qualia are part of the same consciousness or not. Perhaps we could do that by adding a second scale K, and saying that any pair of qualia separated by no more than K cubes, at any given time, form part of the same consciousness, and that belonging to the same consciousness is a transitive relation on qualia. In other words, if a pair of qualia are not joined by a path through the qualia that takes no more than K cubes for each step, then they belong to different consciousnesses.

For anything like this to work, even at the cartoon level, one would have to find properties P_j that tend to produce qualia within human brains and (perhaps) central nervous systems. The P_j also should not produce qualia from (at the very least) most of the matter around our brains and central nervous systems, so that we can speak of separate single consciousnesses associated with each brain. One would also need some plausible description of elementary qualia $j \in J$. And then, much harder still, one would need that the P_j actually produce the right sort of collections of qualia – corresponding to the sort of things we actually consciously experience – for the enormous variety of brains and brain states for which we have experience (direct or reported). It does not matter for the cartoon whether or not the P_j imply that things other than brains – modern computers, large rocks, spiral nebulae – are also conscious.

Obviously, I am not suggesting any of this is actually possible. Whatever the fundamental physical theory of consciousness – if there is one – looks like, I am pretty sure it does not resemble this cartoon. But suppose, just for the sake of the argument, that it were possible to make the cartoon work. We would then have a theory of consciousness, including a classification of qualia and a data selection principle. The description of the P_j would, let's assume, be significantly simpler than just a dictionary of all the brain states and corresponding conscious states that we can identify. In that sense, we would have a significantly better understanding of consciousness. But our theory, as described, would be of consciousness as an epiphenomenon. It could possibly nonetheless represent a very substantial advance in our understanding of consciousness, if it turned out to describe the rich variety of our experiences from a simple set of principles P_j. But it could not explain how and why humans had evolved to produce brains that just happen to produce conditions in which many P_j tend to apply, and in which the corresponding qualia produce the sort of consciousnesses we have. So it could not be fundamentally correct: at best it might be a good approximation.

Improving the cartoon?

Any explanation of why humans and other animals evolved to become conscious has to run one of two ways.

One is that human evolution can be understood purely in terms of the familiar material laws of physics, and it is just a nice property of consciousness that it resides in highly evolved creatures that are continually processing information about their environments and acting on it. If one believes this is a satisfactory definition of, or a self-evident property of, consciousness, one can be happy with this explanation.

The other is that familiar materialist explanations of evolution alone are not adequate and that something about consciousness itself gives an extra evolutionary advantage. This needs an extra mechanism that implies, in some sense, that conscious creatures tend to prevail in competition with unconscious ones. More than that, since a binary division between conscious and unconscious creatures doesn't give enough room for an evolutionary story, it needs to imply, in some sense, that more conscious creatures tend to prevail in competion with less conscious ones.

Since we have (perhaps foolhardily!) chosen to reject the first type of explanation here, we have to try for the second. We can translate "more (less) conscious" into "having more (fewer) qualia" in our cartoon. Then, fortunately for our cartoon narrative, there is an available option, already explored in a different connection [7] as a natural way of defining generalizations of quantum theory. According to our cartoon, we can (in principle) calculate the probability $P_q(D)$ of any distribution D of qualia, from quantum dynamics and from knowledge of the constraints defining the properties P_j. (The suffix q stands for quantum here.) We can do this for any system S, or in principle (given a good enough quantum theory that incorporates gravity and describes cosmology) for the entire universe. As noted, if our cartoon were actually correct, this calculation would give the correct predictions for an epiphenomenal model of consciousness. But we can change the model, and make it *non-epiphenomenally* dependent on quantum theory, if we postulate instead that the true probability distribution $P_{\rm true}(D)$ of distributions of qualia is a modified version of $P_q(D)$.

For instance, following the ideas of Ref. [7], we could postulate that

$$P_{\rm true}(D) = P_q(D)A(D)\,, \tag{1}$$

where $A(D)$ is some weight factor that depends only on properties of the qualia distribution D.[26] To be clear: if we take quantum theory as ultimately a theory for predicting the experiences of observers, this means postulating that quantum theory is at least subtly incorrect. But the deviation could be very small and subtle, if $A(D)$ depends only slightly and subtly on D.

Now, if $A(D)$ is chosen to favour, even very slightly, distributions with more qualia, we have the potential beginnings of an explanation for the evolution of consciousness. For such an explanation to work, we need that the postulated properties P_j somehow just happen to involve relations among density matrices that are useful for, or naturally fit into the context of, information processing. But given that (big [27]) assumption, we can see that there would be selection pressure towards creatures whose information processing capacities use such relations in their information processing systems, and then selection pressure in favour of those whose systems generate more qualia.[28]

One can only make full sense of this cartoon theory as we have phrased it, with a postulate of the form (1), in a block universe picture. That is, the theory defines the probability distribution for all qualia throughout space and time. Block universe theories of this type are logically consistent, but they can have unusual and counter-intuitive implications, including effects that appear to agents within the theory to be reverse causation and spacelike signalling. There *are* good reasons to take some types of block universe theory seriously, in our present understanding of quantum theory. Like the earlier part of the cartoon, our block universe qualia cartoon theory is meant only as an existence theorem, not a serious theoretical proposal.[29]

SUMMARY

Every line of thought on the relationship of consciousness to physics runs into deep trouble. Because of this, we are inclined to place some (albeit weak) credence in the line of thought we have outlined, despite its own evident problems. Of course, none of the details of our cartoons are meant to be taken seriously. What we do take seriously, at a weak level of credence, is the suggestion that we could make some progress on understanding the problem of the evolution of consciousness if we supposed that consciousnesses alter (albeit perhaps very slightly and subtly) quantum probabilities. A further reason for taking this seriously (still at a weak level of credence) is an aesthetic preference for theories in which fundamental quantities (here qualia and quanta) genuinely interact, rather than one being purely dependent on the other. The same point was used to motivate inventing and testing generalizations of quantum theory in a different context in Ref. [7].

What are the implications? Broadly, to add some support to tests of quantum theory that involve conscious observers. For example, perhaps this line of thought adds a little to the motivation for interferometry experiments involving viruses [8], or ultimately bacteria or larger creatures.[30] It perhaps adds a little, too, to the motivation for long range Bell experiments in which human observers make (their best attempt at) free random choices of measurement outcomes, and observe the outcomes directly, with separations large enough that the combined choice processes and observations on the two wings are spacelike separated.[31] In the longer term, if and when quantum technology advances to the point that direct tests of quantum theory (not necessarily interferometric tests) on macroscopic objects are possible, it gives a strong motivation for carrying them out on animals and humans.

To be provocatively quantitative, I would give credence of perhaps 15% that something specifically to do with consciousness causes deviations from quantum theory, with perhaps 3% credence that this will be experimentally detectable within the next fifty years. No doubt many physicists would give much lower figures. Still – as the existential risk community in particular has emphasized [9–11] – if one assigns non-zero probabilities, however small and uncertain, to events with large costs or benefits, one should focus on the expectation values. The potential benefits here include developing the beginnings of a physical theory of consciousness. That would also offer some hope of getting data to guide us in the ethical questions we already face (how rich are the consciousnesses of animals?) and those we likely will (are human-level AI programmes, or human brain emulations?). It could also significantly change our understanding of the physics of computation, with potentially large implications for the future of intelligence. Even if one has very weak levels of credence for any current ideas on the physics of consciousness, the large potential implications should argue for devoting more thought to possible experiments on conscious (or plausibly conscious) observers.

ACKNOWLEDGEMENTS

This work was supported by an FQXi grant and by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research and Innovation. I thank Graeme Mitchison for helpful and enjoyable discussions.

References

* Electronic address: A.P.A.Kent@damtp.cam.ac.uk

[1] David J Chalmers. The conscious mind: In search of a fundamental theory, 1996.
[2] Daniel C Dennett. Consciousness explained, 1993.
[3] Ian Glynn. *An anatomy of thought: The origin and machinery of the mind*. Oxford University Press, 2003.
[4] William James. Are we automata? *Mind*, 4:1–22, 1879.
[5] Clarence Irving Lewis. *Mind and the world-order: Outline of a theory of knowledge*. Courier Corporation, 1929.
[6] Erwin Schrödinger. *What is life?: With mind and matter and autobiographical sketches*. Cambridge University Press, 1992.
[7] Adrian Kent. Beable-guided quantum theories: Generalizing quantum probability laws. *Physical Review A*, 87(2):022105, 2013.
[8] Markus Arndt, Olaf Nairz, and Anton Zeilinger. Interferometry with macromolecules: Quantum paradigms tested in the mesoscopic world. In *Quantum [Un] Speakables*, pages 333–350. Springer, 2002.
[9] Nick Bostrom and Milan M Cirkovic. *Global catastrophic risks*. Oxford University Press, 2011.
[10] Nick Bostrom. Existential risk prevention as global priority. *Global Policy*, 4(1):15–31, 2013.
[11] Adrian Kent. A critical look at risk assessments for global catastrophes. *Risk Analysis*, 24(1):157–168, 2004.
[12] Simon Saunders, Jonathan Barrett, Adrian Kent, and David Wallace. *Many Worlds?: Everett, Quantum Theory, & Reality*. OUP Oxford, 2010.
[13] Maximilian A Schlosshauer. *Decoherence: and the quantum-to-classical transition*. Springer Science & Business Media, 2007.
[14] Christopher A Fuchs. Notes on a paulian idea: Foundational, historical, anecdotal, and forward-looking thoughts on the quantum: Selected correspondence, 1995-2001, 2003.
[15] James T Cushing, Arthur Fine, and Sheldon Goldstein. *Bohmian mechanics and quantum theory: an appraisal*, volume 184. Springer Science & Business Media, 2013.
[16] Giulio Tononi and Gerald M Edelman. Consciousness and complexity. *science*, 282(5395):1846–1851, 1998.

[17] Francis Crick and Christof Koch. A framework for consciousness. *Nature neuroscience*, 6(2):119–126, 2003.
[18] Max Tegmark. Consciousness as a state of matter. *arXiv preprint arXiv:1401.1219*, 2014.
[19] Eugene P Wigner. Remarks on the mind-body problem. In I. J. Good, editor, *The Scientist Speculates*. Heineman, 1961.
[20] Quite what sort of time-dependent story about $C(S)$ should emerge for conscious spatially extended objects S is very unclear. It seems as though $C(S)$ should be associated with the worldtube of S rather than any fixed worldline, and then it seems that neither proper time or any other single time parameter is adequate to characterise the dependence of $C(S)$. Given that our only empirical examples, our own consciousnesses, seem (?) to be associated with a single time parameter, we have no real basis for speculating further about other possibilities. We thus tentatively file this under "questions that might one day be addressable if there is real progress on a physical theory of consciousness".
[21] Or else the laws of physics and the laws of epiphenomenal consciousness were apparently designed together to give us the false and useless but pleasing sense of being in conscious control of our actions. I don't take this possibility seriously: does anyone?
[22] I claim "falsely" is a statement of fact, not opinion. Whether they are right or wrong, the sort of ideas sketched in this paper show that there *could* be a testable scientific theory of consciousness.
[23] Of course, this too is keenly debated. For some reviews and discussions, see for example Refs. [12–15].
[24] Noteworthy suggestions and discussions include Refs. [16–18].
[25] Any readers who want to pay this crude cartoon the honour of worrying about its consistency with special relativity could imagine that this story works only where large enough local collections of matter define a local inertial frame to a good enough approximation, and that the lattice is defined in that frame. We can add the rule that if there is no such frame, then there are no qualia, and no local consciousness.
[26] This is not the most general possibility, but general enough for the present discussion.
[27] Though it *is* a very big assumption to make, it is still much less contrived than the assumption that we must have consciousnesses of *exactly* the sort we have, given the information processing that we do.
[28] Stretching credulity even further, if $A(D)$ were somehow chosen to favour qualia of particular types (which tend to be "pleasant") and disfavour qualia of other types (which tend to be "painful"), we might also have at least the potential beginning of a story about how creatures came to embed in their information processing systems some subsystems that generate pleasant qualia (which are favoured by our hypothetical postulate, and which are located so that they correspond to evolutionarily favourable activities) and some that generate painful qualia (which are disfavoured, and located so that they correspond to unfavourable activities). But our comments earlier apply: there seems no reason why this would not equally well work for evolution – although not so happily for us, its conscious products – with the pleasure-pain polarities reversed. The best I can offer is the thought that that the pleasure-pain problem might somehow look different and less fundamentally threatening if we understood the actual details of the interaction between material states and conscious states. But we don't have a theory of consciousness, and so I don't see how this could work. Maybe, of course, it doesn't: maybe the pleasure-pain problem actually is insoluble in this approach.
[29] Among the very odd features of the block universe rule (1), as stated, is that it implies that the bias towards consciousness in evolutionary selection arises from a calculation that depends on the global distribution of consciousness in space and time. For example, the weight bias between alternatives that would produce either descendent d_1 or d_2, with differing numbers of lifetime qualia does not in general reduce to a ratio of the weights associated with the qualia they produce over their lifetimes, $A(C(d_1))/A(C(d_2))$. It does not even reduce to a ratio of the form $A(T_1)/A(T_2)$, defined by the weights associated with the qualia their entire trees of descedents T_i produce. of their full trees of descendents. One has also to consider the effects of T_1 and T_2 on other future conscious lifeforms and evaluate the ratio of weights $A(D_1)/A(D_2)$ associated with the full qualia distributions D_i over all future space-time arising if d_i is the descendent. A version of the rule in which the weight function A is chosen so that $A(D_1)/A(D_2) = A(C(d_1))/A(C(d_2))$ would probably improve the cartoon.
[30] Among the further uncertainties here are that it is far from clear why we should expect viruses or bacteria to have any sort of consciousness.
[31] Existing intuitions that such experiments might be worthwhile are mostly based on the idea [19] that quantum collapse may be connected to consciousness. Our discussion gives another reason for speculating that the direct involvement of conscious observers might possibly alter something relevant to experiment. It does not motivate looking at Bell experiments in particular, however.