A Mathematical Characterization of the Physical Structure of Observers.

Matthew J. Donald

The Cavendish Laboratory, JJ Thomson Avenue,
Cambridge CB3 0HE, Great Britain.

e-mail: mjd1014@cam.ac.uk

web site: http://people.bss.phy.cam.ac.uk/~mjd1014

October 1993
Revised: January 1994
Appears: Foundations of Physics 25, 529-571 (1995)

Abstract. It is proposed that the physical structure of an observer in quantum mechanics is constituted by a pattern of elementary localized switching events. A key preliminary step in giving mathematical expression to this proposal is the introduction of an equivalence relation on sequences of spacetime sets which relates a sequence to any other sequence to which it can be deformed without change of causal arrangement. This allows an individual observer to be associated with a finite structure. The identification of suitable switching events in the human brain is discussed. A definition is given for the sets of sequences of quantum states which such an observer could occupy. Finally, by providing an a priori probability for such sets, the definitions are incorporated into a complete mathematical framework for a many-worlds interpretation. At a less ambitious level, the paper can be read as an exploration of some of the technical and conceptual difficulties involved in constructing such a framework.

CONTENTS

1. Introduction.

Either there is a physical domain to which quantum theory does not apply (the "classical regime"), or we should be able to define "measurement", or we should be able to define "observer". The boundaries of any "classical regime" are notoriously difficult to draw; especially as quantum theory contains classical theory within its structure[(1)]. Quantum theory may simply break down; but we have already searched for over sixty-five years for manifestations of such a breakdown. Defining "measurement" means deciding, for example, whether in a Stern-Gerlach experiment we are measuring spin directions or atomic positions or both, and deciding just when and how each such measurement occurs. It may be easier to define "observer" merely because observers are less various than measurements. Such a definition is attempted in this paper.

The fundamental definition to be given is of a set of sets of sequences of quantum states specified by an abstract structure of a particular type. The postulate underlying this definition is that the physical structure of any possible observer will correspond to such a set while the abstract structure models the information processing of the observer. An a priori probability for these sets, and thus for the corresponding observers, will also be defined. Although it will be impossible to avoid making any comments about "consciousness" in this paper, "consciousness" is not the central issue. As an epiphenomenalist, viewing mind and matter as distinct types of existence and viewing mind as having no physical role, I see mind as being describable as like a "ghost in a machine". The problem for this paper is to define the family of "machines" which could be "haunted" by such a ghost. I do not claim that all such machines must be haunted, but I do propose that only such machines are.

Let us start by recalling Everett's original many-worlds argument (Ref. 2, pp 65–68) in its simplest form. Everett imagines a universe consisting of an observer with wave function ψ observing a system with wave function φ. The total wave function of the universe is then a sum of tensor products of φ's and ψ's. If, at the beginning of a measurement, with the observer in some fixed initial wave function ψ, the system is in an eigenstate φ_a (respectively φ_b) of the operator being measured, then the observer at the end will be in some definite corresponding wave function ψ_a (resp. ψ_b) and the final total wave function will be $\varphi_a \otimes \psi_a$ (resp. $\varphi_b \otimes \psi_b$). On the other hand, simply by the linearity of the Schrödinger equation, if the initial wave function of the system is a superposition $\lambda\varphi_a + \mu\varphi_b$, then the final total wave function must be

$$\Psi = \lambda\varphi_a \otimes \psi_a + \mu\varphi_b \otimes \psi_b. \tag{1.1}$$

The idea now is that the observer enters into Ψ separately in a form described by wave function ψ_a and in a form described by wave function ψ_b and that each of these separate forms describes independent observers experiencing different experimental results (φ_a and φ_b, respectively).

This argument is intriguing but it raises all sorts of new problems. In particular, it calls for an analysis of classes of suitable observer states. At least for the observer, we need something analogous to Zurek's "pointer" basis[(3)] or to Deutsch's "interpretation" basis[(4)]. I have already commented in section 5 of Ref. 5 on the relevance to my work of analyses like Zurek's which provide arguments to show that interference effects can often be negligible. On the other hand, in my view, even without the immediate problems raised by Foster and Brown[(6)], Deutsch's mathematics, like that of Everett, does not generalize beyond the elementary and unphysical models on which it is developed. Any version of the many-world's interpretation should deal with relativity theory, with the macroscopic nature of observers, with stability under perturbations, and with the temporal development of individual worlds. In the present work, we seek an abstract characterization of sets of sequences of states suitable for the description of individual, localized, warm observers. The theory is adapted to relativistic quantum field theory using the mathematics of states on local operator algebras to describe macroscopic thermal systems. Stability under perturbations is taken account of, particularly by working with a range of states for each observer at each moment. Observer histories are described by sets of sequences of states rather than just by the state, or the set of states, presently attained.

This paper is a sequel to Ref. 7. In that paper, I pointed out that, if quantum theory is correct, then the question of what it is that we are observing when we observe our own brains is by no means straightforward. I postulated that the elementary observation was an observation of a two-status object (a "switch") and proposed that observers exist as families of such switches. I went on to give a definition of a switch as a particular sort of set of sequences of quantum states and I argued that such switches exist in the human brain and are structurally stable. This paper focuses on the idea of patterns of switches; proposing conditions under which two different sets of sequences of quantum states are alternative manifestations of the same observer and then discussing the assignment of probabilities to such sets. In other words, this paper is initially concerned with an index set for something like a basis of observer wave functions, or, more generally, with a classification of observer histories and then with computing probabilities for such histories.

This paper is also a sequel to Ref. 5. There I considered the analysis of quantum probabilities for localized observers. However, the treatment of observers in Ref. 5 is unsophisticated, with no details of their definition being given.

The definition in this paper will be presented in seven parts, labelled A to G. A and B define a pattern of switching events. An observer will correspond to such a pattern. C defines the set of possible spacetime manifestations of the switching pattern; in particular, the possible paths along which the switches can move. D identifies the set of local observables on which, for a given spacetime manifestation, the quantum states of the switches will be defined. E and F define the set of sequences of those states. E requires that the states describe identifiable entities moving along the paths defined in C, and F specifies the switching nature of those entities. Finally, G defines the a priori probability of existence of an observer, allowing for all spacetime manifestations compatible with the given switching pattern.

One major purpose behind this paper is to argue that it is possible to define a probability measure on the futures of an individual observer, given a fixed bound on complexity. The central step towards this will be taken by arguing that there are, in fact, only a finite number of possible futures within a given bounded complexity. It is in order to do this that the extra level of abstraction represented by looking for a set of sets of sequences of states is introduced, whereas in Ref. 5 an observer was taken to be defined by a set of sequences of states.

Fundamental to Refs 7 and 5 was a generalization of the idea of "wave-packet collapse" or "reduction". In working with localized, non-isolated, thermal systems like a human brain, there are strong arguments for using density matrices rather than pure-state wave functions. This means generalizing von Neumann's projection postulate, according to which the wave function of a system during a measurement changes discontinuously to an eigenstate of some measured operator. According to Refs 7 and 5, "measured operators" become switch statuses, quantum states are never assumed to be pure, and a succession of "collapses" gives rise to, and has a priori probability determined by, a set of finite sequences of such states.

Everett developed his theory as a version of quantum mechanics without a "collapse processs". By this he meant that, for example, after the observation discussed above, the total wave-function Ψ of equation (1.1) was assumed to continue to be the true total wave-function of the universe. However, "collapse" creeps back into his theory at the level of appearance for an individual observer. Apparent collapse has also been described as "world-splitting", and, as several authors have warned[(8–11)], it is tempting to take world-splitting as a physical process and thus effectively return to the Copenhagen interpretation. In the present explicit analysis of a many-worlds theory, observer-dependent collapse becomes a central focus. Nevertheless, the temptation mentioned is avoided. It has to be because the states collapsed to are observer-dependent and neither unique nor globally defined. There remains a universal background state (denoted by ω). Individual observers collapse the sequences of locally-defined states that form their "worlds" out of ω, and it is relative to that state that the a priori probability of such sequences is determined.

The probability of a given sequence of states for an individual observer will be defined by a function which has been analysed in detail elsewhere[5,12], and which is a generalization of conventional quantum probability. In particular, it generalizes the idea that the probability of observing a component of a mixture is the coefficient of that component, to the situation where the state which may be observed is not an exact component of the given mixture. This is just what is needed for a spatially localized analysis of Everett's universal wave function. Using these fundamental probabilities, we can define relative probabilities for a given individual observer to observe given future events – at least, in as far as we can define such "events". Empirically observed textbook probabilities calculable in an "objective" physics shared between many observers will be related to these probabilities by arguments along lines developed in Ref. 5. It will also be argued that the frequency that a given type of event will be observed by any observer will be close to the textbook probability for the typical manifestation of that observer, where "typical" is defined using the fundamental probability.

The philosophical objection to the idea of mind as a "ghost in a machine", is that if the mind has no physical power and neither more information nor more computational ability than follow directly from the physical structure of the brain, then one has gained little by the invention of the ghost. This objection, in my view, was reasonable in classical physics as long as one was not thereby led to deny any distinction between subjective existence and the existence of objects – a haunted machine does exist as a subject, an unhaunted machine does not[13,14]. In quantum physics, however, we cannot take the physical structure of the brain as a simple given[7]. The mind-brain distinction becomes interesting. Faced with a quantum mechanical picture of the state of the world as being globally some horrendous superposition of different observer histories and locally some, just as horrendous, approximate mixture, and faced with the problem of defining or extracting the elements of such a superposition or mixture, it is appropriate to use the simplest possible philosophical language.

In such philosophically unsophisticated terms, one can think of this paper as presenting the following picture of reality: There is a fundamental initial state ω of the universe – the horrendous superposition/mixture. There are many possible ways in which this universal state can be experienced by observers. Each possible way corresponds to a certain abstract structure, which is the structure required to define the information processing of a particular observer. These abstract structures each have possible physical manifestations which are patterns of sequences of quantum states. The manifestations have their own a priori probabilities of existing as a part of ω and the a priori probability of an observer is the supremum of the a priori probabilities of his possible manifestations. Observers like us, have an accurate picture of reality. This is, firstly, in the sense that such observers define themselves through a consistent narrative[(15)] with a relatively low rate of loss of a priori probability. Secondly, it is in the sense that there is, in the main, a correspondence between the way a part of the universe appears to such an observer and the states of highest a priori probability assigned to those parts of the universe, given the observer's existence. For each such observer, all his physical manifestations of relatively high a priori probability are closely similar and constitute his "machine", "body", or "brain". The ghosts haunt their machines making sense of them and of the world they interact with entirely through their abstract information processing structures. The a priori probabilities define likelihoods for each of their possible futures.

At the level of speculation, this sort of picture has been around for many years. What I am interested in exploring here is the technical and conceptual difficulties in giving explicit definitions to the ideas involved. As far as I am aware at the time of writing, the definitions given below could be correct, and, at one level, the point of this paper is nothing less than the formulation of a complete set of definitions, with details open to critical analysis and development. At this stage however, I do not expect many readers to have the enthusiasm to follow all of the details. Nevertheless I hope that even those who skim the paper will gain a clearer understanding of what the technical and conceptual difficulties are and of the possibility that they might not be insuperable. I hope also that between the level of vague speculation and the level of technical detail, there is a discernable intermediate level. This is the level at which the importance of finiteness becomes apparent. It is the level of the decision to work with density matrices rather than pure states; the level of the idea of interpreting the functioning of a brain as a pattern of switching between sets of quantum states linked to neural firing; the level of the idea of an abstract characterization of an observer. Developing the technical detail is worthwhile, perhaps, mainly because there is no other way of discovering, understanding, justifying, and revising a coherent intermediate level.

I am not concerned here with the question of whether the sort of picture of the world presented here is necessarily true or not but merely with whether it could conceivably be true. It could conceivably be true if it is possible to provide suitable definitions compatible with our general knowledge of quantum physics. I attempt to do this here. One might begin to believe that it actually is essentially true, if one eventually became convinced that it provided a complete and consistent interpretation of quantum mechanics and one could not find any plausible alternatives.

From the outset, two vital constraints on any postulated physical structure for observers should be noted:

1.2) Humans possess such structure.

1.3) The only entities which can with significant a priori probability possess such structures, at the human level of complexity, are entities which we would be prepared to believe might be physical manifestations of consciousness.

2. The light bulb model and the translatability claim.

Some things are observers and some things are not. Of course, one could deny even that, saying instead that all interactions lead to, or are in themselves, observations. However, I propose that things like books, sand beaches, stars, and vats of liquid helium are not observers while things like humans are. The problem is to decide what in this context makes a thing "like" a human. It might be the possession of some "vital fluid", but anatomists have long since given up their search for this. It might be a level of complexity, but I cannot see how to define complexity in such a way as to rule out stars, sand beaches, and books (even very, very long books). It ought to involve the possession of a functioning brain, but not the facts, for example, that that brain in constructed through a carbon-based biochemistry, is about 1.5kg in weight, and functions at around 310° K. It ought, in other words, to involve something about the brain which can be abstractly characterized, and, I propose, that it is the existence of the brain as an information processor processing a finite amount of information through a finite number of two-status elements or "switches" which have a particular quantum mechanical description.

It is possible to disagree, as Penrose[(16)] does, with the very first step in this proposal. This is the claim that humans as observers could be arbitrarily well modelled by finite information processors. In my opinion, even flashes of mathematical insight have physical structures describable, quantum mechanical refinements apart, by patterns of electric fields across neural membranes, and such descriptions can be approximated by a finite amount of information sufficient to encode the insight. A similar opinion is expounded, at length, by Hofstadter (Ref. 17, chapter 17). Other parts of Ref. 17, in particular chapter 6 on "the location of meaning", provide useful discussions of issues related to this section.

A geometrical pattern of switches is a set of two-status objects which move along defined physical trajectories changing status at defined times. Two distinct concepts are involved in defining such a pattern. On one hand, we need a simple, natural, abstract definition of "two-status" in quantum mechanical terms. This is treated in Ref. 7, and we shall leave the definition arrived at there to one side for the moment. On the other hand, we have the idea of a geometrical array of moving objects which acts as, or can be interpreted as part of the hardware of an information processor. The suggestion is that this part is all that is relevant to the information processing structure of the observer. This can be modelled as follows:

The Geometrical Light Bulb Model *The information processing structure of any observer can be perfectly modelled by a finite set of light bulbs moving through space-time along well defined finite paths and turning on and off a finite number of times.*

Switches are referred to as "light bulbs" in this model, partly to underline that this is not itself a quantum mechanical model, and partly as a metaphor stressing how much of the hardware, which would be necessary for the construction of a functioning model, is being postulated here to be irrelevant to the mirroring by that model of a possible observer. In particular, all the structure which "causes" the switches to switch or the light bulb to flash is viewed as irrelevant. The light bulb model is exemplified by the following model of information processing in the human brain:

The Geometrical Neuronal Model *The information processed by a brain can be perfectly modelled by a three dimensional structure consisting of a family of switches, which follow the paths of the brain's neurons, and which open and close whenever those neurons fire.*

In Ref. 7, this model (referred to there as the "neural model") was ultimately replaced because it was argued that neurons are not elementary quantum switches. The replacement in Ref. 7 has similar information processing content but involves the sodium channel molecules which, according to Ref. 7, can function as such switches:

The Geometrical Sodium Channel Model *The information processed by a brain can be perfectly modelled by a three dimensional structure consisting of a family of switches, which follow the paths of the brain's sodium channels, and which open and close whenever those channels open and close.*

Ref. 7 focuses on sodium channels in order to show that at least some quantum switches do exist in the brain. However, I shall indicate in section 6 of this paper that a switching structure involving every such channel might be of such low a priori probability, in the sense of Ref. 5, that if that level of a priori probability is taken as significant then constraint 1.3 would not be satisfied. Such a structure might, therefore, not be appropriate as a model of an observer. Nevertheless, the techniques developed in Ref. 7 remain relevant and will be used, in section 6, to discover more suitable alternative sets of quantum switches – one candidate being the set of immobilized sodium channels. We shall denote by "neural switches" any of the possible quantum switches:

Definition *A "neural switch" is any entity in the human nervous system which satisfies the definition of a quantum switch given in Ref. 7 and which has switching behaviour determined by the firing of a particular neuron.*

Mobile sodium channels, for example, however improbable the structures to which they give rise, are neural switches, but entire neurons are not.

The Geometrical Neural Switch Model *The information processed by a brain can be perfectly modelled by a three dimensional structure consisting of a family of switches, which follow the paths of a sufficiently large collection of neural switches, and which open and close in unison with those switches.*

The notion of a sufficiently large collection will be discussed below.

None of these models say anything about the meaning with which the observer invests his pattern of switches. That meaning, indeed, does not inhere in the physical structure, it is just that physical structure of a certain kind is needed before meaning can inhere. Return to the light bulb model as the simplest example. No attempt has been made to classify patterns of light bulbs. A small set of light bulbs could not be given a very interesting meaning, and a random sequence of flashing presumably has no meaning. However, the claim made by the model is not affected by the suggestion that some sets of light bulbs are not models of observers.

What the model does suggest is that there should be some method of translating a suitable pattern of flashing lights into a meaning which, at least in outline, should be unique. This is to claim that another observer, a "reader", given arbitrary computing power but no prior knowledge of the life history he is to deduce, would be able, by considering the pattern of flashing from its beginning, to come up with a statement about the current flashing like "Ah, he's complaining because he thinks the train is going to be late again".

If this "translatability" claim were not true, then the characterization of physical structure of observers given in this paper could still be valid, but only if physical structure was only a part of the observer's information processing structure. This is conceivable. I see an observer as needing physical structure only in order to determine the a priori probability of his existence and of his possible futures. Sufficient structure is given in this paper to yield a theory of a priori probability compatible with observation and with relativistic quantum field theory (Ref. 5 and below).

Nevertheless, I think that the translatability claim is plausible when applied to neural switch models. Indeed, the statement that we understand the world purely through the functioning of our brains, seems to me to be the statement that, for some sort of brain model, some version of the translatability claim holds. The novelty here is the suggestion that that model could be as simple as a geometrical light bulb model. Such a suggestion would be of little interest in classical physics because every aspect of the total state of the brain could be assumed to be given. In quantum theory, however, we must choose which variables we wish to observe. Here we are looking for a minimal and finite description of the observer as observer of his own brain. We shall use the translatability claim as a guide to strip even the geometric light bulb model of structure.

The translatability claim can be made with different strengths. The strongest form is:

STC) *Any light bulb model can be translated uniquely and by a natural algorithm which can be found by the exercise of reason.*

Even if this claim were true, it would not reduce mind to matter, nor make any physical exemplification of a quantum switching model necessarily conscious. I view STC as being implausible.

A much weaker form is:

WTC) *Members of the class of light bulb models which could correspond to human neural switching models, could be translated by a reader with a sophisticated knowledge of general human biology but no prior knowledge of the specific life histories to be deduced.*

I believe that the following intermediate version is also plausible:

TC) *Members of the class of light bulb models which could correspond to human neural switching models, could be translated by an extra-terrestial reader with no* *knowledge of Earth's biology, given only one example and, at most, some general hints.*

If a human neurophysiologist was given a pattern of neural firings and asked to find its meaning then he would start with a knowledge of neural anatomy, interconnectivity, and function. He would know, for example, that visual information is carried by the optic nerve to the back of the brain. The extra-terrestial reader is assumed to have none of that knowledge. He must deduce interconnectivity from statistical measures of the extent to which one light bulb flashing tends to follow or proceed another. In fact, if he is to translate the sort of models to be introduced below, he will need to start by identifying groups of neighbouring bulbs, corresponding to groups of switches on particular neurons. Statistical analysis will eventually allow him to differentiate between "input" groups and "output" groups. This, of course, depends on the contingent fact that for a pattern of neural firing such differentiation is possible. Repeated patterns on some of the input groups leading to repeated output patterns could be given names and analysed for interconnections at a higher level.

Does the pattern of neural firing experienced in the first year of life allow the names "mother" and "hungry" to be distinguished meaningfully without a prior labelling of some internal neurons as "happy" and "unhappy"? Certainly, the correlations made to "mother" and to "hungry" are different. "Mother" tends to follow closely after "hungry", but not vice versa.

Our ability to see how such translations should be made is helped enormously by our knowledge of what human brains have evolved to do and how they tend to do it. Initially, the extra-terrestial reader has none of that knowledge. All that he has is correlation. If he is to be able to make a complete translation working purely on the basis of the correlations, then he has to define "train", for example, in terms of correlations to patterns with other correlations. TC states that this can be done, although it allows for the reader to be given some help: for example, by the hint that, in the earlier stages of life, calm, regular patterns of excitation could be translated as "happy" and disquiet, random, violent patterns of excitation as "unhappy". TC is a claim about the existence of sufficient correlations between identifiable patterns of excitation, or, in other words, about richness of structure. Without hints, the reader would have to guess his starting point. As there is a possible translation, it would be possible for him to guess right. It is plausible that no wrong guess could give a coherent story.

3. Minimal structure and the plurality problem.

Although much of the apparent structure has been stripped out of the brain in reducing to the geometrical neural switching model, that model surely retains unnecessary structure. It seems implausible that the detailed path of, for example, an ion channel or a piece of neural membrane, should really be relevant to the information processing done by the brain. Irrelevant structure is a particular problem in a many-worlds interpretation of quantum theory, because it is difficult to avoid the assumption that all the physical structures which can be observers are equivalent and that all may exist with their own a priori probability. A slight alteration in a switch path should not correspond to a new observer. This metaphysical problem is reflected in the physical problem of defining an a priori probability measure on sets of observers. A solution, sufficient at least for the second problem, lies in defining an equivalence relation between instances of a geometrical light bulb switching model to determine when two instances may be assumed to correspond to the same observer.

The metaphysical problem, which I shall refer to as the "plurality problem", is not entirely resolved in this paper. There may be many different physical observer structures, as defined below, all corresponding to what we would (naively?) think of as the same human at the same moment. Nevertheless, this paper does take the essential step of reducing to only a finite number of such structures, at least for any given bound on improbability. Some people find any "many-worlds" interpretation unacceptable because of the suggestion that we may have many different futures. The plurality problem makes that suggestion in the strong form that some of these different futures may be indistinguishable to external observers over extended periods.

I suspect that if one could construct a coherent theory from a many worlds theory like Everett's using a splitting into elementary orthogonal basis vectors, then one would also have a plurality problem. I also suspect that the degree of plurality in such cases, would be much worse than that contemplated here, as one would be using a much finer splitting.

In its most difficult aspect, the plurality problem is a sort of converse of the translatability claim. Instead of asking for the meaning of a pattern of neural switchings, we must ask for a specification of the class of patterns of neural switchings which could underlie physical structures for an observed observer. In particular, dropping a single switch, or altering its switching sequence, should not essentially change the translation of a pattern of switchings on, say, 10^{15} switches. Would dropping 10^{10} switches make a significant difference? Reducing the number of switches tends to increase a priori probability, so there does not seem any way of finding a natural maximal number of switches to associate with "an" observer.

The specific problem that it may often be possible to remove switches from observer structures I shall refer to as the "trimming problem". Although this warrants a name because of its simplicity, I think that it can be left unsolved without destroying the rest of the paper. Once again, it is sufficient as a first step to argue that some switching structures are observers, even if we do not have any theory about which, if any, possible structures are not observers.

Returning to the problem of identifying minimal information processing structures, note that in considering the translatability claim, correlations between repeated patterns of switchings were fundamental. The minimal structures to be assigned to an observer must retain these correlations and patterns. One important source of information is the time-ordering of the switchings both for individual switches and between separate switches. If each switch switches sufficiently often, then this information will contain the correlations, and it may also contain patterns, albeit in a different form. In the geometric model, patterns are naturally found by looking for repeated groupings of switchings on neighbouring switches. If we discard the geometrical information, then we need a new definition of "neighbouring", or, equivalently, of "similarity". Just given a time-ordering on a large set of long-lived switches, we can define degrees of similarity between a pair of switches by counting, for example, the number of times which one switch switches exactly once within a switching cycle of the other switch. Other degrees of similarity can be assigned using the strength of correlations between groups of switches similar in the first sense.

The problems with this approach are that, not only are neural switchings too widely-spaced and too numerous to be strictly time-orderable, but also switches in the human brain of high a priori probability tend to have short lifetimes. Both problems can be solved by using a more general relation than a time ordering to express the space-time arrangements of a sequence of events.

The more general relation, to be presented in the next section, will have equivalence classes called "dockets". The docket of a sequence of events will depend on which pairs of events are timelike and which are spacelike and will also carry some topological information. Once we have the definition of a docket and have dealt with various subtleties connected with the labelling of switches the models of section 2 can be modified. This yields for example:

The Minimal Light Bulb Model *The information processing structure of any observer can be perfectly modelled by the suitably labelled docket of the space-time sets where a finite number of light bulbs flash on and off a finite number of times.*

The Minimal Neural Switch Model *The information processed by a brain can be perfectly modelled by a structure consisting of the suitably labelled docket of the space-time sets of switchings of a family of switches, which follow the paths of a sufficiently large collection of neural switches, and which open and close in unison with those switches.*

The versions of the translatability claim should now be taken to refer to these models. The minimal models provide much less information than the geometrical models, so that the plausibility of TC and WTC depend all the more on contingent facts about human neural switchings.

4. Definition and properties of dockets.

A docket is a geometrical structure in space-time defined as an equivalence class of ordered sequences $(A_i)_{i=1}^M$ of suitable space-time sets. Two such sequences $(A_i)_{i=1}^M$ and $(B_i)_{i=1}^M$ will have the same docket if they have the same space-time, or causal, arrangement – in other words, if, for every pair i, j, B_i is in the past of/spacelike to/in the future of B_j exactly when A_i is in the past of/spacelike to/in the future of A_j – and if one sequence can be continuously deformed into the other while the arrangement is essentially unaltered. Much of the geometry is lost if we only consider space-time arrangements, but readers with no prior contact with algebraic topology may wish at first to concentrate on definition 4.2. There is some interesting mathematics arising from the definitions in this section, but I shall merely outline some results here.

Definition 4.1 *A function r from pairs of path connected space-time sets to $\{p, s, f, p \wedge s, f \wedge s, p \wedge s \wedge f\}$ is defined by*

$r(A_1, A_2) = p$	*if every point in A_1 is in the strict timelike past of every point in A_2*
$r(A_1, A_2) = s$	*if every point in A_1 is strictly spacelike to every point in A_2*
$r(A_1, A_2) = f$	*if every point in A_1 is in the strict timelike future of every point in A_2*
$r(A_1, A_2) = p \wedge s$	*if $A_1 \cap A_2 = \emptyset$ and there is a point in A_1 in the null past of a point in A_2, but there are no points in A_1 in the null future of any point in A_2*
$r(A_1, A_2) = f \wedge s$	*if $A_1 \cap A_2 = \emptyset$ and there is a point in A_1 in the null future of a point in A_2, but there are no points in A_1 in the null past of any point in A_2*
$r(A_1, A_2) = p \wedge s \wedge f$	*in any other case.*

An elementary lemma, using the assumption of path connectedness, shows that $r(A_1, A_2) = p \wedge s \wedge f$ if and only if either A_1 intersects A_2, or there is a a point in A_1 in the null past of a point in A_2 and a point in A_1 in the null future of a point in A_2.

Definition 4.2 *A space-time arrangement for M sets is a map*

$$\gamma : \{(i, j) : 1 \leq i < j \leq M\} \to \{p, s, f, p \wedge s, f \wedge s, p \wedge s \wedge f\}.$$

The sequence $(A_i)_{i=1}^M$ has arrangement γ if $r(A_i, A_j) = \gamma(i, j)$ for $i < j$.

Not all maps γ correspond to possible arrangements of sets. For example, $\gamma(1, 2) = p$ and $\gamma(2, 3) = p$ requires $\gamma(1, 3) = p$. The set of possible arrangements characterizes the dimension of space-time[18,19].

Definition 4.3

i) Let $\mathbb{M}$ denote Minkowski space. $A \subset \mathbb{M}$ is a space-time retract if there exists a continuous map $f : \mathbb{M} \to A$ such that $f(a) = a$ for all $a \in A$.

For example, any closed convex set is a retract, and if A_1, A_2, and $A_1 \cap A_2$ are all retracts then so is $A_1 \cup A_2$.

ii) Let Ξ^M denote the set of ordered sequences $(A_i)_{i=1}^M$ of space-time retracts.

iii) For an arrangement γ let

$$\Xi_\gamma^M = \{(A_i)_{i=1}^M \in \Xi^M : r(A_i, A_j) = \gamma(i, j) \text{ for all } i < j\}.$$

iv) Define $\overline{\gamma}$ by taking $\overline{p} = \{p\}$, $\overline{s} = \{s\}$, $\overline{f} = \{f\}$, $\overline{p \wedge s} = \{p, s, p \wedge s\}$, $\overline{f \wedge s} = \{f, s, f \wedge s\}$, and $\overline{p \wedge s \wedge f} = \{p, s, f, p \wedge s, f \wedge s, p \wedge s \wedge f\}$. Thus if $r(A_i, A_j) = \gamma(i, j)$ then $\overline{\gamma}(i, j)$ is the set of possible relations between path connected subsets of A_i and A_j.

v) For $(A_i)_{i=1}^M, (B_i)_{i=1}^M \in \Xi_\gamma^M$ define $(A_i)_{i=1}^M R_\gamma (B_i)_{i=1}^M$ if and only if there exists a continuous map $F : [0, 1] \times \{1, 2 \ldots, M\} \times \mathbb{M} \to \mathbb{M}$ such that $F(0, i, x) \in A_i$ and $F(1, i, x) \in B_i$ for all $x \in \mathbb{M}$, $F(0, i, a) = a$ when $a \in A_i$, $F(1, i, b) = b$ when $b \in B_i$, and, for $0 < t < 1$ and $i < j$, $r(F(t, i, \mathbb{M}), F(t, j, \mathbb{M})) \in \overline{\gamma}(i, j)$.

Proposition 4.4 *R_γ is an equivalence relation.*

proof
i) Let $(A_i)_{i=1}^M \in \Xi_\gamma^M$ and let $f_i : \mathbb{M} \to A_i$ be a retract of A_i.
Set $F(t,i,x) = f_i(x)$ to show that $(A_i)_{i=1}^M R_\gamma (A_i)_{i=1}^M$.
ii) If $(A_i)_{i=1}^M R_\gamma (B_i)_{i=1}^M$ with the map F then $(B_i)_{i=1}^M R_\gamma (A_i)_{i=1}^M$ with the map $(t,i,x) \mapsto F(1-t,i,x)$.
iii) Suppose that $(A_i)_{i=1}^M R_\gamma (B_i)_{i=1}^M$ with the map G and $(B_i)_{i=1}^M R_\gamma (Ci)_{i=1}^M$ with the map H. Choose $b_i \in B_i$. Define F by

$$\begin{aligned} F(t,i,x) &= G(4t,i,x) && \text{for } t \in [0,\tfrac{1}{4}] \\ &= G(1,i,4(\tfrac{1}{2}-t)x + 4(t-\tfrac{1}{4})b_i) && \text{for } t \in [\tfrac{1}{4},\tfrac{1}{2}] \\ &= H(0,i,4(t-\tfrac{1}{2})x + 4(\tfrac{3}{4}-t)b_i) && \text{for } t \in [\tfrac{1}{2},\tfrac{3}{4}] \\ &= H(4(t-\tfrac{3}{4}),i,x) && \text{for } t \in [\tfrac{3}{4},1]. \end{aligned}$$

Then F has the properties required to show that $(A_i)_{i=1}^M R_\gamma (C_i)_{i=1}^M$. ∎

Definition 4.5 *The docket of $(A_i)_{i=1}^M \in \Xi_\gamma^M$ is its equivalence class under the relation R_γ.*

Once the idea of using arrangements is accepted, the extension to dockets is inevitable because exactly the same kind of barrier is passed in changing docket as in changing arrangement.

Proposition 4.6 *The number of dockets on M sets is finite.*

sketch of proof Let $|\gamma|$ denote the number of dockets with arrangement γ. Call an arrangement γ strict if $\gamma(i,j) \in \{p,s,f\}$ for all pairs i,j. For a non-strict arrangement γ, let

$$\mathcal{N}(\gamma) = \{\gamma' : \gamma' \text{ is a strict arrangement with } \gamma'(i,j) \in \overline{\gamma}(i,j)\ \forall i,j\}.$$

The first step in the proof is to show that $|\gamma| \leq \sum_{\gamma' \in \mathcal{N}(\gamma)} |\gamma'|$. This is done by constructing a map, like those defining R_γ, to relate $(A_i)_{i=1}^M \in \Xi_\gamma^M$ to a point $(u_i)_{i=1}^M \in \mathbb{M}^M$ with strict arrangement $\gamma' \in \mathcal{N}(\gamma)$. The next step is to deal with strict dockets. These correspond to components of

$$\{(u_i)_{i=1}^M \in \mathbb{M}^M : \textstyle\prod_{1 \leq i < j \leq M} (u_i - u_j)^2 \neq 0\}$$

so that the result follows from a theorem due to Whitney (Ref. 20, theorem 4). ∎

By using more sophisticated techniques in real algebraic geometry, in particular, the "Milnor-Thom bound", (Ref. 21, Théorème 11.5.2), it is possible to show that the number of strict dockets on M sets in $3+1$ dimensional space-time is less than or equal to $2^{20M}(M!)^5$.

From a mathematical point of view, the strict arrangements which are possible in a space-time of $s+1$ dimensions, form an interesting class of finite posets(19).

Postulate 4.7 *Two instances of a light model bulb will correspond to the same observer if the sequences of regions occupied by the light bulbs when they flash have the same docket.*

Lemma 4.8 *Let $X = (A_m)_{m=1}^M$ and $Y = (B_m)_{m=1}^M$ be two ordered M-tuples of space-time regions. Then X and Y have the same docket if*

A) $B_m = \alpha(A_m)$ for $m = 1, \ldots, M$, where α is a transformation in the identity component of the Poincaré group or a space-time dilation.

B) X and Y are any pair of M-tuples of space-like separated regions.

C) X and Y are any pair of strictly time-ordered M-tuples. ∎

Example 4.9 *Suppose that (A_1, A_2, A_3) has the unique docket defined by the relations $r(A_1, A_2) = p$, $r(A_3, A_1) = s$, $r(A_3, A_2) = s$. Let $a_i \in A_i$ $i = 1, 2, 3$, and chose co-ordinates so that $a_1 = (0, \mathbf{0})$, $a_2 = (x_0, \mathbf{x})$, $a_3 = (y_0, \mathbf{y})$.*

Then $\mathbf{y}^2 + (\mathbf{y} - \mathbf{x})^2 > c^2(y_0^2 + (y_0 - x_0)^2) \geq \frac{1}{2}c^2 x_0^2$, so that A_3 is spatially distanced from at least one of A_1 or A_2 by a distance depending on the temporal separation between A_1 and A_2. ∎

The proofs of 4.8 and 4.9 are straightforward. In lemma 4.8B, we have a docket essentially without structure. In lemma 4.8C, we have correlations. Example 4.9 begins to indicate that when we have mixed spatial and temporal separations, we are given non-trivial topological information. The use of dockets rather than simply space-time arrangements makes this topological information quite rich. For example, there is a space-time arrangement on 25 sets which is only possible in four (or more) dimensions of space-time. The spatial co-ordinates of 4 of the sets define the vertices of a tetrahedron with strictly positive volume. Pairs of distinct dockets with this arrangement correspond to inversions of the tetrahedron. There is one set in the arrangement which is spatially separated from all the other sets. Different dockets correspond to that set having spatial co-ordinates either inside or outside the tetrahedron. Thus, in the mixed regime, dockets can express such fundamental spatial information as handedness and containment. It will be argued in section six that neural switching comes within this mixed regime. Indeed, with the enormous numbers of mixed-regime switchings in a human neural switching model, there will be a vast amount of detailed topological information available from the switching docket.

There are many ways of quantifying this information. As an elementary example, suppose that we have M switchings with switching docket d and arrangement γ. Suppose that a pair of sets of switches correspond to index sets $A, B \subset \{1 \ldots, M\}$. If there are few pairs $a, a' \in A$ such that $\gamma(a, a') = s$ and few pairs $b, b' \in B$ such that $\gamma(b, b') = s$, while there are many pairs $a \in A$, $b \in B$ such that $\gamma(a, b) = s$, then we have a numerical measure of the extent to which A and B are pairs of geometrically-small spatially-distant sets.

The final definition of this section will allow us to impose a time-arrow on observations. Time ordering of state collapses or of measurements is a crucial problem in reconciling relativity theory with any interpretation of quantum mechanics. In the present theory, we shall avoid making an arbitrary choice of time-ordering by maximizing a priori probability over all re-orderings of sequences of collapses which are future-directed – or, more precisely, are never strictly past-directed. It is only because we are considering individual localized observers that this process will give us an adequate time-arrow.

Definition 4.10 *An arrangement γ is ascending if*

$$i < j \Rightarrow \gamma(i,j) \in \{p, s, p \wedge s, f \wedge s, p \wedge s \wedge f\}.$$

A docket is ascending if the corresponding arrangement is.

Any arrangement can be re-indexed to give an ascending arrangement. It should be noted that this would not be true if $f \wedge s$ was not included in the set of possible values.

5. The formal definition of a minimal switching structure.

The first three steps (A, B, and C) in giving a mathematical definition for the concepts introduced so far, are fairly straightforward; indeed, they do little more than introduce notation.

A A minimal ordered switching structure $SO(M, N, d, \varphi)$ is given by:

A1) Two positive integers M (the number of switchings that have occurred) and N (the number of switches).
A2) An M-component ascending docket d. (This defines the spacetime relations between switchings.)
A3) A function $\varphi : \{1, \ldots, M\} \to \{1, \ldots, N\}$. ($\varphi(m) = n$ is to be interpreted as meaning that the mth switching is a switching of switch n.)
A4) Write $\varphi^{-1}(n) = \{j_n(k) : k = 1, \ldots, K_n\}$, where $j_n(1) < j_n(2) < \ldots < j_n(K_n)$. (Switching number $j_n(k)$ is the kth switching of switch n. We shall write $j(\varphi)_n(k)$ in place of $j_n(k)$ in B to display the dependence on φ.) Then, for each $n \in \{1, \ldots, N\}$, $K_n \geq 4$. (A switch must open and close at least twice if all the constraints imposed below are to be brought into play.)

As it is assumed that each switching is a change of switch status, it is not necessary to refer explicitly to switch status in this definition – one switch status will be labelled by $j_n(k)$ for k odd and the other by $j_n(k)$ for k even.

Next, we consider the various allowable re-orderings of switchings in such a structure.

Definition 5.1 *Let $\pi \in \mathcal{S}_M$ – the permutation group of degree M, and let d be a docket on M sets. Then d^π is the docket defined by*

$$(A_m)_{m=1}^M \text{ has docket } d \iff (A_{\pi(m)})_{m=1}^M \text{ has docket } d^\pi.$$

B Given M, N, d, and φ as in A, the minimal switching structure $S(M, N, [d, \varphi])$ is defined by

$$S(M, N, [d, \varphi]) = \{SO(M, N, d', \varphi') : d' \text{ is ascending and } \exists \pi \in \mathcal{S}_M \text{ with } d' = d^\pi,$$
$$\varphi' = \varphi \circ \pi, \text{ and } \pi(j(\varphi')_n(k)) = j(\varphi)_n(k) \text{ for each } n \text{ and } k\}.$$

(Any re-ordering is allowed which is ascending and which, for each n and k, assigns the same possible sets to the kth switching of switch n.)

The label n attached to the switches is arbitrary, but it is not necessary to consider re-orderings over this label as it has no effect. The switching label m, on the other hand, is used to order the state collapses when a priori probability is computed.

The next step is to consider the idea of a switch as a two-status "object" moving through space-time and changing status in specified regions. An object will be defined to be an entity which follows a path $x(t)$ through space-time. The regions occupied by the object will be a set $\{\Lambda(t) : t \in [0,T]\}$ of Poincaré transformations of some initial region Λ with $x(t) \in \Lambda(t)$. We shall have a path $L(t)$ in $\mathcal{L}_+^\uparrow$ (the restricted Lorentz group) such that $\Lambda(t) = \{x(t) + L(t)(x - x(0)) : x \in \Lambda\}$. We shall assume that $L(0)$ is the identity transformation, so that $\Lambda(0) = \Lambda$. As in Ref. 7, this path will eventually be chosen as being that along which the change in local quantum state is minimal; except, of course, when the object changes status. Nevertheless, even at the single switch level, Ref. 7 is now being extended, as only time-translations were considered in that paper. We shall require that $x(t)$, $L(t)$, and $\Lambda(t)$ are continuous, but we shall allow for discontinuities in the four-velocity of the switch and in the derivative of $L(t)$ when changes in status occur.

A Poincaré transformation $(x, L) \in \mathcal{P}_+^\uparrow$ (the restricted Poincaré group) has two components. x is a space-time translation and L is a Lorentz transformation. (x, L) acts on $y \in \mathbb{M}$ by

$$(x, L)y = x + Ly. \tag{5.2}$$

In particular, a transformation sending $x(0)$ to $x(t)$ is $(x(t) - L(t)x(0), L(t))$. This also sends Λ to $\Lambda(t)$. It is necessary to impose a consistency relation on the action of these components on Λ. Assuming that t is proper time on the path $x(t)$ (i.e. that $\left(\frac{dx}{dt}\right)^2 = -1$), $x(t)$ will have four-velocity $u(t) = \frac{dx}{dt}$. The consistency required is that changes in this four-velocity should determine changes in the velocity, up to spatial rotation about $x(t)$ in the co-moving frame, given to the region Λ by the Lorentz transformation $L(t)$. Suppose that switchings occur at parameter times t_k. Then, for $t \in [t_k, t_{k+1})$ we shall require that

$$u(t) = L(t)u(t_k). \tag{5.3}$$

The transformation $L(t)^{-1}$ makes $\Lambda(t)$ look like a translation of Λ and makes $u(t)$ look like $u(t_k)$.

(5.3) can be integrated to give $x(t) = x(t_k) + \int_{t_k}^{t} L(t')u(t_k)dt'$. The paths $x(t)$ and $\Lambda(t)$ then are determined by $x(0)$, by the path $L(t)$, and by the set of four-velocities $u(t_k)$. $\left(\frac{dx}{dt}\right)^2 = -1$ will be a consequence of (5.3) as long as $(u(t_k))^2 = -1$.

The arbitrariness apparently introduced by the choice of $x(0)$ will be resolved later by maximization of a priori probability over all choices. There is also arbitrariness in the choice of Λ. We shall allow Λ to be any spacetime retract (definition 4.3), again maximizing a priori probability in due course.

C The geometrical manifestations of $SO(M, N, d, \varphi)$ comprise the set $GSO(M, N, d, \varphi)$ of all sequences

$$((t_m)_{m=1}^M, (x^n(t), L^n(t), \Lambda_n, P_n, Q_n)_{n=1}^N) \text{ such that}$$

C1) $t_m \in \mathbb{R}$ for $m = 1, \ldots, M$. We shall write t_{nk} for $t_{j_n(k)}$ and T_n for t_{nK_n}. The t_{nk} satisfy $0 = t_{n1} < t_{n2} < \ldots < t_{nK_n} = T_n$. (t_{nk} is the parameter time at which the kth switching of switch n occurs.)

C2) The $x^n(t), n = 1, \ldots, N$, are continuous paths in $\mathbb{M}$ defined for $t \in [0, T_n]$ and with $x^n(0) \in \Lambda_n$.

C3) The $L^n(t)$, $n = 1, \ldots, N$, are continuous paths in $\mathcal{L}_+^\uparrow$ defined for $t \in [0, T_n]$, having a right derivative $L^{n\,\prime}(t^+)$ for $t \in [0, T_n)$, and with $L^n(0) = 1$.

C4) For $n \in \{1, \ldots, N\}$, $k \in \{1, \ldots, K_n - 1\}$, and $t \in [t_{nk}, t_{n(k+1)})$

$$x^n(t) = x^n(t_{nk}) + \int_{t_{nk}}^t L^n(t')u^n(t_{nk})dt' \text{ where } u^n(t_{nk}) \text{ is}$$

a four-vector. (This implies that $\dfrac{dx^n}{dt}(t) = L^n(t)u^n(t_{nk})$.)

C5) The $u^n(t_{nk})$ are timelike, future directed, and $(u^n(t_{nk}))^2 = -1$. (It follows from 4 that $u^n(t) = \dfrac{dx^n}{dt}(t)$ has the same properties and that the path x^n is timelike, future directed, and parametrized by proper time t.)

C6) The Λ_n, $n = 1, \ldots, N$, are space-time retracts.

Set $\Lambda_n(t) = \{x^n(t) + L^n(t)(x - x^n(0)) : x \in \Lambda_n\}$ for $t \in [0, T_n]$.

Set $A_{j_n(k)} = \Lambda_n(t_{nk})$.

C7) $(A_m)_{m=1}^M$ has docket d.

C8) For $n \neq n'$, $t \in [0, T_n]$, and $t' \in [0, T_{n'}]$, $\Lambda_n(t) \cap \Lambda_{n'}(t') = \emptyset$. (If disjointness of $A_{j_n(k)}$ and $A_{j_{n'}(k')}$ were not imposed then a human neural switching model with infinitely many switches and positive a priori probability could be possible. Disjointness of the paths $\Lambda_n(t)$ and $\Lambda_{n'}(t')$ prevents the same physical switches being used in many different abstract switches, each at slightly different times.)

C9) P_n and Q_n are projections in $\mathcal{A}(\Lambda_n)$ with P_n orthogonal to Q_n. (For a space-time set Λ, $\mathcal{A}(\Lambda)$ denotes the local algebra of Λ – the algebra of all observables measurable within Λ (Ref. 5, section 3, Ref. 23). P_n and Q_n are observables defining switch status. Although they are part of the quantum mechanical structure, it is convenient to introduce them at this point.)

6. Neural switches.

In Ref. 7, I proposed an abstract definition of a quantum mechanical switch. I demonstrated that certain neural ion channel proteins could satisfy that definition. This section returns to the neurophysiological analysis of that paper with some substantial extensions and revisions. Like Ref. 7, the purpose of this section is to raise possibilities: to argue that there are ways of interpreting neural functioning as quantum switching. Nevertheless, again like Ref. 7, this section does not provide a definitive identification of any class of entities as being the precise elements in the human brain which function as neural switches.

One reason for this is that there are, at present, too many open questions about the intimate details of molecular motions in the living brain. At a more fundamental level, however, the goal is to demonstrate that an abstract definition is satisfied, but not to supersede that definition. If there was only one type of entity in the human brain which functioned as a quantum switch, and if there was only one way in which the functioning of that entity could be interpreted as switching, and if that functioning was always perfect, then we would not need the abstract definition. As it is, the aim here is merely to identify some of the mechanisms involved in neural switching and to argue that there is at least one sufficiently probable set of sufficiently many neural switches. There is certainly no reason why all the switches in such a set should be of the same type. Given any such set, variations in its definition will undoubtedly be possible. Such variations will include, in the first place, variations in the P_n, Q_n, Λ_n, and other geometrical elements of definition C. This type of variation is straightforwardly dealt with in the abstract framework of definitions A–G; in particular, the supremum in G4 focuses on the variations of maximal a priori probability. This is a major motivation for such an abstract framework. There is also a finite number of possible variations in the minimal switching structures – the $S(M, N, [d, \varphi])$ – to be associated with a given brain. These variations are part of the plurality problem discussed in section 3.

The topic of thermal and environmental fluctuations in the brain is central in Ref. 7. Because of these fluctuations, no localized neural entity ever returns to precisely the same quantum state. Thus, a quantum switch is defined to be an entity with states moving between two circumscribed neighbourhoods, where a precise, abstract, and natural definition of "circumscribed neighbourhood" is given. A given element of the brain may have two functionally distinct modes which are correlated to the firing and non-firing of a particular neuron. For an ion channel, these modes are, in the simplest analysis, open and closed; for a piece of neural membrane, they are levels of polarization due to ambient electric field. Such an element may or may not satisfy the definition of a quantum switch. In particular, the longer an element is considered, the greater the range of fluctuations which it has encountered, and the smaller it has to be if that range is not to have been sufficient to have made the variations in its quantum state too extreme for the required circumscription.

In Ref. 7, I made the assumption that switch lifetimes should be as long as possible. Ion channels are regularly replaced with a half-life of several days (Ref. 24, p. 514, Ref. 25, §6.5). On that time-scale, the most significant type of environmental fluctuation, not affecting function, which is likely to be encountered by a molecular structure within the brain, will be variation in ambient temperature. Allowing for a maximum temperature range implied a nanometre scale for the quantum switch and, on that scale, while ion channels are quantum switches, arbitrary pieces of neural membrane are not.

There is, however, a problem with some ion channels. Some channels appear to diffuse fairly rapidly in the plane of the neural membrane[26]. If these channels were to be used in the structures being developed, it would be necessary to "collapse out" the diffusions at each switching in order to re-localize the switch. Each such collapse would imply a decrease in probability of the switching structure, depending on the area over which a switch may diffuse between switchings. When the diffusion coefficients (10^3 – 10 $(\mathrm{nm})^2$ $(\mathrm{ms})^{-1}$) of mobile membrane proteins are appropriate, there could be an average loss of probability by a factor of at least 10 per switching. 10 corresponds to a switch with switching activity identifiable over a cross-sectional area of 100 $(\mathrm{nm})^2$, diffusing with coefficient 10 $(\mathrm{nm})^2$ $(\mathrm{ms})^{-1}$ and switching at an average rate of 10 times per second.

The resulting rate of loss of a priori probability would be so high that the test imposed by 1.3 would be failed. It is always possible to find a geometric model of essentially any given minimal switching structure with M switches, which has a priori probability smaller than roughly 2^{-M}. Indeed one can use the details of the theory with its allowance for fluctuations in switch states to construct arbitrary switching structures with a priori probability around $(1.14)^{-M}$. Such models can be constructed by imposing, by fiat, the required sequence of collapses on a spin-$\frac{1}{2}$ system with independent spins which return to the appropriate thermal equilibrium between collapses. A concrete model might be given, for example, by independent nuclear spins in a large rock - for example, a dead planet. There will, of course, be some upper bound of the M for which suitable rocks can be found, but human complexity can probably be encompassed. The scale invariance of a docket is used in finding spins with the right space-time relations. Switching structures of this kind will be referred to as "artificial perturbed-equilibrium structures". I intend to provide an analysis of such structures elsewhere. They are artificial in the sense that there is no correlation either between any of the switchings or, except fortuitously, between the physical structure of the system and any pictures of reality which the pattern of switchings might be interpreted as presenting. When we do have such correlations, they ought to yield higher probabilities. The question for this section is how the manifest correlations of a human brain can be expressed in a high probability switching structure.

It might be possible to circumvent the problem of rapidly diffusing ion channels by altering the definition of a priori probability to allow integrals over diffusion processes. I have considered ways of doing this, but none seem entirely satisfactory. On the other hand, many ion channels do appear to be "immobilized" according to recent experiments (Ref. 24, pp 514–519, Ref. 26). Such immobilization, due to links to the cytoskeleton or to clustering in structures like synapses, (Ref. 24, pp 519–522, Ref. 27), may be sufficient to obviate this problem, particularly if we allow a larger size for each switch.

A larger switch size is allowed if we drop the earlier assumption about switch lifetimes. Indeed, dropping this assumption yields significant progress over the analysis of Ref. 7. Ion channels, certainly, have lifetimes shorter than those of the brain in which they exist. Only the minimal constraint A4 is imposed on the lifetime of a switch by the definition given in section 5. It is, therefore, permissible to consider minimizing the number of switchings made by any given switch.

Suppose, for example, that we take a specific switch with a long lifetime from a given switching structure and re-label its switchings as the switchings of a succession of new switches each with a shorter lifetime. In doing this, we shall lose information. That information could be reconstructed, however, if, for example, we were able to say that all of the new switches were identifiable because of some sort of strong geometrical similarity. In a neural switching model, such a complete reconstruction may not be possible, but it should be possible to identify the neuron to which all our switchings belong, and this will be adequate. In general, if a switching docket is sufficiently complex, then it may be quite rare for any pair of switchings to be "adjacent" in the sense – which can easily be made mathematically precise – that the docket obtained by omitting one of them is equal, after the appropriate re-ordering, to the docket obtained by omitting the other. Then if two switchings from different switches are adjacent, that may be taken as valuable evidence that the switches can be identified.

With this idea in mind, we can return to Ref. 7 and ask what entities might function as neural switches if we consider a typical switch lifetime to be only four switchings, which is only two neural firings (on-off, on-off). This is as short as is allowed by A4. According to Abeles (Ref. 28, p. 120), the average firing rate for cortical neurons is approximately five spikes per second, but the firing pattern of these cells is highly irregular and about half the cells fire at less than two spikes per second. This implies an average switching lifetime of less than one second. As long as we are prepared to ignore some firings – and the whole point of the trimming problem (section 3) is that there is no fundamental reason why we should not – then we could even insist, if necessary, that lifetimes had to be under, say, five seconds. The firings we would be ignoring, in this case, would probably not be carrying significant information.

Over an interval of five seconds, changes in the sheltered environment of the brain will mainly be very small. The most dramatic short-term change possible under normal physiological conditions of which I am aware will be the rise in extracellular potassium concentration due to a sustained period of firing in neighbouring neurons (Ref. 29, pp 358-359, Ref. 30). Because this is extracellular, it is significant mainly in that it places a bound on the degree of transverse, whole-membrane, diffusion which can be allowed within a relatively large switch in such circumstances.

On the other hand,if we look for environmental changes which necessarily affect every switch then, once again, temperature changes are the most relevant. Benzinger[31] has measured temperature at the eardrum in such circumstances as entering or leaving hot and cold baths, eating sherbet ice, and performing muscular exercise. His figures show temperature changes as small as 0.1°K take at least 30 seconds. Using the methods of Ref. 7 tells us that the maximum size of a switch subject to a 0.1°K fluctuation in temperature will be about $(32\ \mathrm{nm})^3$. Rewriting this as about $3\ \mathrm{nm} \times (100\ \mathrm{nm})^2$ – with 3 nm representing membrane thickness – we have a size so large that channel diffusion could become irrelevant because, at least in some systems, an average patch of membrane of area $(100\ \mathrm{nm})^2$ will have a high probability of containing at least one channel (Ref. 24, chapter twelve).

Even without channels, the electric polarization change across a relatively large patch of neural membrane due to neural firing can be used to indicate neural switching.

For example, if we consider the polarization change corresponding to a change in potential of 50 mV across a 3 nm membrane of susceptibility $\chi = 2$, then the lower bound provided by Ref. 7, Lemma 8.10 tells us that a patch of volume 3 nm $\times$ (34 nm)^2 could function as a suitable switch. This is well within the limit imposed by short-term thermal fluctuations.

A membrane patch will undergo the transverse diffusion mentioned above. This is limited by membrane elasticity and by internal and external cellular structures. The quantum state to be assigned to such a patch should allow for that diffusion. For example, probability maximization might assign a constant quasi-equilibrium state over the entire switch lifetime to the transverse degree of freedom. In view of the close packing of nerve cells and the supporting neuroglia, which are separated by distances of only 20 nm (Ref. 29, p.364), it seems plausible that we could find a suitable shape for the switch such that it would not be possible, with significant probability, to find within the switch volume a sufficiently large portion of the exterior of the membrane to encounter difficulties with the concentration fluctuations mentioned above.

We now have two candidates for neural switches: anchored ion channels and patches of neural membrane. The comparatively high probability of artificial perturbed equilibrium structures has led us to drop the assumption made in Ref. 7 that diffusing switches can be constantly re-localized by collapse. For the same reason, we need to be more cautious about the assumption that we can use collapse to take us from a state averaged over a switching cycle to a state at a particular phase of the cycle. Indeed, some of the detailed suggestions in Ref. 7, which rely on this assumption, may well require an excessive amount of collapse. Nevertheless, even in normal circumstances, a neural firing time is by no means determined by the times of earlier firings. In the present framework, this implies that a priori probability must be lost in establishing a firing time. Contradiction can only be avoided, if the probability lost can be shared between many switches. A limited amount of state collapse is certainly permitted at each switching and more in unusual circumstances. After all, some decrease of a priori probability with time is inevitable in a many-worlds theory, not least, because of observations of quantum systems, but, in order to satisfy 1.3, the cumulative average loss per switching must be small.

It follows that we must have many strongly-correlated switches on each neuron. The switch candidates proposed above are therefore assumed to be restricted to parts of the neuron which exhibit all-or-nothing firing. They must switch between an average state for the entity concerned in a membrane around resting potential, and an average state in a membrane around firing potential. Possible difficulties with this for sodium channels are raised in Ref. 7, but are probably irrelevant in the present context, because of the allowance for larger switches. The larger the switch, the more polarization by itself will be sufficient to indicate switching status.

Other candidates would be patches of synaptic membrane – either pre- or post-synaptic. The advantages of these candidates are that synapses are comparatively rigid structures which in functioning undergo more dramatic changes than arbitrary pieces of neural membrane. However, the behaviour of post-synaptic membrane may be too complex for our purposes in that it is affected by the firing both of the pre-synaptic neuron and of the neuron to which it belongs.

It is not necessary to choose between these candidates. As long as the abstract definitions are satisfied with sufficiently high probability, a human observer could exist with different switches corresponding to different candidates.

There are many complex issues involved in finding the switches of highest a priori probability in particular circumstances. Some of these issues have been touched on here, but others have been left in the details of Ref. 7. Many of these issues are open questions. How large is the uncertainty in transverse membrane position? How tightly are channels anchored? What are the details of membrane polarization and of channel functioning? The physical models developed in section 8 of Ref. 7 are simplistic and the numbers derived from them are, at best, orders of magnitude. Nevertheless, these numbers do suggest that definitions A–G can apply to the human brain.

It would be a mistake to become so concerned with neurophysiological detail that we lose sight of the fact that a brain functions through information carried in digital form. Definitions A–G are written precisely in order to accomodate that fact. It is conceivable, for example, that, so far, we have stayed too close to the conventional, non-quantum mechanical, picture of the brain. Even in this case, however, it is still possible that quantum switches exist in the brain and that 1.3 is satisfied. As was explained in Ref. 7, it is not appropriate to imagine ball and stick models of the molecules in a quantum switch. The suggestion now is that we may still be under-estimating the richness of the range of possible quantum states on a local algebra. For example, there are so many observables in a local algebra which would be relevant for a thermal neural membrane state, that when one allows for some imposed collapse, it is quite plausible that one could minic the sort of perturbed-equilibrium structures sketched above but using elements sufficiently correlated to neural firing to allow switching with a priori probability loss rate significantly smaller than the loss rate for any artificial system. The imposed collapses in these perturbed-equilibrium structures would allow the switches to be comparatively small, and thus problems with background changes could be avoided. These structures also, I intend to analyse elsewhere.

Once we accept that neural functioning can be interpreted as quantum switching, it is natural to try and investigate how many switchings occur in a human brain in a given time interval. Beyond the roughest of estimates, however, any such investigation will founder on the plurality problem. Nevertheless, we do need to make such a rough estimate here in order to decide whether or not a human neural switching docket can lie in the "mixed-regime" identified in section 4 where we have many pairs of spatially separated switchings as well as pairs which are timelike separated. Light crosses a substantial fraction of the brain in 10^{-10} seconds, while total neuron numbers are usually given nowadays as 10^{11}. There are many distinct classes of neuron varying in size and behaviour. As mentioned above, Abeles[(28)] gives the average firing rate for cortical neurons as approximately five spikes per second. Thus we would be entering the mixed regime with as few as one switch per neuron. However, whatever kind of switch we choose, we can expect to have many thousands or even millions per cell.

One number which is sufficient to indicate this is the estimate (Ref. 28, p. 58) that an "average" human cortical neuron has 40,000 synapses. Each synapse will have an area of order 0.05 $(\mu \text{ m})^2$ (Ref. 28, p. 57) and will include many switching proteins of various types, some of which are almost certainly anchored. On the other hand, even if we restrict attention to the cell body as the cell region most likely to exhibit all-or-nothing firing, we shall be dealing with a surface area of order at a minimum, perhaps, 100 $(\mu \text{ m})^2$. With thousands of switches per neuron, we should certainly be in the mixed regime and should certainly be able to claim that "adjacency" between switches is sufficiently rare as to suggest strong similarity between the switches involved.

Return again to the extra-terrestrial reader whose task is to translate a human neural switching structure. Suppose that the structure he is given is the minimal switching structure corresponding to 10^{11} neurons firing on average 5 times per second with each neuron carrying an average of 2×10^3 switches at any one time. The first step in his task may be thought of as the reconstruction of a geometrical neuronal model from this minimal structure.

He has, in total, 10^{15} switchings per second. This corresponds in the brain he is trying to reconstruct to a length scale of 3×10^{-7} m. The switching docket will give him topological information corresponding to such a length scale. In particular, "adjacent" switches may be expected to correspond to entities separated by such a scale. As an elementary example, he might assign a "time co-ordinate" $t(m)$ to switching m, by defining $t(m)$ to be the number of switchings in the timelike past of m. Using this co-ordinate, he can distinguish switchings corresponding to neurons at the perifery of the brain for which there are spacelike switchings at more different times than there are for more central neurons at around the same time. Some of the information he has will be statistical, so that, for example, analysis of $t(m)$ only tells him that a switching is probably rather than definitely periferal, but there will be so much information that it is hard to imagine that he could consistently develop two significantly different reconstructions. For example, by themselves, the times $t(m)$ do not provide the geometrical meaning of the central/periferal distinction. Nevertheless, as explained in section 4, "inside" and "outside" are terms definable from sufficiently complex dockets, so our extra-terrestrial can find both the distinction and its meaning.

Using all his information, it will be possible for him to divide switches into spatially distinguishable classes corresponding to at most a few neurons. He can refine his classes to correspond to single neurons by using the correlations between switches on each separate neuron. As his classes are already small, this will even be possible with short lifetime switches. It is an important contingent fact about the human brain is that throughout its lifetime each neuron has a roughly constant geometrical relation to most of the other neurons. Because of this, the individual neurons that he has identified at each instant, can, in fact, be followed through time.

The extra-terrestrial has now completed his first step. He can identify, up to scale and Poincaré invariance, the times and places of each neural firing in his model. All he has left to do is to look at the world through this pattern of neural firing and learn to understand it. This may seem an awesome task, but I suspect that we have all been accomplishing something similar from infancy.

The aim of this paper is the characterization of the physical structure of an observer by a pattern of elementary localized switching events. In this section, suitable events have been discovered in the human brain. This allows the essential claim that humans do satisfy the proposed characterization. However, the choices allowed in this section accentuate the plurality problem. Perhaps as few as 100 switches per neuron could provide a suitable geometrical switching structure. On the other hand, switches cannot be reduced in size beyond a certain point without increasing loss of probability and they cannot overlap by C8. However, because of the strong correlation between switches on the same neuron, there need not be significant probability cost per switch in increasing their number over some considerable range. It is not necessary for all switches to have the shortest allowed lifetime. It follows that, in general, for any given bound on loss of a priori probability, if one minimal switching structure is possible then a large, albeit finite, variety of different minimal switching structures will also be possible.

7. The formal definition of a family of quantum switches.

This section defines the set $\mathcal{N}(W)$ of sequences of quantum states which are manifestations of families of quantum switches with a given switching geometry

$$W = ((t_m)_{m=1}^{M}, (x^n(t), L^n(t), \Lambda_n, P_n, Q_n)_{n=1}^{N}) \in GSO(M, N, d, \varphi).$$

Perhaps the most important aspect of the definition is that it is entirely abstract. In particular, there is no explicit reference to biological concepts such as carbon atoms or to statistical mechanical concepts such as temperature or information. This means that the definition can be part of an axiomatic characterization of the physical structure of an observer. Of course, it might be possible to invoke the existence of carbon in such a characterization, but, in that case, we would need to invoke a specific quantum field theory for the universe, and be able to characterize its low energy structures in considerable detail.

Given $(x, L) \in \mathcal{P}_+^{\uparrow}$, we shall use the same notation ($\tau_{(x,L)}(A)$ and $\tau_{(x,L)}(\sigma)$) for the dual pair of transformations which act on an observable A and on a state σ. If A is in the local algebra $\mathcal{A}(\Lambda)$ then $\tau_{(x,L)}(A) \in \mathcal{A}((x, L)\Lambda)$ where, by (5.2), $(x, L)\Lambda = \{x + Ly : y \in \Lambda\}$. By definition, the identity $\tau_{(x,L)}(\sigma)(\tau_{(x,L)}(A)) = \sigma(A)$ will hold. In E and F, we shall constantly Poincaré transform back to the local algebra $\mathcal{A}(\Lambda_n)$ in order to analyse changes in quantum states along the paths $x^n(t)$.

In D, which is an implementation of ideas discussed in Ref. 5, we identify the local observables accessible to our observer.

D

D1) Let $\mathcal{C}(W)$ be the von Neumann algebra generated by

$$\{\tau_{(x^n(t_{nk})-L^n(t_{nk})x^n(0),L^n(t_{nk}))}(P_n), \tau_{(x^n(t_{nk})-L^n(t_{nk})x^n(0),L^n(t_{nk}))}(Q_n) : k = 1, \ldots, K_n, n = 1, \ldots, N\}.$$

(This is the algebra of correlations of switch projections experienced by the observer.)

D2) Let $\mathcal{B}(W)$ be the norm closure of the linear span of

$$\{A_1C_1 + C_2A_2 : A_1, A_2 \in \mathcal{A}(\Lambda_n(t)), C_1, C_2 \in \mathcal{C}(W), t \in [0, T_n], n = 1, \ldots, N\}.$$

(This is the set of all observables accessible to the observer. Elements of $\mathcal{C}(W)$ are correlated with local observables along the paths of the switches.)

E defines the set of sequences of states for which $x^n(t)$ is the path along which change of state is locally minimized. E1 requires that the states be such that the initial conditions $u^n(t_{nk})$ and $L^{n\,\prime}(t_{nk}{}^+)$ are optimal and E2 requires that the continuation at parameter t is optimal.

E $\mathcal{N}(W, E)$ is the set of all sequences of restrictions to $\mathcal{B}(W)$ of sequences of quantum states $((\sigma_m)_{m=1}^M)$ which satisfy the following requirements for each $n \in \{1, \ldots, N\}$ and each $k \in \{1, \ldots, K_n - 1\}$:
E1) Set

$$\begin{aligned} X^{nk} = \{(L,v) : & \ L \text{ is a } C^1 \text{ path in } \mathcal{L}_+^{\uparrow} \text{ on some interval} \\ & [t_{nk}, t_{nk} + \varepsilon) \text{ with } \varepsilon > 0 \text{ and with } L(t_{nk}) = L^n(t_{nk}), \text{ and} \\ & v \text{ is a future directed four-vector satisfing } (v)^2 = -1\}. \end{aligned}$$

For $(L, v) \in X^{nk}$, define $f_{nk}(s, L, v) = \tau^{-1}_{(y^{nk}(s,L,v),L(s))}(\sigma_{j_n(k)})|_{\mathcal{A}(\Lambda_n)}$ where

$$y^{nk}(s, L, v) = x^n(t_{nk}) + \int_{t_{nk}}^{s} L(t')v dt' - L(s)x^n(0).$$

Then we require that f_{nk} has a right derivative at $s = t_{nk}$ and that

$$\inf\{ \ || \lim_{h \to 0^+} (f_{nk}(t_{nk} + h, L, v) - f_{nk}(t_{nk}, L, v))/h \ || : (L, v) \in X^{nk}\}$$

is attained when $L'(t_{nk}{}^+) = L^{n\,\prime}(t_{nk}{}^+)$ and $v = u^n(t_{nk})$.
E2) For each $t \in (t_{nk}, t_{n(k+1)})$, set

$$X^n_t = \{L : L \text{ is a } C^1 \text{ path in } \mathcal{L}_+^{\uparrow} \text{ on some interval } [t, t + \varepsilon) \text{ with } \varepsilon > 0, \text{ and } L(t) = L^n(t)\}.$$

(X^n_t and X^{nk} could be replaced by finite dimensional sets defined in terms of the Lie algebra of $\mathcal{L}_+^{\uparrow}$.)
For $L \in X^n_t$, define $f_t(s, L) = \tau^{-1}_{(y^n_t(s,L),L(s))}(\sigma_{j_n(k)})|_{\mathcal{A}(\Lambda_n)}$ where

$$y^n_t(s, L) = x^n(t) + \int_t^s L(t')u^n(t_{nk})dt' - L(s)x^n(0).$$

Then we require that f_t has a right derivative at $s = t$ and that

$$\inf\{ \ || \lim_{h \to 0^+} (f_t(t + h, L) - f_t(t, L))/h \ || : L \in X^n_t\}$$

is attained when $L'(t^+) = L^{n\,\prime}(t^+)$.

F is the formal expression derived and explained in section 5 of Ref. 7 for the idea that "A switch is something spatially localized, the quantum state of which moves between a set of open states and a set of closed states, such that every open state differs from every closed state by more than the maximum difference within any pair of open states or any pair of closed states." (Ref. 7, hypothesis III).

F $\mathcal{N}(W)$ is the subset of $\mathcal{N}(W, E)$ consisting of sequences $((\sigma_m)_{m=1}^M)$ such that, setting

$$\sigma_{nk} = \tau^{-1}_{(x^n(t_{nk}) - L^n(t_{nk})x^n(0), L^n(t_{nk}))}(\sigma_{j_n(k)})|_{\mathcal{A}(\Lambda_n)},$$

for each $n \in \{1, \ldots, N\}$ and for $k,\ k' \in \{1, \ldots, K_n\}$,

F1) $\sigma_{nk}(P_n) > \frac{1}{2}$ for k odd. ("a set of open states")

F2) $\sigma_{nk}(Q_n) > \frac{1}{2}$ for k even. ("a set of closed states")

F3) $|\sigma_{nk}(P_n) - \sigma_{nk'}(P_n)| > \frac{1}{2}$ and $|\sigma_{nk}(Q_n) - \sigma_{nk'}(Q_n)| > \frac{1}{2}$ for all pairs k and k' with different parity. ("every open state differs from every closed state")

F4) There is no triple (P, k, k') with $P \in \mathcal{A}(\Lambda_n)$ a projection and k and k' of equal parity such that $|\sigma_{nk}(P) - \sigma_{nk'}(P)| \geq \frac{1}{2}$. ("by more than the maximum difference within any pair of open states or any pair of closed states.")

In Ref. 7, hypothesis V, there was a constraint (V(4)) aimed at avoiding the possibility of ignoring part of a switch's activity. This constraint has been dropped here because the switching structure is assumed to be given in advance. An inclination to see the set of quantum states as defining the switching structure did creep into Ref. 7, but here priority is definitely given to the switching structure.

For a given geometry W, $\mathcal{N}(W)$ may be empty. Such sets will be assigned zero probability (by definition). Section 6 can be thought of as discussing sequences of possible quantum states for a brain. Arbitrary sufficiently-small variations of these states will also be permissible brain states. For appropriate neural switching structures, some of these sequences will belong to $\mathcal{N}(W)$ for some geometry W. The differentiability requirements for E will be satisfied by a dense subset of sequences, and the paths required to define W can be constructed on elements of that subset. The physical properties required in section 6 will then ensure that the conditions in F are also satisfied.

8. The a priori probability of a minimal switching structure.

The object of Ref. 5 is to introduce and analyse a definition of a priori probability which can be directly applied to geometrical switching structures. This definition is based on a function $\mathrm{app}_{\mathcal{B}}((\sigma_m)_{m=1}^M \,|\, \omega)$ of a set of operators $\mathcal{B}$, a sequence of quantum states $(\sigma_m)_{m=1}^M$ on that set, and an initial quantum state ω. ω is the state corresponding to the "universal wavefunction" of Everett. $\mathrm{app}_{\mathcal{B}}((\sigma_m)_{m=1}^M \,|\, \omega)$ measures the a priori probability for $(\sigma_m)_{m=1}^M$ to be observed as a sequence of generalized "wave-packet collapses" starting from ω, by an observer who is given the information available from $\mathcal{B}$. The collapses are generalized because we are dealing with mixed as well as pure states. app has many appropriate properties. The following pair are particularly important.

The first shows that successive collapses are treated as independent events:

$$\mathrm{app}_{\mathcal{B}}((\sigma_m)_{m=1}^M \,|\, \omega) = \prod_{m=1}^{M} \mathrm{app}_{\mathcal{B}}(\sigma_m \,|\, \sigma_{m-1}) \tag{8.1}$$

where $\sigma_0 = \omega$.

The second shows that the function generalizes the idea that the a priori probability of seeing a state σ "collapse" out of a mixture of the form $\rho = p\sigma + (1-p)\sigma_d$ where σ_d is disjoint from σ, is the coefficient p of σ:

8.2) Suppose that $\rho = p\sigma + (1-p)\sigma_d$ for $0 \le p \le 1$ and suppose that there exists a projection $Q \in \mathcal{B}$ such that $\sigma(Q) = 1$, $\sigma_d(Q) = 0$.

Then $\mathrm{app}_{\mathcal{B}}(\sigma \,|\, \rho) = p$.

G By using this function in accordance with Ref. 5, the following value is assigned as the a priori probability $\mathrm{app}(W \,|\, \omega)$ of existence of an individual geometrical manifestation $W \in GSO(M, N, d, \varphi)$:

G1) For $m = 1, \dots, M$, define

$$\mathcal{N}^m(W) = \{(\sigma_i)_{i=1}^m : \exists (\sigma_i)_{i=m+1}^M \text{ with } (\sigma_i)_{i=1}^M \in \mathcal{N}(W)\}.$$

G2) Define, by induction on m, the following a priori probabilities. Start with

$$\mathrm{app}(\mathcal{N}(W), \mathcal{B}(W), 1, \omega) = \sup\{\mathrm{app}_{\mathcal{B}(W)}(\sigma \,|\, \omega) : \sigma \in \mathcal{N}^1(W)\}.$$

Then, for $1 < m+1 \le M$, set

$$\begin{aligned} &\mathrm{app}(\mathcal{N}(W), \mathcal{B}(W), m+1, \omega) \\ &\quad = \sup\{\limsup_{n\to\infty} \mathrm{app}_{\mathcal{B}(W)}((\sigma_i^n)_{i=1}^{m+1} \,|\, \omega) : ((\sigma_i^n)_{i=1}^{m+1})_{n\ge 1} \text{ is a sequence of} \\ &\qquad\qquad \text{elements of } \mathcal{N}^{m+1}(W) \text{ and, for } 1 \le k \le m, \\ &\qquad\qquad \mathrm{app}_{\mathcal{B}(W)}((\sigma_i^n)_{i=1}^k \,|\, \omega) \to \mathrm{app}(\mathcal{N}(W), \mathcal{B}(W), k, \omega)\}. \end{aligned}$$

G3) Define $\mathrm{app}(W \,|\, \omega) = \mathrm{app}(\mathcal{N}(W), \mathcal{B}(W), M, \omega)$.

This definition gives the a priori probability of W as the maximum a priori probability to which a sequence of elements of $\mathcal{N}(W)$ can approximate, given that the sequences of initial portions of those elements also approach maximal a priori probability. The inductive nature of the definition imposes a causal structure according to which the most likely states for a given switching are influenced by states earlier in sequence, but not by later states.

The definition is extended to an a priori probability for the whole of $GSO(M, N, d, \varphi)$ simply by taking a supremum. Finally, an additional supremum takes account of the re-orderings allowed by B, to give the a priori probability for the minimal switching structure $S(M, N, [d, \varphi])$ as

G4) $$\begin{aligned} \mathrm{app}(S(M, N, [d, \varphi]) \,|\, \omega) = \sup\{&\mathrm{app}(W \,|\, \omega) : W \in GSO(M, N, d', \varphi') \\ &\text{where } SO(M, N, d', \varphi') \in S(M, N, [d, \varphi])\}. \end{aligned}$$

The idea behind all the suprema in G2 and G4 is that we are dealing with a range of structures between which the observer cannot distinguish.

One goal in this paper is to formulate complete and explicit definitions. These definitions are intended to be the simplest mathematically-coherent expression of a circle of underlying ideas about what quantum physics may be telling us about the nature of reality. The correctness of the details of the definitions is no more accessible to direct test than is the correctness of the details of string theory. The present theory stands or falls on whether more attractive alternative theories can be found and on whether the definitions are consistent both internally and with the picture of reality being developed. For example, if the general interpretation of quantum mechanics developed throughout this work is indeed consistent both with the axiomatic characterization of the function app, given in Ref. 5, and with the analysis of the human brain, given in

Ref. 7 and in this paper, then it should be the case that, for a sufficiently complex neural switching structure $S(M, N, [d, \varphi])$ for which $\mathrm{app}(S(M, N, [d, \varphi]) \,|\, \omega)$ is comparatively large, the only geometries $W \in GSO(M, N, d', \varphi')$ where $SO(M, N, d', \varphi') \in S(M, N, [d, \varphi])$ for which $\mathrm{app}(W \,|\, \omega)$ is close to $\mathrm{app}(S(M, N, [d, \varphi]) \,|\, \omega)$ are such that $\mathrm{app}(W \,|\, \omega)$ is approached, in the context of G2, by and only by sequences $(\sigma_m)_{m=1}^{M}$ of states which are restrictions to $\mathcal{B}(W)$ of local quantum states describing a human brain. One of the most direct ways of attacking the theory would be to construct a counter-example to a claim like this. The artificial perturbed-equilibrium switching structures of section 6, for instance, are unsuccessful counter-examples. The claim is a technical restatement of contraint 1.2 and a significant part of constraint 1.3.

Showing that neural switching models satisfy the definition in section 7, thus satisfying 1.2, was one aim of Ref. 7 and its revision suggested in section 6. I now make the further claim that constraint 1.3 is fully satisfied. This claim cannot be proved and will also always be to open to falsification by example. Indeed, ruling out various potential falsifications was a significant method in the development of the theory. Ultimately, it can only be argued that it is difficult to imagine how the definition could be satisfied, other than by a human brain, or something like a human brain, without continual loss of a priori probability by the imposition of arbitrary "collapses". It is unlikely for the state of a system to change and return towards the original state and then go back towards the second state in the way required by the definition of a quantum switch. A pattern of such behaviour sufficiently complex to be interpretable as a physical manifestation of consciousness, is all the more unlikely.

With a human brain, according to section 6, suitable switches can be found for which the average loss of probability per switching is comparatively small. This depends not only on the nature of the switches, but also on the correlations between switches which reflect the fact that it is possible to interpret neural switchings as being "caused" by a largely deterministic world "external" to the observer. Definition G looks for the most probable sequence of quantum states on the limited set of observables $\mathcal{B}(W)$. The idea of an external world implies extension to a larger set. Which larger sets are appropriate cannot be specified exactly, but they could include sufficient observables to define the states of macroscopic objects up to the accuracy with which they are being observed. The only fundamental sets of observables are the sets $\mathcal{B}(W)$. The idea of "external world" states in the present sense – states other than the universal state ω – is not a fundamental and defined concept, but merely a heuristic one (cf. Ref. 5 and section 10 of this paper).

The claim that 1.3 is satisfied suggests that if we look at local quantum states on such larger sets, then the most probable sequences of "collapses" out of the universal state ω in which one can identify a family of objects moving through space-time and switching with the complexity of switching pattern of a human brain, are sequences in which those switches are built out of patches of cell membrane and placed in a biologically evolved organism. This is also part of the idea that a human neural switching model has an interpretation which is an accurate picture of reality, and it is a statement which invites an anthropic treatment of cosmology. In section 10, some of the technical details relevant to this suggestion will be considered. These details will prove neither the suggestion nor the weaker form of it stated above, which referred only to states on $\mathcal{B}(W)$. Section 10 will be merely another approach in trying to establish a consistent circle of ideas.

9. Sets of switching structures.

Only one more postulate is needed to complete the theoretical structure. G4 defines "a priori probability" for individual minimal switching structures. All this does to give numerical values measuring the extent to which some such structures are more "likely" than others. Such values have some intrinsic meaning, but are not sufficient as a foundation for physical probabilities without a mechanism for calculating the relative a priori probability of sets of observable events. We therefore postulate that G4 can be treated as defining a classical but unnormalized probability on the space of minimal switching structures, so that the a priori probability of an observer experiencing a set $\{S_1, \ldots, S_Z\}$ of distinct structures should be proportional to

$$\mathrm{Prob}\{S_1, \ldots, S_Z\} = \sum_{z=1}^{Z} \mathrm{app}(S_z \,|\, \omega). \tag{G5}$$

The application of G5 requires the ability to specify suitable sets of minimal switching structures. Of course, it is impossible to imagine an observer specifying his own structure. Indeed, at the human level of complexity, the idea of calculating with G5 is fantasy. It is even doubtful whether the most obvious sets of switching structures which one would want for the use of G5 are precisely definable. For example, how could one possibly define precisely the set of minimal structures which would model the event of an observer seeing a given result for an experiment? Even given the strongest (algorithmic) form of the translatability claim in section 2, one would have to deal with, for example, distractions of the observer and partial failures of the experiment. The event in question is itself not precisely definable. In addition, in applying G5, we are faced again with the plurality and trimming problems of section 3.

Nevertheless, G5 does introduce a well-defined function on sets of minimal switching structures. One could, in theory, make computations by approximating the universe by a sequence of finite lattice models with suitable finite-dimensional Hilbert spaces on them. According to Prop. 4.6, for a bounded number of switchings, there is only a finite number of possible dockets. Thus, even at human complexity, there is only a finite number of different possible future extensions within bounded additional complexity. The complete interpretation of the elements of that finite set may well be an insoluble problem, but a partial classification adequate for accurate approximate calculation is at least conceivable.

10. Quantum probabilities.

Using G5 the observer can theoretically estimate relative probabilities for future events. Indeed, G5 should provide the foundation for a complete theory of physical probabilities. In this section, elaborating on Ref. 5, we shall consider the relation between G5 and textbook quantum probabilities. The probability theory developed in Ref. 5 was designed so that, in the context of observer structures of the type proposed here, we could use the many and various arguments demonstrating that for macroscopic systems interference effects are often negligible in practice. This idea, expounded at length in Ref. 5, is the key to the claims made in this section. The comments made in the previous section about the difficulties of defining events and corresponding sets of switching structures precisely will be relevant throughout this section, but will not be repeated.

Suppose that an observer O has experienced the minimal structure S up to time T and wishes to discover the probability of observing a certain experiment to have outcome a relative to that of it having outcome b. It will be assumed here that a and b are fairly broadly defined events like whether there is a number on the screen, or whether it is twenty, or whether it is between twenty and forty. Suppose that O calculates that he will observe outcome a (resp. b) by time T_2 if and only if his future minimal structure is in the set X_a (resp. X_b). Elements of X_a and X_b will be extensions defined in the obvious way of S. The relative probability sought will be

$$T(a|b) = \mathrm{Prob}(X_a)/\mathrm{Prob}(X_b). \tag{10.1}$$

The analysis of section 6 of Ref. 5 provides some basis for believing that there is agreement between (10.1) and calculations in conventional quantum theory. This has the consequence that (10.1) can be taken to be empirically supported by the empirical evidence for the conventional theory. In Ref. 5 a generalization of definitions G1–G3 was worked with. This allowed the definition of a function $\mathrm{app}(O, T, \mathcal{B}_S, C \,|\, \omega)$ measuring the a priori probability of an observer O, at time T, observing a subsystem, defined on a set of observables $\mathcal{B}_S$, to occupy a set of states C. The idea of an observer, however, was limited to something with a structure like $\mathcal{N}(W)$; although even that was not formally defined. In the context of this paper, the corresponding function will be written as $\mathrm{app}(W, \mathcal{B}_S, C \,|\, \omega)$ and defined for a given geometric switching structure W. We shall assume that W is chosen so that $\mathrm{app}(W \,|\, \omega)$ is suitably close to $\mathrm{app}(S(M, N, [d, \varphi]) \,|\, \omega)$.

10.2) Let $\mathcal{B}_S(W)$ be the norm closure of the linear span of

$$\mathcal{B}(W) \cup \{B_1C_1 + C_2B_2 : B_1, B_2 \in \mathcal{B}_S \cup \mathcal{B}_S^*, C_1, C_2 \in \mathcal{C}(W)\}.$$

Define $\mathcal{N}^m(W)$ as in G1, except that, in the definition of $\mathcal{N}(W)$ we identify the restriction of a state to $\mathcal{B}(W)$ with the set of all possible extensions to $\mathcal{B}_S(W)$ of that restriction (Ref. 5, 3.10–3.12). Then define, by induction on m, the following a priori probabilities. Start with

$$\mathrm{app}(\mathcal{N}(W), \mathcal{B}_S(W), 1, \omega) = \sup\{\mathrm{app}_{\mathcal{B}_S(W)}(\sigma \,|\, \omega) : \sigma \in \mathcal{N}^1(W)\}.$$

Then, for $1 < m + 1 \leq M$, set

$$\begin{aligned}
&\mathrm{app}(\mathcal{N}(W), \mathcal{B}_S(W), m + 1, \omega)\\
&= \sup\{\limsup_{n\to\infty} \mathrm{app}_{\mathcal{B}_S(W)}((\sigma_i^n)_{i=1}^{m+1} \,|\, \omega) : ((\sigma_i^n)_{i=1}^{m+1})_{n\geq 1} \text{ is a sequence of}\\
&\qquad \text{elements of } \mathcal{N}^{m+1}(W) \text{ and, for } 1 \leq k \leq m,\\
&\qquad \mathrm{app}_{\mathcal{B}_S(W)}((\sigma_i^n)_{i=1}^{k} \,|\, \omega) \to \mathrm{app}(\mathcal{N}(W), \mathcal{B}_S(W), k, \omega)\}.
\end{aligned} \tag{10.3}$$

Finally, define

$$\begin{aligned}
&\mathrm{app}(W, \mathcal{B}_S, C|\omega) \\
&\quad = \sup\{\limsup_{n\to\infty} \mathrm{app}_{\mathcal{B}_S(W)}((\sigma_m^n)_{m=1}^{M+1} \,|\, \omega) : ((\sigma_m^n)_{m=1}^{M+1})_{n\geq 1} \subset \mathcal{N}^M(W), (\sigma_{M+1}^n)_{n\geq 1} \subset C, \\
&\qquad \text{and, for } 1 \leq k \leq M, \mathrm{app}_{\mathcal{B}_S(W)}((\sigma_i^n)_{i=1}^k \,|\, \omega) \to \mathrm{app}(\mathcal{N}(W), \mathcal{B}_S(W), k, \omega)\}.
\end{aligned} \tag{10.4}$$

With these definitions, it is argued in Ref. 5 that if the observer observes the outcome of an experiment on a macroscopic subsystem defined on a set of observables $\mathcal{B}_S$ and if two of the possible results of that experiment, a and b, have conventionally calculated probabilities p_a and p_b, then

$$\mathrm{app}(W, \mathcal{B}_S, C_a \,|\, \omega)/\mathrm{app}(W, \mathcal{B}_S, C_b \,|\, \omega) \sim p_a/p_b, \tag{10.5}$$

where C_a (resp. C_b) is the set of states on $\mathcal{B}_S$ modelling outcome a (resp. b). The justification in Ref. 5 of (10.5) depended on the claim that the function app yields a priori probabilities which are such that the most likely states modelling a given situation are states which conventional quantum theorists would allow to be assigned to that situation. This claim is made more plausible now because the extra supremum of G4, which we are implicitly applying in choosing W, makes mathematical pathology less likely.

Using G4, (10.3), and (10.4), we can set about identifying, at a technical level, "likely states modelling a given situation". For example, just before the result of the experiment is seen, the observer will assign to $\mathcal{B}_S$ the restriction to that set of observables of the states σ_M such that for sufficiently small $\delta > 0$, there exists $(\sigma_m)_{m=1}^{M-1}$ with $(\sigma_m)_{m=1}^{M} \in \mathcal{N}(W)$, where W is such that

$$\mathrm{app}(W \,|\, \omega) \geq \mathrm{app}(S(M, N, [d, \varphi]) \,|\, \omega) - \delta, \text{ and, for } 1 \leq k \leq M,$$
$$\mathrm{app}_{\mathcal{B}_S(W)}((\sigma_i)_{i=1}^k \,|\, \omega) \geq \mathrm{app}(\mathcal{N}(W), \mathcal{B}_S(W), k, \omega) - \delta.$$

For each δ, this defines a set of states on $\mathcal{B}_S$. The basis of the argument for (10.5) is then that all these states should be similar and should be close to the states which conventional quantum theorists would assign to the situation in question. This is now a technical version of the statement that a human neural switching model has an interpretation which is an accurate picture of reality. It requires that all the information which an observer would use to predict the state of the world on the set $\mathcal{B}_S$ is expressed in the minimal switching structure $S(M, N, [d, \varphi])$ and that the a priori probability function defined in Ref. 5, correctly correlates that information with the state on $\mathcal{B}_S$.

In order to relate (10.5) to the more fundamental statement that

$$\mathrm{Prob}(X_a)/\mathrm{Prob}(X_b) \sim p_a/p_b, \tag{10.6}$$

note that $\mathrm{Prob}(X_a)$ (resp. $\mathrm{Prob}(X_b)$) can be viewed as a sum over terms like the numerator (resp. the denominator) of (10.5) with the observer making a sequence of observations of his own neural switches. Indeed, in this context, the problem raised in section 6 of Ref. 5 of defining a quantum state over an extended time interval is reduced by considering only local states within the brain. On the other hand, one might object that in the measurement model in Ref. 5, distinct outcomes were distinguished by projection values differing by close to unity, while in the definition of a quantum switch, a switching only required a change in expected value for some projection of more than $\frac{1}{2}$ (F3). However, these two different requirements can be reconciled by noting that, in the situation modelled by (10.6), many switches will be involved, so that it will be easy to achieve near-orthogonality between distinct results. Thus (10.6) follows from the arguments for (10.5) as long as the assumption can be made that there is nothing intrinsic to a or b which affects, for example, the number of terms over which numerator and denominator are summed.

This event independence assumption seems entirely natural. There are two cases in which it might fail. One case is that one of the events might have a significant effect on the observer – the obvious example being Schrödinger's cat experiment considered from the cat's point of view. In this case failure of (10.6) is only to be expected. It is also possible that (10.6) could fail because different criteria (for example, different approaches to the plurality problem) are applied in choosing X_a and X_b. This, of course, would be stupid. It is mentioned only in order to emphasize that in the interpretation of relative probabilities derived from G5, it is necessary to ensure that like is compared with like.

The ratio defined by (10.1) is not directly observable. This brings us back to the problem of the meaning of a priori probabilities about which some brief comments are made in Ref. 5. If the observer wishes to measure the probability of outcome a relative to outcome b, he will repeat the experiment many times and count outcome frequencies. Suppose that he finds these to be f_a and f_b respectively. He will then estimate the relative probability as $F(a|b) = f_a/f_b$. Under suitable circumstances, including, for example, the event independence mentioned above, $F(a|b)$ should be likely to be approximately equal to theoretical relative probabilities such as $T(a|b)$ or p_a/p_b calculated using appropriate physical theories. In conventional probability theory, the justification of such an approximate equality would depend on the laws of large numbers. In the present situation, there remains a close relationship with conventional probability theory expressed by (10.5) and (10.6). Thus it is possible to claim as an example of (10.6) that there will be some ζ close to p_a/p_b and to $T(a|b)$ such that, within the bounds of practicality, for any $\varepsilon > 0$ and any $\eta > 0$, there will exist N_0 such that, if the experiment is repeated N_0 or more times then

$$\mathrm{Prob}(|F(a|b) - \zeta| > \varepsilon)/\mathrm{Prob}(|F(a|b) - \zeta| \leq \varepsilon) \leq \eta. \tag{10.7}$$

"Prob" in (10.7) is, of course, the unnormalized probability defined in G5.

(10.7) is not proposed as the statement of a mathematical theorem. Even leaving aside the question of a practical bound on N_0, it could not be so unless the problems of section 9 could be solved algorithmically. (10.7) is intended as a model, of which a variant of some kind is necessary, to give meaning to G4 and G5. In other words, without arguments, like those given in Ref. 5 and this section, for something like (10.7) – in appropriate circumstances, for suitable reasonably large N_0 and suitable reasonably small ε and η – part G of the definition and the phrase "a priori probability" implicit in its notation would be empty. With such partial confirmation, however, G4 and G5 can be proposed as defining numbers fundamental to any situation.

11. Conclusion.

The physical structure of an observer at a given time is characterized as a minimal switching structure of the form $S(M, N, [d, \varphi])$, which defines the information processing structure of the observer's past, together with the corresponding quantum switching structure, expressed by the set of sets of sequences of quantum states

$$\{\mathcal{N}(W) : W \in GSO(M, N, d', \varphi') \text{ where } SO(M, N, d', \varphi') \in S(M, N, [d, \varphi])\}.$$

A primary motivation for the many worlds interpretation came from the idea of quantizing general relativity, and, in particular, of applying quantum theory to a closed universe with no external observer. The theory proposed here should be satisfactory for this purpose. Indeed, there might even be more relevance for quantum general relativists in the fact that the dockets of section 4, which lie at the heart of this analysis of quantum theory, are both finite and deeply geometrical (cf. Refs 18, 22).

As far as cosmology is concerned, as mentioned at the end of section 8, the present theory is particularly well adapted to an anthropic treatment. Nevertheless, the automatic invocation of Ref. 9 in this context, should be tempered by the comment that if consciousness requires a physical substrate of a particular kind, then it is not purely a functional issue, and the "von Neumannn machines" of Ref. 9 may not be truly alive.

It would be an exaggeration to pretend that applications for the present theory might also be found in the study of history, but Hawthorn[32] has emphasized the importance in that field of imagining alternative worlds. Having an analytical basis for such imaginings might not be entirely without intellectual value.

Lewis[33,34] has insisted that the existence of worlds in which counterfactuals are valid would simplify our understanding of modal logic. The present theory provides him with such worlds. It may not provide for all the possibilities that could be imagined, but it does provide for the appearance of all the possibilities that could be imagined as appearing to observers like ourselves. For example, this is, presumably, a universe in which the fine structure constant is close to $1/137$. There are minimal switching structures possible in this universe (with our ω and our time propagation) for whom experiments have always pointed to a value of $1/142$. Such switching structures, presumably, have smaller a priori probability than us and, presumably, have lost considerable probability every time they have made tests of quantum electrodynamics, but they do exist. The advantages of the present theory over Lewis's are; the specification of a priori probability, the finiteness, and the connection to physics.

One crucical question posed by the many-worlds idea is "How many worlds?". If Schrödinger's cat is both alive and dead at the end of its ordeal, then in how many ways is it alive and dead? This paper presents a formalism according to which the answer is essentially finite, but may be different for the cat and the experimenter and the experimenter's friend. Of course, no specific number of worlds has been given, but we do at least have a framework. We could consider the total number of different worlds observed by all conceivable observers with a given bounded complexity. Or we could consider the total number of different worlds observed by all observers with a bound on complexity and with more than a given minimum a priori probability, given an initial state ω for the universe. Or we could consider the total number of different futures for a given observer at a given time given a bound on complexity and on improbability. In thinking about such numbers it is important not to forget that some futures are much more likely than others. Proposition 4.6 shows that each of these numbers is finite, although, as sections 2, 3, and 9 emphasize, it does this only by showing that corresponding numbers of all possible switching structures are finite. Which switching structures are meaningful, or interpretable, or translatable and so could "really" be observers is left open.

Other central questions tackled in this paper and its predecessors are, "How, in a quantum framework, can one analyse the functioning of the warm wet brain as a fundamental information-carrying entity, while allowing for environmental fluctuations?" and "How should one compute quantum probabilities for localized observers?" Some technical questions have been left open, particularly the neurophysiological questions raised in section 6, but the most obvious open questions are metaphysical. "What meaning inheres in a given pattern of switching?" "What class of such patterns could define structures for a given observer?" "Does the plurality problem matter?" "What is probability?"

Another metaphysical question, already mentioned in the introduction, is that of whether all possible switching structures, or all interpretable switching structures, are "machines haunted by ghosts" (cf. Refs 10, 35, 36). Of particular relevance to this issue, perhaps, is the existence of switching structures of very small a priori probability. As discussed in section 6, when a priori probability of about $\frac{1}{2}$ is lost at every switching, essentially all switching structures become possible. This is a direct result of constructing a theory capable of dealing with fluctuations. At sufficiently low a priori probability, the fluctuations drown the messages. In my opinion, the many-worlds interpretation is not without experimental support, because it is among the simplest and most complete interpretations available for the mass of evidence validating quantum theory. However, the problem of whether all switching structures are "machines haunted by ghosts," is one which is utterly beyond experimental attack. It is closely analogous to the old philosophical problem of solipsism, which asks whether other people are as much haunted machines as we are. Nevertheless, while it seems only reasonable to believe in the consciousness of other people, and while this is neither the most probable nor, even for me, perhaps, the best(37) of all possible worlds, the idea that awareness is also immanent among low probability chaos seems undesirable.

Despite all these remaining problems, I believe that the analysis in this paper should be sufficient to raise as a serious possibility the notion that quantum theory is compatible with each of us being a pattern formed in time of a million billion scintillations in each second.

References.

1. K. Hepp, "The classical limit for quantum mechanical correlation functions," *Commun. Math. Phys.* **35**, 265–277 (1974).
2. H. Everett, III "The theory of the universal wave function". (1957). In B.S. DeWitt and N. Graham, "The Many-Worlds Interpretation of Quantum Mechanics". (Princeton, 1973).
3. W.H. Zurek, "Pointer basis of quantum apparatus: Into what mixture does the wave packet collapse?" *Phys. Rev.* **D 24**, 1516–1525 (1981).
4. D. Deutsch, "Quantum theory as a universal physical theory," *Int. J. Theor. Phys.* **24**, 1–41 (1985).
5. M.J. Donald, "A priori probability and localized observers," *Found. Phys.* **22**, 1111–1172 (1992).
6. S. Foster and H. Brown, "On a recent attempt to define the interpretation basis in the many worlds interpretation of quantum mechanics," *Int. J. Theor. Phys.* **27**, 1507–1531 (1988).
7. M.J. Donald, "Quantum theory and the brain," *Proc. R. Soc. Lond.* **A 427**, 43–93 (1990).
8. M.A.B. Whitaker, "The relative states and many-worlds interpretations of quantum mechanics and the EPR problem," *J. Phys. A: Math. Gen.* **18**, 253–264 (1985).
9. J.D. Barrow and F.J. Tipler, "The Anthropic Cosmological Principle" (Oxford, 1986).
10. D. Albert and B. Loewer, "Interpreting the many worlds interpretation," *Synthese* **77**, 195–213 (1988).
11. M. Lockwood, "Mind, Brain and the Quantum" (Blackwell, Oxford, 1989).
12. M.J. Donald, "On the relative entropy," *Commun. Math. Phys.* **105**, 13–34 (1986).
13. J.-P. Sartre, "L'être et le néant" (Gallimard, Paris, 1943).
14. T. Nagel, "The View from Nowhere" (Oxford, 1986).
15. D.C. Dennett, "Consciousness Explained" (Little Brown, Boston 1991).
16. R. Penrose, "The Emperor's New Mind" (Oxford, 1989).
17. D.R. Hofstadter, "Gödel, Escher, Bach: An Eternal Golden Braid" (Basic Books, New York, 1979).
18. L. Bombelli, J. Lee, D. Meyer, and R.D. Sorkin, "Space-time as a causal set," *Phys. Rev. Lett.* **59**, 521–524 (1987).
19. G. Brightwell and P. Winkler, "Sphere orders," *Order* **6**, 235–240 (1989).
20. H. Whitney, "Elementary structure of real algebraic varieties," *Ann. Math.* **66**, 545–556 (1957).
21. J. Bochnak, M. Coste, M.-F. Roy, "Géométrie Algébrique Réelle" (Springer-Verlag, Berlin, 1987).
22. G. Brightwell and R. Gregory, "Structure of random discrete spacetime," *Phys. Rev. Lett.* **66**, 260–263 (1991).
23. H. Baumgärtel and M. Wollenberg, "Causal Nets of Operator Algebras" (Akademie Verlag, Berlin, 1992).

24. B. Hille, “Ionic Channels of Excitable Membranes” – 2nd Edition (Sinauer, Sunderland, 1992).
25. M.D. Houslay and K.K. Stanley, “Dynamics of Biological Membranes” (Wiley, Chichester, 1982).
26. K.J. Angelides, L.W. Elmer, D. Loftus, and E. Elson, “Distribution and lateral mobility of voltage-dependent sodium channels in neurons,” *J. Cell Biol.* **106**, 1911–1925 (1988).
27. Y. Srinivasan, L. Elmer, J. Davis, V. Bennett, and K. Angelides, “Ankyrin and spectrin associate with voltage-dependent sodium channels in brain,” *Nature* **333**, 177–180 (1988).
28. M. Abeles, “Corticonics – Neural Circuits of the Cerebral Cortex” (Cambridge, 1991).
29. S.W. Kuffler, J.G. Nicholls, and A.R. Martin, “From Neuron to Brain”– 2nd Edition (Sinauer, Sunderland, 1984).
30. G.G. Somjen, “Extracellular potassium in the mammalian central nervous system,” *Ann. Rev. Physiol.* **41**, 159–177 (1979).
31. T.H. Benzinger, “Heat regulation: Homeostasis of central temperature in man,” *Physiol. Rev.* **49**, 671–759 (1969).
32. G. Hawthorn, “Plausible Worlds” (Cambridge, 1991).
33. D. Lewis, “Counterfactuals” (Blackwell, Oxford, 1973).
34. D. Lewis, “On the Plurality of Worlds” (Blackwell, Oxford, 1986).
35. E.J. Squires, “The unique world of the Everett version of quantum theory,” *Found. Phys. Lett.* **1**, 13–20 (1988).
36. A. Sudbery, “Comments on ’The unique world of the Everett version of quantum theory’ by E.J. Squires,” *Found. Phys. Lett.* **1**, 21–23 (1988).
37. Voltaire, “Candide” (1759).