February 25, 1996 LBL-37163new

The Hard Problem: A Quantum Approach. *

Henry P. Stapp

Theoretical Physics Group
Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720

Contents

Prepared for a special issue of the Journal of Consciousness Studies.

*This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U.S. Department of Energy under Contract DE-AC03-76SF00098.

Disclaimer

1. Introduction: Philosophical Setting

In his keynote paper David Chalmers defines "the hard problem" by posing certain "Why" questions about consciousness? Such questions must be posed within an appropriate setting. The way of science is to try to deduce the answer to many such questions from a few well defined assumptions.

Much about nature can be explained in terms of the principles of classical mechanics. The assumptions, in this explanatory scheme, are that the world is composed exclusively of particles and fields governed by specified mathematical laws that refer neither to any individual person, nor to anyone's experiences. These physical laws are supposed to be such that particles and fields, acting in concert, can form causally efficacious real functional entities such as driveshafts and propellers. Similarly, surges of electrical and mechanical activity in appropriately designed material substrates, composed of particles and fields acting in concert, could implement, in the world of matter, complex functional structures and long sequences of logical operations. Thus it is conceivable that all of our behaviour, and all of the internal processing that occurs in our bodies and brains could be *deduced*, at least in principle, from the principles of classical mechanics and appropriate boundary conditions.

There is, however, a problem in principle with 'experience'; i.e., with the streams of consciousness that constitute our psychological selves. Although, according to the principles of classical mechanics, all of our internal processing, and functionally described body/brain activity, should in principle be deducible from the principles of classical physics, and appropriate boundary conditions—namely the presence of a living body/brain in a certain state of readiness—and although we scientists may therefore one day be able to identify a particular functional activity F as the unambiguous sign of the presence of all of the causal and functional properties needed to identify F as the brain correlate of a certain feeling F' that the person calls 'a searing pain in my left index finger', nevertheless, it is impossible to *deduce* simply from the principles of classical mechanics that F *must* be accompanied by the felt *feeling* F'. This is because the principles of classical mechanics never mention 'feels', and hence these principles alone cannot entail that certain implementations of functional or logical structures *must necessarily* be accompanied by 'experiences'.

Since the principles of classical mechanics do not include any notion that some new sort of ontological entities come into being at some level of functional complexity, one seems to be led either to the notion that there just IS an extra kind of beingness, conscious experience, that is not mentioned in classical mechanics, but that is part of the full description of nature, yet plays no efficacious role in classical dynamics; or to the notion that certain implemented functional (or logical) structures ARE conscious experiences. The first of these possibilities, namely that some new kind of beingness just arises, but makes no physical difference, seems too capricious and unnatural to be true. But what about the second possibility: *functionalism*?

The difficulty with functionalism, within the explanatory framework provided by the principles of classical mechanics, is that the two things that are claimed to be the very same thing are, as initially characterized, described differently, and these differently described things are moreover incapable of being causally connected within the framework of classical mechanics, which never mentions one of them. 'Pains' are known to us from childhood, and it therefore does not resolve the problem of explaining their connection to brains either to deny their existence, or say that they are something different from the very feelings that they are defined to be. On the other hand, as one passes from simple thermostats to more and more complex servo-mechanisms with more and more complex self-monitorings, and memories, and decision-making capacities, etc., etc., all implemented ultimately in a material structure assumed to be exactly represented in terms of the primitives of classical mechanics—namely space, time, particles, fields, and local laws of motion—one can never arrive at the point of being able to deduce from the principles of classical mechanics the necessary presence of a 'feeling of pain'. The classical principles are therefore simply too impoverished to serve as a basis for a description for all of nature, including the felt experiences that constitute for each of us an immediately present reality.

The principles of classical mechanics are, of course, unable to explain the properties of the materials from which living brains are made, or the complex chemical reactions that are the basis of brain dynamics. Quantum mechanics is needed for that. But quantum mechanics, according to the the orthodox (Copenhagen) interpretation, involves a huge conceptual shift away from the classical ideal: it brings experiences of observers into the physical theory. The theory is constructed to be fundamentally 'about' our experiences, which thereby become the basic elements of the theory. Thus one need not go beyond the elements of the basic physical theory to accommodate consciousness. Conscious experience is already there, and it is there in a mathematically specified way that is perfectly suited to give it a central and causally efficacious role in mind/brain dynamics. Let me elaborate.

The key idea of the Copenhagen interpretation is encapsulated in two quotations:

"In our description of nature the purpose is not to disclose the real essence of phenomena but only to track down as far as possible relations between the multifold aspects of our experience" [Bohr, 1934]

"Strictly speaking, the mathematical formalism of quantum theory and electro-dynamics merely offers rules of calculation for the deduction of expectations pertaining to observations obtained under well-defined conditions specified by classical physical concepts." [Bohr, 1958][See also Stapp, 1993]

Bohr is emphasizing here that science, in the end, has to do with correlations among our experiences: that our experiences are the ultimate data that science must explain. Hence one can renounce the classical ideal of giving a mathematical description of the objective world itself in favor of constructing a set of mathematical rules that allow us to compute expectations pertaining to certain kinds of experiences. Thus, in contrast to classical mechanics, human experiences occupy a basic primitive place in quantum mechanics: the theory is basically 'about' experiences, even though the mathematical formulation of the "rules of calculation" pertaining to these experiences is based on a quantum mechanical generalization of certain of the "matter-like" properties that occur in classical mechanics. The crucial point here is that quantum theory has a larger base of primitives than classical mechanics, and this base includes experiences. [cf. Stapp, 1972]

The Copenhagen approach is essentially dualistic because the two things that it deals with are, on the one hand, our *experiences* (of a certain special type, namely classically describable perceptions) and, on the other hand, a set of mathematical rules that allow us to compute expectations pertaining to these experiences. These rules are expressed in terms of a generalization of the mathematical structure that occurred in classical mechanics, and which represented, in that idealization, the "objective world of particles and fields".

Bohr's pragmatic approach was revolutionary in its day, and was opposed by some of the most prestigious scientists of that time. In Einstein's opinion:

"Physics is an attempt conceptually to grasp reality as it is thought independently of its being observed" [Einstein, 1951, p.81]

This attitude of Einstein, and of many other scientists, seems so reasonable that one must ask why top scientists interested in atomic physics, which seems so far removed from psychology, should bring 'our experiences' into atomic physics, and why that move should be accepted by the scientific community as the correct way of comprehending atomic phenomena, and why physical theory did not thereby become devoid of objective content.

The answers rest on two points. The first is that the Copenhagen claim is not that physical theory is about *all* of our experiences: it was claimed that physical theory is about what can be called, for purposes of easy identification, our 'classically describable' perceptions of the world about us. The phrase 'world about us' is meant to describe only how we refer to these experiences, not to specify any particular ontological commitment. The phrase 'classically describable' is connected to the fact that visible objects normally appear to us to have a reasonably well defined location: a billiard ball does not appear to us to be, simultaneosly, both at one end of a billiards table and also at the other end. The second point is that the basic equation of quantum mechanics, the Schroedinger equation, applied universally, necessarily leads, in some easily arranged situations, to states in which the state of the billiard ball has one part that is localized at one end of the table and another part that is localized at the other end of the table.

The founders of quantum theory resolved this contradiction between the form of the quantum state and the form of our experience by postulating that the quantum state represents not the full reality itself, but rather the probabilities for, or tendencies for, our perceptions to be various possible specified perceptions. In this formulation, the experience of the observer becomes what the theory is 'about', and this experience enters in a fundamental way, because it is only by bringing these experiences, and their de facto classicality, explicitly into the overall theory that the theory is able to account for the classicality that we always observe, but which the quantum mechanical equation of motion neither entails nor (generally) allows.

The key point, in the context of the mind/brain problem, is that this most orthodox interpretation of quantum theory brings the experiences of the human observers into the basic physical theory on at least a co-equal basis with the 'physical' or 'matter-like' aspects of the description: the matter-like aspects give only half of the dynamical and ontological story.

Under the pressure of diverse goals (e.g., to expand the scope of the theory to include cosmological systems) a number of "ontological formulations" of quantum theory have been created. They attempt to give a rationally coherent description of (what at least could be) the world itself, not just a set of rules that allow us to form expectations about our future experiences. An underlying aim of most of these alternative interpretations is to avoid the explicit occurrence in the theory of the experiences of observers. But all of these interpretation are dualistic in that they have two kinds of entities that obey different, though intertwined, dynamical laws. One of these two parts is the 'wave funtion'. This part is the quantum analog of the 'matter' of classical mechanics, in the sense that it normally evolves in accordance with a local deterministic equation of motion that is the quantum analog of the corresponding classical equation of motion. The other part is associated with mind, in the sense that it 'picks out' from an amorphous mass of potentialities, represented by the first part, a sequence of particular actual experiences of the kind we actually experience.

It will be useful to give a brief description of these alternative formulations, emphasizing these two aspects. This will pave the way to an understanding of a quantum theory of the mind/brain proposed in Stapp [1993] and elaborated upon here. I shall spend what may seem like a inordinate amount of time on the model of David Bohm: this is because I shall treat all the other proposals by contrasting them with Bohm's.

The simplest quantum ontology is that of David Bohm (1952). In the orthodox (Bohr) theory one spoke of the complementary "particlelike" and "wavelike" aspects of a quantum system. That was confusing because particles stay confined to tiny regions while waves spread out: the two concepts contradict each other, physically. This is what forced Bohr into his epistemological stance, and his idea of "complementarity".

For a world consisting of a single quantum entity Bohm's model would have both *a particle* and *a wave*: the particle rides like a surfer on the wave. One easily sees how the puzzling double-slit experiment is explained by this model: the wave goes through both slits and influences the motion of the particle, which goes through just one slit. This model is dualistic in the sense of having both a particle and a wave. But this dualism is basically a mind/matter dualism, because the function of the "particle", or more specifically its generalization to the many-particle universe, is basically to specify what *our experiences* will be. There is a huge gap in quantum theory between the information contained in the "wave" and the information contained in our experience. The purpose of, and need for, the particle, and its generalization to the many-particle universe, is basically to supply the information—not contained in the wave (function)—that specifies *which one* of the many mutually incompatible experiences allowed by quantum theory the observer actually has. If there were no need to describe the *experiential* aspects of reality, which are very different in character from what the deterministically evolving wave (function) describes, there would be no need for the "particle-part" of Bohm's ontology. The critical assumption in Bohm's model is precisely the assumption that even though the "wave" (i.e., wave function of the universe) might describe a superposition of many different brains of some one particular scientist, say Joe Smith, and although each these different superposed "brains" would correspond to Joe's perceiving a different result of some experiment that he is performing, nevertheless, only one of these brains will actually be illuminated by the light of consciousness, and this particular brain—the one that possesses consciousness—is picked out from the host of possibilities by the "particle" aspect of the theory: in Bohm's theory the contents of our consciousness is determined by what the 'particle' part of the universe is doing, not the wave part, and, in fact, the only reason to bring in this particle part is, first, to account for the empirical fact that we 'experience' only one of the branches of the wave, and, second, to determine which branch this is.

To explain how this (and also the other models) work, I shall often use the term "branches of the wave (function)". To visualize these branches, imagine a large pond with an initially smooth surface (no waves). A source of waves is placed at the center, but is surrounded by a barrier that has some gaps.

These gaps allow ripples to spread out only along certain beam-like regions, with most of the surface of the pond remaining smooth. These well separated beam-like regions of propagating ripples I call “branches”, or “branches of the wave (function)”.

The surface of a pond is only two dimensional. But the quantum-mechanical wave that corresponds to a universe consisting of N particles would be a wave in a $3N$-dimensional space. The “branches of the wave (function)” will typically be relatively narrow beams of waves in this $3N$-dimensional space, and each beam will correspond, in a typical measurement situation, to some particular “classically describable” result of the measurement. For example, one beam may describe, at some late stage, a particle detector having detected a particle; *and* a corresponding pointer having swung to the right as a consequence of the detector's having detected the particle; *and* the eye and the low-level processing parts of the brain responding to the light signal from the pointer in the swung-to-the-right position; *and* the top-level neural activity that corresponds to the observer's perceiving the pointer in the swung-to-the-right position: the other branch would describe the particle detector's having failed to detect the particle; *and* the pointer remaining in the center position; *and* the eye and low-level processing parts of the brain responding to the light signals coming from the pointer in the center position; *and* the top-level neural activity corresponding to the observer's perceiving the pointer in the center position. The fact that *both* branches of the wave are present simultaneously is not surprising once one recognizes that the wave represents essentially only a *probability for an experience to occur*: there is, in a typical measurement, a possibility for each of several possible experiential results to occur, and the probability function (or wave function) must therefore have a “branch” corresponding to each possibility.

Of course, the observer, Joe Smith, will see only one of the two possibilities: he will see *either* the pointer swung-to-the-right *or* the or the pointer remaining at the center position. To accommodate this empirical fact Bohm introduces his “surfer” in the $3N$-dimensional space. The surfer is merely a point in the $3N$-dimensional space that move always in a direction defined by the shape of the $3N$-dimensional wave at the place where this point is, and this rule of motion for the surfer ensures that the surfer will end up in one branch or another, not in the intervening “still” part of the $3N$-dimensional space. Each branch corresponds to one of the possible experiences. If the "surfer" (which is just the moving point in the $3N$-dimensional space) ends up in the branch that corresponds to the experience "I see the pointer in the swung-right position" then, according to Bohm's theory, this perception of the pointer "swung-to-the-right" is the experience that actually occurs: only the single branch in which the surfer ends up will be "illuminated"; all others "remains dark". Bohm's rules for the motion of the surfer ensure that if the various possible initial conditions for the surfer are assigned appropriate "statistical weights" then the statistical predictions of his theory about what observers will experiences will agree with the those given by the orthodox (Bohr) rules. In this way Bohm's causal model reproduces the quantum statistical predictions about what our experiences will be.

The two parts of Bohm's ontology, namely the wave in the 3N-dimensional space and the 'surfer', can both be considered 'material', yet they are essentially different because the waves describe all the possibilities for what our actual experiences might be, and therefore has a beingness that is essentially "potential", whereas the trajectory of the surfer specifies the actual choice from among the various alternative possibilities, and therefore has a beingness that represents "actuality" rather than mere "potentiality": the wave generates all the *possible* experiences, whereas the trajectory defined by the surfer specifies which one of these possible experiences actually occurs.

Bohm's model is very useful, but as a model of reality it has several unattractive features. The first is the "empty branches": once two branches separate they generally move further and further apart in the $3N$-dimensional space, and hence if the "surfer" gets in one branch then all of the alternative ones become completely irrelevant to the evolution of experience: the huge set of empty branches continues to evolve for all of eternity, but has no effect upon anyone's experience.

A more parsimonious ontological theory, not having these superfluous empty branches, was described by Heisenberg (1958). It also involves a reality consisting of two kinds of things. His two kinds of things are "actual events", and "objective tendencies for those events to occur". The objective tendencies can be taken to be represented by the wave on the $3N$-dimensional pond, and the actual events can be represented by sudden or abrupt changes in this wave. Each such change "collapses the wave" to one of its branches. Thus Bohm's "surfer", which specifies a *choice* between branches, is replaced by an "actual event", which also specifies a choice between branches. But whereas Bohm's surfer has no back-reaction on the wave, each of Heisenberg's actual events obliterates all branches but one. The big problem with Heisenberg's theory is to find a reasonable criterion for the occurrence of these actual events.

Wigner (1961) and von Neumann (1932), noting that there is nothing in the purely material aspect of nature that singles out where the actual events occur, suggest that these events should occur at the points where consciousness enters: i.e., in conjunction with conscious events. This theory can be regarded as the 'ontologicalization' of the Copenhagen interpretation, in the sense that the change of the state that occurs when a perception generates new knowledge is basically subjective in the Copenhagen interpretation, but is interpreted as an objective change in the Wigner-von Neumann interpretation used here. This ontology is the most parsimonious possibility: all of the verified predictions of quantum theory can be reproduced by limiting the actual events to brain events that correspond to experiential events.

An argument based on survival of the species [Stapp, 1995] provides support for the idea that actual events occurring in human brains will tend to occur at the brain-wide level of activity that corresponds to conscious events, rather than at some microscopic (e.g., molecular, or individual-neuron) level.

This Wigner-von Neumann version of Heisenberg's theory will be discussed presently in some detail. But first a few remarks about the final major interpretation are needed.

In the Everett many-minds theory the basic quantum mechanical equation of motion, the Schroedinger equation, holds uniformly: there are no sudden collapses of the wave function; all branches continue to exist. Moreover, it is assumed that, because all of the branches exist, all of the corresponding streams of conscious must also occur.

Since the various branches propagate into different parts of the $3N$ dimensional space they will evolve independently of each other: the physical "memory banks", associated with one branch will not affect, or be affected by, the brain activities specified by another branch. Hence each different branches can be considered to define a different "self", or "psyche", with each of these selves continually dividing into different extentions of itself into the future.

At first sight this idea seems to allow the whole theory to be reduced to just one entity, the evolving wave, with the different psychological persons being just "aspects" of corresponding brain activities on different branches. But that is not correct. The branches of the wave function appear as parts of a *conjunction* of branches: all branches on the 'pond' exist simultaneously, even though they evolve independently. But the predictions of quantum theory are an essential part of the theory, and these statistical predictions pertain to experiences that are 'this experience' *or* 'that experience', not 'this experience' *and* 'that experience'. To speak of probabilities one needs something with an *or* character: something that can become associated with *either* this branch *or* that branch, not both simultaneously. Just as the different branches of a wave on a pond are conjunctively present and hence do not, by themselves, provide any ontological basis for assigning different probabilities to these simultaneously present things, so also is the quantum wave, by itself, insufficient for this task.

In Bohm's theory this extra element of the theory was the 'surfer', which determined the experiences of the observers; in Heisenberg's theory the extra things were the actual events, which also determined the experiences of the observers. In the Everett interpretation the only existing things besides the waves are our experiences, and there is supposed to be a separate experience associated with each branch. Thus we end up again with a dualistic theory; with a world that is composed of the one "material" universe represented by the wave function, which evolves always according to the the Schroedinger equation, plus, for each named person, an great profusion of many minds, or streams of consciousness: the stream of consciousness of Joe Smith must be continually splitting into different separate branches, with at least one for each of the perceptibly different results of any experiment that he performs. Consequently, the proponents of the theory need to develop, in order to complete this interpretation, some coherent dualistic ontology involving, for each of us, a profusion of branching minds, each known only to itself, and a theory that assigns to each of these 'independent' (but generally overlapping) branches a well defined 'subjective probability to occur', even though these branches all occur together in the full 'objective' description of reality.

In summary, all the major ontological interpretations of quantum theory are dualistic, in the sense that they have one aspect or component that can be naturally identified as the quantum analog of the *matter* of classical mechanics, and a second aspect that is associated with *choices* from among the possible *experiences*. All interpretations are, in this sense, basically similar to the Wigner-von-Neumann interpretation to be explored here, but are less parsimonious, in that they involve either existing but unobserved branches (Bohm), or existing but unobserved actual events (Heisenberg), or existing but inaccessible and unverifiable other branches of reality (Everett).

2. A Quantum Model of the Mind/Brain

The main features of the mind/brain theory proposed in Stapp [1993] are briefly described in the following fourteen points:

1. **Facilitation**: The pattern of neurological activity associated with any occurring conscious thought is "facilitated", in the sense that the activation of this pattern causes certain physical changes in the brain structure, and these changes *facilitate* subsequent activations of this pattern.

2. **Associative Recall**: The facilitation of patterns mentioned above is such that the excitation of a part of a facilitated pattern has a tendency to excite the whole. Thus the sight of an ear tends to activate the pattern of brain activity associated with a previously seen face of which this ear was a part.

3. **Body-World Schema**: The physical body of the person in its environment is represented within the brain by certain patterns of neural and other brain activity. Each such pattern has *components*, which are sub-patterns that represent various parts or aspects of the body and its environment, and these components are normally patterns of brain activity that have been facilitated in conjunction with earlier experiences.

4. **Executive-Level Template for Action:** A main task of the alert brain at each moment is to construct a template for the impending action of the organism. This template is formed from patterns of neural and brain activity that, taken together, represent a coordinated plan of action for the organism. This representation is implemented in the brain by means of an automatic causal spreading of neural excitations from the executive level to the rest of the nervous system. This subsequent activity of the nervous system causes both motor responses and lower-level neural responses.

The executive-level templates are based on the body-world schema, in the following sense. There are two kinds of templated actions: attentions and intentions. Attentions *up date* the body-world schema: they bring the brain's representation of the body in its environment up to date. Intentions are formulated in terms of a *projected* (into the future) body-world schema: they are expressed in terms of an image of how the body in its environment is intended to be at a slightly future time. (Thus, for example, the tennis player imagines how he will strike the ball, or where the ball he is about to hit will land in his opponent's court).

5. **Beliefs and other Generalizations**: The simple Body-World Schema, with attentional and intentional templated actions, is the primitive level of brain action: it gives the general format. However "beliefs" can be added to the landscape. Also, each templated action has both intentional and attentional aspects.

6. **Quantum Theory**: The features mentioned above are key elements of this theory. But they are aspects that hold at the level corresponding to a particular classically described 'branch'. But classical mechanics cannot account for the essential properties of the materials (such as tissues and membranes) from which the brain is made. Hence, an adequate basic theory of nature must show how the classically describable aspects of nature that seem important, and that we seem to experience, arise in rational way from the quantum underpinning.

7. **Superposition of Templates**: An analysis [Stapp, 1993, 1996] of processes occurring in synapses shows that if there were no quantum collapses occurring in brains then a brain evolving according to the quantum laws *must* evolve, in general, into a state that contains a superposition of different "branches", with each of these branches specifying the template for a different macroscopic action. Each of these different templates for action will evolve into a different response of the nervous system, and consequently into a difference macroscopic response of the organism. Thus without collapses the body/brain would evolve into a superposition of macroscopically distinguishable possibilities, just as a measuring device normally does.

8. **The Reduction Postulate**: Following the Wigner-von-Neumann approach, I postulate that the quantum collapse of the brain state occurs at the high level of the template for action. The (Heisenberg-picture) state (of the

universe) undergoes the collapse

$$\Psi_i \rightarrow \Psi_{i+1} = P_i \Psi_i,$$

where P_i is a projection operator that acts on appropriate macroscopic variables associated with the brain. It picks out and saves, or "actualizes", *one* of the alternative possible templates for action, and eradicates all others. The organism will then proceed to evolve automatically in accordance with this one particular plan of action, rather than evolving (á la Everett) into a superposition of states corresponding to *all* of the different possible macroscopically distinguishable courses of action that were formerly available to it. Thus the "quantum event", or "collapse of the wave function", *selects* or *chooses* one of the alternative possible coherent plans of action—previously generated by the purely mechanical functioning of the brain—by actualizing the executive-level pattern of brain activity that constitutes *one* of the alternative possible templates for action.

This collapse of the wave function is to be understood not as some anomalous failure of the laws of nature, but rather as a natural consequence of the fact that wave function does not represent actuality itself, but rather, in line with the ideas of Heisenberg, merely "objective tendencies" for the next actual event.

Each such event is represented, within the Hilbert space description, as a sudden shift in the wave function, or state Ψ_i, to a new form that incorporates the conditions or requirements imposed by the new actual event.

These 'collapse' events in the Hilbert space are not introduced willy-nilly: they are needed to block what will otherwise automatically occur, namely the evolution of the wave function to a form that directly contradicts collective human experience: all of us who see the pointer agree that the pointer does not both swing to the right *and* also remain motionless. Under the conditions of the measurement it does one thing *or* the other, and all of us who witness what it does, and are able communicate our findings to each other, agree about which one of these two possible things actually occurs.

9. **The Psycho-Physical Postulate**: Adhering to the Wigner-von-Neumann approach, I postulate that the physical brain event, namely the collapse of the wave function to the branch that specifies one particular template for action, is the brain correlate of a corresponding psychological or experiential event. Thus the occurrence or emergence of the psychological command " raise the arm" is

represented in the physical description of nature by the occurrence or emergence of the physical command "raise the arm". The causal relationship is discussed in point 15.

Attending is a special kind intending: the intention, in the case of attending, is to up-date the body-world schema.

Notice an important difference between the context in which the Wigner-von Neumann idea is applied here, and the context in which they themselves applied it. They were considering a large system consisting of some atomic system plus a device that measures some property of that system plus the human observer's body and, finally, his brain. In that context the collapse at the level of the observer's consciousness determined the outcome of the earlier distant experiment on the atomic system. I am focussing, however, on the mind/brain system itself, and the role of the collapse event in determining principally the internal dynamics of this system. This role is more in line with what seems, intuitively, to be the role of conscious. For the present purpose one may assume that collapses occur already at the level of all quantum devices, so that the principal effect of the conscious event is then to determine only the internal dynamics of the brain, not the results of any measurements performed externally to the brain.

10. **The Efficacy of Consciousness**: In this model the choices associated with conscious events are dynamically efficacious: each such event *effects* a decision between different templates for action, and these different templates for action lead on to different distingushable responses of the organism.

11. **Consciousness and Survival**: It is often thought that consciousness emerges because it aids survival. For this to be so consciousness must be efficacious. Yet in the Bohm and Everett models (just as in classical physics) consciousness is not efficacious: all behaviour is completely determined with no mention of any causal role for consciousness. Consciousness would be nonefficacious also in the Heisenberg model if we did not follow Wigner-von-Neumann in associating actualizing events with conscious events. (The survival issue is addressed in Stapp [1996])

12. **Conscious Events and Unconscious Processing**: The general temporal development in the brain proceeds by periods of unconscious processing

punctuated by conscious events. A conscious event actualizes a template for action that, by the automatic spreading of top-level neural activity to the rest of the nervous system, controls: 1) motor action, 2) the collection of new information (including the monitoring of ongoing processes), and 3) the formation of the next template for action.

13. **Overall Picture**: Classically only a single "next template" would be formed. This could be achieved either by the formation of a resonant state that sucks energy from competing possibilities, or by inhibitory signals, or by dropping into the well of an attractor. But in any of these cases the quantum uncertainties entail that the quantum brain will, if no collapse occurs, necessarily evolve into a superposition of branches corresponding to the different alternative possible classical templates for action. One of these virtual templates for action will then emerge as the actual one, and the automatic (unconscious) neural processes will proceed to carry out the instructions encoded in the template. Thus there is an alternation between discrete conscious events and periods of unconscious activity controlled by the local deterministic laws. Each of the discrete events chooses from among the alternative possible allowed templates for action generated by the automatic action of the local deterministic laws of quantum mecanics, and hence between the different associated macroscopic responses of the organism.

14. **The Generality of Psycho-Physical Connection**: My focus above has been on the human mind/brain system, and on the interplay between the experiential and physical aspects of nature in this particular system, for which we have both verbal reports and personal knowledge pertaining to the form and content of the experiential aspect. These special characteristics make the human mind/brain unique as a subject of scientific study by human scientists. However, it is unreasonable to suppose that events of this kind are confined to human beings. This suggests, therefore, that there is in nature a general 'second process' that includes as a special case the one that is needed, according to (Wigner-Von Neumann) ontologicalization of the Copenhagen interpretation, to bring the quantum process into accord with human experience. In line with the ideas of Wigner and von Neumann, each such event in nature is assumed to be an 'experiential reality' with physical ramifications. But the specific qualities of the experiences associated with different systems can be vastly different from

human conscious experiences, although all are considered to belong to the same ontological category.

To support this suggestion of a universal connection of collapse events to experience I note, first of all, that, as Arthur Eddington observed, the quantum world is more like a 'giant mind' than like the 'giant machine' described by classical mechanics. For, the evolving state vector represents not 'substance', but rather a 'probability' for something to happen, and probability is normally considered to be a subjective or mental sort of thing, not a material reality. The second part of the quantum reality is the 'actual' event, which Heisenberg contrasts with the 'potentia' from which the event arises. The 'actual' specifies what is able to be experienced: only the actualized branches can be experienced. This connection of the actual to experience is tightened by the Wigner-von Neumann proposal, which is essentially to *identify* the actual with experience. This proposal is the quantum analog of the dictum 'esse est percipi': to be is to be perceived.

Technically there is a specific need for a tie-in of the actualization event with features characteristic of experience. This tie-in has two aspects. The first is that the actualization singles out a 'classically describable' structure. This is a feature of experience that is not inherently present in the physical aspect of nature represented by the (Schroedinger-directed) evolving wave (function). The great and essential move of the Copenhagen interpretation was precisely to realize that although no classical aspect naturally pops out from the quantum physical reality, represented by the normally evolving wave, (certain of) our experiences are, in fact, classically describable, and hence the empirically observed classical aspect of nature can be brought consistently into physical theory by introducing our (classically describable) experiences, per se, directly into the theory as the very thing that the theory is about. The whole history of efforts to go beyond the Copenhagen interpretation is essentially the history of attempts to find some *other* rationally coherent way of explaining why our perceptions of the quantum universe are classically describable. The present suggestion is essentially to accept, following the Copenhagen approach, that this classical aspect does not come from the physical side of nature, but comes instead from the experiential side. This makes the experiential aspect of the actualization events the *cause* of the classical character of the collapse events, which otherwise appears to have no natural explanation.

15. **Causation**

This reference to 'cause' raises the general question of the cause of both the occurrence of the actualizations and of the form (e.g., the classical form) of the actualizations. Much can be said without delving into the question of the cause of the event. Indeed, all of contemporary quantum mechanics is covered without delving into that question, by just accepting the statistical rules as 'given': then the cause of the events is effectively 'pure chance'.

This occurrence of pure chance is quite acceptable in an admittedly pragmatic interpretation of quantum theory, such as the Copenhagen interpretation. But it is not acceptable, I think, in an account that represents itself as ontological, i.e., as a putative description of reality itself. Thus if one wishes, in an ontological context, to discuss the 'cause' of the occurrence of an actual event then *something* is needed to replace the 'irrational'—that's Pauli's word for it—element of pure chance: to embark upon an ontological discussion of the 'cause' of the actualizations entails assuming that the element of 'pure chance' that occurs in contemporary quantum theory is merely a mask for our ignorance of the true cause, which must necessarily be highly nonlocal [Mermin 1994, Stapp 1994].

The part of the dynamics that is matter-like is represented by the Schroedinger (directed) evolution, which is the contrary of the collapses. This fact, together with the general mind-like quality of the quantum world mentioned above, suggests that the cause of the actualizations must come from the experiential aspect of things.

The orthodox (Copenhagen) interpretation is based on the fact that certain features of the quantum state can be regarded as a superposition of states that represent different possible experiences. This connection becomes more direct when the quantum system in question is the brain itself. The process of selection of the actual experiential state from the superposition of possible ones must, of course, depend upon this decomposition into the possible experiential states. So the decomposition of the Hilbert space state vector into the possible experiential components must play a central role in the dynamical processes, at least if one accepts the naturalistic hypothesis that the possible experiences are represented in the same mathematical structure that represents the matter-like aspects of nature. This suggests that the selection process is essentially a competition among the experiential possibilities, with the winner becoming the actual experience, which is the reality whose coming into being is represented by the transition to the new quantum state.

This conception of the process of selecting the actualized state entails that the experiential aspect of the actualization event is, effectively, the cause of this event, in the sense that this experience, in its potential form, has raised itself above its competitors to become the actual experience: each actualization is a self-actualization.

In this conception, the experienced reality is the coming into being of a psychologically felt command: "Do X!" This experiential reality, in its potential form, is represented (or embodied) in Hilbert space, and the coming into actual being of this representation (or embodiment) constitutes a coming into actual being of both the psychologically felt command "Do X!" and the physically implemented command: "Do X!". This connection is somewhat like a dual-aspect theory, but now in the context of a physical theory that properly accounts for *all* brain processes, and that, following the ideas of the quantum physicists Bohr, Heisenberg, von Neumann, Wigner, and others, incorporates human experience into the basic fabric of physical theory in order fulfill the scientific demand for a closed rational account of physical phenomena.

3. Person and Self

According to William James:

"Such a discrete composition is what actually obtains in our perceptual experience. We either perceive nothing, or something that is already there in a sensible amount. This fact is what is known in psychology as the law of the 'threshold'. Either your experience is of no content, of no change, or it is of a perceptible amount of content or change. Your acquaintance with reality grows literally by buds or drops of perception. Intellectually and on reflection you can divide these into components, but as immediately given they come totally or not at all."[James, 1910, p. 1062]

"... however complex the object may be the thought of it is one undivided state of consciousness."[James, 1890, p. 276]

"The consciousness of Self involves a stream of thought, each part of which as 'I' can (1) remember those that went before, and know the things they knew; and (2) emphasize and care paramountly for certain ones among them as *'me'*, and *appropriate* to these the rest... This *me* is an empirical aggregate of things objectively known. The *I* that knows them cannot itself be an aggregate. Neither for psychological purposes need it be considered to be an unchanging metaphysical entity like the Soul, or a principle like the pure Ego, viewed as "out of time". It is a *Thought*, at each moment different from that of the last moment, but *appropriative of the latter*, together with all that the latter called its own ... *thought is itself the thinker*, and psychology need not look beyond..." (James. 1890, p. 401)

In line with these ideas of James, and those of the preceding section, the conception of a 'person' that emerges here is that of a sequence of discrete psychological (i.e., experiential or conscious) events bound together by a matter-like structure, namely the brain/body, which evolves in accordance with the local deterministic laws of quantum mechanics. Each conscious event is a new entity that rises from the 'ashes' of the old, which consists of the propensities for its occurrence carried by the brain/body.

A felt sense of an enduring 'self' is experienced, and hence it must, within this theory, be explained as an aspect of the structure of the *individual* discrete conscious events. The explanation is this: each conscious event has a "fringe" that surrounds the central image, and provides the background in which the central image is placed. The slowly changing fringe contains the consciousness of the situation within which the immediate action is taking place; the historical setting including purposes (e.g., getting some food to eat). The sense of feeling of self is in this fringe. It is not an illusion, because the physical brain/body is providing continuity and a reservoir of memories that can be called upon, even though each thought is, according to this model, a separate entity. As explained by James—see also Stapp (1993)—each thought, though itself a single entity, has components that are sequentially ordered in a psychological time, and hence each thought has within its own structure an aspect that corresponds to the flow of physical time.

4. Free-Will

Our experience includes the feeling that we are 'free'. That feeling is, in some sense, accurate. The whole organism is free to make high-level choices in which the various perceived possible consequences enter as whole experiential units. One's fate is not controlled exclusively by mechanical local deterministic laws, or by an avalanche microscopically entering chance elements that would make a mockery of the idea of personal choice.

It might be objected that we are not free because, according to quantum theory, our choices are determined by blind chance. That misses the point. In the first place the choices are not blind. If the quantum events in the brain occurred at the level of the neurons then the choices would be blind, for the consequences of each individual choice would be screened from view by the inscrutable outcomes of billions of similar independent random choices. But the choices being made by the organism, acting as a unit, are choices between plans for actions that have clear and distinctive consequences for the organism as a whole, in terms of its future behavior. The choice is made at the level of the organism as a whole, and the event has a distinctive 'feel' that accurately portrays its consequences for the organism as a whole. The conditioning for this event is an expression of the values and goals of the whole organism, and the choice is implemented by a unified action of the whole organism that is normally meaningful in the life of the organism. And this meaning is felt as an essential aspect of the act of choosing.

The final 'random' decision between the alternative possible distinctive actions of the organism is not some wild haphazard stab in the dark, unrelated to the needs or goals of the organism. It is a choice that is governed essentially by the number of ways in which the mechanistic aspect of the organism, which has been honed to construct templates for action concordant with the needs of the organism within its environment, can come up with that particular template. Thus the choice is not like the throw of an unconditioned die. It is a carefully crafted choice that tends to be the "optiminally reasonable" choice under the conditions defined by the external inputs, and the needs and goals of the organism. Each of the alternative possible templates for a coherent and well-coordinated action of the organism emerges from the quantum soup, and is given, by the quantum mechanism, a weighting that reflects the interests of the organism as a whole, within the context in which he finds himself. The choice is conditioned by these personally molded weights, and therefore tends to be a decision that is optimally reasonable from the point of view of the organism. This arrangement avoids both the Scylla of a fate sealed at the birth of the universe by a microscopically controlled blind mechanism, and also the Charybdis of a haphazard wild chance that operates at a microscopic level, and is therefore blind as regards likely consequences. The intricate interplay of chance and determinism instituted by quantum mechanics effectively frees the organism to pursue, in an optimal way, its own goals based on its own values, which have themselves been created, from a wealth of open possibilities, by its own earlier actions. Each human being, though never in full control of the situation in which he finds himself, does create both himself and his actions, through a process of a microscopically controlled deterministic evolution punctuated by organic meaningful choices that are top-down in the sense that each one is instituted by an actualization event that selects as a unit, and feels as a unit, an entire top-level plan of action.

Within the contemporary framework of quantum theory that I have been adhering to in the above three paragraphs there remains, in the end, an element of 'pure chance' that selects one of the templates for action 'randomly'. Whether this occurrence of pure chance is a permanent feature of basic physical theory, or merely a temporary excursion, no one knows. In my own opinion this occurrence of pure chance is a reflection of our state of ignorance regarding the true cause, which must in any case be nonlocal, and hence both difficult to study and quite unlike the local causes that science has dealt with up until now. In another place [Stapp, 1996] I have described in more detail the technicalities of the actualization process, and also the possibility of replacing the element of pure chance by a nonlocal causal process that makes the felt psychological subjective 'I', as it is represented within the quantum-theoretic description, rather than pure chance, the source of the decisions between one's alternative possible courses of action.

5. The Hard Problem

'The Hard Problem' is the problem of *conscious experience*: What is it? Why is it present at all? Why is it so different from the other part of Nature, namely the objective aspect of reality? Why is it personal, or subjective? Why is it so fleeting, whereas matter is permanent and conserved? Can it be 'reduced' to matter? Can any purely physical account explain it? Is the material of which the brain is made crucial, or is it only the functional aspect that is critical? Why is it so closely connected to function? How do functional aspects become ontological aspects, i.e. how does function become being? How can anything, and in particular consciousness, be added to the already closed laws of physics? Is experience a fundamental element of nature, or derivative, or emergent? What are the *bridging laws* that connect mind to matter?

Chalmers asks these questions, and says that right now we have no candidate theory that answers these questions. But we do!

Chalmers suggests that perhaps there is a small loop-hole in quantum theory that might provide an opening for consciousness. But there is not just a small loop-hole: there is a gigantic gap, which consists of *fully half of the theory*, and this hole provides an ideal home for consciousness. For, quantum dynamics consists not only of the mechanical process that is governed by the Schroedinger equation, which controls the matter-like aspect of nature, but also an entirely different 'second process', which involves a beingness of an entirely different kind. This second process is needed to bring the experiential aspect of nature into concordance with the physical part of the theory, and to specify or determine what our experiences will actually be.

The core idea of Bohr was to recognize that physics is basically about our experiences, and to introduce our experiences directly into the physical theory in order to resolve a deep problem with the theory that otherwise arose. Bohr's stance was cautious and pragmatic, and avoided ontological commitment. But von Neumann, and more unambiguously Wigner, went the next step and brought consciousness into the theory as a causal agent that actively did what needed to be done to make the theory work at the ontological level in the same way that it worked at the practical level. The ontology then included conscious experience as the co-equal partner of a more shadowy world of 'possibilities', or, in Heisenberg's terminogy, of 'objective tendencies' for transitions from the possible to the actual.

The new ontology is fundamentally different from the classical-mechanics ontology in many ways. The first is the physical 'possibilities' acquire an ontological reality that is mathematically represented or embodied. The second is that possible conceptual/experiential realities are represented or embodied in this same structure. Thus the actualization of a conceptual/experiential possibility is simultaneously an actualization of a corresponding physical reality. In this way concept evolves into implementing functional structure in a mathematically described way.

Each actualization event has its physical side, which is just the 'collapse' of the wave function itself, and also its experiential side. In a rational causal theory the collapse must have a cause. This cause is not to be found in that physical part of the quantum ontology: considered from the purely physical standpoint the collapse seems to come from nowhere, as an unpredictable and undetermined 'bolt from the blue'. And the collapse represents merely a sudden 'change in tendencies' for a 'change in tendencies' for a 'change in tendencies' for ...etc., etc.: there is no closure, no basic reality. For the technical reasons mentioned earlier the theory needs something that will bring 'classicality' into the dynamics, and it needs a 'cause' for the collapse, and it needs a reality to complement the 'potentia'. This missing element is not present in the physical part of the theory. It must be something that exists, and the only thing that we know exists, besides the physical part of reality, (which perhaps we do not really 'know' at all, directly) is the experiential part. This part fills the various needs perfectly, and, of course, needs to be included in a complete description of nature. So we put the two parts of nature together in the natural way that fits all these needs simultaneously.

How does this theory answer the questions, listed above, that Chalmers raised about consciousness?

What is consciousness? It is the part of nature that: 1) complements the the background 'physical' aspect by being the reality that the quantum potentiality is the potentiality for; 2) specifies the classical character of reality that is not determined by the background physical aspect of the quantum, which represents only an amorphous potentiality, with no inherent preference for a unique and specifically classical reality; and 3) causes the physical state to collapse to a new form that incorporates the conditions specified by the new experiential new reality.

Why is consciousness present at all? Because the local-reductionistic laws of physics, regarded as a causal description of nature, are incomplete: something else is needed to cause the collapse events to occur, to inject 'classicality' into nature, and to provide the reality that the quantum state is the potentiality for.

Why is consciousness so different from the other part of Nature, namely the physical aspect of reality? The physical part of reality represents merely the possibilities for an actual experience, not the actually experienced reality itself.

Why is consciousness subjective? An actualization event has many components, all of which are integral parts of the whole. The totality contains the slowly changing fringe of the experience that constitutes the 'I', or 'psyche', which is felt as the experiencing subject and actualizer. The experiencing subject is *part of the thought*, not an outside observer of the thought: if the 'I' were not *part of the thought* then there could be in the thought no awareness of 'I' as the background relative to which the focus of the thought is the foreground. Thus it not that the thought belongs to an 'I', but rather that an 'I' belongs to the thought.

Why is a thought so fleeting, whereas matter(energy) is conserved? Because a thought is an element in a discrete sequence of events, whereas 'matter' is the continuously evolving 'potentia' for such an event to occur.

Can consciousness be 'reduced' to matter? "Matter" is the ground from which the experiential event springs: the whole process is represented in the Hilbert space in which the quantum analogue of matter is represented. But rising out of the matter-like aspects of nature lies another dynamics governed by the experiential aspects of nature.

Can any purely physical account explain it? That depends on what 'physical' means. The account given here is physical in that it is rooted in quantum mechanical description of nature. But there is something decidedly unphysical about the idea of 'pure chance'. The present account allows the orthodox(?) idea that pure chance is an essential part of nature's process to be replaced by the more 'physical' idea that the collapse process, which is needed to bring the quantum mechanical representation of the physical world into concordance with our experience of it, is a causal process that is representable in the Hilbert space of quantum theory. This proposal is based on the fact (which is the key element upon which the orthodox Copenhagen interpretation is based) that certain aspects of the quantum mechanical superposition of the physical possibilities can be interpreted also as a superposition of experiential possibilities. The dynamics that controls the emergence of one experienced reality from the superposed comglomerate must, of course, depend upon the set possible experiential realities.

Is the material of which the brain is made crucial, or is it only the functional aspect that is critical? Neither! What is important is the presence in the physical substrate of potentialities for quantum actualizations of experiential structures.

Why is consciousness so closely connected to function? Because the event is an actualization of a template for action. The biological reason for this link of actualization to function is undoubtedly the survival advantage it confers. [Stapp, 1996]

How do functional aspects become ontological aspects? The actualization events actualize functional structures.

How can consciousness be added to the already closed laws of physics? Nothing efficacious could be added if the laws were already complete! But the quantum laws are grossly incomplete before consciousness, or some stand-in for consciousness, is added.

Is experience a fundamental element of nature, or is it derivative, or emergent? It is fundamental because the fundamental realities are experiential. On the other hand, the particular sort of consciousness that we human beings experience is emergent, because it represents a highly evolved form of the general ontological type. The complexity of a human experience is a consequence of the complexity of the body/brain that supports the physical activity that supports. The complexity of the physical carrier has undoubtedly co-evolved with the complexity of the associated experiential reality.

What are the bridging laws that connect mind to matter? They have been described here.

Acknowledgements. This paper has greatly benefited from useful correspondences with David Chalmers, Gordon Globus, Jonathan Shear, and Aaron Sloman.

References

Bohm, David (1952), 'A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables' I and II, *Phys. Rev.* **85,** pp. 166-93; The Undivided Universe: An Ontological Interpretation ofQuantum Theory, with Basil Hiley, Routledge, London, 1993.

Bohr, Niels (1934), *Atomic Physics and Human Knowledge*, Cambridge Univ. Press, Cambridge, p. 18.

Bohr, Niels (1958), *Essays 1958/1962 on Atomic Physics and Human Knowledge*, (Wiley, New York), 1963, p. 60.

Einstein, Albert (1951), in *Albert Einstein: Philosopher-Scientist*, P. A. Schilpp (ed.), (Tudor, New York).

Everett III, Hugh (1957), 'Relative State Formulation of Quantum Mechanics', *Rev. of Mod. Phys.* **29**, pp. 454-62

Heisenberg, Werner (1958), *Physics and Philosophy*, (Harper and Row, New York) Chap. III.

James, William (1890), *The Principles of Psychology*, (Dover, New York, 1950) Vol. 1.

James, William (1910), *William James: Writings 1902-1910*, (The Library of America, Viking, New York, 1987), p. 1061. (From: Some Problems of Philosophy, Ch X).

N. David Mermin (1994), 'Quantum Mysteries Refined', *American Journal of Physics* **62**, 880-887.

Stapp, Henry P. (1972), 'The Copenhagen Interpretation', Amer. J. Phys. **40**, 1089-1116; reprinted in Stapp (1993) Ch.3

Stapp, Henry P. (1990), 'Quantum Measurement and the Mind-Brain Connection', *Symposium on the Foundations of Modern Physics 1990*, P. Lahti and P. Mitterstaedt eds., (World Scientific, Singapore, 1991).

Stapp, Henry P. (1993), *Mind, Matter, and Quantum Mechanics*, (Springer-Verlag, Heidelberg, Berlin, New York). p. 62, and Chapter 6: A Quantum

Theory of the Mind-Brain Interface LBL-28574 (1990)
http://www-physics.lbl.gov/~stapp/stappfiles.html

Stapp, Henry P. (1994), 'Strong Versions of Bell's Theorem, Phys. Rev. **A49**, 3182-3187. (See also Stapp 1993, pp. 5-9)

Stapp, Henry P. (1995), 'Quantum Mechanical Coherence, Resonance, and Mind', To appear in the Proceedings of the Norbert Wiener Centenary Congress (Edited by V. Mandrekar and P.R. Masini) to be published in the American Mathematical Society Series Proceedings of Symposia in Applied Mathematics (PSAPM) Also to appear in PSYCHE. Lawrence Berkeley Laboratory Report LBL-36915mod. http://www-physics.lbl.gov/~stapp/stappfiles.html

Stapp, Henry P. (1996) 'Chance, Choice, and Consciousness: A Causal Quantum Theory of the Mind/Brain, LBL-39744mod, (Tucson II) http://www-physics.lbl.gov/~stapp/stappfiles.html

von Neumann, John (1932), *Mathematical Foundations of Quantum Mechanics*, (Princeton University Press, Princeton, 1955), Chap. VI.

Wigner, Eugene (1961), 'Remarks on the Mind-Body Problem' *The Scientist Speculates*, I.J. Good, ed., (Heineman, London)