The Journal of Mind and Behavior
Spring and Summer 1997, Volume 18, Numbers 2 and 3
Pages 171 [69] – 194 [92]
ISSN 0271–0137
ISBN 0–930195–08–6

Science of Consciousness and the Hard Problem

Henry Pierce Stapp
University of California, Berkeley

Quantum theory can be regarded as a rationally coherent theory of the interaction of mind and matter, and it allows our conscious thoughts to play a causally efficacious and necessary role in brain dynamics. It therefore provides a natural basis, created by scientists, for the science of consciousness. As an illustration it is explained how the interaction of brain and consciousness can speed up brain processing, and thereby enhance the survival prospects of conscious organisms, as compared to similar organisms that lack consciousness. As a second illustration it is explained how, within the quantum framework, the consciously experienced "I" directs the actions of a human being. It is concluded that contemporary science already has an adequate framework for incorporating causally efficacious experiential events into the physical universe in a manner that: (1) puts the neural correlates of consciousness into the theory in a well defined way, (2) explains in principle how the effects of consciousness, per se, can enhance the survival prospects of organisms that possess it, (3) allows this survival effect to feed into phylogenetic development, and (4) explains how the consciously experienced "I" can direct human behavior.

"The Hard Problem" has dominated recent discussions of consciousness. Defined by Chalmers (1996), the problem basically is this: Why do our conscious experiences, which seem so totally different from the matter of which our brains are made, and which, according to the principles of classical mechanics can have no effect upon processes in our brain and bodies, exist at all. In this paper I address this problem within the context of the need to create an adequate theoretical foundation for the science of consciousness. Such a foundation must accommodate in a rational and useful way our

This paper is based on a plenary talk at the Conference Toward a Science of Consciousness 1996 held at the University of Arizona, April 8–13, 1996. I thank Pat Hayes and Aaron Sloman for helpful e-mail dialogs about functionalism and W. Swiatecki for useful comments. 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. Requests for reprints should be sent to Henry Pierce Stapp, Ph.D., Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720.

knowledge of (1) our conscious experiences, (2) the physical processes in our brains, and (3) the tight relationship that exists between these seemingly disparate parts of nature.

The hard problem of consciousness arose at the dawn of modern science when Descartes suggested that the world be conceived to consist of two different kinds of stuff, mind and matter, which interacted in human brains. The difficulty was this: How one can rationally comprehend an interaction between two such different kinds of stuff. The problem was compounded by the development of classical mechanics, which entailed that the matter part of nature, by itself, is dynamically complete: classical mechanics asserted that the dynamical evolution of matter is completely controlled by matter alone. Thus mind, contrary to the idea of Descartes, was reduced to an impotent witness to the flow of material events: mind became epiphenomenal.

The problem was solved around 1930 by scientists led by Niels Bohr, Werner Heisenberg, and Wolfgang Pauli, and the mathematician John von Neumann. They went beyond classical mechanics, and created the practically useful and empirically successful quantum theory. This theory (in the Neuman/Wigner elaboration of the Bohr/Heisenberg formulation) is essentially a rationally coherent framework for the interaction of mind and matter and, as such, is the natural foundation for the science of consciousness.

The problem of the relationship of mind and matter is basically a physics problem: it is the problem of how two different aspects of nature are related to each other. Many of the problems about consciousness debated by philosophers fade away once the tacit physics assumptions are made manifest.

Take, for example, the zombie question that is being hotly debated at the moment (Sutherland, 1995). In those discussions a "zombie" is typically defined to be a creature that is just like a normal human being in every detail of behavior, down to the submicroscopic level, but has no conscious experiences. Can such a thing exist?

To avoid working in a theoretical vacuum, one should specify the underlying physics assumption. If one is referring to real human beings, then one is speaking about quantum systems, and in a quantum-mechanical description of a human being the interaction between mind and brain is a crucial part of the dynamics. Hence mind cannot be left out. But if one is tacitly assuming that a classical-mechanics conceptualization of the brain is adequate, then, since mind is epiphenomenal within classical mechanics, consciousness can be eliminated without affecting behavior at any level. The natural conclusion to be drawn from this is simply that the counterfactual classical-physics assumption is incorrect.

Purpose

My aim here is first to describe a quantum-mechanical approach to the mind/matter problem, and then show how this approach can resolve in principle some basic problems that arise from trying to employ counterfactual classical mechanics instead of empirically adequate quantum mechanics as the theoretical foundation of the science of consciousness. These problems are to understand how consciousness can aid survival — epiphenomenal consciousness certainly cannot do so! — and can evolve during phylogenetic and individual development, and to see how a satisfactory theory of "free will" pops out.

Experience Within Science

Classical mechanics purports to describe the physical world and how it functions, and claims to achieve this goal without bringing in thoughts, feelings, or any other experiential aspect of nature. For centuries this restriction to the purely physical was regarded as an important virtue of science: science had banished the primitive superstition that spirits were lurking everywhere, and causing things to happen. Instead, the physical world was asserted to be built out of nothing but quantifiable properties that could be localized in a space–time, and whose functioning was completely determined by rigid mathematical laws that referred to nothing but these physical properties themselves. Thus when the creators of quantum theory introduced "our experience" and "our knowledge" into the theory of atomic phenomena their move was initially opposed by the scientific community. Soon, however, this important enlargement of the scientific conception of basic physical theory came to be accepted, at least nominally, by most workers in the field.

Recently some quantum theorists have been trying to exorcise "the observer" from quantum theory. These attempts encounter difficulties that I shall mention below. But in any case the important point in our quest for a science of consciousness is not that our basic physical theory *might conceivably* some day be able to be formulated without introducing observers — it is rather that our basic theory of matter, in its contemporary orthodox form, has an explicit and dynamically efficacious place for conscious experiencings. Science already has in place the basis of a rationally coherent and practically useful theory of the interaction of mind and matter: there is no need to invent another one.

The focus of orthodox quantum theory on the experiential aspects of nature was emphasized in the opening words of Niels Bohr's (1934) principal book on the subject, *Atomic Theory and the Description of Nature*: "The task of science is both to extend the range of our experience and reduce it to order"

(p. 1), and "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 experiences" (p. 18). An analogous statement by Heisenberg (1958a) is:

> The conception of the objective reality of the elementary particles has evaporated in a curious way, not into the fog of some new, obscure reality concept, but into the transparent clarity of a mathematics that represents no longer the behavior of the elementary particles but rather our knowledge of this behavior. (p. 99)

As these quotations indicate, the original formulation of quantum theory was essentially about "our experienced knowledge" of the physical world. This original formulation was subsequently extended by von Neumann, and it is this formulation, as elaborated by his close colleague Eugene Wigner, that I shall use here. Wigner called it the orthodox interpretation, and I think this terminology is justified by the fact that all other proposed interpretations are compared to it, or to the Copenhagen interpretation that it elaborates, to determine whether the predictions of the proposed new interpretation agree with the usual predictions.

But how can a physical theory rationally encompass two things so different as matter and our experiencings, and accommodate an interaction between them?

To see how this works, let it first be recalled how it is done in classical mechanics. All statements in science must be transcribed into statements about "our possible experiences" before they can be tested by human beings, or used to make predictions about what our future experiences will be. All such predictions are based upon some prior knowledge of the world about us. But this prior knowledge never determines the state of the world exactly. In classical statistical mechanics this prior knowledge, call it K, is represented by a "probability density function," $D(x,p;K)$. Here the argument x represents the positions of all of the particles of the system being examined, and p represents the momenta (or velocities) of these particles. The function $D(x,p;K)$ defines the probability density in phase space (i.e., in (x,p)-space) corresponding to the prior knowledge K.

For example, one might know only that some set of particles of interest lie in a certain box, and have a certain temperature. This knowledge K can be represented by a particular probability density function $D(x,p;K)$: this function will vanish when any coordinate x_i lies outside the box, and the momentum dependence will be specified in a well-known way (i.e., by the Boltzmann factor).

Suppose I now ask the following question: Given the statistical information represented by the probability function $D(x,p;K)$, what is the probability $P(m,e;K)$ that if I observe the system in manner m I will have experience e?

Let $E(m,e;x,p)$ be the probability (density) that if the system is in the state specified by the point (x,p) in phase space, and I observe it in manner m, then I will have experience e. Folding together these two probabilities one obtains a basic formula of classical statistical mechanics:

$$P(m,e;K) = \int dxdpE(m,e;x,p)D(x,p;K). \qquad (1)$$

This same formula holds in quantum mechanics. But in quantum mechanics the quantities $D(x,p;K)$ and $E(m,e;x,p)$ are not positive numbers, and hence a classical probability interpretation is ruled out. Also, the equation of motion in quantum theory is such that the different members of what in classical mechanics would be a "statistical ensemble" of independently moving points (x,p) in phase space do not evolve independently: in quantum theory these "independent components" are influenced by their neighbors. Hence, they cannot be conceived of as members of a classical statistical ensemble. Thus, the physical significance of the variables x and p becomes obscure: our intuitive conception of the physical part of nature fails. But the phenomenal variables e and m are just the same as before: these are the realities that we can hang onto. The commitment in quantum theory is to hang onto this basic formula, and to the *experiential* quantities, even though this means abandoning our classical notion of what the physical part of nature is like.

In the original Copenhagen interpretation of Bohr and Heisenberg, and their colleagues, the variables x and p referred to an external system that was being examined by the scientists. Thus in the function $E(m,e;x,p)$ the pair of variables (x,p) referred to one system (some small "observed" part of the universe) and the pair of variables (m,e) referred to things associated with a different part of the universe, namely the brain of the observer plus his body, extended to include his measuring devices.

Thus in the original interpretation the mapping function $E(m,e;x,p)$ connects two different parts of the universe. It therefore depends upon a separation of the physical universe into parts. But this separation of the physical world into parts was not well defined within the theory. Thus the theory, as originally formulated, was not fully satisfactory.

John von Neumann (1932) and Eugene Wigner (1967) extended the original Copenhagen form of the theory by identifying "the system" with the entire universe. Then no mysterious — and probably impossible to coherently implement — separation of the universe into two completely different kinds of matter was needed: the variables x and p become the variables needed to describe the entire universe, including, in particular, the brains of the observers. The function $E(m,e;x,p)$ then relates an "experiential space," whose elements are labeled by the variables m and e, to a physical space that includes, in particular, the brain of the relevant observer. The mapping

$E(e,m;x,p)$ from experiences to physical variables defines in this theory the neural correlates of consciousness, which thus enter the theory in a fundamental and well-defined way. This injection of the neural correlates of consciousness into the basic equations of physics was troubling to some physicists, who wished to return to a more classical type of physics in which consciousness was kept out. They endeavored (not successfully I think) to reformulate quantum theory in some way that gets consciousness out of physics. But, regardless of whether they can succeed or not, for the deeper issues before us the orthodox formulation is ideal.

A key difference between the classical and quantum theories concerns the nature of the interaction between mind and matter. When the new experience labeled by (m,e) occurs, the prior knowledge K is augmented: $K \rightarrow (K,(m,e))$. If the observation can yield only the answers YES = (m,e) and NO = $(m,-e)$ then the transformations to the new probability functions are, classically:

$$D(x,p;K) \rightarrow D(x,p;K,(m,e)) = E(m,e;x,p)D(x,p;K), \qquad (2a)$$

and

$$D(x,p;K) \rightarrow D(x,p;K,(m,-e)) = E(m,-e;x,p)D(x,p;K), \qquad (2b)$$

where,

$$E(m,e;x,p) + E(m,-e;x,p) = 1. \qquad (2c)$$

[Here I have normalized the probabilities $D(x,p;K,(m,e))$ and $D(x,p;K,(m,-e))$ relative to the original condition K.]

These equations entail the equality

$$D(x,p;K,(m,e)) + D(x,p;K,(m,-e)) = D(x,p;K). \qquad (3)$$

This identity means that in classical mechanics the observation itself is "passive": if one adds together the probabilities corresponding to the alternative possible outcomes, YES and NO, then the result is the same as the original probability function: nothing is changed if one makes the observation but does not discriminate between the results.

In quantum mechanics the functions $D(x,p;K)$ and $E(m,e;x,p)$ can be regarded as the matrix elements $\langle x | D(K) | p \rangle$ and $\langle p | E(m,e) | x \rangle$ of *operators* $D(K)$ and $E(m,e)$. [This change of notation also accommodates the generalization from ordinary quantum mechanics to quantum field theory, which is what actually must be used.] Then a huge difference between quantum mechanics and classical mechanics is that the occurrence of the experience (m,e) is an event that acts dynamically back on the physical world. The physical world before this event, which is represented by $D(K)$, is transformed as follows:

$$D(K) \Rightarrow D(K,(m,e)) = E(m,e)D(K)E(m,e), \qquad (4a)$$

or

$$D(K) \Rightarrow D(K,(m,-e)) = E(m,-e)D(K)E(m,-e), \qquad (4b)$$

where

$$E(m,-e) + E(m,e) = 1. \qquad (4c)$$

The analogue of (3) does not hold here, in general, because (4), unlike (2), is not linear in the quantities E: the observation changes the physical system, even if no discrimination is made between the two possible results.

The transformation (4a) represents a "reduction of the state": the state $D(K)$ prior to the actual experiential event that is represented by $E(m,e)$ is transformed to a new state $D(K,(m,e))$ that incorporates the new conditions labeled by (m,e). In keeping with this meaning, the operator $E(m,e)$ satisfies the (idempot) condition (i.e., that P times P = P)

$$E(m,e)E(m,e) = E(m,e) :$$

a single experience acting twice has the same effect as its acting once.

This description shows how our experiencings become woven into the fabric of the quantum mechanical description of nature: they are the identifiers of events that are the comings into being of these experiencings, and that also act efficaciously upon the mathematical structure that represents the physical aspect of nature. In this new picture of nature the physical aspect constitutes the more subtle aspect of reality: it acts merely as a substrate of *propensities* for experiential events to occur. These experiential events are the more robust basic realities.

Human experiences are presumed to be very high-level forms of experiential events. They are the foundation of the human scientific enterprise: they constitute the data upon which our science is based. Quantum mechanics, as formulated by Bohr and his colleagues, is predicated on the fact that our experiences of the physical world — our immediate phenomenal knowledge of it — can be described in the language of classical mechanics, considered as an extension of ordinary every-day language. Quantum dynamics itself, in the von Neumann/Wigner form, involves the fact that in many situations classical mechanics can provide very accurate approximations. Thus quantum theory is a unified and seamless theory that accurately describes the quantum features of nature, but also justifies the use of classical concepts in situations where those concepts are applicable.

Problems with the Classical-Mechanical Description of the Brain

Classical mechanics cannot account for the empirically observed properties of matter in general, or of the matter out of which brains are made, in particular. However, because consciousness seems to be associated with large-scale brain activities, and with chemical processes that seemingly can be simulated by quasi-classical processes that mock up the atomic-level processes, it is widely assumed that the crucial role of quantum theory in correctly describing brain dynamics is merely a technical complication that has no basic ontological or physical significance in understanding the dynamical effects of consciousness, per se, upon brain activity. That is, it is assumed that as far as the effects of consciousness on brain dynamics is concerned, it is legitimate to imagine the brain to be the sort of thing that it is imagined to be in classical mechanics rather than the radically different sort of thing that quantum mechanics says it is. But this simplistic assumption immediately encounters problems associated with the epiphenomenal character of mind within the classical-mechanical conception of nature.

The classical-mechanical description of the physical world (although empirically false) is, logically speaking, dynamically complete, even though it never mentions the experiential (i.e., phenomenal) aspects of nature. To account, within this framework, for the factual occurrence of these experiential aspects some scientists and philosophers have been led to suppose that certain brain activities simply "elicit" corresponding experiences without the latter reacting back on the brain. According to this idea, the experiential world is merely an epiphenomenal add-on to a physical world that, in the context of the mind–brain problem, can be imagined to be described by classical mechanics. This scenario might be logically possible, but it seems preposterous that nature should create a whole extra world that is totally unlike the physical world, and in no way entailed by the laws that govern the physical world, and then give this add-on world no dynamical role to play.

The unnaturalness and non-parsimoniousness of this (classical dualistic) notion has led to an opposing (classical identity/functionalist) claim that experiencings simply *are* certain functional activities of the brain, described in a phenomenal rather than physical/functional language: that is, that all that there is in nature (at a level that is adequate to cope with the mind–brain problem) are the classically described physical/functional activities, but that certain of these activities can be *identified* as *also* our conscious experiences. I say "also" because if classical mechanics is accepted, then one has, in principle, the physical description that it entails: any other description is then something beyond what classical mechanics itself entails.

To examine the identity/functionalist claim, suppose brain science has finally evolved to the point where it can give a complete description of brain process: suppose it can provide a detailed understanding of how "memory tracks" are laid down in the brain, and how these memory tracks are accessed by later brain activities. And suppose that the brain scientist could even wire up a brain and map out its various patterns of activity in sufficient detail to be able to follow through what happens in the brain when the person is asked "What did you eat for breakfast this morning?" Suppose the brain scientist is able to follow through the progression of patterns of excitations and see how the physical memory tracks laid down during breakfast come to be accessed, and to see how the content of those memory tracks feeds into the process that finally produces the spoken reply "Ham and eggs!"; and even the reply to the follow-up question "What color were the eggs?": "They were a chalky-whitish kind of yellow, rather than an orange-ish shade of yellow!" And suppose the brain scientist's description is detailed enough to see even the laying down of certain memory tracks that will allow the person to respond to later queries about his sequence of thoughts as he was formulating his answers to the questions. Suppose further that the brain scientist is able to construct a mapping from the physical space of certain kinds of patterns of neural activity to corresponding "phenomenal events" described in a phenomenal language, and that this mapping is such that it fits perfectly with all the responses that the person makes to questions about his "experiencings" of pain, of color, and of every other kind of experience that he says he has.

Within this context we may consider, in the framework of a classical-mechanical conceptualization of the brain, the two alternative claims:

1. The phenomenal activity *is elicited by* a neural/functional activity.

2. The phenomenal activity *is* the corresponding neural/functional activity.

The advocate of claim 2, which I call (classical) functionalism, can claim parsimony, and can point to the unnaturalness of the existence, asserted by claim 1, of a whole world that is fundamentally different from the physical world, and has no effect upon the physical world. The existence of such a world would seem to require a whole new machinery in nature, a machinery that would somehow "cause" the phenomenal events to occur, or "elicit" them, even though nothing in the classical physical laws requires the existence of any such extra machinery.

On the other hand, the advocate of claim 1, which I call (classical) dualism, can insist that any claim that two different descriptions describe one single thing must be supported by some explanation of how the one thing acquires these two descriptions. For example, the claim that temperature in thermodynamics *is* the same as mean kinetic energy in statistical mechanics is supported by the fact one can *deduce* this correspondence from the laws of physics. The claim that "the morning star *is* the planet Venus," is also explained on the basis of the laws of physics, by noting that the phrase "the morning star" has an original meaning that refers to a certain kind of "experiencing" (an experiencing of a certain brightness in the morning sky not too far from the horizon), and by the fact that this experiencing can be deduced, on the basis of the laws of physics, to be caused by sunlight reflected off of the planet called Venus, provided such "experiencings" of an observer can be assumed to be evoked by corresponding activities in the brain of this observer. But in the case of the claim that "the pain P *is* the functional brain activity F," there is no possibility of deducing this connection from the orthodox principles of classical mechanics.

The functionalist can reply that he has in fact, on the basis of the principles of classical mechanics, provided a detailed causal account of the very activity in the brain that *constitutes* experiencing. He can claim that his phenomenal knowings are, precisely, his brain's accessings and monitorings of certain aspects of itself. He can claim that his experiencings *are* his brain's functional activities of laying down and retrieving certain kinds of memory tracks that (1) contain all of the information that he feels that he is becoming aware of; and (2) initiate all of the actions that he feels he is initiating.

The dualist can reply that if one takes the principles of classical mechanics as the basic principles, then one can *prove* that the identity/functionalist hypothesis is false, provided one does not simply assume it to be true on the basis of phenomenal experience. The point is that from the classical principles and initial boundary conditions one can *deduce* the presence of the physical activities but cannot deduce the presence of the phenomenal activities. Hence, the two things are distinguishable within that framework. Thus, the functionalist hypothesis, though perhaps logically possible, contradicts what the classical principles by themselves entail. This looseness in the logical situation arises precisely because there is something that is known first-hand to be ontologically real and present, but whose presence is not implied by the physical theory that is being used to describe the system.

This egregious omission arises, it appears, from a faulty conceptualization of the situation, namely from the use, in the conceptualization of the mind/brain system, of a physical theory that has already been found by science, on other grounds, to be inadequate precisely at the point at issue, namely the relationship between our physical description of nature and our phenomenal knowledge of it. If one is going to go beyond the principles of classical mechanics, and add some extra hypothesis regarding mind, then the most reasonable procedure, within science, is to try first to use the theory that science itself has already discovered in this connection.

Within the Bohr/von Neumann/Wigner formulation of quantum theory, the phenomenal facts are introduced from the very beginning as the basic actualities that the theory is about. And these actualities are causally efficacious. Hence, within this quantum description of nature there is no need to introduce into the theory any element not clearly entailed by the original basic principles, or any element that is dynamically inert: we are not forced by science to accept an unreasonable stance on either point, provided we accept what science itself has been telling us for seventy years.

Quantum Mechanics as the Solution to the Scientific Aspect of the Hard Problem

The "hard problem" has several aspects. From the perspective of science the question "Why does consciousness exist?" can be compared to the question "Why does the electromagnetic field exist?" A physicist can answer this question by giving an account of the important function that the electromagnetic field plays in workings of nature, as they are represented in his physical theory. Of course, consciousness plays no role at all in the classical mechanics account of nature, and hence, no functional answer is possible within the classical-mechanics conceptualization of nature. Since it is unreasonable for nature to have such a nonefficacious component, the question of "Why consciousness exists" becomes essentially a plea for a more adequate conceptual understanding of nature, one in which consciousness plays an essential role.

Two essential roles of consciousness in the quantum formulation are:

1. Our conscious experiencings are what both science in general and quantum theory in particular are about. One cannot eliminate our experiences from the theory without eliminating both the connection of the theory to science and also the basic realities upon which the theory itself rests: experiences are the basic realities that the more subtle "physical" aspects of nature are propensities for.

2. Technically, experiences are used to solve the so-called basis problem in quantum theory. That is, within the physical domain itself there is no natural foundation or basis for deciding which special states are the ones into which the quantum state can "collapse." The core idea according to Bohr is that special states correspond to our experiences, and this idea is carried by the von Neumann/Wigner formulation into equations (1) and (4). This amounts to the idea that the body/brain processes generate possibilities that correspond

to possible experiences, and then nature selects, in accordance with the basic quantum statistical rule, one of these possible experiences, and actualizes it, and its body/brain counterpart. But this means that our experiences are not only the basic realities of the theory, and the link to science, as noted in 1, but also play a key role in specifying the "set of allowed possibilities" that enter into the causal chain of mind/brain events. These allowed possibilities must be just the ones that correspond to our possible experiences or the theory would lose its tight connection to science: the events in the theory would no longer correspond to the experiential realities.

The point here is that the physical aspects of nature alone do not define the preferred states, and hence something that is either consciousness itself, or a stand-in for consciousness that plays a role essentially indistinguishable from it, is needed. Since consciousness is known to exist, and it is indistinguishable from what seems needed to solve the basis problem in a satisfactory way, it would be both nonparsimonious and extremely perverse to introduce something as yet unknown to simulate something that is both known to exist, and in need of a role to play.

Macroscopic Quantum Effects In Brain Dynamics

Brain dynamics is controlled by chemical processes. Eventually, we will want to have a coherent account in which chemical processes of the brain fit seamlessly into the whole process. Ultimately, for these basically chemical reasons, a quantum description will be needed. But I wish to focus here on more macroscopic quantum effects: effects that would distinguish the quantum treatment from a classical model in which the currents flowing along neurons are described in classical terms.

Brain process is essentially a search process: the brain, conditioned by earlier experience, searches for a satisfactory response to the new situation that the organism faces. It is reasonable to suppose that a satisfactory response will be programmed by a template for action that will be implemented by a carefully tuned pattern of firings of some collection of neurons. This executive pattern would be a quasi-stable vibration that would commandeer certain energy resources, and then dissipate its energy into the initiation of the action that it represents. [See Stapp (1993) for a more detailed description of this process.]

If the programmed action is complex and refined, then this executive pattern must contain a great deal of information, and must, accordingly, be confined to a small region of phase space. Stated differently, the relative timings of the pulses moving along the various neurons, or groups of neurons, will have to conform to certain ideals to within very fine levels of tolerance. How does the hot, wet brain, which is being buffeted around by all sorts of thermal and chaotic disturbances, find its way to such a tiny region in a timely manner?

The problem, reduced to its basic dynamical form, is this: How, in a 3n-dimensional space (where n represents some huge number of degrees of freedom of the brain), does a point that is moving in a potential well that effectively blocks out those brain states that are not good solutions to the problem (i.e., that do not represent templates for satisfactory actions, under the conditions at hand), but that does not block the way to the good solutions, find its way in a short time to a good solution, under chaotic initial conditions? Classically, the point in the 3n-dimensional space must just follow the deterministic equation of motion until it eventually wiggles its way out of the potential well. But the quantum system has the advantage of being able to explore simultaneously all possible ways to get out. This is because the quantum state corresponds, essentially, to a superposition of all the allowed possibilities. Moreover, this 3n-dimensional cloud of virtual possibilities satisfies an essentially hydrodynamic equation of motion: it acts like a single glob of water, rather than like a collection of independently moving droplets (Feynman, Leighton, and Sands, 1965). That is, the motion of each point in the cloud is influenced by its neighbors, as was emphasized earlier. But then when some parts of the glob find their way out of the potential well, and thus flow out of the confined region leaving a partial void, the nearby parts of the glob will flow in to take their place, and will then in turn flow out. Thus, all of the glob will tend to flow out quickly, like water flowing out of a leaky bucket.

The brain is a quantum system, and will automatically use this hydrodynamical property, and hence will presumably operate faster in searching for an acceptable template for action than its classical counterpart can. Thus, the need to use quantum theory is not just a philosophical matter: it will probably be needed to account for the speed of the (analog) search processes.

Decoherence

It has often been observed that the coupling of a system to its environment has a tendency to make interference phenomena that are present in principle within quantum systems difficult to observe in practice. Phase relationships, which are essential to interference phenomena, get diffused into the environment, and are difficult to retrieve. The net effect of this is to make a large part of the observable phenomena in a quantum universe similar to what would be observed in a world in which certain collective (i.e., macroscopic) variables are governed by classical mechanics. This greatly diminishes the realm of phenomena that require for their understanding the explicit use of quantum theory.

These decoherence effects will have a tendency to reduce, in a system such as the brain, the distances over which the idea of a simple single quantum system holds. This will reduce the distances over which the simple hydrodynamical considerations described above will hold. However, the following points must be considered.

(a) A calcium ion entering a bouton through a microchannel of diameter x must, by Heisenberg's indeterminancy principle, have a momentum spread of h/x, and hence a velocity spread of $(\hbar/x)/m$, and hence a spatial spread in time t, if the particle were freely moving, of $t(\hbar/x)/m$. Taking t to be 200 microseconds, the typical time for the ion to diffuse from the microchannel opening to a triggering site for the release of a vesicle of neurotransmitter, and taking x to be one nanometer, one finds the diameter of the wave function to be about 0.04 centimeters, which is huge compared to the 1/100000000 centimeter size of the calcium ion. Of course, this free-particle spreading will be greatly reduced by the multiple scatterings of the ion as it moves through the aqueous medium. But this quantum spreading of wave packets constitutes a counterforce to the mechanisms that tend to diminish quantum coherence effects. If the wave functions of the centers of mass of these calcium ions are not confined to regions that are small compared to the size of the ion itself, then if no collapses occur, the brain must evolve into an amorphous superposition of states corresponding to a continuum of different possible macroscopic behaviors.

(b) The normal process that induces decoherence arises from the fact that a collision of a state represented by a broad wave function with a state represented by a narrow wave packet effectively reduces the coherence length in the first state to a distance proportional to the width of the second state. But in an aqueous medium in which all the states of the individual systems have broad packets this mechanism is no longer effective; coherence lengths can remain long.

(c) Even if the coherence length were only a factor of ten times the diameter of the atom or ion involved in some process, the cross section involved would be a hundred times larger. The search processes under consideration here involve huge numbers of atoms and ions acting together, and the cross-section factors multiply. Thus, even a small effect at the level of the individual atoms and ions could give, by virtue of the hydrodynamical effect, a large quantum enhancement of the efficiency of an essentially aqueous macroscopic search process.

Everett and Consciousness

Einstein (1952) illustrated the central logical problem in contemporary quantum theory with a simple example. It involves a radioactive source, a detector of some product of the decay, a pen that draws a line on a moving strip of paper and makes a blip when the decay is detected, and a human observer of the blip. If one uses the Schrödinger equation, then one finds that the system evolves into a continuous superposition of states corresponding to all possible positions of the blip on the strip of paper. But when the human observer looks, he sees the blip in one well defined place. Thus, the Schrödinger equation is not telling the whole story. If one wants to have an account of what is actually happening, then something else needs to be added, namely Heisenberg's "transition from possible to actual" (or some substitute for it) that allows the many possibilities generated by the Schrödinger dynamics to be reduced to the single actually experienced reality. The von Neumann/Wigner form of quantum theory accepts the Heisenberg transitions from possible to actual as real events.

The strangeness of Heisenberg's idea (1958b) of transitions from "possible" to "actual," naturally has led scientists to explore diligently the possibility that these transitions never happen: that the Schrödinger equation never fails. This possibility was examined in some detail by Everett (1957). The consequence of that work, and of many later efforts to clarify it, is to focus attention even more strongly than ever on the problem of our consciousness experience. For if the Schrödinger equation never fails then there is a huge disparity between the objective world, which is represented by the evolving state of the universe, and our subjective experiences of it.

The basic problem with this interpretation is that the needed psychological, i.e., experiential, properties of brains do not follow from the Schrödinger equation. The latter can perhaps generate independently evolving "branches" of the wave function of the brain, with different branches corresponding to different streams of consciousness, but these branches are *conjunctively* present. However, in order to obtain the statistical predictions of quantum theory, which pertain to our experiences, the experiential streams that correspond to these different physical branches must be *disjunctive*. That is, the objective physical state will contain branch A *and* B, and so on, whereas to get statistical statements about our subjective experiences, one needs the logical structure of experience A *or* experience B, and so on. This means that "mind" needs an ontology and dynamics that does not logically follow from the Schrödinger equation that controls the "brain." This need for a second level of reality, controlled by a dynamic that does not follow logically from the first, appears to nullify the advantage that the Everett interpretation seemed at first to provide. In any case, a spotlight becomes focused more strongly than ever on the problem of the connection between objective aspects of a mind/brain represented in the conjunctively present branches of the wave function and the disjunctive subjective aspects.

Looking at the evaluations by physicists who are pursuing environmental decoherence effects, and other essentially "Everett" ways of approaching the problem of quantum measurement, we find Zurek (1986) saying, of these approaches, that they do not allow us to understand how we as "observers" fit in, and hence they appear to him to be merely "a hint about how to proceed rather than the means to settle the issue quickly" (p. 96). Joos (1986) says "Of course, the central problem remains unsolved: Why are there local observers?" (p. 12). Gell–Mann and Hartle (1991) emphasize that: "If history dependence can be properly introduced into the explicit treatment of quantum mechanics, then we may be able to handle individuality [of observers] with the care that it deserves" (p. 14). Omnes (1994), who gives perhaps the most comprehensive description of these Everett-type theories says, about the Everett proposal, that he feels "it impossible to accept as a satisfactory answer to the problem of actuality" (p. 348). So, almost forty years after the Everett paper appeared, it is acknowledged by these workers, and I think by all others who have examined the matter with sufficient care, that the problem of the observer has not been solved by that approach. This issue, namely the problem of how our individual experiences fit into nature, becomes emphasized as the central unsolved problem: the Everett approach makes this problem even more pressing and glaring than before. The Everett approach tries to resolve the problem of the observer without going beyond the Schrödinger equation, but it certainly has not succeeded in doing so.

There are also some proposed interpretations that try to leave consciousness out altogether. One is the pilot-wave model of David Bohm (1952). Einstein rejected this model as the real solution on the basis of its being "too cheap," and Bohm agreed with this assessment. In the last chapter of his book with Basil Hiley (Bohm and Hiley, 1993) on this subject, he tried to go beyond it and to bring in consciousness. But the effort was not carried very far.

The other most highly developed theory is that of Ghirardi, Rimini, and Weber (1986) and Pearle (1989, 1996). In this model, the free parameters have to be very finely adjusted (Collett, Pearle, Avignone, and Nussinov, 1995) in order for the predictions to be compatible with experiment, and it may be ruled out altogether in the not too distant future. It seems that nature would have to be very malicious to have finely tuned the parameters so that we could not distinguish between this state of affairs and the orthodox theory, which fits all known data without any fine tuning. In view of these considerations it is, I think, completely reasonable to take the orthodox

(Bohr/Heisenberg/von Neumann/Wigner) interpretation as the one best suited as the basis for the science of consciousness.

Causality and Chance

One can ask what causes a particular phenomenal event to occur. Orthodox quantum theory says "statistical cause": the quantum state of the brain specified the "propensities" for the various phenomenal possibilities. Thus, the cause of the phenomenal events is not the local deterministic "mechanical" sort of cause that occurs in classical mechanics. If one insists on naming what it is that picks out the one particular possibility that actually occurs in a given situation, orthodox quantum theory can only answer: pure chance!

As regards the role of chance, Bohr (1958) says this:

> The circumstance that, in general, one and the same experimental arrangement may yield different recordings is sometimes picturesquely described as a "choice of nature" between such possibilities. Needless to say, such a phrase implies no allusion to a personification of nature, but simply points to the impossibility of ascertaining on accustomed lines directives for the course of a closed indivisible phenomenon. Here, logical approach cannot go beyond the deduction of the relative probabilities for the appearance of the individual phenomenon under given conditions. (p. 73)

Bohr carefully avoids affirming that there actually is in nature herself an irreducible element of chance. He says, rather, that the entry of chance is due to difficulties that arise from trying to apply customary (local-reductionistic) thinking to closed indivisible phenomena. This suggests that nature herself may have some [necessarily nonlocal (Mermin, 1994; Stapp, 1993, 1997)] way of determining which event will actually occur: i.e., that we do not have to accept the absurdity of something definite arising out of absolutely nothing at all.

Be that as it may, quantum theory, in its contemporary form, separates the dynamics into two parts: the Schrödinger-directed evolution of the quantum state, and the quantum selection process, which appears to operate according to some specified rules of chance. This second process surveys the quantum state in terms of the possible experiences that it (the selection process) could extract from that state, and actualizes one of the possibilities in accordance with the quantum statistical rules.

Science may aspire to probe more deeply into this selection process, but doing so within the scientific paradigm would appear to require data indicating some deviations, under certain conditions, from the quantum statistical rules. Here I stay strictly within the bounds of contemporary orthodox science in accepting the quantum statistical rules as primitive elements of our basic theory.

General Description of Brain / Mind Dynamics

Before going further, I give a general overview of the conception of brain/mind dynamics that seems best to fit the quantum process. This is a very brief synopsis of the description of mind/brain dynamics given by Stapp (1993, chapter 6).

Body–World Schema

It is accepted here (or postulated) that there is in a person's brain a high-level representation of his body and its environment: a person's body and its environment are represented in the brain by patterns of neurological and sub-neurological brain activity. This representation in the person's brain of his/her body and its environment is called the "body–world schema." It is expanded to include representations of beliefs, and hence is sometimes called the body–world–belief schema, but I shall stick to the shorter name. Each phenomenal quantum event is assumed to actualize a body–world schema. An *attentional* event up-dates the body–world schema; an *intentional* event actualizes a body–world schema that is an image of an intended state of the body–world. This projected (into the future) image serves as a template for action: the automatic unfolding, in accordance with the Schrödinger equation, of the pattern of neural activity that constitutes the body–world schema, tends to evolve into the intended action.

Facilitation, Associative Recall, and Control

The persistence of a pattern of neural excitation etches this pattern into the physical structure of the brain, in the sense that this pattern is facilitated (made easier to activate), and that a later activation of part of the pattern tends to spread to the whole. This facilitation and spreading effect provides the basic mechanism for an explanation of associative recall, and of the control aspect of the body–world schema.

The Effect of Quantum Theory

The effect of quantum theory is essentially the same as it was in the Einstein example described earlier: the evolution controlled by the Schrödinger equation will produce, instead of one single body–world schema, rather a continuum, consisting of a superposition of all the possibilities, with no one possibility singled out as the one that is actually experienced. Thus, for example, for every possibility in which a "synaptic event" occurs — the release of a vesicle of neuro-transmitter — there will be other superposed possibilities in which this event does not occur; and for every situation in which an action potential spike exists at one place along an axon there will be other superposed possibilities in which the spike is a little earlier, or a little later, and still others in which it is much earlier or much later. To extract the actually-experienced reality from this amorphous conglomerate of superposed possibilities, one needs, according to the Heisenberg ontology accepted here, a transition from possible to actual. This transition is called an actualization event — it selects and actualizes one of the alternative possibilities generated by the Schrödinger–equation-controlled evolution.

Many people, even many scientists, suppose that the quantum events (i.e., the collapse events) occur at a microscopic level. However, there is no reason for this to be so, and no empirical evidence that it is so. Indeed any evidence for microscopic quantum events would be evidence *against* the correctness or completeness of contemporary quantum theory, and no such evidence has ever been found. A core idea of the quantum model is that each actual event is a phenomenal event that is essentially an integrated picture of the body–world that expresses the information contained in the body–world schema that is actualized by the event.

Survival Value of Consciousness, and Phylogenetic Development

A major difficulty with the classical physics approach to consciousness is the problem of the connection of consciousness to survival. Consciousness is not entailed by the principles of classical mechanics, and hence it cannot, per se, have any effect on the dynamics — within that framework it is epiphenomenal. Postulating that the functional activities *are* the conscious activities does not help, because this postulate is, logically, an addition to the classical principles, which already give a dynamically complete framework. The physical evolution would be exactly the same even if the extra postulate were not added, and consciousness were not present. But, then, within the classical framework consciousness cannot enhance the survival prospects of creatures that possess it. Hence, it is not understandable in naturalistic terms why we should possess it.

In the quantum world, consciousness *can be* causally efficacious, and in the orthodox theory being discussed here consciousness *is* causally efficacious. Thus, creatures possessing consciousness could in principle have enhanced survival prospects. In this section I shall show how, in the quantum model, the survival prospects of a creature could be enhanced by the possession of consciousness.

It has been mentioned above that high-level brain process is essentially an analog search process, and that the hydrodynamical character of the quantum law of evolution can speed up this process. This speed-up does not require any brain-wide quantum coherence — short-distance coherence suffices. If one imagines a big potential well that is being explored for a route into a little pocket that represents the appropriate template for action, then we immediately see an additional speed-up arising from the physical action associated with conscious experiences. In a model with no collapses — and this includes both classical models and Everett-type quantum models and also the Bohm pilot-wave model — once the little pocket gets filled it may tend to stay filled, in equilibrium with other parts of the system. But in the orthodox model, this little pocket can be repeatedly probed by the selection process. If the result is "yes," then the experience occurs and the state is collapsed into this pocket, and the resulting action proceeds. But if the answer is "no," then no experience occurs. But then, in accordance with (4b) and (4c), the pocket is cleared out, in accordance with the new fact that the system is now "known" to be not in this state. That is, the probability function will have a "hole" where the probability drops to zero. But then the hydrodynamic effect will tend to pull more probability into the pocket where the selection process can have another crack at it. So, at least in principle, the orthodox model would tend to be more efficient at implementing effective action. If a similar species, otherwise on a par, but organized so that its templates for action do not mesh with possible conscious experiences, then for these creatures there could be no such dynamical effect and they would tend to act more slowly, and hence be less likely to survive.

Bats and dogs and other animals are probably conscious. So we do not expect all possible experiences in nature to be just like our own. Indeed, the demands of phylogenetic development would entail that there be a continuum or closely spaced spectrum of possible experiences extending back down to a very primitive level. Then the survival advantage mentioned above could enter at a very low level and work its way by natural selection up the phylogenetic ladder. A similar chain would probably be operative in the embryonic and subsequent development of the individual organism.

I stress that the enhancement of the survival prospects of conscious creatures in this quantum theory of consciousness arises simply from following through the logic of the orthodox (Bohr/Heisenberg/von Neumann/Wigner) quantum theory, as represented, basically, by equations (1) and (4), and the Schrödinger equation. It does not depend upon exotic effects such as brain-wide quantum coherence, which would be very difficult to achieve, or upon the difficulties involved with reconciling quantum theory and general relativity.

At the present stage of empirical technique it is probably not possible to confirm the existence or nonexistence of the dynamical effects described in this section. But they do follow rationally from the quantum principles, and provide a basis for understanding, in a naturalistic way, the occurrence of consciousness in connection with human brain process.

Free Will

The quantum framework leads naturally to the everyday concept of free will and personal responsibility. The key point is the concept of "I." In classical mechanics the personally experienced "I" is not entailed by the (dynamically complete) physical principles, and it thus lies impotently, and hence without responsibility, outside the causal chain of physical events. In the quantum picture, the experienced quality of I-ness is experienced, and is therefore *part* of the stream of conscious events: the experienced I-ness belongs to the experience, not vice versa. It belongs to what William James calls the fringe of experience. Surrounding the central focus of our experience is a slowly changing background part that keeps us vaguely aware of who we are and what we are doing, both in the long and short terms. The way the dynamics works is that the state of the brain (and universe) evolves under the control of the Schrödinger equation, and then a collapse occurs. This collapse actualizes a template for action that is the physical counterpart of the corresponding experienced thought. Thus, the effect of the thought gets injected into the causal chain of events.

The overall guidance part of the thought comes from the slowly changing fringe part that is the experienced "I." This part is carried over time by the memory structure in the brain, and reflects both genetic input, educational training, and the effects of all earlier conscious experiences, which have likewise had their effects injected into the causal chain of brain/mind events by means of collapses to templates for action that are the physical images of these thoughts. In this way, the experienced "I" feeds into one's behavior in essentially the way that we intuitively feel that it does, and is in turn being created in its forward development by the combined effect of its own input into the physical process and the action of the environment upon the body and brain.

There is, of course, some "static" injected into this process of personally controlled behavior by the quantum selection process, but this static is limited to selections between options to which our own personal process has assigned significant statistical weight. Thus, although the quantum selection process gets the final say at the level of the individual selections, the statistical weights are controlled by the personal process that is itself controlled basically by the experienced "I" (for more details see Stapp, 1993).

In this paper I have adhered to the orthodox position that the quantum selection process is bound by the quantum statistical rules of contemporary quantum mechanics. Any suggestion that this law fails in certain cases should be supported by powerful data or reasoning, for the failure of this law would open a Pandora's box of theoretical difficulties and uncertainties.

Open Questions

The major research problem is to determine the neural correlates of consciousness. These correlates are represented in the theory by the quantities $E(m,e;x,p)$. Determining them is a major part of what Chalmers (1996) called the "easy" problem. But there are also deeper questions. What are the principles and mechanisms that determine the set of possible experiences? Since this set must be large and varied, there must be some process that determines them. From a naturalistic perspective, this process must be a natural process. However, the basic equations of quantum theory, namely the Schrödinger equation, and equations (1) and (4), do not immediately shed much light on this deeper question.

This difficulty emphasizes the fact that quantum theory is basically a pragmatic theory (Stapp, 1972): it is a way of making progress toward some practically useful understanding of nature without knowing how everything really works at the fundamental level. This is perhaps a humbling admission for science. But the fact is that we still have a long way to go. The creators of quantum theory did provide us, however, with a rational theoretical framework that allows progress to be made.

References

Bohm, D. (1952). A suggested interpretation of quantum theory in terms of "hidden" variables, I and II. *Physical Review*, *85*, 166–193.

Bohm, D., and Hiley, B. (1993). *The undivided universe*. London: Rutledge.

Bohr, N. (1934). *Atomic theory and the description of nature*. Cambridge: Cambridge University Press.

Bohr, N. (1958). *Atomic physics and human knowledge*. New York: Wiley.

Chalmers, D.J. (1996). *The conscious mind: In search of a fundamental theory*. New York: Oxford University Press.

Churchland, P.S. (1989). *Neurobiology: Toward a unified science of the mind–brain*. Cambridge, Massachusetts: MIT Press.

Collett, B., Pearle, P., Avignone, F., and Nussinov, S. (1995). Constraint on collapse model by limit on spontaneous X-ray emission in Ge. *Foundations of Physics*, *25*, 1399–1412.

Einstein, A. (1952). In P.A. Schlipp (Ed.), *Albert Einstein: Philosopher–scientist* (pp. 667–673). New York: Tudor.

Everett, H., III (1957). Relative state formulation of quantum mechanics. *Reviews of Modern Physics*, *29*, 463–562.

Feynman, R.P., Leighton, R.B., and Sands, M. (1965). *The Feynman lectures in physics* (Volume III, Chapter 21). New York: Addison–Wesley.

Gell–Mann, M., and Hartle, J.B. (1991). *Classical equations for quantum systems* UCSBTH-91-15, University of California, Santa Barbara.

Ghirardi, G.C., Rimini, A., and Weber, T. (1986). Unified dynamics for microscopic and macroscopic system. *Physical Review*, *D 34*, 470–491.

Heisenberg, W. (1958a). The representation of reality in contemporary physics. *Daedelus*, *87(3)*, 95–108.

Heisenberg, W. (1958b). *Physics and philosophy* (Chapter III). New York: Harper Row.

Joos, E. (1986). Quantum theory and the appearance of a classical world. In D.M. Greenberger (Ed.), *New techniques and ideas in quantum measurement theory, Annals of the New York Academy of Science, 480* (p. 12). New York: New York Academy of Science.

Mermin, N.D. (1994). Quantum mysteries refined. *American Journal of Physics, 52*, 880–887.

Omnes, R. (1994). *The interpretation of quantum theory*. Princeton: Princeton University Press.

Pearle, P. (1989). Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A 39*, 2277–2289.

Pearle, P. (1996). Relativistic collapse model with tachyonic features. *Hamilton College Physics Preprint*. Clinton New York: Hamilton College.

Pearle, P., and Squires, E. (1995). Bound state excitation, nucleon decay experiments, and models for wave function collapse. *Physical Review Letters, 73*, 1–5.

Stapp, H.P. (1972). The Copenhagen interpretation. *American Journal of Physics, 40*, 1098–1116. [Reprinted by permission, this issue]

Stapp, H.P. (1993). *Mind, matter, and quantum mechanics*. Heidelberg: Springer–Verlag.

Stapp, H.P. (1997). Nonlocal character of quantum theory. *American Journal of Physics, 65*, 300–304.

Sutherland, K. (Ed.). (1995). Conversations with zombies. *Journal of Consciousness Studies, 2*(4), 312–72.

von Neumann, N. (1932). *Mathematical foundations of quantum mechanics* (Chapter VI). [English Translation 1955.] Princeton: Princeton University Press.

Wigner, E. (1962). Remarks on the mind–body problem. In I.J. Good (Ed.), *The scientist speculates* (pp. 171–184). New York: Basic Books.

Zurek, W.H. (1986). Reduction of the wave packet and environment-induced superselection. In D.M. Greenberger (Ed.), *New techniques and ideas in quantum measurement theory: 480. Annals of the New York Academy of Science* (p. 96). New York: New York Academy of Science.