Interpreting the Many Worlds Interpretation
Author(s): David Albert and Barry Loewer
Source: *Synthese*, Vol. 77, No. 2 (Nov., 1988), pp. 195-213
Published by: Springer
Stable URL: https://www.jstor.org/stable/20116589

DAVID ALBERT AND BARRY LOEWER

INTERPRETING THE MANY WORLDS INTERPRETATION

INTRODUCTION

Here is a partial list of the elementary principles of Quantum Theory:[1]

(I) Any isolated quantum mechanical system is characterized by a *state function* $\$_S(t)$.

(II) So long as S remains isolated, $\$_S(t)$ evolves deterministically in accordance with the Schrodinger equation (or in accordance with one of the relativistic generalizations of the Schrödinger equation).

(III) For any complete compatible set of observables O of S, $\$_S$ can always be expressed as a sum or a "superposition" of eigenstates of O, as follows:

(1) $\$_S = c_1 O_1 + c_2 O_2 + \cdots$

where the c_i are complex numbers, and the O_i represent quantum states (eigenstates of O) of S in which O has the particular value o_i such that if $i \neq j$ then $o_i \neq o_j$.

(IV) When a measurement of O is carried out on S in state $\$_S$ the probability of obtaining $O = o_k$ is equal to the absolute square of the amplitude c_k of O_k in the state function $\$$.

(V) When a measurement of O is carried out with the result that $O = o_k$ then the state of S "collapses," or is "reduced" instantaneously into the eigenstate O_k.

This list is sufficient to show why quantum theory is philosophically perplexing. The first perplexity is that quantum mechanical superpositions are not like any classical state. When textbook writers attempt to explain what it is for an electron to be in, for example, the spin state $c_1{\uparrow}_x + c_2{\downarrow}_x$ they are reduced to saying things like "it neither has x-spin up nor x-spin down but is in some sense in both states and in neither". ("${\uparrow}x$" represents the eigenstate of x-spin $= +\frac{1}{2}$(up) etc.) So one problem is to "interpret" superpositions. Part of an interpretation is given by IV which connects quantum states with the results of

Synthese **77** (1988) 195–213.

measurements. This brings us to the second difficulty, which is known as the "measurement problem". Suppose that M is a device for measuring O on S. This means that there is an observable Q with values q_i such that after M and S interact $P(O = o_j / Q = q_k) = \delta_{jk}$. (Thus M's measurements of O are perfect. We make this assumption throughout our discussion.) It follows from the linearity of the dynamical equations of quantum theory (as mentioned in II) that if the composite system M and S is isolated, then at the conclusion of the measurement interaction the final state of $M + S$ is given by:

(2) $$s_{M+S} = c_1 Q_{1M} \otimes O_{1S} + c_2 Q_{2M} \otimes O_{2S} + \cdots$$

where Q_{1M} are the eigenstates of Q with eigenvalues q_i.

The problem is that equation (2) entails that at the conclusion of the measuring process the measuring device is not characterized by an eigenstate of Q; rather, the final state of $M + S$ is a superposition of products of the eigenstates of Q and O. But the usual account of measurement (principles IV and V) stipulate that the state of M at the conclusion of the measurement will be some eigenstate of Q, and that the state of S at the conclusion of the measurement will be one of the eigenstates of O. The two treatments are flatly inconsistent with one another; and moreover there are, in principle, *physically testable differences* between the final states that these two different accounts entail.[2]

A great deal of ingenuity has been expended over the years in attempting to say precisely how and when collapses might occur, and in attempting somehow to reconcile principles I, II and IV, V. Typically, such accounts involve claiming that "measurements" do not conform to I and II but instead produce a collapse of the observable measured in accordance with V. We shall not pursue any of these lines of thought here except to remark that (as is well known) they all face very serious difficulties.[3] In particular, such accounts require a characterization of those interactions which are measurements and an account of how it is that measurements fail to satisfy the Schrodinger equation and instead instantaneously produce a collapse of the state of the system being measured.

THE MANY WORLDS INTERPRETATION

Here we want to discuss a response to the measurement problem called "the many-worlds" interpretation, first suggested by Everett and later developed by DeWitt, Wheeler, and others.[4] The fundamental idea of the many worlds interpretation is that measurements, and indeed all physical processes, must take place in accordance with the Schrodinger equation (or one of its relativistic variants). On this view, the state at the conclusion of a measurement process such as was described above *is*, the uncollapsed state described by the right hand side of (2).

The many worlds interpretation resolves the measurement problem simply by *denying* that measuring O on S when the state of S is a superposition of O-eigenstates brings about a collapse of the state-function of S into one of those O-eigenstates; that is, by denying principle V. (The status of principle IV within the many-worlds interpretation is, as shall presently be clear, a more complicated matter.) Consequently, there is simply no need, within this interpretation, to find an "explanation" of the collapse. One of the attractions of the many worlds interpretation is that since it does not require a division between measured system and measuring apparatus it is in principle possible to characterize the entire universe at any given time by a state function like (1), a "universal wave function", the time evolution of which is governed by a Schrödinger equation. Thus the account has special appeal to cosmologists who may want to consider the quantum state of the universe.[5]

Although it may "solve" the measurement problem, the many-worlds interpretation faces difficulties of its own. The most obvious is that it is difficult to square its claim that (in the measurement described above) the final, post measurement state of $M+S$ is *not* an eigenstate of Q & O but a superposition of such states. How can that be? That the pointer on a measuring device is at *some* particular position at the conclusion of a measurement is not, after all, a theoretical hypothesis which is open to question, bur rather (or so it would seem), a straightforward empirical *fact*. When we look at the pointer (look to see the value of Q) it manifestly has a particular position.[6] Moreover, when a human observer interacts with a system like $M+S$ in a state like the one on the right hand side of (2) by reading M's dial, then the Schrödinger equation (for the system Observer +M +S) entails that she (her brain, that is) evolves into a superposition of *belief* states concerning the position of the pointer! But what can *that* mean? We certainly do not experience ourselves as in superpositions of belief states, or as evolving into them, when we make observations!

There is another difficulty with the many-worlds claim that (2) correctly describes the measurement of O. In (1) c_k is (according to IV) the probability that a measurement of O on S will yield the result $O = o_k$. These probabilities are the points at which quantum theory makes contact with the outcomes of experiments; they are absolutely essential to our understanding of "superposition" and to the application and testing of the theory. But if the state at the end of a measuring process is with *certainty* the one in (2), then what sense can it make to say that when O is measured the *probability* of finding $O = o_k$ is c_k? And if such statements should turn out not to make any sense at all wouldn't that simply pull the rug out from under the theory?

In addition to making the claim that all physical processes are described by a Schrödinger equation, the many worlds interpretation is supposed to explain how that claim can be true while avoiding the problems just mentioned. It is just this which makes it an interpretation. We want to inquire here whether or not the interpretation can deal with these problems; but this task is complicated at the outset by the fact that it isn't entirely clear just what the many-worlds interpretation is. In an illuminating article, Richard Healy remarked that "the interpretation itself needs interpreting".[7] We very much agree with him.

THE SPLITTING WORLDS VIEW

So perhaps it will be best to start off by clearing up a certain confusion. There is a way of understanding the many worlds interpretation (a way one especially finds in the popular literature on quantum theory) which, we will argue, is unsatisfactory. On this interpretation, which we call "the splitting worlds view", (SWV) when a quantum measurement occurs, the measuring device and indeed the whole literally *splits* into two or more (depending on the number of possible outcomes of the measurement) worlds.[8]

The following passage by DeWitt suggests the SWV:

> The universe is constantly splitting into a stupendous number of branches, all resulting from the measurement like interactions between its myriads of components. Moreover, every quantum transition taking place on every star, in every galaxy, in every remote corner of the universe is splitting our local world on earth into myriads of copies of itself. (p. 161)

One way of understanding DeWitt's remarks is as asserting that the components of the universal state function of the universe at a given time themselves each represent a world at that time. The time-dependence of the universal state function is taken to characterize sequences of such worlds or world histories. If at t_o @ is such a world-component of the overall state-function, and if an interaction takes place within the world described by @, and if that interaction results in $@_1$ and $@_2$ being world-components at the completion of the interaction, then the SWV instructs us to conceive of the world corresponding to @ as having "split" into the two worlds, $@_1$ and $@_2$.[9]

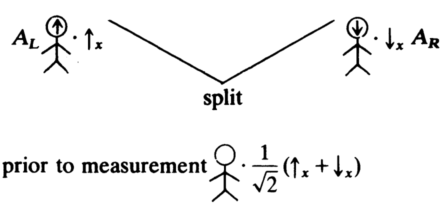
DeWitt says of this view:

> I still recall vividly the shock I experienced on first encountering this multiworld concept. The idea of 10^{100+} slightly imperfect copies of oneself all constantly splitting into further copies, which ultimately become unrecognizable, is not easy to reconcile with common sense.

DeWitt attempts to reconcile the SWV with common sense by arguing that "to the extent to which we can be regarded simply as automata and hence on a par with ordinary measuring apparatuses, the laws of quantum mechanics do not allow us to feel the splits" (p. 161). His reason for this claim is as follows: Suppose that an observer A measures the x-spin of an electron in state $c_1\uparrow_x + c_2\downarrow_x$. Assuming that A's measurements are perfectly accurate (i.e., $P(A$ believes spin $=$ /spin $=) = 1$) it follows that at the conclusion of the measurement $A +$ electron is in the state:

(3) $$c_1\uparrow_x \otimes \text{BEL}``\uparrow_x" + c_2\downarrow_x \otimes \text{BEL}``\downarrow_x"$$

(BEL"$\uparrow_x$" is the state of A's believing that the x-spin is up, etc.) On the SWV this is interpreted as meaning that when A makes the measurement the world splits into two worlds and that A *literally* splits into two successors, A_L and A_R. This is illustrated in the following diagram:



In one world (the one depicted on the left) A_L correctly believes that the state of the electron is $\uparrow_x$ (since x-spin is up in that world) and in the other world A_R correctly believes that the state of the electron is $\downarrow_x$. A_L doesn't know about the split since she only remembers making the measurement and obtaining the result that $x\text{-spin} = \uparrow_x$. ($A_R$ remembers making the measurement and obtaining $x\text{-spin} = \downarrow_x$.) Furthermore, it follows from the Schrödinger equation that subsequent measurements by A_L and others (same for A_R) will confirm the result obtained. So from A_L's point of view the state of the electron appears to be $\uparrow_x$. It *is* $\uparrow_x$ in her world. And she has no knowledge of her twin A_R and the result of her measurement.[10] Everett and DeWitt discovered, surprisingly, that even though the Schrödinger equation is inconsistent with collapses, it predicts that it will *appear* to observers (in the sense just explained) that collapses occur. On the SWV, splitting plays the role of a collapse by producing successors of measuring devices (including observers) which record unique values of measurements. It is this which leads one, and indeed led Everett in the first place, to think that there is something to the many worlds idea.

There remains the problem of making sense of the probability interpretation of the amplitudes on the SWV. The problem is a formidable one because the dynamical equations are completely deterministic. So how does probability enter the picture? DeWitt shows that if we confine our attention to situations in which, e.g., N independent measurements of the x-spin of electrons initially prepared in state $1/\sqrt{2}(\uparrow_x + \downarrow_x)$ are made then as N goes to infinity the amplitude of the sum of states in which sequence with a frequency different from 1/2 appear in the total superposition converges to 0. DeWitt suggests identifying 0 amplitude with 0 probability. It follows that the "probability" of a sequence with a frequency different from 1/2 is 0. DeWitt seems to think that this identification solves the difficulty concerning probability and remarks that the argument shows that "the conventional probability interpretation of quantum mechanics thus emerges from the formalism itself" (p. 163).

There are many things wrong with the SWV. Our criticisms will advance us toward an interpretation which, we will argue, is a more satisfactory way of understanding Everett's fundamental idea; that all physical processes are governed by the dynamical equations as stipulated by II. The first problem is the least subtle but perhaps the most destructive to the SWV. It is that the SWV is actually inconsistent with the dynamical equations. For example, according to the Schrödinger equation the total mass-energy of $M+S$ before and after the interaction are the same. This follows again from the linearity of the equations and the fact that in each of the states in the superposition the mass energy is the same before and after interaction. Similarly, the total number of particles comprising $M+S$ is conserved. But on the SWV a measurement literally results in an astronomical increase of the number of particles and of the total mass-energy of $M+S$. Of course it will be the case that this increase in the number of particles (or mass-energy) will go undetected by observers. But this in no way alleviates the main point that the splitting process is literally inconsistent with the dynamical equations and so cannot be taken as an interpretation in which all physical processes are described by them. No adequate many-worlds interpretation can countenance any such incompatibility with the dynamical equations.

The second difficulty concerns the understanding of probability within the SWV. Whatever merits DeWitt's argument for his claim that the probability interpretation emerges from the quantum mechanical formalism may have, it doesn't address the really difficult problem that the dynamical equations are deterministic.[11] Since, according to the SWV, it is certain that all outcomes of the measurement will occur and will be observed by successors of A, what can be meant by saying that the probability of a particular outcome $= c^2$? If probability is to be introduced into the picture, it must necessarily be by *adding* something to the interpretation. For example, we might say that some of the worlds into which (2) splits are more "actual" than others (or that one of the successors and not the other is *really* A). Then probability can be identified with the probability that the actual world will follow a particular branch (or that A will find himself in a particular world). The trouble with this suggestion is not only that it is mysterious (what distinguishes the more actual worlds from the less actual ones?), but also that it gives up the central feature of the SWV, that the state-function entirely exhausts what there is to be said about the physical world.

The third difficulty is called "the problem of the democracy of basis". It concerns the way in which the worlds of the SWV are specified. As stipulated by principle I the state of any system can be written as a linear combination of some complete set of *basis* vectors. According to the SWV the *worlds* present when the system is in this state correspond to these basis vectors. For example, if A measures the x-spin of an electron in the state $1/\sqrt{2}(\uparrow_x + \downarrow_x)$ then the post-measurement state of A + electron will be

(4) $$\frac{1}{\sqrt{2}}(\uparrow_x \otimes \text{BEL``}\uparrow_x\text{''} + \downarrow_x \otimes \text{BEL``}\downarrow_x\text{''})$$

wherein there are, according to the SWV, two worlds, one in which A believes that spin is up and the spin is up and one in which A believes that spin is down and the spin is down. The problem is that there are an *infinity* of distinct complete sets of basis vectors in terms of which (4) (and any such state) can be written. For example, (4) can also be written thus:

(5) $$\frac{1}{\sqrt{2}}(((\rightarrow_A) \otimes \rightarrow_x) + ((\leftarrow_A) \otimes \leftarrow_x))$$

where

$$(\rightarrow_A) = \frac{1}{\sqrt{2}}(\text{BEL}(\uparrow_x) + \text{BEL}(\downarrow_x)),$$

$$(\leftarrow_A) = \frac{1}{\sqrt{2}}(\text{BEL}(\uparrow_x) - \text{BEL}(\downarrow_x)),$$

$$\rightarrow_x = \frac{1}{\sqrt{2}}(\uparrow_x + \downarrow_x),$$

$$\leftarrow_x = \frac{1}{\sqrt{2}}(\uparrow_x - \downarrow_x).$$

In (5) the basis vectors, and consequently the *worlds*, are $\rightarrow_A \otimes \rightarrow_x$ and $\leftarrow_A \otimes \leftarrow_x$. These are worlds in which the x-spin has no definite value and A has no definite belief about the value of x-spin. This splitting is certainly not what the proponents of the SWV have in mind. Rather, they suppose that the lines along which the world splits when A measures the x-spin are the ones that correspond to (4) and not the ones that correspond to (5). At least in those situations involving observers, they seem to suppose that the splitting process is along lines determined by brain states associated with definite beliefs about the measured system. The problem is that in quantum theory the choice of a set of basis vectors in which to express a state has absolutely no physical significance. So there is nothing in quantum theory itself that supports the representation (4) over (5). The MWV must add something if it is to support its position that the worlds split along a preferred basis.

The upshot of our discussion is that the SWV is certainly not an adequate interpretation of the quantum mechanical formalism. The postulated splitting is not fully characterized (the democracy of basis problem); it is incompatible with the dynamical equations it is supposed to interpret (the conservation of mass problem), and it leaves us completely in the dark about the meaning of probability claims (the determinism problem).

THE MANY MINDS VIEW

The fundamental idea of the many worlds interpretation is *that all physical processes whatsoever are governed by the Schrödinger equation*. We will construct an interpretation which entails precisely this while avoiding the problems that we found in the SWV. Our procedure will be to build up to our interpretation in stages, motivating its features by considering solutions to problems that face the SWV.

As we saw, the main difficulty that must be overcome to implement the fundamental idea is that this idea entails that macroscopic measuring devices and indeed observers themselves can be in superpositions. What we want is an "interpretation" which explains how it is that we always "see" (mistakenly so, if the many worlds interpretation is correct) macroscopic objects as not being in superpositions and never experience ourselves as in superpositions. The heart of the problem is that the way we conceive of mental states, beliefs, memories, etc., it simply makes no sense to speak of such states or of a mind as being in a superposition. When we introspect following an x-spin measurement we never, as apparently predicted by the theory, find ourselves in a superposition of thinking that spin is up and thinking that spin is down. If introspection is to be trusted, and it seems part of our very concept of mental states that it is trustworthy at least to this extent, then we are never in such superpositions.

Let's make this point sharper. Suppose:

(i) A is an observer who can perfectly measure x-spin. (That is: subsequent to an x-spin measurement by $M\uparrow x$ iff M believes that $\uparrow_x$ and $\downarrow_x$ iff M believes that $\downarrow_x$.)

(ii) A can correctly report some of her mental states. Specifically, when A sincerely reports that she has a definite belief about the value of x-spin (i.e., she reports: "It is either the case that I believe that spin is up, or it is the case that I believe that spin is down, but it is not the case that my beliefs about the value of x-spin are in any way uncertain, or ill-defined, or superposed"), then A does believe that the x-spin has a definite value.

(iii) The state wherein A believes that $\uparrow_x$ and the state wherein A believes that $\downarrow_x$ are identical with certain physical states of A's brain. We will call those states $B\uparrow$ and $B\downarrow$.

It turns out (and this is the sharper form of our problem) that (iii) is inconsistent with (i) and (ii) and quantum mechanics. The argument goes like this: let A measure the x-spin of an electron which is initially in the state $c_1\uparrow_x + c_2\downarrow_x$. Subsequent to that measurement, if quantum theory and supposition (i) are correct, the state of A + electron is

(8) $$c_1 B\uparrow_x \otimes \uparrow_x + c_2 B\downarrow_x \otimes \downarrow_x.$$

It follows from supposition (ii) that A will answer the question "Do you believe that the x-spin of the electron has a definite value?" affirmatively whenever either of the states $B\uparrow_x \otimes \uparrow_x$ or $B\downarrow_x \otimes \downarrow_x$ obtains. And now it follows from the linearity of the quantum mechanical equations of motion, that A will answer this question affirmatively whenever *any* linear superposition of $B\downarrow_x \otimes \uparrow_x$ and $B\downarrow_x \otimes \downarrow_x$ obtains. So, in particular, it follows from (i) and (ii) and quantum mechanics that when state (8) obtains, A will affirm that she believes that the x-spin of the electron has a definite value, either up or down. But if supposition (iii) is true then when (8) obtains it is not the case that A believes that x-spin is up, and it is not the case that A believes that x-spin is down since A's brain is in a superposition of states. So we have derived a contradiction from (i), (ii), (iii), and quantum theory.

If we are to maintain the fundamental idea of the many worlds interpretation and also maintain our ordinary conception of mental states as accessible to introspection then we must give up (iii). Because

(iii) identifies mental states (e.g., the state of believing that spin is up) with physical states it results, given quantum theory and (i), in the conclusion that belief states can be in superpositions. If we deny that belief states can be in superpositions (as (ii) seems to require us to do), then we are denying (iii). This response together with the assumption that all physical states are quantum mechanical commits us to a modest (so far) non-physicalism. This non-physicalism is inevitable in the present context. As long as we maintain both that we have introspective access to our mental states (i.e., a person who says he believes that the spin has a definite value is, other things being equal, correct concerning his belief) and that all physical states are quantum mechanical, our argument shows that we must accept this much non-physicalism.

Here is a way that we can give up (iii). Consider A again about to measure the x-spin of an electron. The electron is initially in state $c_1\uparrow_x + c_2\downarrow_x$ and A is initially in a state of readiness. At the conclusion of the measurement A + electron is the state described by (8). We will postulate that the evolution of A's mental state in the course of the measurement is not deterministic (as is the evolution of A's body + electron) but is ***probabilistic***. A's mind starts out in a state in which she has no beliefs about x-spin and ends up either in the state corresponding to $B\uparrow_x$ or in the state corresponding to $B\downarrow_x$ (that is, ends up either believing that spin is up or believing that spin is down) but never is in a superposition of belief states. It is postulated, and this is a principle that is added to quantum theory, that

(P) the probability that A's mind ends up believing that spin is up $= c_1^2$ and the probability that A's mind ends up believing that spin is down $= c_2^2$.

(P) generalizes to any perfect measurement by an observer.

This view, which we call the single mind view (SMV), solves some of the problems that confront the SWV. On the SMV the entire physical world is a thoroughly quantum mechanical system as described by principles (I) and (II). Principle (IV) is satisfied by interpreting quantum mechanical probabilities as the probabilities that an observer (a mind) will have certain beliefs after making a measurement. The world does not split upon measurement; rather, the mind associated with a brain ends up being in the mental state associated with one of the brain states in the superposition of B states that describes her brain. The probability that the mind will end up in a particular state is completely determined by the physical state of the observer + system measured. Everett's and DeWitt's original argument to explain why systems appear to collapse on measurement is easily adapted by the SWV. After measurement the minds assocaited with the left hand part of (6) will all believe that x-spin is up. Further measurements by the observer will result in "confirmation" since the brain states that are part of the overall state all perceive the same value. There is no worry about the fact that we never feel ourselves to be in or introspect superpositions since mental states are never in superpositions. An observer's sincere reports about her beliefs in circumstances like (8) are true. She correctly reports that she has a definite belief about spin because she does in fact believe it to be either up or believe it to be down.

Of course the startling feature of the SMV is its non-physicalism. On the SMV, all but one of the elements of a superposition like (8) represent, as it were, mindless brains; and which element represents a mind is not determined by the physical nature of the underlying brain state and cannot be deduced from the quantum state or from any physical experiment. It is clear from the dilemma generated by (i), (ii), and (iii) that *any* many worlds interpretation which respects (ii) will be committed to some form of non-physicalism, but the non-physicalism of the SMV seems especially pernicious. It entails that mental states do not even *supervene* on brain states (or physical states generally) since one cannot tell from the state of a brain what its single mind believes.

Here is a way to remedy the situation. Suppose that every sentient physical system, every observer, has associated with it not a single mind but rather an infinite set of minds. (The reason for the set's being infinite will soon be apparent.) We will call this the "many minds view" (MMV). We suppose that when an individual's brain (or brain + environment) is in state B_k^* the individual is in a certain mental state M_k, i.e., has certain beliefs, intentions, memories, experiences, etc. This supposition is *physicalistic* in that it says that mental states are determined by or supervenient on brain (or brain + environment) states. The B_k^* do not form a complete basis but they are a subset of one. So any brain can be represented as in a state:

$$(9) \qquad B = c_1 B_1^* + c_2 B_2 + c_3 B_3^* + \cdots$$

where the B_k^* are the states associated with mental states and the unstarred Bs represent states not associated with mental states. Now our proposal is to associate with each B_K^* in (9) an infinite set of minds in the corresponding mental state M_K and a measure P on the totality of minds associated with all the B^* such that the P of the set of minds in state $M_k = c_k^2$. P is a measure of the "proportion" of minds in state M_k. Finally, each individual mind is supposed to evolve probabilistically in accordance with postulate (P).

The individual minds, as on the SMV, are not quantum mechanical systems; they are never in superpositions. This is what is meant by saying that they are non-physical. The time evolution of each of the minds on the MMV is, just as on the SMV, probabilistic. However, unlike the SMV, there are enough minds associated with the brain initially so that minds will end up associated with each of the elements of the final superposition. An infinity of minds is required since a measurement or a sequence of measurements may have an infinite number of outcomes. Furthermore, although the evolution of individual minds is probabilistic, the evolution of the set of minds associated with B is deterministic since the evolution of the measurement process is deterministic and we can read off from the final state the proportions of the minds in various mental states.

Here is how the account works in the example of x-spin measurement. Suppose an observer is about to measure the x-spin of an electron is in a brain state B wherein no mind has a belief that the value of x-spin is up (or that it is down) and that the electron is in state $c_1\uparrow_x + c_2\downarrow_x$: at the completion of the measurement the observer + electron is in the familiar state

$$(10) \qquad c_1 B\uparrow_x \otimes \uparrow_x + c_2 B\downarrow \otimes \downarrow_x$$

wherein the proportion of minds that have observed spin up (down) is c_1^2 (c_2^2). Before making the measurement, each observer's mind can know that (8) will be the post-measurement state and so can assign a probability of c_1^2 to *its* observing spin up.

We have purchased supervenience of the mental on the physical at the cost of postulating an infinity of minds associated with each sentient being. No doubt this talk of infinitely many minds sounds *crazy*. Perhaps it will seem less so after we see what it does for the interpretation of quantum mechanics and how it compares with other interpretations.

First, and most important, the MMV (unlike the SWV) is in accord with the fundamental idea of the many worlds interpretation—that the entire physical universe, and every physical system, is quantum mechanical in the sense of principles I and II. There is no need to postulate collapses or splits or any other non-quantum mechanical *physical* phenomena. And so there arises no conflict with conservation laws as we saw on the SWV.

Second, the MMV entails that the time-evolution of the whole physical world is completely deterministic, and that the "global mental state" of every sentient physical being (that is: the *distribution* of mental states among the infinity of that being's minds) is uniquely fixed by the physical state of that being. Unlike the abandoned SMV, the global mental state is unambiguously determined by its physical state and consequently the time-evolution of the global state is, likewise, deterministic.

Third, the MMV is in accord with our very deep conviction that mental states never superpose; consequently it is in accordance with the claim that competent sentient beings can accurately report their mental states.

Fourth, the MMV (unlike the SWV) entails that the choice of basis vectors in terms of which the state of the world is expressed has no physical significance. There is always but *one* physical world in but *one* quantum mechanical state on this account; and that state can be equally well written in terms of any complete set of basis vectors. As long as a brain is in a state which *can* be represented as a superposition of B states it will have minds associated with it.

Fifth, there is no special problem on the MMV of interpreting probability. Probabilities are completely objective although they do not refer to physical events but always to sequences of states of individual minds. The assertion, "the probability that A will obtain result $\uparrow_x$ when measuring the x-spin of an electron in state $c_1\uparrow_x + c_2\downarrow_x = c_1^2$" means simply that the probability that a mind associated with A will observe that x-spin is up $= c_1^2$ (recall that A's measurement is "perfect.") Although probabilities do not emerge from the quantum mechanical formalism we can adapt DeWitt's argument to show from our postulate connecting the proportion of minds with amplitudes that the probability that a mind will believe that it resides in a "maverick" world (i.e., has beliefs different from those in accord with quantum mechanical probabilities) converges to 0 (but "maverick minds" will still be present).

Sixth, the account is *realist* in the sense that it entails that there is a uniquely correct state for the whole universe and in the sense that it does not suppose that the state of the universe in any way depends on a consciousness or on what observables an observer decides to measure. In this it contrasts markedly with some "idealist" interpretations which entail that consciousness, by bringing about a collapse or in choosing to measure certain observables, in some mysterious way makes reality (perhaps different realities for different observers).[11] This realism, however, does have the consequence that a mind's beliefs about the state of a system after measurement are typically false. Thus, a mind associated with A that measures the x-spin of an electron in a superposition will at the conclusion of the measurement believe, say, that the x-spin is up (of course some of A's other minds will believe that spin is down). In fact, spin is neither up nor down but rather the system A's brain + electron (and of course the intermediary measuring devices, etc.) is in a superposition. So A's belief is strictly incorrect. However, it is, we might say "pragmatically correct", in the sense that subsequent measurements of the x-spin by A will, from the perspective of that mind, yield results which agree with its initial measurement.[12]

Not only is the MMV a *realist* interpretation of quantum mechanics, it also provides an account in which all interactions are *local*. This is surprising because Bell's theorem is widely thought to rule out realist, local interpretation of quantum mechanics.[12] To see how the MMV allows for a local realist interpretation without violating Bell's theorem, consider the sort of situation that Bell considers in proving his theorem: A system consisting of two electrons is in the EPR state

$$\text{(EPR)}\quad \frac{1}{\sqrt{2}}(\uparrow_1 \otimes \downarrow_2 - \downarrow_1 \otimes \uparrow_2).$$

The electrons separate to distinct points 1 and 2 and an observer A_1 measures a component of the spin of electron 1 and an observer A_2 measures a component (not necessarily the same component) of the spin of electron 2. Bell's theorem says that there is no consistent description of the system which (a) assigns a state to a system in which the measured components have definite values (realism) and (b) in which the value obtained by one observer is independent of which component of spin the other observer decides to measure (locality), and which yields the correlations predicted by quantum theory for a system in EPR. The MMV denies Bell's version of realism since according to it the *real* state of the electron + observer after a measurement of spin on the system is still in a superposition. More specifically, when A_1 measures the x-spin of electron 1 his brain will go into a superposition correlated with the EPR state. In this superposition, half his minds will believe that spin is up and half will believe that spin is down. Similarly for A_2's measurement on electron 2. In fact the MMV entails that at the conclusion of A_1's and A_2's measurements, no matter which two spin components they may have measured (even if they measure different components), 1/2 of A_1's minds will believe that the result of the measurement *he* carried out was "up" and 1/2 of his minds will believe that the result was "down". Precisely the same holds for A_2. No "correlations" will have arisen at this point between the results recorded in A_1's minds and those recorded in A_2's minds, and indeed no talk of correlations is even intelligible. Those correlations come into being only when the observers communicate the results of their respective measurements to one another. This communication is, of course, a *local* dynamical process governed by the Schrödinger equation. It is then guaranteed that if A_1 and A_2 measured the same spin component they will end up (after communication) believing that they obtained opposite values. If they measured different spin components, say x and $x + \theta$, then the probability that they will end up believing that they obtained opposite values is $\cos^2 \theta$. These are the usual quantum mechanical probabilities predicted by the theory. In this way the MMV permits a *realist*, since the system $A_1 + A_2 + \text{electron}_1 + \text{electron}_2$ is always in some particular state, and *local*, since interactions are all local, description.

The preceding list seems to us sufficiently impressive to give the MMV a hearing in spite of its dualist commitments. In general it can be said that unlike other accounts it allows one to tell completely explicit and intelligible stories, without employing non-classical logics or mysterious collapses, about any quantum mechanical process, including measurement; stories which are entirely consistent with principles I and II and hence are locally realistic. At the same time the account preserves our ordinary (and perhaps a priori) conception of mental states as accessible to introspection. The same cannot be said for any other "interpretation". Of course, what allows for all of this is what must be admitted to be an extravagant dualism. One may wonder whether it may be possible to get away with less dualist commitments.

One assumption, perhaps the most disturbing aspect of the dualism, is that there is a matter of fact concerning whether or not a mind associated with a brain at one time is the *same* mind as one associated with the brain at a different time. This assumption violates the supervenience of the mental on the physical since the evolution of the physical state of the universe does not determine such identities. (On the other hand, as we previously saw, the proportion of minds in various mental states associated with a brain at a given time does supervene on the physical state of the brain, or brain plus environment.) We could effect a partial retreat from this dualism by postulating that associated with a brain at any one time is a set of "momentary minds" while not postulating that there is a matter of fact that a mind (associated with a particular brain) at one moment is trans-temporal identical with a mind (associated with the same brain) at another moment. However, the cost of surrendering the "trans-temporal identity of minds" would seem to be that we can no longer make sense of statements like "the probability that *I* will observe spin up on measurement is *p*" since such statements seem to presuppose that it makes sense to talk of a single mind persisting through time. Perhaps, however, it could be argued (though we will not do so here) that the conception of a mind persisting through time is an illusion; one that results from the fact that "most" momentary minds will be associated with brain states which record a "personal" history and so it will seem to such a mind that it has existed as a persisting entity.

Could we go further and eliminate all reference to minds? Such an eliminativism would entail a radically different way of talking about quantum theory and about ourselves. To do so would, since quantum theory is completely deterministic, give up any way of making sense of quantum mechanical probabilities. Further, on such an account there would be no mental states as we ordinarily understand them. There will still be states, the brain states B^*, which register information and so may resemble mental states, but these states will not be introspectible in the way that (iii) requires. Whether it might be possible for us to learn such a language, that is, to learn to describe the world entirely in quantum mechanical terms, is a question that we must leave for another occasion.[13]

NOTES

* The authors would like to thank Yakir Aharanov and Davis Baird for discussions of the many worlds interpretation.

[1] For example, J. von Neumann: 1955, *Mathematical Foundations of Quantum Mechanics*, Princeton.

[2] When $M+S$ is in the state described by (2) there is an observable R which is in an eigenstate R_k. On the other hand if $M+S$ is in one of the eigenstates in the superposition (2) then R is in a superposition of R_i states. So in principle one can tell the difference between the two post measurement states by measuring R. If the state is (2) we will obtain the same value every time we measure R on a system prepared as $M+S$ is. If the state is one of the eigenstates of (2) we will obtain various values for R. However, the measurement of (R) is beyond the means of current technologies because R is very unstable; for example it would be disturbed if $M+S$ interacted with a single air molecule. Some authors appeal to the inability to determine whether the state of $M+S$ is (2) or one of its eigenstates to argue that the measurement problem is of no import. But this just misses the point. The two states are different.

[3] A discussion of various attempts to deal with the measurement problem can be found in B. d'Espagnat: 1971, *Conceptual Foundations of Quantum Mechanics*, Benjamin.

[4] The main papers on the Many Worlds Interpretation are collected in DeWitt and Graham (eds.): 1973, *The Many Worlds Interpretation of Quantum Mechanics*, Princeton University. Page references are to DeWitt's papers in this volume.

[5] According to the orthodox interpretation every measurement presupposes a distinction between the system being measured (which is described quantum mechanically) and the measuring apparatus (which is described classically). On this account it is difficult to see how quantum theory can be applied to the whole universe. In contrast the many worlds interpretation allows one to speak of the quantum state of the entire universe and so seems more suited for cosmological applications. For an example of such an application see D. Albert, 'Possibility that the Present Quantum State of the Universe is the Vacuum, forthcoming.

[6] It is not clear what it would be to see a macroscopic observable in a superposition although some authors imagine that it is like seeing a blur.

[7] Richard Healy: 1984, 'How Many Worlds', *Nous* **18**.

[8] Heinz Pagels: 1984, *The Cosmic Code*, Bantam, seems to understand the many worlds interpretations as the SWV. Bryce DeWitt in his papers in the book footnoted in 4 also understands the interpretation this way. However, Everrett's views are much more ambiguous, at times suggesting aspects of the view we later advance.

[9] The worlds of the SWV may remind one of the "possible worlds" which are discussed in philosophical logic. They differ at least in two respects. All the worlds of the SWV are equally real and none are abstract. In this it differs from the common view that possible worlds are abstract entities of some kind, sets of propositions or properties. It is more like David Lewis' realism about possible worlds except that on the SWV worlds split. Second, their worlds of the SWV are "quantum" worlds in that some observables are in superpositions at a world.

[10] For an observer to discover that she is in a superposition like (7) involves her making a *self-measurement* on an observable which is incompatible with state (7). It follows from the quantum mechanical formalism that such an observable exists. See Albert, David 'Self-Measurement' *Physics Letters*, Vol. 98A and 'A Quantum Mechanical Automaton', ***Philosophy of Science***, **forthcoming.**

[11] DeWitt's argument falls short of demonstrating that the probability interpretation emerges from the formalism. In order to secure a probability interpretation he must

identify an amplitude converging on 0 and a probability converging on 0. This is a substantial assumption, and no part of the quantum mechanical formalism. Indeed, DeWitt's argument, as he himself seems to recognize, is circular. At one point he suggests that "maverick worlds" (worlds with the "wrong" frequencies) are "simply absent for the grand superposition." That is, not only is the amplitude (probability) of such worlds 0 but they are physically impossible. However, this suggestion in conflict with the guiding idea of the many worlds interpretation that the Schrodinger equation describes all physical processes, since according to it here will be elements of the overall superposition which correspond to maverick worlds. There is no escape from the fact that any sequence of measurements such as was described above will invariably produce maverick world histories (e.g., some in which the results of every x-spin measurement is $\downarrow_x$).

[12] Hilary Putnam has put forward a view according to which quantum mechanical entities "are real ... but they are relative to an observer". See 1983, Vol. III of *Philosophical Papers*, Cambridge.

[13] J. S. Bell: 1966, *Rev. Mod. Phys* **38**.

[14] Loewer and Albert continue this speculation in 'The Epistemology and Metaphysics of the Many Worlds View', in preparation.

Department of Philosophy
Columbia University
New York, NY 10027

Department of Philosophy
Rutgers University
New Brunswick, NJ 08903