Brit. J. Phil. Sci. **47** (1996), 159–188

SYMPOSIUM

'Many Minds' Interpretations of Quantum Mechanics

Michael Lockwood

> Whatever may be the truth about the universe, it is bound to be astonishing.
>
> (Bertrand Russell)

> The only 'failure' of quantum theory is its inability to provide a natural framework for our prejudices about the workings of the universe.
>
> (Wojciech H. Zurek)

Seventy years after the discovery of modern quantum mechanics, there is still no consensus as to how the theory should be understood. The philosophies of Niels Bohr and of the logical positivists, which once served to deter physicists and philosophers from asking embarrassing questions about the observer-independent *reality* underlying the observational predictions of quantum mechanics, have long ceased to satisfy. Yet the foundational work of John Bell and others has now revealed that any realist construal of quantum mechanics, if it is to reproduce the predictions of the conventional theory, must inevitably conflict with deeply rooted and intuitively appealing principles of classical physics or of common sense. There are no conservative options.

In quantum mechanics, possible states of a physical system are represented by vectors in an abstract space called *Hilbert space*. These vectors can be multiplied by numbers, known as *coefficients*, and added together (by way of a generalization of the familiar *parallelogram law*) so as to yield new vectors which then stand for states which are said to be *superpositions* of the original states. It is a feature of quantum mechanics that *any* such *linear combination* of *any* set of distinct state vectors of a physical system corresponds to a possible state of this system. This is known as the *superposition principle*. From a classical perspective it is a bizarre proposition. Consider, for example, a set of distinct states of the gears of my car. The superposition principle implies that, since being in first gear is a state, and so are being in second, third, and reverse gear, there is a possible state of my car in which it is, say, $\frac{1}{2}$ in first gear plus $\frac{1}{2}$ in second, $\frac{1}{2}$ in third, and $\frac{1}{2}$ in reverse gear. This is a state which appears to defy interpretation in common-sense terms; it certainly does *not* mean that the car is in neutral! The implication would be, however, that when I inspect the car to see which gear it is in, I am equally likely to observe it as being in first, second, third, or reverse. (The precise significance of the '$\frac{1}{2}$' will emerge shortly.) To what extent quantum mechanics should be regarded as being applicable, in this fashion, to the macroscopic world, is a question which lies at the heart of this article.

The Alice-in-Wonderland logic of quantum mechanics can be illustrated, relatively simply, by the phenomenon of electron *spin*. Spin is an intrinsic angular momentum which, like ordinary angular momentum, can be oriented in any spatial direction. Think of the electron as being pierced by an arrow aligned with the axis of spin. Then the electron is said to be *spin-up* if, as viewed by an imaginary observer looking along the arrow from its tail, the spin is clockwise, and *spin-down* if it is anticlockwise.

Suppose we have a pair of oppositely oriented spin states: for example, spin-up and spin-down in the z direction. The vectors representing any such pair of states will be mutually *orthogonal*, or 'at right angles' to each other, in the corresponding (two-dimensional) Hilbert space, and constitute a *basis* for this space. This means that any arbitrary spin state can be expressed as a superposition of these two states. For example, spin-up in the x direction can be represented as $1/\sqrt{2}$ times z spin-up, plus $1/\sqrt{2}$ times z spin-down. If x spin-up, z spin-up, and z spin-down are represented, respectively, by the *Dirac kets*, $|\uparrow_x\rangle$, $|\uparrow_z\rangle$ and $|\downarrow_z\rangle$, then we can write this as:

$$|\uparrow_x\rangle = 1/\sqrt{2}|\uparrow_z\rangle + 1/\sqrt{2}|\downarrow_z\rangle. \tag{1}$$

If an electron is spin-up in a given direction, and we *measure* the component of spin angular momentum in that same direction, then we are bound to get the result $\frac{1}{2}\hbar$ (where $\hbar$, or *h-bar*, is *Planck's constant*, h, divided by 2π). Similarly, if the electron is spin-down, in this direction, we are bound to get the result $-\frac{1}{2}\hbar$. For that reason the results $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ are themselves referred to as spin-up and spin-down, respectively. If spin behaved like classical angular momentum, then we should expect a spin measurement in a direction which was neither identical nor opposite to that of the spin *state*, to yield some result *intermediate* in value between $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$. Not so, however. Amazingly, we are guaranteed, still, to get either $\frac{1}{2}\hbar$ (spin-up) or $-\frac{1}{2}\hbar$ (spin-down), with probabilities which can be determined by expressing the electron's spin state as a superposition of the states spin-up and spin-down in the direction of measurement. This is the *quantization* that gives quantum mechanics its name. If, for example, we are measuring, in the z direction, the spin of an electron which is initially in the state x spin-up, then (1) above tells us that there is a fifty–fifty chance of our getting spin-up or spin-down; this corresponds to the fact that the squares of the two coefficients are each $\frac{1}{2}$. In general, the coefficients are *complex* numbers; and the corresponding probabilities are equal to their *square magnitudes*.[1]

This illustrates one aspect of the quantum-mechanical *statistical algorithm* (or *Born rule*). Spin, as measured in some specific direction, is an example of a quantum-mechanical *observable*, something one can measure or observe. And $\frac{1}{2}\hbar$ and $-\frac{1}{2}\hbar$ are its so-called *eigenvalues*: the possible *results* of the measurement or observation. Associated with these two results are, as we have seen, those possible spin-up and spin-down *states* which, were the electron to be in them, would *ensure* that the measurement yielded the results in question. These are known as the *eigenstates* of this spin observable. It is true quite generally that the probability of getting a given result or eigenvalue, when measuring an observable (or at least, a so-called *complete*[2] observable), is the square magnitude of the coefficient attached to the corresponding eigenstate, when the state of the system is expressed as a superposition of the eigenstates of this observable.

A second aspect of the statistical algorithm is this: if one measures a given (complete) observable on a quantum-mechanical system, and the measurement yields a given eigenvalue, then one is entitled, for most practical purposes, to regard the system as having been *projected into* the corresponding eigenstate.

This too is puzzling, on the face of it. The most straightforward way of understanding it is that the effect of measurement is *indeed* to effect a physical transformation of the system, whereby it is carried into the eigenstate corresponding to the measured value. This is referred to as *state-vector reduction* or the *collapse of the wave-function*. But this *collapse* interpretation of measurement ostensibly conflicts with the predictions of quantum mechanics, since any such physical transformation would represent a departure from the so-called *unitary evolution* prescribed by the *Schrödinger equation*. From the standpoint of physics, after all, measurement is just a certain kind of physical *interaction*; as such, it can hardly be subject to different rules from other interactions. So perhaps we should *change* the rules: modify the Schrödinger equation in such a way as to provide for such dynamical collapse. Well, there have been a number of ingenious suggestions as to how this might be done; but no one, as yet, has succeeded in coming up with a really satisfactory proposal.

---

[1] The square magnitude (or *square modulus*) of a complex number of the form $a+bi$ is a^2+b^2.

[2] A complete observable is one which, when measured on a given system, yields *maximal* information about the system's state: information which cannot be enhanced by any further measurements.

There is, however, another possible line one can take here, and that is the so-called *hidden variable* approach, exemplified by the Bohm theory (Bohm [1952]). Hidden variable theorists take the 'reduction' of the state-vector to be essentially *epistemic*: a reduction, merely, of our *ignorance* about the prior state. The result of any measurement, so they claim, is determined by what is antecedently true of the system being measured; and the fact that these results stand in a probabilistic relation to the state-vector description is due to the *incompleteness* of this description. The hidden variables are whatever might serve to flesh out the description, by providing a specification of those alleged *additional* features of the state, which the state-vector leaves out of account. This approach, as we shall now see, has its own problems, which arise from what is known as quantum *entanglement*.

Once again, we can use spin to illustrate this phenomenon. Certain interactions have the effect of generating pairs of electrons in an entangled state known as the *singlet state*. Each electron, in a system consisting of two electrons, will have its own Hilbert space for spin; but so also will the *composite* system associated with the *overall* spin-state of the electrons, taken together. Let $|\uparrow_{z1}\rangle$ and $|\downarrow_{z1}\rangle$ be the states z spin-up and z spin-down of electron 1, and $|\uparrow_{z2}\rangle$ and $|\downarrow_{z2}\rangle$ be the states z spin-up and z spin-down of electron 2. Suppose, first, that we wanted to represent a state of the electrons, in their joint Hilbert space, where their spins are oppositely aligned in the z direction. Then we could do so as follows:

(2) $$|\uparrow_{z1}\rangle|\downarrow_{z2}\rangle.$$

This is known as the *tensor product* of the two states $|\uparrow_{z1}\rangle$ and $1\downarrow_{z2}\rangle$. Saying that the state of the two electrons is (2), in their *joint* Hilbert space, is equivalent to saying that electron 1 and electron 2 are in the states z spin-up and z spin-down, in their *individual* Hilbert spaces.

But now recall the superposition principle, and consider the singlet state, which can be represented as

(3) $$1/\sqrt{2}|\uparrow_{z1}\rangle|\downarrow_{z2}\rangle - 1/\sqrt{2}|\downarrow_{z1}\rangle|\uparrow_{z2}\rangle.$$

The key point about (3) is that it depicts a state which is *irreducibly* a state of the two electrons considered as a *composite* system; it cannot be equated with *any* combination of spin states of the two electrons, considered individually, which would always be a product of the form $|\psi_1\rangle|\psi_2\rangle$. We have here an example of what is sometimes referred to as *quantum holism*, in which the whole transcends the sum of its parts. It is not merely that the state (3) cannot be *reduced* to the states of the individual electrons. These two electrons do not even *possess* individual spin states, if by 'state' one means a *pure* state, one that corresponds to a particular spin direction. But we have what is known as a quantum *correlation* of the two systems; in the singlet state each possible spin-up state of electron 1 is said to be *correlated* with the corresponding spin-down state of electron 2, and conversely. Entanglement, yet again, is a phenomenon with no classical antecedent; and it is *ubiquitous* in quantum mechanics. Virtually any interaction will result in an entanglement of some kind.

Imagine, now, that a pair of electrons is prepared in the singlet state, and that the electrons then fly apart, arriving at locations where their respective spins are measured by two observers, the proverbial Alice and Bob, in certain chosen directions. In whatever direction a given observer measures the spin, there is a fifty–fifty chance of obtaining the value spin-up or spin-down. This reflects the *spherical symmetry* of the state (3), which is somewhat obscured by its having been expressed in terms of z spin-up and z spin-down. In fact, one can replace the z, here, with x, y or indeed *any* spatial axis, and end up with an expression mathematically equivalent to (3). But now consider the probability that Alice and Bob will get the same result: both getting spin-up or both getting spin-down. This turns out to be equal to $\frac{1}{2}(1 - \cos\theta)$, where θ is the angle between the two directions in which they measure the spin. It follows that, when the directions are the same, so that $\theta = 0°$, the probability of Alice and Bob getting the same result is zero. When $\theta = 90°$, however, the probability of their getting the same result is 0.5; and when $\theta = 120°$, the probability is 0.75.

Suppose (following Mermin [1985]) that we have a sequence of pairs of electrons in the singlet state which fly off in opposite directions, and then have their spins measured by Alice and Bob respectively. We assume that the allowed directions of measurement consist of just three directions, which lie in the same plane at 120° to each other. For every successive measurement, Alice and Bob independently select one of these directions on some random basis. (For example, they may each roll a die, having previously associated the three directions of measurement with the pairs of numbers, 1 and 2, 3 and 4, and 5 and 6, respectively.) What I have just been saying then implies (a) that whenever Alice and Bob happen to measure the spin in the same direction, they are bound to get different results, and (b) that for a sufficiently long sequence, agreements and disagreements between the results obtained should occur with approximately equal frequency.

Think of the two electrons as like a pair of individuals, John and Susan. They are told that they will be placed in separate rooms, and repeatedly and simultaneously asked certain yes–no questions—each of which will be selected randomly from a set of just three, this being done over and over again, and independently for the two individuals. John and Susan are also told in advance what the three questions will be, and are challenged to devise an agreed strategy which, when they are in their separate rooms and hence unable to confer, will ensure (a) that whenever they are simultaneously asked the same question, they will invariably give opposite answers, and (b) that, in the long run, they will tend to give the same, and different, answers with equal frequency. It is not difficult to show (though I leave this as an exercise for the reader) that the challenge is one which it is logically impossible for John and Susan to meet: there is simply no such strategy.

The analogous conclusion, for a hidden variable construal of measurement, is that the hidden variables associated with the two electrons must be *non-local* in character in order to explain all the theoretically predicted correlations between spin measurements. It is in principle impossible to reproduce all these correlations by ascribing *independent* hidden variables to the two electrons, such that what happens to either electron can affect the hidden variables associated with the other only by way of influences propagating at the speed of light or less. This is what Bell [1964] demonstrated.

Bell's theorem has equally unwelcome consequences for a dynamical collapse theory. Such a theory, if it is to be viable, must entail that a spin measurement carried out on either of two electrons prepared in the singlet state will precipitate a collapse involving both, thereby projecting them into opposite spin eigenstates. And once again, Bell's theorem tells us that it is impossible, in general, to square this account with the predictions of conventional quantum mechanics, if we assume that the collapse would have to propagate from the directly measured electron to the other at a speed no greater than that of light.

It seems widely to be believed, even amongst physicists, that in view of Bell's result (based, as it is, on quantum-mechanical predictions which have since been experimentally confirmed, in respect of photon polarization, by Aspect *et al.* [1982], *any* realist interpretation of quantum mechanics must be committed to non-local interactions of some kind. But that is untrue. One can avoid this implication by rejecting the assumption that when a measurement is carried out, *one* of the possible outcomes occurs *to the exclusion* of all the others.

Bear in mind that it is already implicit in the superposition principle, that states of affairs which common sense would regard as mutually exclusive, can nevertheless occur together in linear combination. The best-known example is that of an electron encountering a double slit. Here, according to the conventional wisdom, the electron—in the absence of a measuring device which can tell us which path it takes—will be in a superposition of going through one slit and of going through the other. What justifies us, then, in assuming that the (unobserved) micro- and (observed) macroworlds march to different drummers in this respect? Schrödinger raised exactly this question in a talk given in Dublin in 1952:

> Nearly every result [a quantum theorist] pronounces is about the probability of this *or* that or that ... happening—with usually a great many alternatives. The idea that they be not alternatives but *all* really happen simultaneously seems lunatic to him, just *impossible*. He thinks that if the laws of nature took *this* form for, let me say, a quarter of an hour, we should find our surroundings rapidly turning into a quagmire, or sort of a featureless jelly or plasma, all contours becoming blurred, we ourselves probably becoming jelly fish. It is strange that he should believe this. For I understand he grants that unobserved nature does behave this way—namely according to the wave equation (Schrödinger [1995], p. 19).[3]

Let us now pursue this idea that there are, in fact, no hidden variables and no objective state vector reduction, but *only* unitary evolution according to the Schrödinger equation. What does such a view have to say about measurement? Well, it tells us that the effect of measuring an observable is to create an *entanglement* between the states of the system being measured, the measuring apparatus, and the mind of the observer. Imagine, to begin with, that we have a single electron prepared in the state z spin-up. And suppose the observer, Alice, measures the spin in the x direction, using a Stern–Gerlach apparatus so adjusted that it magnetically deflects the electron in a left or right direction, depending on whether it is measured as being spin-up or spin-down.[4] Detectors on the two paths are connected to a dial, which accordingly shows an 'L' or an 'R'. And corresponding to these two states of the dial, there will be two states of Alice's conscious mind after she has inspected the dial. We can represent the two final states of the apparatus as $|\mathrm{L}\rangle$ and $|\mathrm{R}\rangle$, and the states of consciousness to which they respectively give rise as $|\blacktriangle\rangle$ and $|\blacktriangledown\rangle$. Unitary quantum mechanics then tells us that the final state of the composite system comprising electron, apparatus, and Alice will be:

$$(4) \qquad 1/\sqrt{2}|\uparrow_x\rangle|\mathrm{L}\rangle|\blacktriangle\rangle + 1/\sqrt{2}|\downarrow_x\rangle|\mathrm{R}\rangle|\blacktriangledown\rangle.$$

It was Everett [1957] who first had the audacity to suggest (in print, at least)

---

[3] I am grateful to Michel Bitbol for bringing this passage to my attention. Another passage, a sentence or so later on, is also worth quoting here: 'The compulsion to replace the *simultaneous* happenings, as indicated directly by the theory, by *alternatives*, of which the theory is supposed to indicate the respective *probabilities*, arises from the conviction that what we really observe are particles—that actual events always concern particles, not waves. Once we have decided for this, we have no choice. But it is a strange decision' (Schrödinger [1995], p. 20). For Everett's openly acknowledged debt to Schrödinger, see fn. 5.

[4] In reality, it would be atoms, rather than electrons by themselves, that were passed through the Stern–Gerlach apparatus: for example, silver atoms, with the spins of their valency electrons prepared and measured as described in the text.

that evolution into a state such as (4) is *all* that happens in measurement[5]—the establishment of a quantum *correlation* between states of the electron, the apparatus, and Alice. But if this suggestion is correct, how are we to reconcile the state (4) with what Alice actually *experiences*, having measured the spin of an electron? Well, we can approach this in stages.

Given an entangled state of a composite system, and a possible pure state of one of the systems so entangled, one can define what Everett calls a *relative state* of the remainder of the composite system (where this relative state is also a pure state). Take, by way of illustration, the singlet state (3), and consider the state, for either electron, of being spin-up in some arbitrary direction; the relative state of the other electron will then be spin-down in that same direction. Now let us apply this concept to the state (4). With respect to the state $|\blacktriangle\rangle$, of Alice's mind, the relative state of the electron and apparatus is $|\uparrow_x\rangle|\mathrm{L}\rangle$. And likewise, with respect to the state $|\blacktriangledown\rangle$, of Alice's mind, the relative state of the electron and apparatus is $|\downarrow_x\rangle|\mathrm{R}\rangle$.

In what follows, I shall use the term *Everett branch* for the tensor product of an observer state and the corresponding relative state of the remainder of the composite system of which the observer is a part. Thus, we here have the two Everett branches

$$(4(\mathrm{i})) \qquad |\uparrow_x\rangle|\mathrm{L}\rangle|\blacktriangle\rangle$$

and

$$(4\mathrm{ii})) \qquad |\downarrow_x\rangle|\mathrm{R}\rangle|\blacktriangledown\rangle.$$

Everett's hypothesis, in effect, is that if (in Nagel's felicitous but, by now, well-worn phrase) one asks *what it is like to be* Alice, when she is caught up in the entangled state (4), the answer is that it is like remembering seeing the dial read 'L' and nothing else, *and* like remembering seeing it read 'R' and nothing else. And unless Alice has been converted to the Everett point of view, these recollections will be accompanied by beliefs about the state of her world which respectively coincide with (4(i)) and (4(ii)). Alice, we must conclude, is *literally* in two minds here!

A remarkable conclusion, to say the least. But dispassionately regarded, no *more* remarkable, I would argue, than the already utterly mysterious fact that, at a given time, there is even *one* 'what it is like to be' associated

---

[5] In his doctoral dissertation, Everett describes his theory as a *wave interpretation*, in Schrödinger's sense of that term (referring to Schrödinger [1952]), and says of his view that it 'corresponds most closely with that held by Schrödinger', but goes on to point out that 'this picture only makes sense when observation processes themselves are treated within the theory . . . The "quantum jumps" exist in our theory as *relative* phenomena . . . while the absolute states change quite continuously' (Everett [1973] p. 115).

with my brain. The very existence of consciousness is, after all, a complete enigma from the standpoint of physics. Everett's approach does, however, face a number of prima-facie difficulties.

First, the state expressed by (4) can be represented as a superposition of states in an infinity of different ways, corresponding to different vector bases which are rotated with respect to each other in Hilbert space. For example, it can be represented as

$$(4') \quad \begin{aligned} &1/\sqrt{2}|\uparrow_y\rangle\{1/\sqrt{2}|\mathrm{L}\rangle|\Delta\rangle + 1/\sqrt{2}|\mathrm{R}\rangle|\nabla\rangle\} \\ &\quad + 1/\sqrt{2}|\downarrow_y\rangle\{1/\sqrt{2}|\mathrm{L}\rangle|\Delta\rangle - 1/\sqrt{2}|\mathrm{R}\rangle|\nabla\rangle\}. \end{aligned}$$

Why, then, should this postulated schism within consciousness mirror the decomposition of the state-vector displayed in (4), as opposed, say, to that displayed in (4′)? To be sure, (4′) looks more complicated than (4). That, however, is only because the very manner in which I have formulated (4) and (4′) reflects the way in which things *in fact* appear to us, and is therefore inherently biased in favour of (4). *Why* things appear that way remains to be explained. This is known as the *preferred basis* problem. I must emphasize, however, that it is no *more* of a problem for the Everett approach than it is for the collapse theory. For one can similarly challenge a proponent of objective state vector reduction to explain why the composite system is projected into one or other of the two superposed terms of (4)—namely, (4(i)) or (4(ii))—instead of being projected, for example, into one of the two superposed terms of (4′).

What also remains to be accounted for, on this approach, is how *probabilities* arise. I shall shortly return to this issue, as also to the preferred basis problem. For the moment, though, I want to explain precisely how the Everett approach avoids having to postulate non-local interactions.

Suppose, once again, that we have two electrons, 1 and 2, in the singlet state, and that their spins are being measured, respectively, by Alice and Bob (A and B for short). For simplicity, let us ignore the measurement apparatus and focus merely on the post-measurement states of the electrons, and of Alice and Bob. Assuming, to begin with, that Alice and Bob both measure the spin in the x direction, the resulting state of the composite system, including both electrons and both observers, will be:

$$1/\sqrt{2}|\uparrow_{x1}\rangle|\Delta_{\mathrm{A}}\rangle|\downarrow_{x2}\rangle|\nabla_{\mathrm{B}}\rangle - 1/\sqrt{2}|\downarrow_{x1}\rangle|\nabla_{\mathrm{A}}\rangle|\uparrow_{x2}\rangle|\Delta_{\mathrm{B}}\rangle. \qquad (5)$$

Now it is crucial, here, to distinguish between non-local *states* and non-local *interactions*. (5) is certainly a non-local entangled state; but the point is that it has evolved, in two stages, by dint of purely local interactions. First the entangled pair of electrons is generated in a local interaction, and the resulting entanglement is spatially propagated by way of the two electrons flying apart. Then the two observers, Alice and Bob, become caught up in the entanglement through their respective local measurement interactions.

As a rough analogy, think of the situation where two spatially separated systems are in an entangled state, on the model of two people being linked by a length of rope, and consider how this might come about. One way would be if the rope starts out coiled, the two people grasp the two ends, and they then walk in opposite directions; it is somewhat in this fashion that the non-local entanglement of the two electrons is brought about. Another way would be if the rope has already been spread out on the ground, so that its two ends are now far apart, and two spatially separated individuals, who are standing adjacent to the two opposite ends, then take hold of them; this is analogous to the way in which Alice and Bob come to be in an entangled state *vis à vis* each other.

Thus far I have quite deliberately avoided any reference to 'many worlds'. This may seem strange, seeing that Everett is usually credited with having invented the many worlds theory. The reader will search in vain, however, for any explicit reference to 'worlds' or 'universes' in Everett's published writings; such terminology seems first to have been introduced by Bryce DeWitt [1970]. So if Everett's theory is, by implication at least, a many worlds theory, what exactly is it in his theory that corresponds to these worlds?

Well, the best candidates for 'worldhood' in Everett's own writings would appear to be the Everett *branches*, which I talked about earlier. The universe, as Everett views it, is clearly to be thought of as one vast, deterministically evolving entangled system. Suppose, now, that Alice has read Everett and is persuaded of the correctness of his interpretation of quantum mechanics. Given some conscious state in which she finds herself, Alice might think of this state as defining a relative state of the entire remainder of the universe. She might then regard the tensor product of this cosmic relative state and her own conscious state as the current state of a 'world' or 'universe', in the sense intended by this talk of 'many worlds'. For accepting, as she does, Everett's theory, she acknowledges that there are parallel states of her mind, with respect to which distinct such 'worlds' would be defined. Let us call these *Everett worlds*.

What must now be pointed out is that these Everett worlds are actually very poor candidates for DeWitt worlds; for they are irreducibly *egocentric*. Imagine that Alice and Bob both go through the procedure of defining such Everett worlds, having measured their respective electrons. And suppose, first, that they have each measured the spin in the x direction. Here their procedures will essentially converge on the same result. Both will end up with the same pair of Everett worlds. In one of these worlds, electron 1 is in the state x spin-up, with Alice remembering seeing an 'L', while electron 2 is in the state x spin-down, with Bob remembering seeing an 'R'; in the other, electron 1 is in the state x spin-down, with Alice remembering seeing an 'R', while electron 2 is in the state x spin-up, with Bob remembering seeing an 'L'.

But such convergence *cannot* be expected in general. For now consider what the same procedure will yield when Alice measures the spin in the x direction while Bob measures it in the y direction. This will result—once again by way of purely local interactions—in a more complicated entangled state. Here tensor products, containing states of Alice, her Stern–Gerlach apparatus, and electron 1, which correspond to a measurement of x spin-up and to a measurement of x spin-down, will be correlated, respectively, with two different *superpositions* of tensor products of states of Bob, his apparatus and electron 2, the components of which correspond to measurements of y spin-up and y spin-down. (And the relationship is, of course, a symmetrical one: an equally correct and formally equivalent description of this entangled state can be arrived at by interchanging, in the previous sentence, 'Alice' and 'Bob', 'her' and 'his', 'electron 1' and 'electron 2', and 'x' and 'y'.) Accordingly, Alice and Bob find themselves in *distinct* pairs of Everett worlds in each of which *they alone* have determinately got the result spin-up *or* spin-down.

That, at least, is the situation before Alice and Bob *communicate* their results to each other: yet another local interaction. The effect of their comparing notes will be to restore approximate convergence by creating, for each of them, four Everett worlds: tensor products which respectively incorporate the four possible outcomes, $|\uparrow_{x1}\rangle|\blacktriangle_{\mathrm{A}}\rangle|\uparrow_{y2}\rangle|\blacktriangle_{\mathrm{B}}\rangle$, $|\uparrow_{x1}\rangle|\blacktriangle_{\mathrm{A}}\rangle|\downarrow_{y2}\rangle|\blacktriangledown_{\mathrm{B}}\rangle$, $|\downarrow_{x1}\rangle|\blacktriangledown_{\mathrm{A}}\rangle|\uparrow_{y2}\rangle|\blacktriangle_{\mathrm{B}}\rangle$ and $|\downarrow_{x1}\rangle|\blacktriangledown_{\mathrm{A}}\rangle|\downarrow_{y2}\rangle|\blacktriangledown_{\mathrm{B}}\rangle$.

Now there is, in fact, a way of ensuring convergence here throughout; and that is to define the Everett worlds with respect to tensor products of Alice's *and* Bob's conscious states. Indeed, one could define Everett worlds with respect to tensor products of conscious states of *all* sentient beings which figure somewhere in the universal state vector. The resulting Everett worlds would of course remain, if not *ego*centric, then at least *psycho*-centric. So perhaps one should focus instead on the states of some more all-encompassing set of systems. But whatever states of whatever class of systems one decided to take as the basis for defining a set of Everett worlds, it could reasonably be asked why *those* things should be regarded as metaphysically privileged.

From the standpoint of what I take to be the *many worlds* theory, however, the question is readily answered. Many worlds theorists do not *arbitrarily* assign a metaphysically privileged status to a certain class of states of a certain class of systems. Their starting point is, instead, the not unreasonable hypothesis that certain things *appear* to be in certain definite states because they *are*, when regarded from the right perspective—albeit that common sense is mistaken in supposing that they are in *unique* such states at any given time. The purpose of the preferred basis is, accordingly, to capture this perspective, thereby providing a mode of description which 'slices reality at its joints'. It follows, then, that this way of expressing the universal state vector must, if it is to do its job, represent macroscopic objects, as they figure in the terms of the superposition, as being in eigenstates of the macroscopic observables, and observers as being in determinate states of mind. This, it seems to me, is a perfectly intelligible motivation for adopting a many worlds approach. But I would challenge its initial premiss. For, in the context of the Everett approach, I can see no good reason for supposing that the apparent macroscopic definiteness of the world is anything other than an artefact of our own subjective point of view.

That brings me to the class of theories which, by contrast with many worlds theories, I shall refer to as *many minds* theories. The label was, as far as I know, first coined by Albert and Loewer [1988] to describe their own interpretation of quantum mechanics. But I here use the term for a *genus* of theories, of which theirs and mine (Lockwood [1989, 1992]), and also those of a number of other authors,[6] represent different species.

A many minds theory, as I understand it, is a theory which takes completely at face value the account which unitary quantum mechanics gives of the physical world and its evolution over time. In particular, it allows that, just as in special relativity there is a fundamental democracy of Lorentz frames, so in quantum mechanics there is a fundamental democracy of vector bases in Hilbert space. In short, it has no truck with the idea that the laws of physics prescribe an *objectively* preferred basis. For a many minds theorist, the *appearance* of there being a preferred basis, like the *appearance* of state vector reduction, is to be regarded as an illusion. And both illusions can be explained by appealing to a theory about the way in which *conscious mentality* relates to the physical world as unitary quantum mechanics describes it.

Finally, a many minds theory, like a many worlds theory, supposes that, associated with a sentient being at any given time, there is a multiplicity of distinct conscious points of view. But a many minds theory holds that it is these conscious points of view or 'minds', rather than 'worlds', that are to be conceived as literally dividing or differentiating over time—or (as is

---

[6] I have in mind the following: Zeh's [1981] *multi-consciousnesses interpretation* (prefigured in Zeh [1970, 1973, 1979] and recently drawn to my attention by Euan Squires); Squires' [1987] *many-views* theory, which he has since abandoned; a theory proposed by Stapp [1991], which he also no longer holds; the work of Donald [1990, 1992, 1995]; and finally, Page's [1995] *many-perceptions* interpretation.

possible in principle, though unlikely in practice)[7] fusing or converging. The many minds theorist does not deny, of course, that 'multiplicity', within the world at large, which is an inescapable consequence of allowing superpositions of what classical physics would regard as mutually exclusive alternatives. Indeed, one *could* interpret this refusal to acknowledge a preferred basis as adding an extra dimension of multiplicity. One could see it as replacing a many worlds theory by what Page [1995] wryly describes as a 'many-many-worlds theory': one which posits a distinct multiplicity for each of the infinity of different bases.

It will be helpful now if we consider, yet again, the state evolution associated with a measurement of spin—but one, this time, where the two possible outcomes have *different* probabilities. Let us suppose the initial state of electron, apparatus, and observer, our faithful Alice, to be:

$$(6) \qquad \{\sqrt{2/3}|\uparrow_x\rangle + \sqrt{1/3}|\downarrow_x\rangle\}|\mathrm{B}\rangle|\mathrm{O}\rangle.$$

$\sqrt{2/3}|\uparrow_x\rangle + \sqrt{1/3}|\downarrow_x\rangle$ represents a state of spin-up in a direction lying between the positive x and z axes, at an angle of just under 70° $32'$ to the former. $|\mathrm{B}\rangle$ ('B' for a blank dial) is the state of the Stern–Gerlach apparatus, before the electron has been passed through and the dial has displayed an 'L' or an 'R'; and $|\mathrm{O}\rangle$ is the corresponding state of Alice's mind at this stage. When the electron has passed through the apparatus, but before Alice inspects the dial, the composite system will have evolved into the state.

$$(7) \qquad \{\sqrt{2/3}|\uparrow_x\rangle|\mathrm{L}\rangle + \sqrt{1/3}|\downarrow_x\rangle|\mathrm{R}\rangle\}|\mathrm{O}\rangle.$$

And when Alice looks at the dial, it will evolve further into the state

$$(8) \qquad \sqrt{2/3}|\uparrow_x\rangle|\mathrm{L}\rangle|\blacktriangle\rangle + \sqrt{1/3}|\downarrow_x\rangle|\mathrm{R}\rangle|\blacktriangledown\rangle.$$

Suppose one thinks, as does the many worlds theorist, that Nature herself has an objective preference for a set of basis vectors in which only determinate states of Alice's consciousness and the Stern–Gerlach apparatus' dial feature. Then one can regard (6), (7), and (8) as describing successive stages of the evolution of the composite system, the first stage of which involves division or differentiation of the world or worlds corresponding to (6), into worlds that respectively incorporate the product states

$$(7(\mathrm{i})) \qquad |\uparrow_x\rangle|\mathrm{L}\rangle|\mathrm{O}\rangle$$

---

[7] Deutsch ([1985], pp. 32–7) presents a thought experiment in which interference between two Everett branches, corresponding to two different measurement results, brings about a reconvergence of the observer's minds, such that the observer can no longer remember which result was obtained. Albert [1983, 1993] has developed an intriguing concept of *self-measurement* which involves a combination of reconvergence and divergence (see Albert [1993], Ch. 8). In both thought experiments, we end up, in my terms, with minds that should regard themselves as having multiple, qualitatively distinct, *pasts*.

and

(7(ii)) $|\downarrow_x\rangle|R\rangle|O\rangle.$

These worlds then evolve, respectively, into ones which incorporate the outcomes

(8(i)) $|\uparrow_x\rangle|L\rangle|\Delta\rangle$

and

(8(ii)) $|\downarrow_x\rangle|R\rangle|\nabla\rangle.$

A many minds theorist will view this evolution differently. *Objectively*, the many minds theorist will insist, the state of the composite system, or of the universe as a whole, is no *more* split or differentiated at the end of the measurement than at the beginning. For a basis state, with respect to one basis, is invariably a superposition with respect to others. We have here merely a smooth evolution of an essentially seamless state—albeit, to repeat, one in which states of affairs that we should ordinarily deem to be incompatible are superposed. But what *is* objectively true, in the sense of not being relative to a choice of basis, is that there is a division or differentiation within consciousness. That, however, occurs only when Alice looks at the dial, and thereby effects a division or differentiation of her own points of view. Thus, at stages (6) and (7), she seems to herself to be in a state corresponding to $|O\rangle$; while at the end (8) she *both* seems to herself to be in (and only in) a state corresponding to $|\Delta\rangle$ *and* seems to herself to be in (and only in) a state corresponding to $|\nabla\rangle$. In these terms, if one were to judge merely by the evidence of his published writings, one might be tempted to classify Everett himself as a many minds, rather than a many worlds, theorist. For he never speaks of dividing or differentiating worlds or universes, but only of the 'branching' and 'splitting' of 'observer states'.

But there is a problem with this branching or splitting model, which is common to both the many worlds and the many minds approaches. For intuitively, if the worlds merely divide in two, with the passage from (6) to (7), or if the minds do so, in the passage from (7) to (8), the subjective probabilities of getting, or observing, spin-up and spin-down should be *equal*. But that is not what quantum mechanics predicts in this case. Instead, it predicts that one is twice as likely to get spin-down as to get spin-up. (That, indeed, was the whole point of choosing this example.) As a solution to this problem, in the context of a many worlds theory, Deutsch [1985] suggests that we postulate the existence, at all times, of a continuous *infinity* of worlds. The idea then is that there is a natural measure, defined on subsets of these worlds, which satisfies the probability calculus; and the evolution of the state vector corresponds to a differentiation of the worlds over time, of just such a kind as to generate subsets whose measures mirror the quantum-mechanical probabilities. The quantum-mechanically predicted and experimentally confirmed probabilities, governing the above measurement, are accounted for by supposing that of the original set of worlds, corresponding to (6), 'twice as many', so to speak, evolve into states corresponding to (7(i)) and (8i)) above, as evolve into states corresponding to (7(ii)) and (8(ii)).

The observed probabilities here are thus to be explained along essentially the same lines by which one might explain the empirical finding that one is twice as likely to draw a black as a white ball when drawing blind from a jar whose contents have been thoroughly mixed: namely that there are twice as many black as white balls in the jar. In short, each quantum-mechanical observation, as experienced in a given world, is to be regarded as a *sampling* of the *multiverse*, the multi-faceted reality corresponding to the universal state vector.

What Deutsch says about worlds clearly admits of being transposed into a many minds context. Instead of postulating a continuous infinity of worlds, we could credit every sentient being with a continuous infinity of simultaneous minds or conscious points of view, which differentiate over time. That, indeed, is one of the key ideas underlying my own favoured interpretation of quantum mechanics. (As I shall explain later, there is a way of understanding these 'minds', which makes this hypothesis sound rather less extravagant than it at first appears.) In the spin measurement just described, these minds will, according to this theory, undergo their key differentiation during the transition from (7) to (8), when two-thirds of them will come to correspond to the state $|\Delta\rangle$ and one-third to the state $|\nabla\rangle$.

This feature of positing a continuous infinity of differentiating minds is one that my version of the many minds view shares with that presented by Albert and Loewer [1988]. Albert and Loewer, however, arrive at these minds by an entirely different route from my own, and assign them a different metaphysical status from that which they have in my version of the theory. Albert and Loewer's explanation of probability is also completely different from that just sketched, in relation to Deutsch's approach.

Albert and Loewer are influenced by the following two considerations. First, conventional unitary quantum mechanics—quantum mechanics without state vector reduction or hidden variables—is *deterministic*. This by itself constitutes an insuperable obstacle, they think, to the idea that one might find something *within* unitary quantum mechanics which corresponds to the probabilities prescribed by the quantum-mechanical statistical algorithm. (As Albert and Loewer rightly observe, DeWitt's [1970] claim to have derived probabilities from within the formalism, based on a calculation of the limiting values of certain amplitudes associated with sequences of measurements, is inherently circular. They might have added that exactly the same can be said of Everett's [1957] attempt to explain probabilities in terms of the limiting values of certain measures, defined on sets of *memory sequences* associated with repeated measurements.)

Secondly, mental states—by which Albert and Loewer appear to mean *conscious* mental states—cannot, as they see it, be identified with *physical* states of, let us say, the brain. For assuming the universal applicability of quantum mechanics to the physical world, all bona fide physical states are quantum states, and hence can be *superposed.* But, they argue, 'the way we conceive of mental states, beliefs, memories etc., it simply makes no sense to speak of such states of a mind as being in a superposition' (Albert and Loewer [1988], p. 203). This is because it is part of our concept of a mental state that it is 'accessible to introspection'. Accordingly, Albert and Loewer insist that an acceptable account of mental states should respect the *principle of charity*—so that when Alice reports, after the spin measurement, that she now, for example, has a recollection of getting the result spin-up (and nothing else), she can be regarded as saying something *true.* If, however, we adopt the Everett interpretation of quantum mechanics and identify mental states with physical states, we are obliged to conclude that what she is saying is *false.* For, given that what we have here, physically speaking, is an entangled superposition, it follows that there is *no* actually occurring physical state which corresponds to any such determinate recollection.

These considerations lead Albert and Loewer, first, to explore the following idea. Consider, once again, the evolution represented by (6) to (8), above, but this time written with certain elements highlighted.

$$\{\sqrt{2/3}|\uparrow_x\rangle + \sqrt{1/3}|\downarrow_x\rangle\}|\mathrm{B}\rangle|\underline{\underline{\bigcirc}}\rangle \tag{6'}$$

$$\{\sqrt{2/3}|\uparrow_x\rangle|\mathrm{L}\rangle + \sqrt{1/3}|\downarrow_x\rangle|\mathrm{R}\rangle\}|\underline{\underline{\bigcirc}}\rangle \tag{7'}$$

$$\sqrt{2/3}|\uparrow_x\rangle|\mathrm{L}\rangle|\underline{\underline{\blacktriangle}}\rangle + \sqrt{1/3}|\downarrow_x\rangle|\mathrm{R}\rangle|\blacktriangledown\rangle. \tag{8'}$$

The states $|\bigcirc\rangle, |\blacktriangle\rangle$ and $|\blacktriangledown\rangle$ are now to be understood differently from before, as denoting brain states, rather than mental states. And the highlighting is intended to indicate the evolution of a mind which *tracks* this evolution of the universal state vector, and does so in a doubly selective fashion. First, it latches on merely to a single brain, or to that part or aspect of it on which the contents of the corresponding mind depend. Secondly, whenever, as in the transition from (7′) to (8′), the mind is faced with a 'choice' as to which element of the state vector to follow, it follows one or the other on an irreducibly *random* basis, with probabilities, here $\frac{2}{3}$ and $\frac{1}{3}$, defined by the square magnitudes of the coefficients. The highlighting above thus indicates an evolution which follows the higher of the two probabilities. What is equally possible, though only half as *likely*, to happen, here, is the alternative evolution of (7′) into

$$(8'') \qquad \sqrt{2/3}|\uparrow_x\rangle|\mathrm{L}\rangle|\Delta\rangle + \sqrt{1/3}|\downarrow_x\rangle|\mathrm{R}\rangle|\underline{\underline{\nabla}}\rangle.$$

Thus we have a stochastically evolving mind harnessed to a deterministically evolving brain.

As it stands, this view strikes Albert and Loewer as deeply unattractive—and for two reasons. First, it generates what has come to be known as the *mindless hulk* problem. If I embark on what I take to be a conversation with my wife, how would I know, on this view, that there was really 'anyone at home'? For it is entirely possible, on the single mind view, that at some decision point in the past, her mind and mine went their separate ways, and that, in this Everett branch, I am addressing an empty shell. Not only is this deeply counterintuitive (not to say disconcerting); it is also in gross violation of the principle of the *supervenience of the mental on the physical.* For there will presumably be absolutely nothing about the physical character of the brain, as it manifests itself in different Everett branches, to distinguish those cases where it is 'inhabited' by a mind, and those where it is not.

The possibility of one's actually *meeting* mindless hulks can perhaps be ruled out by *fiat*. For one could stipulate that the stochastic evolution takes place under the constraint that living bodies, as they manifest themselves within a single Everett branch, are required either all to be ensouled, or all to be mindless hulks. That, however, would call for a *non-local* coordination of the various mind evolutions, and thus undercut what is one of the major motivations for adopting a no collapse, no hidden variables view, in the first place. And in any case, there would still *be* mindless hulks under this proposal, even though they went through the motions of carrying on conversations and so on *only* with their fellow hulks.

What Albert and Loewer actually suggest, of course, is that, associated with each living brain, there is a continuous infinity of minds, each independently evolving according to the same stochastic law. That we are dealing with a continuous infinity, here, is sufficient to ensure that each brain is not only certain to be inhabited, in every Everett branch, but certain to be inhabited by a continuous infinity of minds. Moreover, there is no problem about understanding the quantum-mechanical probabilities, in the context of this theory. For these probabilities are *put in by hand* simply by stipulating that each mind obeys an irreducibly probabilistic law of evolution, which mirrors the predictions of the quantum-mechanical statistical algorithm.

This approach plainly succeeds in reconciling the universal occurrence, within the physical universe, of unitary evolution, with the appearance of state vector reduction in accordance with the usual statistical rules. But the appeal to dualism, in order to make sense of quantum mechanics, strikes me as a rather desperate expedient. Admittedly, the dualism in question is of a modest variety, inasmuch as the *proportion* of minds which are in different states at any given time supervenes completely on the physical world, as characterized by the universal state vector. But supervenience nevertheless fails in relation to the question which minds end up tracking which terms of the state vector as new superpositions arise; this transcends anything that could be inferred from the quantum-mechanical state description.

So can we do better? Can we reap the benefits of Albert and Loewer's approach, without having either to embrace dualism or to introduce any irreducible element of randomness into the picture? I believe we can.

In what follows, I shall make use of the concept of a *mixed state*, in the sense of what is generally called an *improper mixture*. (This is a mixed state that does *not* admit of being understood in accordance with the *ignorance interpretation*: not, in other words, as the system's being in some pure state but our having merely probabilistic information as to which that is.) Thus far, we have been representing pure states by Hilbert space *vectors*. But we can equally well represent them by the corresponding *projection operators*. The projection operator, $|\psi\rangle\langle\psi|$, corresponding to the state vector, $|\psi\rangle$, can be thought of, concretely, as an observable, with the two eigenvalues, 1 and 0. If $|\psi\rangle\langle\psi|$ is measured on a system, S, it is guaranteed to yield 1 if the system is in the state $|\psi\rangle$, and guaranteed to yield 0 if S is in a state which is orthogonal to $|\psi\rangle$. If S is in a superposition of $|\psi\rangle$ and certain other states, then a measurement of $|\psi\rangle\langle\psi|$ will yield 1 with probability, p, equal to the square magnitude of the coefficient associated with $|\psi\rangle$, as it figures in that superposition, and will yield 0 with probability $1 - p$.

These projection operators can be added together, with coefficients that are required to sum to one, so as to produce new states which cannot be represented by Hilbert space vectors. In these terms, the mixed state of an electron, prepared in the state z spin-up, and then subjected by Alice to a measurement of spin in the x direction, will, according to unitary quantum mechanics, be $\frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x| + \frac{1}{2}|\downarrow_x\rangle\langle\downarrow_x|$ after the measurement is complete. And correspondingly, Alice herself will be in the mixed state $\frac{1}{2}|\Delta\rangle\langle\Delta| + \frac{1}{2}|\nabla\rangle\langle\nabla|$. $\frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x| + \frac{1}{2}|\downarrow_x\rangle\langle\downarrow_x|$ is a spherically symmetric state, in which a measurement of spin in any direction is equally likely to yield spin-up or spin-down. (No pure state of an individual electron could possibly have this property). Given this spherical symmetry, the state can be written as a mixture of spin-up and spin-down in any direction, for example as $\frac{1}{2}|\uparrow_y\rangle\langle\uparrow_y| + \frac{1}{2}|\downarrow_y\rangle\langle\downarrow_y|$. Whilst the components of an entangled system cannot individually be regarded as being in pure states, they can invariably be regarded as being in mixed states. Moreover, I take it that, strictly speaking, given the ubiquitousness of entanglement, no real-life macroscopic object or observer will ever, on the Everett view, be in a pure state.

In general, if we think of pure states as limiting cases of mixed states, the mixed state of a physical system embodies everything about the system which affects the probabilities associated with the outcomes of measurements on the system, considered in isolation. The respective mixed states of our two electrons, which are prepared in the singlet state and then fly apart, will thus embody everything about the two electrons which can manifest itself *locally*. In these terms, another way of seeing that, on the Everett interpretation, there is no need to postulate any non-local interactions is to appreciate (i) that *regardless* of the direction in which the spin is measured, each electron, both before and after the measurement, will be in the same mixed state $\frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z| + \frac{1}{2}|\downarrow_z\rangle\langle\downarrow_z|$ (where x or y can be substituted for z); and (ii) that each observer is bound to end up in the corresponding mixed state $\frac{1}{2}|\blacktriangle\rangle\langle\blacktriangle| + \frac{1}{2}|\blacktriangledown\rangle\langle\blacktriangledown|$. Thus, measuring the spin of either electron can have no effect, *locally*, on the state of the other. (In fact, on *no* interpretation of quantum mechanics which reproduces the observational predictions of the conventional theory, is there any local observation which enables one to *tell* that a spin measurement has been made on the other electron: this is known as the *no signalling*, or 'no Bell telephone', theorem.)

Some further terminology will be useful. The use of the term 'world' or 'universe', in the context of a so-called 'many worlds' or 'many universes' view, introduces an ambiguity. For there is clearly *a* sense in which even an advocate of a 'many universes' view would concede that there is just one universe: what, as we saw earlier, is sometimes referred to as the *multiverse*. A similar distinction is called for in the context of a many minds view. There is *a* sense in which I can regard myself as having just one mind. I could call this my *multimind*; but the word is not a very euphonious one, and in what follows I shall therefore mark the distinction by writing 'Mind' when I mean multimind. I shall also employ the expression *maximal experience* for a complete state of consciousness. In these terms a given *mind*, as opposed to the *Mind* of which it is an aspect, will have only a single maximal experience at any given time. This distinction enables us, by the way, to accommodate Albert and Loewer's wish to respect the principle of charity. Alice can, after all, correctly report that she is having a maximal experience which includes a recollection of having seen an 'L', and nothing else, on the dial of her Stern–Gerlach apparatus, provided she does not attempt to *equate* this maximal experience with the current state of her Mind as a whole. We need, here, to allow for states of *un*consciousness, such as occur in dreamless sleep and coma; I shall therefore use the term 'maximal experience' in a way that includes states of unconsciousness as limiting cases—as *null* experiences, so to speak. I shall say of a (type)[8] maximal experience E_1 and a (type) maximal experience E_2 that they are *subjectively identical*, if and only if what it is *like* to have the experience E_1 is precisely the same as what it is like to have E_2; failing that, E_1 and E_2 will be said to be *subjectively distinct*. (It follows from this definition that all null experiences are subjectively identical.)

I am now in a position to summarize my own favoured version of the many minds approach. It consists, in essence, of the following three interpretative assumptions, the first of which commits me to a materialist view of the Mind (albeit one that is not intended to gloss over the profound problem posed by consciousness).

(I) My Mind is a subsystem of my brain (or at least, of my body). There is, in other words, some subset of my brain's vast number of degrees of freedom that is *constitutively*, and not just causally, involved in my conscious mentality.

(II) There exists a set of mutually orthogonal pure states, comprising a basis for my Mind, which I call the *consciousness basis*. Each of these basis states is such that, were my Mind (*per impossibile*) to be *in* that state at time t, then the maximal experiences which I was having at t, in different Everett branches, would be subjectively identical. Moreover, for every maximal experience (type), E, which my Mind is capable of generating, there is an associated state $|\varphi\rangle$, belonging to the consciousness basis of my Mind, such that, were my Mind to be in the state $|\varphi\rangle$, I should have (a token of) E in every Everett branch. Where the maximal experiences E_1 and E_2 respectively correspond, in this way, to two distinct basis states, I regard E_1 and E_2 (*qua* types) as themselves distinct, even if they happen to be subjectively identical. In my usage, therefore, *subjective* distinctness, for maximal experiences, implies, but is not implied by, distinctness *tout court*.

(III) Let $|\varphi\rangle$ be a pure state belonging to the consciousness basis of my

---

[8] I here employ a convention whereby 'E', with or without a subscript, stands for a maximal experience *type*, and 'e' stands for a *token* maximal experience.

> Mind, and let E be the corresponding maximal experience. And suppose that the mixed state of my Mind, at time t, is represented (as it invariably can be) by a weighted sum of projection operators corresponding to mutually orthogonal pure states. Each of these pure states can, of course, be expressed as a superposition of elements of the consciousness basis; as a limiting case, it may *be* a basis state. Now consider every term, $w|\psi\rangle\langle\psi|$, in this canonical representation of the mixed state of my Mind, for which $|\psi\rangle$, *qua* superposition of basis states, includes the state $|\varphi\rangle$ with the non-zero coefficient c. Assign this occurrence of $|\varphi\rangle$ a numerical value equal to the product of w and the square magnitude of c. (If $|\psi\rangle$ *is* $|\varphi\rangle$, the occurrence will simply be assigned the value w.)[9] Let s be the sum of the values assigned to all these occurrences of $|\varphi\rangle$. Then, I contend, my Mind at t will contain a continuous simultaneous infinity of tokens of E, with a measure that is proportional to s.

Assumption (III) requires spelling out in considerably greater detail; in so doing, I shall, for the next paragraph alone, use 'mind' in its ordinary sense, rather than to mean an element of the multimind.

In one context, we are familiar with the idea that there can be mutually incompatible experiences both of which are equally *our* experiences; this is possible, provided that they occur at different *times*. Moreover, one thinks of one's mind (in the usual sense of that term) as being *wholly* present at each of these times. One doesn't think of *part* of one's mind as existing at one time and *part* at another. I now want to propose a *second* context in which it is appropriate to speak this way. If one starts by thinking of time as being defined by a vertical dimension, then I want to supplement this picture by introducing a horizontal *superpositional* dimension, along which, once again, distinct (and sometimes incompatible) experiences are distributed. One should now think of one's mind as being *wholly present* at each of the points on a line lying at right angles to any given point on the time axis, where these points represent *simultaneous* maximal experiences. The picture which thus emerges is that of a two-dimensional *array* of maximal experiences, which I call an *experiential manifold.* This manifold must include, of course, regions corresponding to states of

---

[9] The mixed states of my Mind engendered by such *ideal* measurements as we have hitherto been discussing—that is, measurements which establish *perfect* quantum correlations between distinct eigenstates of the measured observable and distinct elements of the consciousness basis—can be expressed as weighted sums of projection operators each of which corresponds to a different element of the consciousness basis. But the mixed states produced by real-life measurements—which my account is also intended to cater for—will only ever, at best, approximate to this form.

unconsciousness. These one can think of as being 'blank', by contrast with the remaining regions, which are 'coloured in' with qualia and so forth. In terms of this picture, what one would ordinarily think of as *the* history of one's mind, over a given time interval, will correspond to a sequence of maximal experiences, forming a continuous line running in an upwards direction in the manifold. But the common-sense assumption that this history is *unique* is, I suggest, a figment of *memory*, which confines the gaze of consciousness to a kind of 'tunnel vision', directed downwards in the experiential manifold. We cannot look 'sideways' through the manifold, any more than we can look 'upwards', into the future.

Each time-slice of the manifold corresponds to the mixed state of the Mind at the associated time. It can be pictured as a horizontal line, divided into different segments, which correspond to distinct occurrences of various different basis states within this mixed state (when expressed as in (III) above). The width of a given segment will be proportional to the numerical value attached to the corresponding occurrence of the associated basis state, and the maximal experiences it contains will accordingly be of the type which supervenes on this state. Such a segment constitutes a cross-section of a two-dimensional homogeneous region of the experiential manifold. Suppose, now, that we have two time-slices, corresponding to an earlier and a later time, and consider a particular segment of the earlier time-slice. The way in which mixed states evolve over time then allows us to identify a specific set of segments of the later time-slice which represent *descendants* of the original segment. Likewise, if we consider a particular segment of the later time-slice, we can identify a specific set of segments of the earlier time-slice which represent *antecedents* of that segment of the later time-slice (a set which in practice is likely to consist of segments all of which correspond to occurrences of the same basis state).

A corresponding pair of concepts can now be introduced in relation to (token) maximal experiences. Suppose that we have a maximal experience, *e*, belonging to a given segment of a given time-slice. The instantaneous mind associated with *e* is, on the present view, entitled to regard as *predecessors* of *e* *all* the maximal experiences on an earlier time-slice which are located within some segment that is an antecedent of the segment containing *e*. Likewise, it is entitled to regard as *successors* of *e*, at the time in question, *all* the maximal experiences on a later time-slice which are located within some segment that is a descendant of the segment containing *e*.[10] If, in an upwards direction in the manifold, the homogeneous region containing *e* gives way, on a later time-slice, to several new and distinct regions, then the mind corresponding to *e* (assuming that it knows this) should regard maximal experiences belonging to *all* these regions as lying in store for it, and at the very same future time. This, accordingly, is what each of Alice's minds is entitled to think, before the spin measurement is complete, in light of the future differentiation of the region containing her pre-measurement maximal experiences into distinct regions corresponding to her seeing an 'L' and seeing an 'R'.

By a *continuous sequence* of maximal experiences, I mean a time-ordered set of maximal experiences, the elements of which are in one–one correspondence with the elements of some given time interval, and in which, for any pair of distinct elements of the sequence, one member of the pair is either a predecessor or a successor of the other. A maximal such sequence I call a *biography*. But I need to explain 'maximal sequence'. To accommodate our coming into and going out of existence, one must suppose the experiential manifold to have an upper and a lower *edge*. The lower edge will be essentially straight, if one takes the view that a given Mind comes into existence at virtually the same time in all branches of the universal state vector in which it exists at all. By contrast, however, the upper edge will be ragged in the extreme, with clefts (corresponding to premature death) that can extend arbitrarily far down towards the lower edge. This allows me to assume that any continuous sequence of maximal experiences which is extended sufficiently far in both temporal directions will be bounded, at each end, by an edge; this is what I mean by a maximal sequence. A biography thus begins, as it should, at the lower edge of the

---

[10] By saying '*all* maximal experiences', as opposed to '*all and only* maximal experiences', I have left it open whether the prescription given in this paragraph succeeds in capturing *every* bona fide successor or predecessor of *e*. I strongly incline, however, to the view that it does not. Suppose, for example, that Alice performs a measurement of spin in the x direction, on an electron prepared in some spin state which is neither x spin-up nor x spin-down, and follows this up with a measurement of spin in the y direction if and only if the result of the first measurement is spin-down. Then a time-slice of Alice's experiential manifold, immediately after the first measurement, will feature a segment embodying a recollection of having got x spin-up and a segment embodying a recollection of having got x spin-down (with each segment's being associated with an occurrence, within the canonically represented mixed state of Alice's Mind, of the corresponding basis state). On a time-slice subsequent to the second measurement, the segment embodying a recollection of having got x spin-down will, as one would expect, have as descendants a pair of segments: one corresponding to a recollection of having got x spin-down followed by y spin-up, and the other to a segment embodying a recollection of having got x spin-down followed by y spin-down. But likewise, the segment embodying a recollection of having got x spin-*up* will have as descendants a pair of segments—albeit, in this case, ones containing identical maximal experiences. This reflects the fact that, as a result of the second measurement, the original single occurrence of the corresponding basis state, within the canonically represented mixed state of Alice's Mind, will give way to two occurrences. If we reject the idea of a measurement's effecting a real change on an Everett branch in which it does not take place, then we are obliged, surely, to *pool* these two segments—to rule, in other words, (a) that a maximal experience contained in a descendant or antecedent of either segment is to count as a successor or predecessor of the maximal experiences contained in both segments, and (b) that there can be no fact of the matter as regards which of the two segments a given maximal experience belongs to. What is needed, clearly, is a general criterion of when such pooling is called for; but I shall not here attempt to formulate such a criterion.

experiential manifold and ends at the upper edge. It follows from what I have been saying that every (token) maximal experience will belong to a continuous infinity of biographies—some subjectively distinct from each other and some subjectively identical.

I offer this picture for the sake of vividness, and do not intend it to be taken too literally. I do, however, wish to insist on two things here: first, the simultaneous existence of distinct (and indeed, of *subjectively* distinct) maximal experiences, each of which exists in a continuous infinity of identical copies; and second, the existence of a naturally preferred *measure* on sets of simultaneous maximal experiences, which plays a role closely analogous to that which elapsed *proper time* plays in respect of *successive* maximal experiences. For in order to make sense of probabilities, in this context, I want to be able to say something with respect to the superpositional dimension(s) that, intuitively speaking, is parallel to saying, for example, that this pain lasted twice as long as the last one: I want to be able to say that this pain is, superpositionally speaking, twice as *extensive* as that. And I also want the two measures to have similar moral implications. How bad, overall, a pain of a given, constant intensity is depends on the overall *area* which it occupies in one's experiential manifold; and this will be a function both of temporal 'length' and of superpositional 'width'.

In discussing these ideas with colleagues, I have met widespread scepticism as to whether any real *meaning* can be given to the existence of such a natural measure, which could intelligibly be supposed to manifest itself in experience in the form of probabilities. After all (so people have objected), if we're postulating continuous sets of maximal experiences here, then presumably there is an infinity of distinct measures which *could* be defined on these sets, each of which would obey the probability calculus. By what criterion, then, is any one of them to be regarded as alone constituting a genuine *physical* probability measure? (Indeed, one could ask the same question, even if there were only one such measure.)

This is a line of argument which troubled me for a long time (both before and after these views were first published). But having now become acquainted with the Albert–Loewer version of the many minds view, it seems to me that there is a simple and, once seen, obvious answer to this question. As I say, Albert and Loewer's theory clearly succeeds, in my estimation, in explaining why quantum-mechanical systems appear to obey the statistical laws that they do. And I regard the following principle as obviously true. Given a theory, T, which explains (whether or not correctly) why things appear thus and so, and given that another theory, T', predicts (and explains) the occurrence of precisely the same *experiences* as does T, it must follow that T' also explains why things appear thus and so.

Now Albert and Loewer and I all suppose that, associated with every Mind, there is, at every moment, a continuous infinity of simultaneous maximal experiences. Albert and Loewer would have us see each such continuously infinite set, and its continuously infinite subsets, as being generated by an irreducibly probabilistic law governing the evolution of the individual minds whose experiences these are. I reject that idea. But there is nothing to prevent my stipulating that the content and structure of these successive continuously infinite sets of maximal experiences are indeed *precisely* what they would be, if they actually *were* being generated in the manner that Albert and Loewer suppose, and that the 'natural measure' which I require is just the measure which would reflect the operation of such a law, were there to be one. Then it surely follows that *my* theory must succeed in accounting for the appearance of things obeying the quantum-mechanical statistical predictions, provided that Albert and Loewer's does. For the two theories are, by hypothesis, experientially equivalent.

They are not, however, metaphysically equivalent, even if one ignores the fact that Albert and Loewer's theory is dualist. Consider those of Alice's instantaneous minds which correspond to her being about to measure the spin, in the x direction, of an electron prepared in the state which figures in (6), above, where, in conventional terms, she is twice as likely to get spin-up as to get spin-down. For Albert and Loewer, none of these minds is in principle able, before the measurement, to *know* what it will subsequently experience, because they all evolve stochastically. But still, since the minds possess transcendental identity over time, any particular mind will definitely evolve in one way or the other. On my view, by contrast, it is already certain what will happen. After the measurement there will be a continuous infinity of maximal experiences, all of which will stand in an equal relation of succession to the given instantaneous mind, or maximal experience, with which we started. This infinity of maximal experiences will consist of two sets, one corresponding to her having seen an 'L', and the other to her having seen an 'R'; and the first will have a measure twice the size of the second.

Interestingly, Albert and Loewer themselves consider the possibility of restoring supervenience of the mental on the physical, by repudiating (in my terms) the assumption that in a case such as this, there is a uniquely correct way of linking earlier and later maximal experiences of the same Mind together to form persisting minds, or biographies. This would mean treating the minds somewhat as quantum mechanics treats identical particles. Albert and Loewer's objection to this move is that, if it doesn't make sense to ask which of the minds end up in which state, then it doesn't make sense, either, to ask what the *probability* is of their doing so. Well I agree that this follows. But Albert and Loewer are, I believe, mistaken in supposing, as they seem to, that this would be fatal to their project. For all that is necessary, in an interpretation of quantum mechanics, is that it be capable of explaining the *appearance* of stochastic evolution. And how things appear can depend only on *what* appearances there are, not on the presence or absence of additional (and indiscernible) relations of transcendental identity. It follows from my view that the overall character of Alice's experience will be just what one would expect were it indeed, as Albert and Loewer propose, a matter of each mind ending up in a determinate such state, with a probability that is proportional to the 'width' of the corresponding region of the experiential manifold.

In my theory, or in my development of Everett's theory, there is thus complete supervenience of the mental on the physical. To be sure, I have to *postulate that certain subsystems of the brain are associated with experiential manifolds*. But this is merely an assumption about what it is like, in total, to *be* in certain brain states. The assumption no more carries any dualistic implications than the conventional assumptions, which even physicalists allow themselves, about what it is like to be in such states.

There remain, however, a number of important residual issues, which I do not have the space to address properly, but which I must briefly mention. First, since, for ease of exposition, I have been associating sets of maximal experiences with Schrödinger states, I must make it clear that I am not really committed to the apparent implication that maximal experiences are associated with *instantaneous* states of the relevant brain subsystem (see Lockwood [1989], Ch. 15). Indeed, given that the states underlying maximal experiences are presumably spatially extended, they could not possibly be instantaneous with respect to all frames of reference. For that reason, the ideas I have been presenting here would probably fit more comfortably into a formal framework which represented states of the relevant brain subsystem by appropriate projection operators on regions of space-time, or something similar (as in Donald [1990, 1992, 1995] and Page [1995]).

Secondly, is not my *consciousness basis* a *preferred* basis, in just the sense that I earlier insisted that I did not *want* to postulate a preferred basis? Well, no. For as far as the laws of physics are concerned, it is a matter of indifference in what basis states of the Mind (*qua* brain subsystem) are expressed. From our subjective standpoint, it *does*, to be sure, make a difference which basis we choose. But that is only because from a subjective standpoint it is, of course, the subjective, i.e. the conscious states that matter. So a basis which includes all those pure states which, as they figure in mixed states, carry experiences in their train is, naturally enough, *subjectively* preferred.

On the conception I am offering, we all view the world from the perspective of the consciousness basis of our own Minds. Consequently, we think of external macroscopic objects as being, all the time, determinately in those types of state with which elements of the consciousness basis are perpetually becoming correlated through the mechanisms of perception. For that, of course, is how things are bound to appear to us. To appreciate that consciousness (which in a sense is the *primary* observable) is associated with a specific basis for a specific subsystem of the brain is to understand, *a fortiori*, why the rotational symmetry of Hilbert space is subjectively broken in the world as we perceive it.

Still, it is a fair question *why* consciousness and perception should favour the states that they do. And here, it seems to me, currently fashionable talk about *decoherence* may come into its own. There are good theoretical grounds for thinking that interactions between macroscopic objects and their environment will serve constantly to corral these objects into states which, to an extraordinarily good approximation, will resemble mixed states of a kind that would admit of an *ignorance interpretation* with respect to such macroscopic observables as approximate position. Specifically, the effect of these interactions is to *erode*, with incredible rapidity, all correlations between entangled systems which involve states *other* than eigenstates of certain macroscopic observables.[11] The shared eigenstates of these observables, which comprise what (loosely speaking) we can call the *decoherence basis*, have the property that they are *stable* under the ambient interactions. Macroscopic objects are thus channelled into (and maintained in) states in which, for all practical purposes, we may regard them as being, with certain probabilities, determinately in one or other of a number of narrowly defined spatial locations.[12]

That being so, it would make very good sense for the mechanisms of consciousness and perception to have evolved in such a way as to exploit the stability which this phenomenon entails. (Indeed, the very occurrence

---

[11] An immediate corollary, of which much is made by advocates of decoherence, is that we are entitled to discount the possibility of significant quantum interference between Everett branches corresponding to distinct eigenvalues of such macroscopic observables. But it is hardly surprising that it is difficult to detect quantum interference in large complex systems, precisely *because* they are large and complex. And besides, there do exist large-scale systems in which quantum interference *is* observed, such as *superfluids* and *superconductors*; indeed, we could even include light beams here. (There is a regrettable tendency in the literature to *define* 'macroscopic' in such a way that any system which exhibits interference is *ipso facto* not to be counted as macroscopic.)

[12] The fact that we *perceive* macroscopic objects as possessing well-defined locations, states of motion, and the like cannot, however, be accounted for on the strength of decoherence alone. To explain this perceived definiteness, any interpretation of quantum mechanics is obliged, at some point, to appeal—as do I in my assumptions (I)–(III) above—to some explicit or implicit theory about the way in which *consciousness* maps on to physical states (specifically, of the brain). This point is rightly emphasized by Harvey Brown, in his accompanying commentary.

of biological evolution presumably requires such stability.)[13] So it is a very plausible speculation that elements of the consciousness basis, if they are not themselves elements of the decoherence basis of the Mind, are, at any rate, constantly becoming correlated with states belonging to the decoherence bases of other brain subsystems. For as Zurek ([1992] p. 18) observes:

> If 'awareness' or 'consciousness' involves processes in which one part of the brain uses the data stored in the other, 'memory' part of the brain—as seems natural to assume—our analysis of the environment-induced decoherence in a detector applies directly: Only the states of the preferred [i.e. decoherence] basis of neurons are still correlated with the states of the relevant observables of the 'rest of the Universe', and therefore, contain reliable information.

But I insist, yet again, that even this decoherence basis does not admit of being made wholly precise, and is certainly not to be thought of as preferred in any *deep* sense, such as would justify our regarding the formation of superpositions with respect to it as amounting to a genuine fissioning of reality. Compare the way in which the expansion of the universe defines, in a rough-and-ready fashion, a distinguished time coordinate: one corresponding to a set of ideal clocks which simply 'go with the flow' of universal expansion, and relative to which all competent observers will agree on, say, the temperature of the cosmic microwave background. It would clearly be a philosophical error to think that this in any way compromises the fundamental democracy of space-time *foliations* ('slicings up' of the continuum into simultaneity surfaces) which is implicit in the covariance of the underlying laws, or that it entitles us to postulate an *absolute* cosmic time. Likewise, the deep fact of the democracy of bases in Hilbert space is in no way compromised by the phenomenon of decoherence, even if, as some would argue, it helps to explain how Nature succeeded for so long in concealing from us the astounding truth which Schrödinger suspected—and which Everett first had the courage wholeheartedly to embrace.

Acknowledgements

I am indebted to David Albert, Michel Bitbol, Harvey Brown, Jeremy Butterfield, David Deutsch, Don Page, David Papineau, Simon Saunders, and Euan Squires for some very illuminating discussions; and to Harvey Brown, David Deutsch, and David Papineau, in particular, for their

---

[13] See Saunders [1993].

valuable comments on earlier drafts of this paper and suggestions for improvements.

Department for Continuing Education
University of Oxford
1 Wellington Square
Oxford OX1 2JA
UK

References

Albert, D. Z. [1983]: 'On the Theory of Quantum-Mechanical Automata', *Physics Letters*, **A98**, pp. 249–52.

Albert, D. Z. [1993]: *Quantum Mechanics and Experience*, Cambridge, MA, Harvard University Press.

Albert, D. and Loewer, B. [1988]: 'Interpreting the Many Worlds Interpretation', *Synthese*, **77**, pp. 195–213.

Aspect, A., Dalibard, J. and Roger, G. [1982]: 'Experimental Test of Bell's Inequalities Using Time-Varying Analyzers', *Physical Review Letters*, **49**, pp. 1804–7.

Bell, J. S. [1964]: 'On the Einstein Podolsky Rosen Paradox', *Physics*, **1**, pp. 195–200.

Bohm, D. [1952]: 'A Suggested Interpretation of the Quantum Theory in Terms of Hidden Variables, I and II', *Physical Review*, **85**, pp. 166–93, reprinted in Wheeler and Zurek [1983].

Deutsch, D. [1985]: 'Quantum Theory as a Universal Physical Theory', *International Journal of Theoretical Physics*, **24**, pp. 1–41.

DeWitt, B. S. [1970]: 'Quantum Mechanics and Reality', *Physics Today*, **23**, 9, pp. 30–5, reprinted in DeWitt and Graham [1973].

DeWitt, B. S. and Graham, N. (*eds*) [1973]: *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton, Princeton University Press.

Donald, M. J. [1990]: 'Quantum Theory and the Brain', *Proceedings of the Royal Society of London*, **A427**, pp. 49–93.

Donald, M. J. [1992]: 'A Priori Probability and Localized Observers', *Foundations of Physics*, **22**, pp. 1111–72.

Donald, M. J. [1995]: 'A Mathematical Characterization of the Physical Structure of Observers', *Foundations of Physics*, **25**, pp. 529–71.

Everett, H. [1957]: '"Relative State" Formulation of Quantum Mechanics', *Reviews of Modern Physics*, **29**, pp. 454–62, reprinted in DeWitt and Graham [1973] and Wheeler and Zurek [1983].

Everett, H. [1973]: 'The Theory of the Universal Wave Function', Ph.D. thesis, Princeton University, in DeWitt and Graham [1973], pp. 1–140.

Lockwood, M. [1989]: *Mind, Brain and the Quantum: The Compound 'I'*, Oxford, Basil Blackwell.

Lockwood, M. [1992]: 'What Schrödinger Should Have Learned from His Cat', in

M. Bitbol and O. Darrigol (*eds*), *Erwin Schrödinger, Philosophy and the Birth of Quantum Mechanics*, Gif-sur-Yvette Cedex, France, Éditions Frontières, pp. 363–84.

Mermin, D. [1985]: 'Is the Moon There When Nobody Looks? Reality and the Quantum Theory', *Physics Today*, **38**, 4, pp. 38–47.

Page, D. N. [1995]: 'Sensible Quantum Mechanics: Are Only Perceptions Probabilistic?' University of Alberta preprint.

Saunders, S. [1993]: 'Decoherence, Relative States, and Evolutionary Adaptation', *Foundations of Physics*, **23**, pp. 1553–85.

Schrödinger, E. [1952]: 'Are There Quantum Jumps?' *British Journal for the Philosophy of Science*, **3**, pp. 109–23, 233–42.

Schrödinger, E. [1995]: 'July Colloquium, 1952', in E. Schrödinger, *The Interpretation of Quantum Mechanics: Dublin Seminars (1949–1955) and other unpublished essays*, edited and with introduction by M. Bitbol, Woodbridge, CT, Ox Bow Press, pp. 19–37.

Squires, E. J. [1987]: 'Many Views of One World: an Interpretation of Quantum Theory', *European Journal of Physics*, **8**, pp. 171–3.

Stapp, H. P. [1991]: 'Quantum Measurement and the Mind–Brain Connection', in P. Lahti and P. Mittelstadt (*eds*), *Symposium on the Foundations of Modern Physics, 1990*, Singapore, World Scientific, pp. 403–24.

Wheler, J. A. and Zurek, W. H. (*eds*) [1987]: *Quantum Theory and Measurement*, Princeton, Princeton University Press.

Zeh, H. D. [1970]: 'On the Interpretation of Measurement in Quantum Theory', *Foundations of Physics*, **1**, pp. 69–76.

Zeh, H. D. [1973]: 'Towards a Quantum Theory of Observation', *Foundations of Physics*, **3**, pp. 109–16.

Zeh, H. D. [1979]: 'Quantum Theory and Time Asymmetry', *Foundations of Physics*, **9**, pp. 803–18.

Zeh, H. D. [1981]: 'The Problem of Conscious Observation in Quantum Mechanical Description', *Epistemological Letters* (Ferdinand Gonsieth Association), 63.0.

Zurek, W. H. [1992]: 'Quantum, Classical and Decoherence', Los Alamos preprint.