10

The Interpretation of Quantum Mechanics

1. Two Mysteries

The problem of quantum mechanics is almost as hard as the problem of consciousness. Quantum mechanics gives us a remarkably successful calculus for predicting the results of empirical observations, but it is extraordinarily difficult to make sense of the picture of the world that it delivers. How could our world be the way it has to be, in order for the predictions of quantum mechanics to succeed? There is nothing even approaching a consensus on the answer to this question. Just as with consciousness, it often seems that *no* solution to the problem of quantum mechanics can be satisfactory.

Many people have thought that these two most puzzling of problems might be intimately linked (e.g., Bohm 1980; Hodgson 1988; Lockwood 1989; Penrose 1989; Squires 1990; Stapp 1993; Wigner 1961). Where there are two mysteries, it is tempting to suppose that they have a common source. This temptation is magnified by the fact that the problems in quantum mechanics seem to be deeply tied to the notion of observership, crucially involving the relation between a subject's experience and the rest of the world.

Most often, it has been suggested that quantum mechanics may hold the key to a physical explanation of consciousness. But as we have seen, this project will always fall short of its goal. At the end of the day quantum "theories" of consciousness suffer from the same sort of explanatory gap as classical theories. Either way experience must be taken as something over and above the physical properties of the world. Perhaps quantum mechanics might play a role in characterizing the psychophysical link, but quantum theory alone cannot tell us why consciousness exists.

But the problems may be linked in a more subtle way. Even if quantum mechanics does not explain consciousness, perhaps a theory of consciousness might shed light on the problems of quantum mechanics. After all, it is widely agreed that these problems have something to do with observership and experience. It is natural to suppose that a theory of experience might help us come to grips with the issues. Some have proposed an active role for consciousness in quantum theory—suggesting that consciousness brings about the "collapse of the wave function," for example—but I will argue for a more indirect role for consciousness in dealing with these questions. In particular, I will argue that we can reconceive the problems of quantum theory as problems about the relationship between the physical structure of the world and our experience of the world, and that consequently an appropriate theory of consciousness can lend support to an unorthodox interpretation of quantum mechanics.

2. The Framework of Quantum Mechanics

The basic framework of quantum mechanics consists in a calculus for predicting the results of experimental measurements. I will describe a version of that calculus here, glossing over a number of technical details in order to provide a simple description that covers the most crucial features. In this section, I present the framework merely as a calculus for empirical predictions, leaving open the question of whether it provides a direct description of physical reality. The deep problems of interpretation are discussed in the next section.

Within a classical framework, the state of a physical system can be expressed in very simple terms. The state of a particle, for example, is expressed by giving determinate values for each of a number of properties, such as position and momentum. We can call this sort of simple value a *basic value.* Within the quantum framework, things are not so simple. In general, the state of a system must be expressed as a *wave function,* or a *state vector.* Here, the relevant properties cannot be expressed in simple values, but instead must be expressed as a kind of combination of basic values. A quantum state can be seen as a *superposition* of simpler states.

The simplest example is a property such as *spin,* which has only two basic values.[1] These basic values can be labeled "up" and "down." In quantum mechanics, the spin of a particle is not always up or down, however. Instead, a particle's spin must in general be expressed as a *combination* of up and down, each with a different complex magnitude. The spin of a particle is therefore best regarded as a vector in a two-dimensional vector space. It is most naturally visualized as a superposition of a spin-up state and a spin-down state, with different magnitudes corresponding to each.

The same goes for position and momentum, except that each of these has an infinite number of basic values. The position and the momentum of a classical particle can each take on any of an infinite number of values in a continuum. The position of a quantum particle, correspondingly, must be expressed as a kind of infinite-dimensional vector with a different magnitude for each of these locations. This vector is best regarded as a *wave,* with different amplitudes at different locations in space; the function that takes a location to the corresponding amplitude is the wave function. Similarly, the momentum of a quantum particle can be regarded as a wave with different amplitudes at different basic values of momentum. Again, we can think of the position or momentum of such a particle as a superposition of basic values of position or momentum.

Because these states are just vectors, they can be decomposed into components in many ways. While it is often useful to see a two-dimensional spin vector as a sum of an "up" component and a "down" component, it can be decomposed in many other ways, depending on the basis chosen for the vector space. All of these bases are equally "natural"; none is preferred by nature. In fact, it turns out a single vector represents both the position and the momentum of a particle. Decomposed according to one basis, we get the "position" amplitudes; decomposed according to a different basis, we get the "momentum" amplitudes. In general, the decomposition that is relevant in a given case depends on which quantity we are interested in, and in particular on which quantity we choose to *measure,* as I discuss shortly.

The states of systems consisting of more than a single particle are somewhat more complex, but the basic idea is the same. Take a system consisting of two particles, *A* and *B.* The state of the system cannot generally be expressed by combining a wave function for *A* and a wave function for *B* in any simple way; the states of the two particles will often be *nonseparable.* Rather, the state of the system must be expressed as a wave function in a more complex space. This wave function can be seen as a kind of superposition of simpler states of the two-particle system, however, so the general picture still applies. The same goes for more complex systems, in which a state is still best represented as a wave function corresponding to a superposition of states.

All this is counterintuitive, but it is not yet paradoxical. If we take this formalism at face value as a description of reality, it is not *too* hard to make sense of. Some have supposed that it is incompatible with an "objective" view of the world, as it implies that entities in the world do not have an objective, determinate state. But this does not follow. On this picture, the state of an entity is best expressed by a wave function rather than by discrete quantities, but it is a perfectly determinate state. The picture simply tells us that on the basic level reality is wavelike. This requires a new way of thinking, but we can get used to it. After all, the basic level of microscopic reality is very far from the macroscopic level we usually deal with, and it is not entirely surprising that it should have some unfamiliar properties. Any problems that arise stem from *further* properties of quantum mechanics.

The core of quantum mechanics consists of two principles that determine the *dynamics* of the wave function: the *Schrödinger equation* and the *measurement postulate.* Between them, these two very different principles determine how the wave function of a system evolves with time.

Most of the substance of quantum mechanics is found in the Schrödinger equation. This is a differential equation that determines how the wave function of a system evolves under *almost* all circumstances. The detailed structure of the equation is not important for our purposes. The most important feature here is that it is a *linear* differential equation: given two states A and B such that A evolves into A' and B evolves into B', then a state consisting in a superposition of A and B will evolve into a superposition of A' and B'. It is also worth noting that under the dynamics of the Schrödinger equation, relatively discrete states usually become more spread out over time. A state that starts as a superposition of values in a limited range will generally evolve into a superposition of values in a much wider range. Finally, the Schrödinger equation is entirely deterministic.

The Schrödinger equation is relatively straightforward and well understood. It is here that the meat and potatoes of quantum theory resides. In applying quantum theory to a practical or experimental problem, the bulk of the work consists in calculating how various states evolve according to the Schrödinger dynamics.

The Schrödinger equation cannot be *all* there is to say, however. According to the equation, the vast majority of physical states will soon evolve into a superposition of a wide range of states. But this does not square with our observations of the world. When we measure the position of a particle, we find a definite value, not the superposition of values that the Schrödinger equation would predict. If the Schrödinger equation were all there is to quantum dynamics, then even at the macroscopic level the world would evolve into a wildly superposed state. But in our experience it does not. Pointers have definite locations, moving objects have a definite measurable momentum, and so on. So there must be more to the story: something that leads us from the equation to the sorts of discrete events that characterize our experience.

The second part of the story in the standard formalism is the *measurement postulate* (also known as the collapse or projection postulate). This asserts that under special circumstances, the Schrödinger dynamics do not apply. Specifically, it says that when a *measurement* is made, the wave function *collapses* into a more definite form. The way that it collapses depends on the property that is being measured. For example, if we measure the spin of a particle, then even if it is in a superposed state beforehand, it will collapse into a state in which the spin is either up or down. If we measure the position of a particle, its wave function will collapse into a state with a definite position.[2] The resulting state still corresponds to a wave function, but it is a wave function in which all the amplitude is concentrated at a definite position; the amplitude everywhere else is zero. To every quantity that we might measure there corresponds an operator, and upon measurement the state will collapse into a state which is an *eigenstate* of that operator. An eigenstate of an operator is always a state in which the corresponding measurable quantity has a definite value. It follows that when we make a measurement of a quantity, a definite value for that quantity always results, which squares precisely with our experience.

The dynamics of collapse are probabilistic rather than deterministic. If a particle is in a state that is a superposition of positions, then when position is measured we know that it will collapse into a state with definite position, but we do not know what that position will be. Rather, for each potential collapsed state, the measurement postulate specifies the *probability* that the system will collapse into that state. This probability[3] is given by the square of the amplitude of the wave function, at the location corresponding to the definite value in question. For example, if the spin of a particle is a superposition of spin up (with amplitude $\frac{1}{2}$) and spin down (with amplitude $\frac{\sqrt{3}}{2}$), then when spin is measured it will collapse into a spin-up state with probability $\frac{1}{4}$, and into a spin-down state with probability $\frac{3}{4}$. The amplitudes in a wave function always have the property that the corresponding probabilities add up to 1.

3. Interpreting Quantum Mechanics

Together, these two principles constitute an extremely powerful calculus for predicting the results of experimental measurements. To predict the results of an experiment, we express the state of a system as a wave function, and calculate how the wave function evolves over time according to the Schrödinger equation, until the point where a measurement is made. Where a measurement is made, we use the amplitudes of the calculated wave function to determine the probability that various collapsed states will result, and to calculate the probability that the measurement will yield any given quantity. Experimental results have been unwavering in their support for the predictions of the theory; few scientific theories have been as successful at a predictive task. As a calculus, the theory is all but watertight.

The problems arise when we ask *how* it could be that the calculus works. What could be happening in the real world to make the predictions of the calculus so accurate? This is the problem of the *interpretation* of quantum mechanics. There are many different options available in grappling with this problem, none of which is wholly satisfactory.

Option 1: Take the calculus literally

The natural first reaction is to take the formalism of quantum mechanics at face value, as we do with most scientific theories. The calculus involves a wave function governed by the dynamics of the Schrödinger equation and the measurement postulate, and the calculus works, so we should suppose that it gives us a direct picture of what is going on in the world. That is to say, the state of a system in reality is precisely the wave state expressed by the wave function, evolving according to the dynamics expressed by the two basic principles. Most of the time, the state evolves according to the Schrödinger equation, but when a measurement is made, the state evolves according to the measurement postulate. On this view, the world consists of waves that usually evolve linearly in a superposition, and that occasionally collapse into a more definite state when a measurement is made.

But it is not easy to make sense of this picture. The problems all stem from the measurement postulate. According to this postulate, a collapse occurs when a measurement is made, but what counts as a measurement? How does *nature* know when a measurement is made? "Measurement" is surely not a basic term in the laws of nature; if the measurement postulate is to be remotely plausible as a fundamental law, the notion of measurement must be replaced by something clearer and more basic. If wave function collapse is an objectively existing process in the world, then we need clear, objective criteria for when it occurs.

One solution that is obviously unsatisfactory is to say that a collapse occurs whenever a quantum system interacts with a *measuring apparatus.* The problem here is that it is just as implausible that the notion of "measuring apparatus" should appear in the basic laws as it is that the notion of "measurement" should. Before, we needed criteria for what counts as a measurement; now, we need criteria for what counts as a measuring apparatus.

A suggestion popular in the early days of quantum mechanics was that a measuring apparatus is a *classical* system, and that a measurement occurs whenever a quantum system interacts with a classical system. But this is clearly unsatisfactory. Quantum theory is meant to be a universal theory, and it should apply to processes within a measuring instrument just as much as it applies to processes elsewhere. Unless we are to suppose that there are two fundamentally different kinds of physical objects in the world—a supposition that would require the development of an entirely new theory—then "classical system" cannot be a term in a fundamental law of nature any more than "measurement" can.

A related suggestion is that a measurement occurs whenever a quantum system interacts with a *macroscopic* system. But it is just as clear that "macroscopic" is not a notion that can figure in a basic law. It must be replaced by something more precise: something like "system with mass one gram or greater." It would be extraordinarily arbitrary for something like this to figure in a basic law, however.

There is no physical criterion for collapse that seems remotely acceptable. A criterion cast at the microscopic level—suggesting that collapse takes place when a system interacts with a proton, for example—is ruled out by experimental results. The alternative is that the criterion must involve a higher-level physical property, so that collapse takes place when systems take on a certain high-level configuration. But any such higher-level property would seem arbitrary, and no plausible candidate has ever been proposed. There is also something very odd about the supposition that the Schrödinger dynamics of microscopic systems should be suddenly overridden when those systems happen to find themselves in the context of certain special configurations.

The only remotely tenable criterion that has been proposed is that a measurement takes place when a quantum system affects some being's *consciousness*. Unlike the previous criteria, this criterion is at least determinate and nonarbitrary.[4] The corresponding interpretation of the calculus is reasonably elegant and simple in its form, and it is the only *literal* interpretation of the calculus that has any wide currency. This interpretation was first suggested by London and Bauer (1939), but it is most closely associated with Wigner (1961).

Note that this interpretation *presupposes* mind–body dualism. If consciousness were just another physical property, then all the previous problems would arise. The view would turn into another "high-level property" view, on which the wave functions of physical systems just happen to collapse in the context of certain complex physical configurations. If dualism holds, on the other hand, then the criterion for collapse can be truly fundamental. Further, the fact that the cause of collapse is external to physical processing allows for a much simpler theory. All purely physical systems are now governed by the Schrödinger dynamics alone, and the very different measurement dynamics have an independent source.

The interpretation has some counterintuitive consequences, though. Take a measuring apparatus such as a pointer that measures the state of an electron, and suppose that the state of the electron is initially superposed. If there is no consciousness in the vicinity, the whole system is governed by the linear Schrödinger dynamics: given that different discrete electron states would produce different discrete pointer states, it follows that a superposed electron state will produce a *superposed* pointer state. That is, the theory predicts that the pointer is pointing to many different locations simultaneously! It is only when I *look* at the pointer that it points to a definite position.

The scenario of Schrödinger's cat brings on even stranger consequences. In this scenario a cat is locked inside a cabinet, an electron's spin is measured by an instrument, and an apparatus is set up so that the cat is killed if and only if the electron's spin is "up." (Assume that the cat is anesthetized, so that its consciousness does not enter the picture.) If the electron is initially in a superposed state, then the cat will move into a state that is a superposition of life and death! Only when a conscious being looks inside the cabinet will it become determinate whether the cat is dead or alive.

In this picture, *any* macroscopic system will usually be in a large-scale superposition if there is no consciousness in the vicinity. Before consciousness evolved, the entire universe was in a giant superposition, until presumably the first speck of consciousness caused its state to suddenly collapse. This may sound crazy, but it is a direct consequence of the only tenable literal interpretation of the principles of quantum mechanics. I hope this helps to bring out just how strange quantum mechanics is, and how serious the problems posed by its interpretation are.

The counterintuitive consequences could perhaps be accepted, but I nevertheless do not advocate this interpretation. For a start, it is incompatible with the view that I have advocated on which consciousness is ubiquitous. If consciousness is associated even with very simple systems, then on this interpretation collapse will happen at a very basic level and very frequently. This is inconsistent with the physical evidence, which requires that low-level superpositions often persist uncollapsed for a significant time. A second problem is that there is nothing approaching a good theory of what *sort* of effect on consciousness brings about collapse, or of what form the resulting collapse will take. There are many different ways this might be spelled out, but no single way of working out the details looks especially compelling.

Other problems stem from the very notion of collapse. For a start, collapse must be *nonlocal*: when two particles have entangled states, measuring the first particle will cause the state of the second to collapse simultaneously. This leads to some tension with relativity theory. For example, it seems that nonlocal collapse requires an appeal to a privileged reference frame. Without such a reference frame, the time of collapse of the second particle will be underdetermined, as simultaneity across locations is not well defined.

More generally, the whole process of collapse sits uneasily with the rest of physics. Taken literally, it is an instantaneous, discontinuous, temporally asymmetric, nonlocal process that is entirely unlike every other process that physical theory gives us reason to believe in. It seems odd that such a strange process should exist alongside the straightforward, continuous, temporally symmetric, local Schrödinger equation. Indeed, compared to the elegance and power of the Schrödinger equation, which is at the heart of quantum theory, collapse seems almost to be an arbitrary, tacked-on element. There is something very awkward about the idea that the world has two such entirely different sorts of dynamics at its basic level.

These are far from knockdown arguments, of course, and the interpretation on which consciousness collapses the wave function deserves to be taken very seriously. Nevertheless, I think there is good reason to look for another interpretation, one that gives us a simpler and more straightforward view of nature's basic processes.

Option 2: Try to get the measurement postulate for free

The problems with the literal interpretation all stem from taking the measurement postulate as a fundamental law. It is tempting to suppose that instead the postulate might be *nonbasic,* a consequence of more fundamental principles. There are two ways this might go. We might try to introduce *further* basic principles, less problematic than the measurement postulate, that have the same effect. This is the strategy of option 4. Or we might try to derive the effects as a consequence of known basic principles, such as the Schrödinger equation. That is, we might try to get the measurement postulate for free.

It is easy to see the intuitive motivation for this strategy. There is an intuition that superposition effects apply primarily at a microscopic level and might somehow "cancel out" at the macroscopic level. Perhaps when there are many microscopic superpositions, they interact in such a way to produce a macroscopic state that is relatively definite. Because of some mathematical properties of complex configurations, we might be able to see how an *effective* collapse could be the consequence of microscopic indefiniteness. A fundamental probabilistic collapse would then be replaced by an emergent statistical process in a complex system.

There have been numerous attempts to work out the mathematics of this, often appealing to the statistical principles of thermodynamics (e.g., Daneri, Loinger, and Prosperi 1962). Unfortunately, all these attempts have failed, and it is now widely accepted that they *must* fail. Because the Schrödinger dynamics are linear, it is always possible to construct situations in which microscopic superpositions lead to macroscopic superpositions. If an "up" electron leads to one macroscopic state, and a "down" electron leads to another, then a superposed electron must lead to a superposed macroscopic state (Albert 1992, p. 75, gives a very straightforward argument for this point). Unless further basic principles are introduced, then, we have to expect superposition on the macroscopic level.

These strategies can offer something. This sort of appeal to statistics, as well as more recent work on "decoherence" by Gell-Mann and Hartle (1990) and others, suggests that a superposed wave function will often resolve into a relatively clearcut superposition of distinct macroscopic states, rather than being a jumbled mess. These macroscopic states "decohere" from each other, with only minimal interference effects between them. This at least helps us find some element of the familiar classical world in the superposed wave function. But the wave function is still a superposition, and nothing in this sort of work tells us why only one element of the macroscopic superposition should be actual. So more work is required in order to solve the basic problem. This sort of work is perhaps most useful when combined with one of the other options, such as option 5.

Option 3: Whereof one cannot speak . . .

Perhaps the dominant view among working physicists is that one simply should not ask what is going on in the real world, behind the quantum mechanical calculus. The calculus works, and that is that. There are two versions of this view. According to the first version, maybe *something* is going on in the world, but we can never know what it is. The calculus gives us all the empirical information that we will ever have, so that anything further is pure speculation. We might as well stop worrying and continue to calculate. This view makes sense for practical purposes, but it is unsatisfying for anyone who wants physics to tell us about the basic level of reality. Given that the calculus works, we want to have at least some idea of *how* it could possibly work. Perhaps we can never know for sure, but it makes sense to ask.

The second version takes a harder line, and says that there is no fact of the matter about what is going on in the world. According to this view, the facts are exhausted by the fact that the calculus works. This view is often not put forward quite as explicitly as this, perhaps because put so straightforwardly the view is almost impossible to believe. It offers us a picture of reality that leaves out the world! It leads to a version of idealism, on which all that exists are our perceptions, or to something very close to this. Before we open the cabinet containing Schrödinger's cat, it is not in a dead state, it is not in an alive state, and it is not in a superposed state; it is simply in no state at all. By giving up on a fact of the matter about what lies behind our measurements, this view gives up on an independently existing reality.

The "Copenhagen interpretation" due to Bohr and his colleagues is often taken to be a version of this view, although Bohr's writings are somewhat ambiguous and interpretation is not easy. These writings also sometimes suggest elements of the first option, and of the epistemological version of this option. Bohr put great emphasis on the "classical" nature of a measuring apparatus, and his views can be read as suggesting that only classical (or macroscopic) objects have an objective state. Questions about the real state of an object described by a superposition are simply proscribed. But this relies on a division between classical and quantum systems that is difficult to draw on any objective grounds, and it is hard to imagine that reality simply "fades out" as we descend from the macroscopic to the microscopic level. It has seemed to many that if Bohr's view is taken seriously, it leads to the strong operationalism discussed in the last paragraph. Like that view, it offers a picture of the basic level of reality that is no picture at all.

Option 4: Postulate further basic physical principles

Given that the literal interpretation of the measurement postulate is unacceptable, and that it cannot be derived from existing physical principles, it is natural to suppose that something more must be going on. Perhaps if we postulate *further* basic physical principles, we might be able to explain the effectiveness of the quantum-mechanical calculus in a less problematic way.

The first way to do this is to retain the idea of collapse, but to explain it differently. Such a strategy retains the assumption that basic physical states are wave functions governed by the Schrödinger equation, but introduces new principles to explain how microscopic superpositions turn into macroscopic discreteness.

The best-known example of this strategy is the "GRW" interpretation due to Ghirardi, Rimini, and Weber (1986; see also Bell 1987a).[5] This interpretation postulates a fundamental law according to which the position state vector of any elementary particle may undergo a microscopic "collapse" at any moment, with some very small probability (the chance that a particle will collapse in a given second is about one in 10^{15}). When such a collapse occurs, it will generally lead to a collapse of the state of a macroscopic system in which it is embedded, due to nonseparability effects. There are many such particles in any macroscopic system, so it follows that any given macroscopic system at any given time will usually be in a relatively discrete state. It is possible to show that this comes very close to reproducing the predictions of the measurement postulate.

The alternative is to eliminate the need for collapse by denying that the basic level of reality is represented by a superposed wave function. If properties such as position have determinate values even at the basic level, then collapse never needs to happen. Such a theory postulates "hidden variables" at the basic level, which directly explain the discreteness of reality at the macroscopic level. The cost of this suggestion is that new principles are needed to explain why the principles of wave function evolution and collapse *seem* to work so well.

The most prominent example here is the theory developed by Bohm (1952). On this theory, the position of basic particles is always determinate. The wave function retains a role as a kind of "pilot wave," guiding the evolution of a particle's position, and the wave function itself is governed by the Schrödinger equation. The probabilistic predictions of the measurement postulate are reinterpreted as *statistical laws.* It turns out that on this theory we can never know the exact position of a particle before measuring it, but only its wave function. The measurement postulate tells us the *proportion* of particles with a given wave function that will have a given position. It therefore yields the best statistical predictions we can expect, given our ignorance.

All of the proposals in this class have problems. Both the GRW interpretation and the Bohm interpretation give a special determinacy to *position,* thus breaking the symmetry between position and momentum in the quantum mechanical calculus. This makes sense for predictive purposes, as it is arguable that determinate positions always underlie our judgments of macroscopic determinacy (think of the position of a pointer, for instance), but it makes for a more awkward theory. For related reasons, there are serious difficulties reconciling these approaches with relativity theory.

The GRW theory has some further difficulties, perhaps the most serious of which is that it does not strictly imply that the macroscopic world is discrete at all. A macroscopic state is still represented by a superposed wave function: although most of its amplitude is concentrated in one place, the amplitude is nonzero wherever the amplitude of the uncollapsed wave function is nonzero. So the problems of superposition recur. The pointer is still pointing to many locations, even after a measurement. It is true that the amplitude for most of these locations is very small, but it is hard to see why a low-amplitude superposition is any more acceptable than a high-amplitude one.

Bohm's theory has fewer technical problems than the GRW interpretation, but it has some strange consequences. Most strikingly, it is *nonlocal* to an extraordinary degree. (Any hidden-variables theory satisfying the predictions of the calculus must be nonlocal, for reasons given by Bell 1964.)[6] It is not just that the properties of a particle can instantly affect the properties of a particle some distance away. It turns out that in determining the trajectory of a particle, one may have to take into account the wave functions of particles in other galaxies! All of these things play a role in composing the global wave function, and that wave function simultaneously governs the trajectories of particles all over the universe.

Perhaps the most basic reason to be suspicious of these interpretations, however, is that they postulate *complexity behind simplicity.* Whatever its problems, the quantum-mechanical calculus is extraordinarily simple and elegant. These interpretations, on the other hand, introduce complex and relatively *ad hoc* further principles to replace and explain this simple framework. This applies slightly less to the GRW interpretation, whose further complexity consists only in introducing two new fundamental constants and in breaking the symmetry between position and momentum; but it remains the case that it is extraordinarily "lucky" that the values of the constants just happen to be such as to almost reproduce the predictions of the standard framework. The extra complexity of the Bohm interpretation is worse: it postulates determinate positions *and* a wave function, a complex new fundamental principle by which the wave function determines the position of particles, and it breaks the symmetry of the original framework.

We might say that these interpretations make it look like the world was constructed by Descartes's evil demon, as they lead us to believe that the world is one way when really it is another. As Albert and Loewer (1989) put it, the God of the Bohm view does not play dice, but he has a malicious sense of humor. The scenario in which the complex Bohm interpretation happens to duplicate the predictions of the simple framework differs only in degree from the case in which the inputs to a brain in a vat are manipulated to produce the appearance of a straightforward external world. It is reminiscent of an "interpretation" of evolutionary theory according to which God created the fossil record intact a few thousand years ago and ensured that the predictions of evolutionary theory would be duplicated. The simplicity of an explanatory framework has been sacrificed for a complex hypothesis that happens to reproduce the results of the original theory.

The framework of quantum mechanics is so simple and elegant that a basic theory that does not replicate that simplicity and elegance can never be satisfying or fully plausible. If there were a few anomalies in quantum theory, some experimental results that the framework did not predict perfectly, it might be more plausible to think that this simplicity is the tip of a complex iceberg. As it stands, though, the framework is so robust that it seems extraordinary that we should need to postulate a complex apparatus to explain its simple predictions.

Given the problems with *every* interpretation of quantum mechanics, these interpretations need to be taken seriously. But it is natural to look for a simpler picture of the world.

Option 5: The Schrödinger equation is all

The centerpiece of quantum mechanics is the Schrödinger equation, and it is present in some form in every interpretation of quantum mechanics. The various interpretations we have considered all add something to the Schrödinger equation, in order to explain the macroscopic discreteness of the world. But the simplest interpretation by far is the one that says that the Schrödinger equation holds, and that is all. That is, the physical state of the world is completely described by a wave function, and its evolution is completely described by the Schrödinger equation. This is the interpretation given by Everett (1957, 1973).

A strategy canvased earlier (option 2) also held that the Schrödinger equation is all, but argued that this is compatible with discreteness at the macroscopic level. We have seen that this must fail for straightforward mathematical reasons. The Everett interpretation is much more radical. On this view, the Schrödinger equation is taken at face value, and the state of the world at every level is described by a wave function. It follows that contrary to appearances, the world is in a superposed state even at the macroscopic level.

4. The Everett Interpretation

The motivation for this interpretation is obvious. The heart of quantum mechanics is the Schrödinger equation. The measurement postulate, and all the other principles that have been proposed, feel like add-on extras. So why not get rid of them? The problem with this interpretation is equally obvious. If the Schrödinger equation is all, then the world is superposed at every level. But it does not *look* superposed: we never perceive pointers that are in a superposition of two states. Why not?

At the very least, this interpretation is highly counterintuitive. According to this view, not only is the state of an electron best described by a superposition, but so is the state of a pointer that measures it! Objectively, it is not strictly true to say that the pointer is pointing up, or pointing down. Rather, it is in a superposition of the states of pointing up and down. The same goes for the macroscopic state of almost everything, which is in a state described by a wave function that will almost never correspond to a single "discrete" state. Superposition, on this view, is everywhere. Why then does the world appear discrete?

Everett's answer to this question is to *extend superposition all the way to the mind.* If we take Schrödinger's equation seriously, then if the pointer measuring an electron is in a superposition of states, the brain of a person perceiving the pointer will itself be in a superposition. The state of the brain will be described as a superposition of one state in which it perceives a pointer pointing upward, and another state in which it perceives a pointer pointing downward. Everett's key move is to suppose that each of these two states should be associated with a separate observer. What happens after such a measurement is that two observers are produced. One of them experiences an "up" pointer, and the other perceives a "down" pointer. It follows that *each* observer will experience a discrete state of the world.

Everett goes on to show that according to this framework, these observers will have most of the properties that we expect observers to have, and that most of the predictions of the quantum-mechanical calculus can be derived. For example, it is not hard to see that each of the two superposed states here will have no access to the other superposed state, so that the superposition of the mind will not be betrayed in any single state. It is even possible to show that when an observer making a measurement perceives another observer measuring the same quantity, the perceived results of the measurements will be in accord, so that the world will seem quite coherent. In short, any single observer will experience the world in largely the way that we expect, even though the world itself is in a superposed state.

This interpretation should not be confused with the *splitting-worlds* interpretation, according to which the world literally splits into many separate worlds every time a measurement is made. There is one world in which the pointer is pointing up, and an entirely separate world in which the pointer is pointing down. Taken this way, the view is the furthest thing from simple. For a start, it requires a new and extraordinary basic principle to describe the "splitting" process. It is far from clear just when "splitting" should take place (the "measurement" problem revived in a new form), and it is very unclear what the worlds resulting from a split should be. For a literal split to happen, a wave function has to "divide" into numerous components; but there are many ways to decompose a wave function, and quantum mechanics delivers no preferred basis for the decomposition. This interpretation seems even more complex and *ad hoc* than the various "collapse" interpretations, and there is little reason to accept it.

The splitting view is frequently attributed to Everett (largely due to the expositions of Everett's work by DeWitt (1970, 1971), but it cannot be found in his writing. Everett's own view is not entirely clear, but it is interpreted much more naturally along the lines I have suggested; this interpretation is also recommended by Albert and Loewer (1988) and by Lockwood (1989). On this view, there is no objective "splitting." Rather, the wave function evolves into a superposition of states, where the superposed states are best regarded as components of a single world. Everett's view is sometimes called a *many-worlds* interpretation (thus suggesting the splitting-worlds view), but the view I am discussing is more accurately a *one-big-world* interpretation. There is only one world, but it has more in it than we might have thought.[7]

On this view, if there is any splitting, it is only in the minds of observers. As superpositions come to affect a subject's brain state, a number of separate minds result, corresponding to the components of the superposition. Each of these perceives a separate discrete world, corresponding to the sort of world that we perceive—call this a *miniworld,* as opposed to the *maxiworld* of the superposition. The real world is a maxiworld, and the miniworlds are merely in the minds of the subjects. Everett calls his view a *relative-state* interpretation: the state of a miniworld, in which pointers point to discrete positions, only counts as the state of the world *relative* to the specification of an observer. The objective state of the world is a superposition.

A key element is left unanalyzed in this interpretation, however. Why is it legitimate to identify each component of an associated brain state with a separate observer? Why is there not instead a single observer with a confused, superposed mental state? Why indeed does such an incoherent brain state give rise to any minds at all? Everett's treatment skates over these crucial questions. Indeed, it may seem that in associating the wave function of a brain state with a number of minds each perceiving a discrete state, Everett is making an illegitimate appeal to a *preferred basis,* just as the splitting-worlds interpretation did. A wave function does not come with an objective division into components, but can be decomposed in many ways, depending on the choice of a basis for the corresponding vector space. It is often natural for our purposes to decompose a wave function one way, according to a particular basis, but such a decomposition does not reflect an objective property of the wave function. Where the brain state can be decomposed into a "perceiving up" and a "perceiving down" state, it can equally be decomposed into two states each of which have confused perceptions. In postulating an objective decomposition, Everett seems to go beyond the resources that the Schrödinger equation provides.

The crucial element omitted from Everett's treatment is an analysis of the relationship between mind and body. Everett assumes that a superposed brain state will have a number of distinct subjects of experience associated with it, but he does nothing to justify this assumption. It is clear that this matter depends crucially on a theory of consciousness. A similar suggestion is made by Penrose (1989):

> In particular, I do not see why a conscious being need be aware of only "one" of the alternatives in a linear superposition. What is it about consciousness that demands that one cannot be "aware" of that tantalizing linear combination of a dead and a live cat? It seems to me that a theory of consciousness would be needed before the many-worlds view can be squared with what one actually observes. (p. 296)

Indeed, it is possible to see the central question in quantum mechanics as a question about the relationship between physical processes and experience. The centerpiece of quantum mechanics is the picture in which microscopic reality is described by a superposed wave function evolved according to the Schrödinger equation. But we *experience* the world as discrete. The central question is, how is this so? Different interpretations give different answers. Some (such as Bohm's) deny the first premise, positing that reality is discrete even at the basic level. Some posit basic principles (the measurement postulate, or the GRW collapse law) to mediate a transition from the superposed to the discrete. Some theories (those of option 2) try to explain how superposed microscopic states can statistically produce a discrete macroscopic reality.

These last three strategies are all *indirect* strategies, attempting to explain the discreteness of experience by explaining an underlying discreteness of macroscopic reality.

An alternative strategy is to answer the question about experience *directly*. If we take the primacy of the Schrödinger equation seriously, the central question is why, given that the physical structure of the world is like *this*, do we experience it like *this*? This is precisely a question about the way that certain physical structures give rise to experience. That is, it is the kind of question that I have been discussing throughout this book, and it is the kind of question that a theory of consciousness ought to be able to answer.

If we have to postulate an *ad hoc* theory of consciousness to answer this question, the attractiveness of the Everett interpretation is diminished significantly. Its best feature was always its simplicity, but new and arbitrary psychophysical laws would make it as *ad hoc* as the Bohm interpretation. If on the other hand an *independently motivated* theory of consciousness can answer the question, then the Everett interpretation begins to look attractive indeed.

The theory of consciousness that I have advocated can answer this question, and can give the right sort of answer. It turns out that the theory *predicts* that a superposed brain state should be associated with a number of distinct subjects of discrete experience. To see this, let a *maximal phenomenal state* be a phenomenal state that characterizes the entire experience of a subject at a given time. Let a *maximal physical state* be a physical state that fully characterizes the intrinsic physical state of a system at a given time. To establish the conclusion, it suffices to establish the following *superposition principle*:

> If the theory predicts that a system in maximal physical state *P* gives rise to an associated maximal phenomenal state *E*, then the theory predicts that a system in a superposition of *P* with some orthogonal physical states will also give rise to *E*.

If this principle holds, then a superposition of orthogonal physical states will give rise to at least the maximal phenomenal states that the physical states would have given rise to separately. This is precisely what the Everett interpretation requires. If a brain is in a superposition of a "perceiving up" state and a "perceiving down" state, then it will give rise to at least two subjects of experience, where one is having an experience of a pointer pointing upward, and the other is experiencing a pointer pointing downward. (Of course, these will be two *distinct* subjects of experience, as the phenomenal states are each maximal phenomenal states of a subject.) The same holds in the general case. A superposition will always give rise to the ensemble of subjects that the Everett interpretation requires.

So we need to establish that the theory I have outlined implies the superposition principle. The easiest way to see this is to appeal to the framework of Chapter 9, and in particular to the claim that consciousness arises from implementation of an appropriate computation. To use this to establish the principle, we need to establish that if a computation is implemented by a system in maximal physical state *P*, it is also implemented by a system in a superposition of *P* with orthogonal physical states.

Accordingly, assume that the original system (in maximal physical state *P*) implements a computation *C*. That is, there is a mapping between physical substates of the system and formal substates of *C* such that causal relations between the physical substates correspond to formal relations between the formal substates. Then a version of the same mapping will also support an implementation of *C* in the superposed system. For a given substate *S* of the original system, we can find a corresponding substate *S′* of the superposed system by the obvious projection relation: the superposed system is in *S′* if the system obtained by projecting it onto the hyperplane of *P* is in *S*. Because the superposed system is a superposition of *P* with orthogonal states, it follows that if the original system is in *S*, the superposed system is in *S′*. Because the Schrödinger equation is linear, it also follows that the state-transition relations between the substates *S′* precisely mirror the relations between the original substates *S*. We know that these relations in turn precisely mirror the formal relations between the substates of *C*. It follows that the superposed system also implements *C*, establishing the required result. By the principle of organizational invariance, if the original system gives rise to a subject of experience, the superposed system will give rise to a qualitatively indistinguishable subject of experience.

It may also be possible to argue for the superposition principle by applying the double-aspect theory of information and arguing that the relevant information embodied in the original physical state is also present in the superposition. Because of the underdetermination of that theory, however, this argument is less clear than the previous one, so I will not go into it here. What matters is that one way or another, the theory of consciousness that I have partially developed *predicts* the result that the Everett interpretation requires. That is, it predicts that even if the world is in a giant superposition, there will still be subjects who experience a discrete world.

If there are no other problems, it follows that a combination of the Schrödinger equation with an independently motivated theory of consciousness can predict our manifest image of the world. That is, the only physical principle needed in quantum mechanics is the Schrödinger equation, and the measurement postulate and other basic principles are unnecessary baggage.

To be sure, we need psychophysical principles as well, but we need those principles in any case, and it turns out that the principles that are plausible on independent grounds can do the requisite work here. This adds up to a powerful argument for taking the Everett interpretation seriously.

5. Objections to the Everett Interpretation

Everett's interpretation has come under frequent attack in the literature, with some objections more powerful than others. I will group these objections into a number of classes.

Objections based on "splitting"

Many objections arise from interpreting or misinterpreting Everett's view as a "splitting-worlds" view. This is understandable, given that it is often called the "many-worlds" interpretation. For example, Bell (1976) objects that it is unclear when a "branching" event should take place, due to unclarity in the notion of measurement, and that there is no preferred basis for the division into worlds. It is clear that these objections do not apply to the present interpretation, which requires no objective "branching" and no preferred basis. Similarly, Hughes (1989) objects to the "ontological cloudburst" in the splitting process, and Healey (1984) notes that the creation of new worlds violates the conservation of mass-energy! It is a pity that the "splitting" interpretation of Everett's view has gained such wide currency, for its obvious difficulties have meant that the more interesting interpretation has not received the attention it deserves.

Objections to a preferred basis

Some of the objections to the splitting-worlds interpretation arise from its need for a preferred basis, but so also do some objections to the single-world version. In particular, the question arises: Why do the only minds associated with a superposed brain state correspond to its decomposition along the preferred basis? Why are there not minds that arise from other decompositions, or indeed from the superposed state as a whole? This is a reasonable objection to Everett's own version, which seems to require such a canonical decomposition. No such objection arises for the version I have outlined, however, which entails that a superposition gives rise to the associated subjects of discrete experience without any need to postulate a preferred basis. And I have had no need for the assumption that these are the *only* minds that the superposed system gives rise to.

What about superposed minds?

The question arises, "Are there other minds associated with a superposition?" To this the answer is "perhaps." If the double-aspect theory of information is accepted, then we already know that there may be experiences associated with lower-level processes in such a system. It may also be that there are subjects of experience associated with the structure of processing in a superposition. Perhaps there are minds associated with other decompositions of the system. Perhaps there is a big superposed mind associated with the whole superposed system. The existence of such minds depends on the details of a theory of consciousness, but it is hard to see how their existence is a problem.

One might try to parlay the possibility of superposed minds into an objection to the theory. Objection: Why is *my* mind not superposed? Answer: Because I am who I am. The theory predicts that nonsuperposed minds exist, and my mind happens to be one of them. To ask why my mind is not one of the superposed minds is like asking why I am not a mouse. It is simply part of the brute indexicality of my existence. Mouse minds exist, and superposed minds may exist, but my mind is not one of them. Objection: Why don't I have any access to superposed minds, such as memories of superposed experiences? Answer: The theory predicts that the discrete minds in question will experience the world as entirely discrete, and they will have no direct access to other parts of the superposition. All their memories will be of discrete observations, for example.

It is arguable, in any case, that the only *interesting* minds associated with a superposed system are the familiar sort of discrete minds. These minds are complex and coherent, with experience reflecting the structure of rational processes. Any further minds that are associated will be relatively incoherent, without much in the way of interesting structure. This conclusion is lent support by the "decoherence" framework of Gell-Mann and Hartle (1990) and others. According to this framework, the interesting structure in a wave-like, complex adaptive system is generally found within the components of a "natural" decomposition; the system "decoheres" naturally along certain lines. In rational systems, then, coherent cognitive structure may be found only in the components of this natural decomposition, and only these will give rise to complex, coherent minds. Any other subjects of experience in the system will not be the sort of subjects that qualify as persons.

Objections based on personal identity

There is a cluster of intuitive worries based on the identity of the observer. Take the mind M_1 that I remember being around this time yesterday. Today, there will be a large number of minds descending from that mind, in different "branches" of the superposition. My mind M_2 is only one of them. I might well ask: Why did I end up *here,* rather than in one of the other branches? As Hofstadter (1985b) puts it:

> Why is my unitary feeling of myself propagating down *this* random branch rather than down some other? What *law* underlies the random choices that pick out the branch I feel myself tracing out? Why doesn't my feeling of myself go along with the other me's as they split off, following other routes? What attaches *me-ness* to the viewpoint of this body evolving down this branch of the universe at this moment in time?

To this, we must again invoke brute indexicality: my mind is *this* one, and that is that. There is feeling that something deeper must be going on, and that it is somehow a deep fact about the world that yesterday's mind M_1 has evolved into today's mind M_2 and not one of the others. But from an objective point of view, there is nothing especially privileged about this branch. Even from the point of view of M_1, all of today's minds are equally privileged. None of them is the single rightful heir of M_1; all of them carry M_1's "me-ness" to the same degree. It is only from *this* point of view, the point of view of M_2, that M_2 seems privileged (of course, my counterparts elsewhere in the superposition have the same feeling about themselves). This privileged role of M_2 is just another indexical phenomenon, like the fact that I am David Chalmers rather than Rolf Harris. *This* mind is here rather than there. It is as puzzling as indexical facts usually are, but there is no further asymmetry in the world.

There is a strong intuition that there must always be a fact of the matter about personal identity: if there are numerous minds descending from my current state, there must be a fact about which one of them will be *me.* But this idea has been subjected to a powerful critique by Parfit (1984), who argues persuasively that there is no more to the fact of personal identity than facts such as psychological continuity, memory, and the like. If we accept this analysis, then each of tomorrow's minds are equal candidates to count as *me,* and there is no fact to distinguish them. There is something disturbing about this conclusion, which reduces the determinate "flow" of personal identity to an illusion, but Parfit's analysis gives reason to believe that this determinate flow was an illusion all along.

The interpretation of probabilities

The most substantial objection to the Everett interpretation is that it cannot make sense of the *probabilities* that the measurement postulate delivers.[8] In a given case the measurement postulate may tell us that on making a certain measurement, there will be a 0.9 chance of finding an "up" pointer and a 0.1 chance of finding a "down" pointer. According to the Everett interpretation, what really happens is that both the pointer and the brain state of an observer go into a superposition, with (at least) two subjects of experience resulting. One of these has an experience of an "up" pointer, and another experiences a pointer pointing down. Exactly the same would have happened if the probabilities had been 50:50. It is true that in the 90:10 case, most of the *amplitude* of the superposed wave function is concentrated in the area of the "up" brain state, but what does this have to do with probabilities?

Everett deals with this question by placing a *measure* on the space of observers, corresponding to the probabilities delivered by the measurement postulate (i.e., corresponding to the square of the amplitude of the corresponding part of the wave function). Using this measure, he argues that in the limit, *most* observers (that is, a subset of observers with measure one) will have memories of observations that accord with the frequencies predicted by the probabilities in the measurement postulate. For example, among observers who have made a measurement like the one described above many times, most of them will remember finding an "up" pointer 90 percent of the time and a "down" pointer 10 percent of the time. Thus a role is found for the probabilities. The question arises, however: What justifies this measure on the space of observers? If we measured the space differently, then very different frequencies might arise. For example, if we assigned equal measures every time two observers arise from a superposition (regardless of amplitude), then most observers would recall an "up"–"down" ratio of 50:50. Neither the Schrödinger equation nor the psychophysical laws ensures that either of these measures is the "correct" one.

Albert and Loewer (1988) respond to this worry by dispensing with measures. Instead they postulate more radical psychophysical laws, according to which there is an infinity of minds associated with every brain state. For every mind postulated by the previous view, this theory postulates an infinite ensemble of qualitatively identical minds. Further, wherever the Everett theory predicts that a mind will diverge into two minds, this theory says that any given mind will go in one direction or the other, with the probabilities given by the measurement postulate. So, if we take an arbitrary mind associated with the brain state before the measurement above, it will have a 90 percent chance of evolving into a "perceiving up" state and a 10 percent chance of evolving into a "perceiving down" state. This way the probabilistic predictions of the quantum-mechanical calculus are preserved.

There is clearly a loss in simplicity here. The new psychophysical laws have no independent motivation, and the theory also needs extra "intrapsychic" laws governing the *evolution* of minds. By making these *ad hoc* postulates, the theory sacrifices some of the key virtues of the Everett interpretation. It is also arguable that the intrapsychic laws are problematic, in that they postulate deep irreducible facts about personal identity over time. It is hard to know what to make of these facts. Accepting them would require discarding Parfit's analysis of personal identity, for example. They fail to supervene even naturally on physical facts, and so complicate the metaphysical picture. This interpretation should be kept in mind as a possibility, but it comes at a significant cost.

The alternative is to do without the extra apparatus, and to see if the probabilities can be recovered some other way. It is tempting to see this as a problem about indexicality. Why is it that of all the places in the wave function that I could have ended up, I ended up in a region where my memories match the predictions of the calculus? One possibility is simply to take this as a brute indexical fact: *some* minds are in this area, and I happen to be one of them. But this seems unsatisfying, as the remarkable regularity of the calculus turns out then to be a huge fluke. What we need is some way to argue that it is not such a fluke.

Even in noting that it is a fluke that I ended up *here,* the idea is implicit that there is some kind of measure on the space of minds. The suggestion is that it is antecedently more likely that I should end up being a mind of one type rather than another, perhaps because of the relative abundance of those classes. This sort of implicit measure is present in much of our reasoning about the world. When I reason inductively from some evidence to a conclusion, I know that for *some* observers in a similar epistemic position the conclusion will not hold, but I assume that for *most* such observers the conclusion will hold, even if there are an infinite number in each class. That is, I assume that it is antecedently more likely that I will turn out to be in one class rather than another. This sort of reasoning implicitly supposes some kind of measure on the space of minds.

Perhaps we can justify the probabilities, then, by explicitly introducing this sort of measure. Certainly the bulk of the amplitude of the wave function is concentrated in areas where the memories of observers match the predictions of the calculus. Maybe it is more likely that my mind should turn out to be in a high-amplitude area than in a low-amplitude area. In particular, if we assume that the antecedent likelihood that I will turn out to be one mind rather than another is proportional to the squared amplitude of the associated part of the wave function, then it follows that I will almost certainly turn out to have memories in the frequencies predicted by the quantum-mechanical calculus.

But to what does this measure objectively correspond? Does it need to be taken as a basic fact about the distribution of selves? Can it somehow be justified as the canonical measure on this space? These are difficult questions that are closely tied to the mystery of indexicality itself—why did I turn out to be *this* person rather than someone else? This is one of the basic mysteries, and it is very unclear just how it should be answered. Nevertheless, the idea of a measure on the space of minds seems to have some promise, and may even be needed for some other purposes, such as the justification of induction. In the meantime, the interpretation of the probabilities remains the most significant difficulty for the Everett interpretation.

6. Conclusion

It must be admitted that the Everett interpretation is almost impossible to believe. It postulates that there is vastly more in the world than we are ever aware of. On this interpretation, the world is really in a giant superposition of states that have been evolving in different ways since the beginning of time, and we are experiencing only the smallest substate of the world. It also postulates that my future is not determinate: in a minute's time, there will be a large number of minds that have an equal claim to count as *me.* A minute has passed since I wrote the last sentence; who is to know what all those other minds are now doing?

On the other hand, it is clear by now that *all* interpretations of quantum mechanics are to some extent crazy. That is the fundamental paradox of quantum mechanics. The three leading candidates for interpretation are perhaps Wigner's interpretation on which consciousness brings about collapse, Bohm's nonlocal hidden variables interpretation, and the Everett interpretation. Of these, Wigner's interpretation implies that macroscopic objects are often in superpositions, until a casual look from an observer causes them to collapse. Bohm's view implies that the trajectory of every particle in the universe depends on the state of every other. And the Everett view implies that there is much more in the world than we ever would have thought.

Of these, perhaps Bohm's view is the least crazy and Everett's the most, with Wigner's in between. Ranked in order of theoretical virtue, on the other hand, the sequence is reversed. Bohm's view is unsatisfying due to its complex, jury-rigged nature. Wigner's view is quite elegant, with its two basic dynamical laws mirroring the quantum-mechanical calculus, if all the details can be worked out. But Everett's view is by far the simplest. It postulates only the Schrödinger equation, the principle that is accepted by all interpretations of quantum mechanics. It also has the virtues of being an entirely local theory, and of being straightforwardly compatible with relativity theory, virtues that the other interpretations lack.

It is also worth noting that both of the other interpretations contain elements of what is counterintuitive about the Everett interpretation. On the Wigner view, we must accept that the universe evolved in an Everett-style giant superposition—perhaps with superposed stars and superposed rocks, if not with superposed cats—at least until the first conscious entity evolved to collapse the wave function. On the Bohm view, Everett's uncollapsed wave function remains present as the "pilot wave" that guides the position of the various particles. All the structure that is present in other components thus remains present in the state of the world, even though most of it is irrelevant to the evolution of the particles. Given that these views, too, require an uncollapsed wave function in central roles, one might argue that the relative implausibility of the Everett view is diminished.

Of course, it is always possible that a new theory might be developed that surpasses all of these in plausibility and theoretical virtue. But it does not seem especially likely. The complete absence of experimental anomalies suggests that the quantum-mechanical calculus is here to stay as a predictive theory. If so, we cannot expect empirical developments to solve the problem. Perhaps conceptual developments could lead to a new and improved interpretation, but it may be that by now the most promising niches in conceptual space have already been exploited. If so, we may be stuck with something like the current range of options—perhaps with significant refinements, but with advantages and disadvantages of a qualitatively similar kind. Of these options, the Everett interpretation seems in many ways the most attractive, but at the same time it is the hardest to accept.

I have advocated some counterintuitive views in this work. I resisted mind–body dualism for a long time, but I have now come to the point where I accept it, not just as the only tenable view but as a satisfying view in its own right. It is always possible that I am confused, or that there is a new and radical possibility that I have overlooked; but I can comfortably say that I think dualism is very likely true. I have also raised the possibility of a kind of panpsychism. Like mind–body dualism, this is initially counterintuitive, but the counterintuitiveness disappears with time. I am unsure whether the view is true or false, but it is at least intellectually appealing, and on reflection it is not too crazy to be acceptable.

The craziness of the Everett interpretation is of another order of magnitude. I find it easily the most intellectually appealing of the various interpretations of quantum mechanics, but I confess that I cannot wholeheartedly believe it. If God forced me to bet my life on the truth or falsity of the doctrines I have advocated, I would bet fairly confidently that experience is fundamental, and weakly that experience is ubiquitous. But on the Everett interpretation I would be torn, and perhaps I would not be brave enough to bet on it at the end of the day.[9] Maybe it is simply too strange to believe. Still, it is not clear whether much weight should be put on these intuitive doubts in the final analysis. The view is simple and elegant, and it predicts that there will be observers who see the world just as I see it. Is that not enough? We may never be able to accept the view emotionally, but we should at least take seriously the possibility that it is true.