Quantum mechanics and the soul

Hans Halvorson*

June 3, 2008

Contents

*Department of Philosophy, Princeton University, Princeton NJ 08544

1 Introduction

The twentieth century saw several significant developments in our understanding of the physical world. One of the most significant of these developments was the introduction of what has become known as "quantum mechanics." Quantum mechanics is a replacement for the classical physics of Newton, Maxwell, and Einstein, which proved inadequate to handle the full range of experimental data that had been collected by the end of the nineteenth century.

In science, theoretical revisions come in varying degrees of significance. At one end of the spectrum, there are minor revisions, such as the change of a numerical parameter. A more significant sort of revision occurs when we predict or discover the existence of a new entity, for example a new planet, or a new type of elementary particle. But the move from classical to quantum physics is a more profound sort of revision: the replacement of classical physics by quantum physics requires modifications of our worldview at the deepest levels of ingression; or as philosophers might say, it requires a modification of our fundamental metaphysics.

That much is clear. But philosophers and physicists disagree quite widely about which modifications of our worldview are required by quantum mechanics. For example, some claim that quantum mechanics proves that we live in an indeterministic world, with an open future. Others make the more radical claim that quantum mechanics shows that there are a multitude of parallel universes, and that each time a measurement is made, our universe branches again. Still others claim that the lesson of quantum mechanics is that science does not describe an objective world independent of the observations we make of it.

The main goal of this chapter is to put forward an alternative proposal for how to integrate quantum mechanics into our worldview. Let me be clear about my methodology: I do not think that it is feasible that one can approach quantum mechanics from a standpoint of metaphysical neutrality, and expect it to yield a system of metaphysics. Rather, I think that we must approach quantum mechanics with a stock of background beliefs, and then ask whether or not it is consistent with these beliefs, and whether or not it suggests modifications of these beliefs.[1] For example, these background beliefs might include the belief that there is an external world, or the the belief that the universe did not come into existence (along with all of my memories) one second ago, and the belief that there are conscious persons besides myself.

One of the more controversial background beliefs that I bring to this investigation is the "soul hypothesis" — namely the belief that human beings are more than just their bodies, but are also "living souls." I will argue that, given the soul hypothesis, we can reject some of the more wild and implausible metaphysical speculations based on quantum mechanics.

The remainder of this chapter will proceed as follows. In Section 2, I give an informal sketch of quantum mechanics; in particular, I isolate three central features of the theory that give rise to various "paradoxes." In Section 3, I discuss a less publicized, but much more troubling paradox, the so-called "measurement problem." The measurement problem supposedly shows that an embodied observer (like you or me) could not ascertain facts about the physical world by making observations. Finally, in Section 4, I show how the soul hypothesis leads to a solution of the measurement problem, and I discuss generally how to understand quantum mechanics in the light of the soul hypothesis.

[1]For a discussion or background, or "control" beliefs, see (Wolterstorff, 1984).

2 Basic principles of quantum mechanics

Galileo famously said that, "mathematics is the language with which God has written the universe." Indeed, Galileo and successors such as Descartes, Newton, Maxwell, and Einstein, showed the power of using quantitative descriptions of the physical world. But what is a quantitative description? For understanding the metaphysical lessons of quantum mechanics — at least in the current context — it is not important that quantities are numerical, and it makes no difference whether they take on continuous or discrete values. For present purposes, a **quantity** can be thought of as a homogeneous collection of properties (i.e. they are all properties of the "same type", whatever that means), such that the collection is both mutually exclusive (an object can only have one property in the collection at a time) and exhaustive (at each time, an object has at least one property in the collection). Let's call such a collection a **determinable**, and let's call a specific element in this collection a **determinate**. For example, **position** on the earth is a determinable, and a specific position on the earth (for example, latitude 40.348022, longitude -74.655704) is a determinate value of that determinable. Similarly, **color** is a determinable, and **blue**, **red**, and **green** are determinates of this determinable.

2.1 Classical physics

"Classical physics" is a catch-all phrase for a number of different theories developed roughly between the time of Galileo Galilei (1564–1642) and James Clerk Maxwell (1831–1879). Now, it's not really plausible to say that there is a single "worldview of classical physics." Indeed, attempts to state the central points of such a worldview — e.g. deterministic, atomistic — invariably fail to capture much of the nuance of the development of physics. However, for the purposes of this discussion, we can abstract away from many of the details of the individual theories, and describe some common structural elements. In particular, I would like to focus on what might be called the "logical structure of quantities" in classical physics. For concreteness, I will suppose that we are dealing with a theory with a fixed number of elementary objects (or "particles"). The details can be adjusted for theories with a variable number of particles, or theories (e.g. electromagnetism) where the fundamental objects are fields rather than particles.

According to classical particle mechanics, there is a collection of objects, and each object x has a collection D_x of determinables. In general, an object will have several logically independent determinables. By "logically independent" here I mean that for any two determinate values that these determinables can have, the conjunction of those determinates is also logically possible. For example, the **position** and the **velocity** of a baseball are logically independent in this sense: if it's possible for the baseball to be at a certain location in the air above home plate, and if it's possible for the baseball to travel at sixty miles per hour, then it's possible for a baseball to be simultaneously above home plate and traveling at sixty miles per hour. In contrast, classical statistical physics tells us that the determinable **temperature** is logically dependent on the determinables **position** and **velocity**.

Borrowing terminology from the philosopher Rudolf Carnap (1950), let's say that a **state description** is a specification of a determinate value for each determinable of every object in the universe. In other words, a state description tells us which determinate value each determinable possesses. In classical physics, the state of the world at a time is given by one of these state descriptions.

For future reference, we here catalog two important operations on classical state descriptions (which I will usually just call "states"). By an "operation" on states, I mean a mathematical operation that requires one or more input states, and returns a single output state. The first operation on classical states is the operation of **mixing**. For example, consider the state ψ_H in which a coin lands heads, and the state ψ_T in which a coin lands tails. Then an equal-proportion mixture of these states is the state $\frac{1}{2}\psi_H \circ \frac{1}{2}\psi_T$ in which there is a 50% chance that the coin lands heads, and a 50% chance that the coin lands tails. Such a state description would be appropriate either in cases where we lack precise knowledge of the actual state of the coin, or in cases where there is genuine indeterminism. More generally, for any number λ between 0 and 1, the state $\lambda\psi_H \circ (1-\lambda)\psi_T$ represents a $(\lambda \times 100)\%$ chance that the coin lands heads, and a $((1-\lambda) \times 100)\%$ chance that the coin lands tails. (Note that we have here expanded the meaning of the word "state" so that states do not necessarily fix the values of quantities, but only assign probabilities to various values.)

The second operation on classical states only applies to classical waves (e.g. waves in water, or soundwaves, or electromagnetic waves): given two waves ψ_1 and ψ_2, there is a wave $\psi_1 \star \psi_2$, called the **superposition** of ψ_1 and ψ_2, which results from summing the amplitudes of the waves ψ_1 and ψ_2 at each point in space. Thus, if ψ_1 is a tsunami at 25 feet above sea level, and ψ_2 is a tsunami at 30 feet above sea level, then $\psi_1 \star \psi_2$ is a tsunami at 55 feet above sea level. Note that the operations of mixing and superposing are distinct. On the one hand, a mixture of ψ_1 and ψ_2 would describe our ignorance of which tsunami is actual — either the 25 foot or 30 foot tsunami. In either case, the tsunami is at most 30 feet tall. On the other hand, superposition describes an actual physical combination of the two waves — the superposition of ψ_1 and ψ_2 is definitely over 30 feet tall.

Of course, physics is not just concerned with describing the states of objects at one time: physics is fundamentally concerned with dynamics, i.e. with how things change over time. Thus, a theory of classical physics includes **dynamical laws** that specify how the state of a system at one time is related to the state of that system at past and future times. Now, most theories of classical physics have **deterministic** dynamical laws in the sense that the state of an object at any time completely determines the state of that object at all subsequent times. However, the deterministic nature of classical dynamical laws is not required either by the logical structure of quantities, nor by the requirement that there be dynamical laws. It would make perfect sense to have a classical physical theory according to which the present state does not determine the future state.

2.2 Superposition

Now we have said just enough about the structure of classical physics to see which parts of this structure are preserved by, and which parts of this structure are overturned by quantum mechanics. First, quantum mechanics does of course claim that physical objects have quantitative properties (such as position, momentum, or electric charge). But quantum mechanics — at least on the most common interpretation — entails that quantities sometimes fail to have any value at all. For example, quantum mechanics claims that electrons can get themselves into states where they are not definitely located at any particular position in space.

Why does quantum mechanics say that electrons sometimes fail to have any determinate position? Doesn't this just tell us that quantum mechanics is an incomplete theory, since it leaves out details that seem to be required for a complete description of the physical facts? In order to understand why quantum mechanics makes such a bizarre claim, we need to see that this claim is the consequence of postulating additional structure in the world.

The additional structure in quantum mechanics is an operation on states which behaves in some ways like a mixture of states, and in other ways like a superposition of waves. We will illustrate this new operation by looking at a famous example: the two slit experiment.

Suppose that a stream of electrons is shot at a barrier with two openings, one on the left (L) and one on the right (R). Let ψ_L denote a state where the electron is going through the left slit, and let ψ_R denote a state where the electron is going through the right slit. Then quantum mechanics postulates the existence of a state $\psi_L + \psi_R$, which is traditionally (but misleadingly) called the "superposition" of ψ_L and ψ_R.[2] But what does this "+" operation mean? On the one hand, $\psi_L + \psi_R$ behaves like a superposition of classical waves: If we run the experiment several times, then the fluorescent screen will display the characteristic interference pattern for two waves, one of which emanates from the left slit, and one of which emanates from the right slit. On the other hand, $\psi_L + \psi_R$ *also* behaves like a mixture of ψ_L and ψ_R. In particular, if detectors D_L and D_R are placed, respectively, on the left and right slits in the barrier then: (i) on each run of the experiment, either D_L or D_R, but not both, will register a particle passing through the slit, and (ii) the interference pattern does not appear, but rather the sort of pattern we would expect if each electron was a discrete particle.

But a classical state cannot be both a mixture and a superposition! A classical superposition of two state descriptions is another state description, namely the state description obtained by adding the values that the

[2] Experts might be troubled by the fact that if ψ_L and ψ_R are orthogonal unit vectors, then $\psi_L + \psi_R$ is not a unit vector. For purposes this simplified discussion, I will systematically suppress the leading coefficient of $2^{-1/2}$.

states assign to quantities. Thus, a classical superposition involves no probabilities. In contrast, a mixture assigns probabilities to different possible values of the quantities.

Now, it might not seem metaphysically unreasonable to suppose that electrons sometimes lack a determinate position. Indeed, we are taught – even as high school students — to think of the electrons orbiting the nucleus of an atom as a cloud, where each electron is smeared out in space. Similarly, it does not seem crazy to think that an electron can pass both through the left slit L and the right slit R of a barrier simultaneously — so long as we think that electrons might really be more like waves than like discrete particles. But the notion of a superposition of properties becomes truly puzzling when we realize that, in principle, macroscopic properties can also be superposed. In order to see why macroscopic superpositions are possible, we need to recall how quantum mechanics describes the composition of non-elementary objects.

2.3 Holism

Most people — and even most philosophers — do not spend much time thinking about what quantum mechanics tells us about the sort of world we live in. One reason for this neglect may be a sense that quantum mechanics describes only a very limited domain, namely the subatomic realm. But there is no principled reason to think that the laws of quantum mechanics are restricted to subatomic objects. Indeed, quantum mechanics predicts that if an object's parts obey its laws, then that object will also obey its laws. Thus, anything built out of subatomic particles should also obey the laws of quantum mechanics; the theory does not make any special exceptions for large or heavy objects such as rocks, stars, or the bodies of animals or human beings.

However, quantum mechanics also says some very strange things about the relation between an object and its parts. For example:

> *If $A + B$ is the composite of A and B, then there are states of $A + B$ in which neither A nor B has any determinate properties.*

This fact is interesting and surprising in itself. But it will become really problematic later, when we see that such states arise naturally in situations — such as measurement interactions — where they cause the most trouble. In order to see how these strange holistic states can arise, we need to spell out some more details.

First, if ψ^A is a state of an object A and if ψ^B is a state of an object B, then we let $\psi^A \otimes \psi^B$ denote the state of the composite $A + B$ in which A is in state ψ^A and B is in state ψ^B. For example, if ψ^A is the state in which Adam weighs 170 pounds, and if ψ^B is the state in which Bob weighs 170 pounds, then $\psi^A \otimes \psi^B$ is the state in which both of them weigh 170 pounds. (Supposing that such a state exists is a mild independence assumption, namely that states of A can be freely combined with states of B.) Now for $i = 0$ or $i = 1$, let ψ_i^A be a state of A in which one of its quantities Q^A has value i, and similarly for ψ_i^B. Thus, $\psi_0^A \otimes \psi_0^B$ is the state in which Q^A has value 0 and Q^B has value 0. (For concreteness, you might suppose that A and B are "quantum quarters", and ψ_0 is the state in which the quarter is heads, and ψ_1 is the state in which the quarter is tails. Then in state $\psi_0^A \otimes \psi_0^B$ both quarters are heads.) Then quantum mechanics claims that any two states of the composite object $A + B$ can be superposed; moreover, the following fact holds:

> **Holism:** If $A + B$ is in state $\psi_s = \psi_0^A \otimes \psi_0^B + \psi_1^A \otimes \psi_1^B$, then neither A nor B has any determinate properties.

What is this entangled state ψ_s? Using our earlier interpretation of superposition states in quantum mechanics, ψ_s has dual-aspect behavior: one the one hand, if we measure the quantities Q^A and Q^B, then ψ_s behaves like a probabilistic mixture of the states $\psi_0^A \otimes \psi_0^B$ and $\psi_1^A \otimes \psi_1^B$. In other words, 50% of the time, the measurement would return the results "0 and 0", and the other 50% of the time, the measurement would return the result "1 and 1". On the other hand, ψ_s says that the system has a determinate value for some other property besides Q^A and Q^B, and this property's having a determinate value is inconsistent with Q^A and Q^B having determinate values. Thus, neither Q^A nor Q^B has a value in the state ψ_s.

How about other properties of A and B? It turns out that the state ψ_s cannot be expressed in the form $\psi \otimes \phi$ for any states ψ of A and ϕ of B.[3] In other words, ψ_s is a genuinely holistic state that does not attribute values to any properties of the constituent parts A or B; it only ascribes properties to the whole $A + B$.

The holistic state ψ_s is constructed using only two mathematical assumptions: first, the assumption that there are conjunctive states such as $\psi \otimes \phi$, and second, the assumption that the states of $A + B$ can be superposed. Both of these assumptions are taken on board by quantum mechanics without reservation. Thus, if quantum mechanics is true, then *all* states of *all* physical objects can be superposed.

As a variation on a classic example, consider a cat Tibbles. We can, in thought, build Tibbles up piece by piece from elementary particles. Furthermore, at each stage in the construction, we should — if we believe that quantum mechanics is true — follow the recipe for composition that is provided by quantum mechanics. So, beginning with the first two elementary particles A and B, which obey the superposition principle, we put them together to form a composite particle $A + B$, which then also

[3]The proof of this claim requires a non-trivial result called the Schmidt decomposition theorem.

obeys the superposition principle. We then add a third elementary particle C, which obeys the superposition principle, and the result is an object $A + B + C$ that also obeys the superposition principle. Proceeding in this manner, we finally end up with Tibbles, the composite $A + B + C + \cdots$ of elementary particles, who also is subject to the superposition principle. That is, for any two states that Tibbles can be in, he can also be in the superposition of those two states.

Consider, for example, the state ψ_{alive} in which Tibbles is alive, and the state ψ_{dead} in which Tibbles is dead. Then Tibbles can also be in the superposition state $\psi_{\text{alive}} + \psi_{\text{dead}}$, in which he is neither definitely alive nor definitely dead. Or consider the states ψ_{mouse} in which Tibbles sees a mouse in front of him, and $\psi_{\text{no-mouse}}$ in which Tibbles sees no mouse in front of him. Then Tibbles can also be in the state $\psi_{\text{mouse}} + \psi_{\text{no-mouse}}$ in which he neither sees a mouse, nor sees no mouse. These consequences of quantum mechanics are not so easily passed off as curiosities that can be left for physicists to sort out.

2.4 Dynamics

The fact that cats can be in indeterminate states of life might be considered something of a paradox. However, the most profound paradox of quantum mechanics arises only after we combine superpositions with the dynamics of quantum mechanics.

As is the case with almost all modern physical theories, quantum mechanics is governed by mathematically precise and *deterministic* dynamical laws of the form:

> If the state at some initial time t_0 is ψ, then the state at a later time t_1 will be ψ'.

To simplify things, let's fix the two times t_0 and t_1, and let's use the notation $\psi \Rightarrow \psi'$ to indicate that if the state at time t_0 is ψ, then the state at time t_1 will be ψ'. Then the crucial structural feature of the dynamical laws of quantum mechanics is that they preserve superposition relations.

If $\psi_0 \Rightarrow \psi'_0$ and $\psi_1 \Rightarrow \psi'_1$, then $(\psi_0 + \psi_1) \Rightarrow (\psi'_0 + \psi'_1)$.

The assumption that the superposition relation is preserved through dynamical change is most frequently called the assumption of "linear dynamics" or "unitarity."[4]

To summarize, quantum mechanics is based on the following postulates:

1. Superposition principle: if ψ_0 and ψ_1 are possible states, then $\psi_0 + \psi_1$ is a possible state. If a quantity Q has a determinate value in state ψ_0 and a distinct determinate value in state ψ_1, then the quantity Q has no determinate value in the superposition state $\psi_0 + \psi_1$.

2. Composition principle: If two objects satisfy the superposition principle — i.e. their states can be superposed — then so does the composite object built out of them. Moreover, composite objects have entangled states in which their components have no determinate properties whatsoever.

3. Linear dynamics: the superposition relation is preserved through dynamical changes.

[4]Not every linear mapping is unitary, but we will not need to draw that distinction in this discussion.

3 The measurement problem

Let us now ask what it takes for a dynamical process to count as a "measurement." Suppose that we intend to measure the value of a quantity on some object. For example, the object might be a radioactive atom, and the quantity might have two values according to whether the atom has or has not decayed. Let ψ_d be the a state of the atom in which it has decayed, and let ψ_n be a state of the atom in which it has not decayed. In order then to measure whether or not the atom has decayed, we would need something like a pointer, and we need to implement a process that reliably correlates the state of decay of the atom with the direction of the pointer. In particular, if our pointer is a reliable indicator of the state of the atom, then the following counterfactual statements should hold:

> If the atom is in state ψ_n, then the pointer will be in state ϕ_n.
> If the atom is in state ψ_d, then the pointer will be in state ϕ_d.

For concreteness, let's suppose that the pointer always starts in some neutral state ϕ_*. Then the reliability requirement can be formulated in our previous notation as:

$$\begin{aligned} \phi_* \otimes \psi_n &\implies \phi_n \otimes \psi_n, \\ \phi_* \otimes \psi_d &\implies \phi_d \otimes \psi_d. \end{aligned}$$

But even this mild reliability requirement appears to have disastrous consequences.

> **Measurement Problem:** There are states of the atom such that at the end of the measurement process, the pointer is not definitely pointing in any direction at all. In fact, for almost every state of the atom, the measurement process results in the

pointer not having any determinate properties.

To prove this claim, we need one further fact about how quantum mechanics describes the composition of objects. Recall that if each component object permits superpositions of its states, then quantum mechanics says that the composite object also permits superpositions of its states. But how does the state

$$\phi \otimes (\psi_0 + \psi_1), \tag{1}$$

which is a "conjunction with a superposition," relate to the state

$$\phi \otimes \psi_0 + \phi \otimes \psi_1\,? \tag{2}$$

which is a "superposition of conjunctions." According to quantum mechanics, these two states are the same — a fact that we call the "distribution equation." But why should we believe the distribution equation? Unfortunately, the motivation for the distribution equation is not simple, and so we must ask the reader to believe that physicists have tried to build theories in which this equation is false, but none of these theories has been able to make predictions as accurate as quantum mechanics.

We now complete the derivation of the measurement problem; i.e. we show that at the end of a measurement, the measuring device (e.g. pointer) typically has no properties. Suppose that at the beginning of the measurement process, the atom is in the superposition $\psi_d + \psi_n$ of having decayed, and having not decayed. Then the composite atom-pointer is initially in the state

$$\phi_* \otimes (\psi_n + \psi_d) \;=\; \phi_* \otimes \psi_n + \phi_* \otimes \psi_d,$$

where we have used the distribution equation. But now the assumption of the reliability of the pointer, in conjunction with the preservation of superposition under dynamical evolution, entails that at the end of the measurement process the atom-pointer composite is in the entangled state

$$\phi_n \otimes \psi_n + \phi_d \otimes \psi_d,$$

in which the pointer has no determinate properties.

3.1 Interpretations of quantum mechanics

An "interpretation" of quantum mechanics is an explanation of how the world could possibly be the way that the theory describes it to be. So, an interpretation should resolve the paradoxes in such a way that we are able to understand and believe what the theory is telling us.

It is convenient to classify interpretations of quantum mechanics according to how they solve the measurement problem. First, some interpretations solve the measurement problem by denying that some quantities fail to have determinate values. Second, some interpretations solve the measurement problem by rejecting the dynamical laws of quantum mechanics. Third, some interpretations solve the measurement problem by denying that measurements end with unique outcomes. In the remainder of this section, I will give a brief overview of some popular interpretations of quantum mechanics. In the following section, I will argue that the soul hypothesis suggests a novel and distinct interpretation of quantum mechanics.

First, some philosophers and physicists solve the measurement problem at its very root by denying the idea that there are quantities that fail to have determinate values. Recall that the motivation for asserting quantities can fail to have a determinate value is a sort of democracy of quantities: it is a mathematical fact that not all quantities can simultaneously possess determinate values, and there seems to be no good reason to prefer some quantities over others. To consider a very important case, it is not possible for both position and momentum to have determinate values. If ψ_L is a state of an electron passing through a slit on the left, and ψ_R is a state of an electron passing through a slit on the right, then $\psi_L + \psi_R$ is a state in which the electron has some determinate momentum, but no determinate position. But it is the very fact that $\psi_L + \psi_R$ is a state of determinate momentum that prevents us from interpreting it as a mixture of states (e.g. ψ_L and ψ_R) in which position is determinate. (If two states fail to ascribe a determinate value to a quantity, then no mixture of these states can ascribe a determinate value to that quantity. For example, suppose that ψ is a state that assigns 50% chances to heads and tails, and ϕ is a state that assigns 25% chance to heads, and 75% chance to tails. Then no mixture of ψ and ϕ is a state in which there is a 100% chance of heads.) However, if we deny that there is a quantity (usually called "momentum") that is determinate in the state $\psi_L + \psi_R$, then we can interpret $\psi_L + \psi_R$ as a mixture. More generally, by reducing the number of quantities that we take seriously, we can maintain that all quantities have determinate values in every state.

The strategy outlined above is sometimes misleadingly called giving a "hidden variable interpretation" of quantum mechanics, and the most famous example of this strategy is Bohmian mechanics (see Albert, 1992, chap. 7). Although this strategy works to fix the measurement problem, and moreover provides a pleasingly classical metaphysical picture, we will not pursue it further here.

Second, some physicists have proposed rejecting quantum mechanics' account of change over time — in particular, the linearity of dynamics. The idea behind this strategy is that quantum mechanics makes a false prediction for the outcome of the measurement, i.e. it predicts the entangled state

$$\phi_n \otimes \psi_n + \phi_d \otimes \psi_d,$$

when it should predict *either* the state $\phi_n \otimes \psi_n$ *or* the state $\phi_d \otimes \psi_d$. Under the general rubric of this strategy, there are two distinct approaches. On the one hand, Ghirardi, Rimini, and Weber (see Albert, 1992, chap. 5) claim that there are laws of nature more fundamental than the dynamical laws of quantum mechanics, and that these laws of nature predict that the final state is one in which the pointer succeeds to point in the right direction. On the other hand, Eugene Wigner (1961) famously claimed that when a *conscious* observer looks at the measured object, then the state collapses — in violation of the standard dynamical laws of quantum mechanics — onto one of the two terms of the superposition. (In the following section, I will follow up on Wigner's suggestion, although I will argue that observations do not violate the laws of nature.)

Finally, perhaps the most widely accepted — at least among physicists and philosophers — solution of the measurement problem is the one first conceived by Hugh Everett (1957), which is now sometimes called the "many worlds interpretation of quantum mechanics," and which been greatly elaborated and sophisticated by philosophers such as Simon Saunders and David Wallace of Oxford University. In short, the Everett interpretation accepts that at the end of a measurement, the composite of pointer and observed object is in an entangled state. However, Everettians claim that their subtle analysis shows that the pointer *has* done its job; in fact, the measurement has resulted in the universe dividing into two separate (mostly non-interacting) branches; in one branch, there is a decayed atom and a pointer that reads "decayed", and in the other world, there is an undecayed atom and a pointer that reads "not decayed". If the Everett interpretation is correct, then it is our judgment that there is a unique measurement outcome that is at fault, and not quantum mechanics' failure to predict that there will be a unique outcome.

Each of these interpretations of quantum mechanics agrees that there is a problem that needs to be solved. The first two responses solve the problem by rejecting a piece of quantum mechanics — either the assumption that a literal reading of quantum mechanics gives the correct account of what properties objects can have, or the assumption that quantum mechanics gives the correct account of how these properties change over time. The third response solves the problem by employing revisionary metaphysics, in particular by denying the unity of the subject who experiences the measurement results.

The first two responses to the measurement problem are consistent with a more traditional metaphysical picture of mind-body relations — although I will soon argue that they are unmotivated. In contrast, the third response to the measurement problem is not only encouraged by a reductionist metaphysic of the experiencing subject, but it is also inconsistent with traditional metaphysical stories about the unity and indivisibility of the soul.

3.2 Pushing the problem into the brain

At this stage, one still has the option to "bite the bullet" by declaring that a measured object and a measuring device (e.g. pointer) might just end up in one of those states where the pointer is not determinately pointing in any direction. In fact, if we forget for just a moment our intense interest in where the pointer is pointing, then we realize that the measurement problem is little more than a restatement of quantum indeterminacy. Indeed, calling the pointer a "measuring device" does not change the fact that it is a physical object; and if quantum mechanics is true, then physical objects never have determinate values for all their determinables. So, we already have independent reason to think that the pointer can be in states where it is not pointing in any particular direction; why should it surprise us that this is the case at the end of a "measurement"? What is special about measurement processes that makes indeterminacy more of a problem than it is otherwise?

The measurement problem presents itself with the most force when we place *ourselves* in the picture; for when *we* try to perform measurements, we invariably find that the pointer is either pointing at "not decayed" or "decayed". So, unless our perception is radically misleading, the pointer always ends up either in state ϕ_n or in state ϕ_d. Thus, putting ourselves in the picture gives both a possible escape route from the measurement problem; but it also raises the frightening possibility that we ourselves are subject to the fate of the non-pointing pointer.

To describe the latter scenario, we quote at length from the classical exposition by David Albert (1992, pp. 76–77).

> Suppose, then (just as we did before), that literally every physical system in the world (and this now includes human beings; and it includes the brains of human beings) always evolves in accordance with the dynamical equations of motion ...
>
> Being a "competent observer" is something like being a measuring device that's set up right: What it means for Martha to be a competent observer of the position of a pointer is that whenever Martha looks at a pointer that's pointing to "decayed", she eventually comes to *believe* that the pointer is pointing to "decayed"; and that whenever Martha looks at a pointer that's pointing to "not decayed", she eventually comes to believe that

the pointer is pointing to "not decayed" (and so on, in whatever direction the pointer may be pointing). What it means (to put it more precisely) is that the dynamical equations of motion entail that Martha (who is a physical system, subject to the physical laws) behaves like this [where we use μ_i to denote the fact that Martha believes that the pointer is at "i", and μ_* to denote the fact that Martha is alert and is intent on looking at the pointer]:

$$\mu_* \otimes \phi_n \implies \mu_n \otimes \phi_n \tag{3}$$

$$\mu_* \otimes \phi_d \implies \mu_d \otimes \phi_d. \tag{4}$$

...

Let's get back to the story. The state of the atom and the measuring device (at the point where we left off) is the strange one $[\phi_n \otimes \psi_n + \phi_d \otimes \psi_d]$. And now in comes Martha, and Martha is a competent observer of the position of the pointer, and Martha is in her ready state, and Martha looks at the device. It follows from the linearity of the dynamical equations of motion (if those equations are right), and from what it means to be a competent observer of the position of the pointer, that the state when Martha's done is with certainty going to be

$$(\mu_n \otimes \phi_n \otimes \psi_n) \ + \ (\mu_d \otimes \phi_d \otimes \psi_d). \tag{5}$$

That's what the dynamics entails.

...That state described in (5) is at odds with what we know of ourselves by *direct introspection*. It's a superposition of one state in which Martha thinks that the pointer if pointing to "decayed" and another state in which Martha thinks that the pointer is pointing to "not decayed"; *it's a state in which there is no matter of fact about whether of not Martha thinks the pointer is pointing in any particular direction.*

And so things are turning out badly.

Thus, according to Albert, quantum mechanics entails a fact — that at the end of a measurement, a person will typically fail to be in a determinate state of perceptual belief — that we know, by direct introspection, to be false. Therefore, quantum mechanics is false.

4 Quantum mechanics on the soul hypothesis

A careful reader will note that Albert's derivation of the measurement paradox begs the question against the soul hypothesis. That is, although Albert never says "a person is their body, and no more," he still identifies Martha's mental state (i.e. whether or not she believes the pointer is at "decayed" or "not decayed") with the physical state of her brain. This sort of facile identification should be considered highly suspicious from the point of view of the soul hypothesis.

In the remainder of this chapter, I reconsider the measurement paradox from the point of view of the soul hypothesis. First, I argue that mental states, unlike physical states, do not permit superpositions; and I will show that if mental states cannot be superposed, then a person's mental states can never enter into entangled states in which she has no determinate perceptual beliefs. Then I propose a generalization of the dynamical laws of quantum mechanics that governs not only physical states, but also the interaction of mental (perceptual) states with brain states, and relative to which perceptual states reliably track the state of the external world.

4.1 The two state space hypothesis

The soul hypothesis is, of course, a "pre-theoretical" idea in the sense that the statement "human beings are more than just their bodies" is not yet precise enough to bring to bear directly on the question of how we should describe Martha's mental states when she is performing measurements on objects like atoms. So, if we are to make any progress discussing the measurement problem, then we must — with all due humility! — try to translate the soul hypothesis into something like a precise metaphysical thesis.

Philosophers have made many attempts over the years to put forward and defend precise versions of the soul hypothesis. I will not, in this chapter, try to survey the various proposals, or elaborate on my choice of a proposal. Suffice it to say that the following version of the soul hypothesis has found widespread support among philosophers of a dualistic bent.[5]

> *Logical Independence Hypothesis:* Mental states are logically independent from physical states. That is, if μ is a possible mental state, and ψ is a possible physical state, then their conjunction "ψ and μ" is a logically possible state.

To restate this independence condition in theological language, God could consistently combine any physical state with any mental state.

[5] For example, Descartes' famous dream argument (in the first Meditation) is supposed to show that his mental state is logically consistent with the physical facts being quite different than they appear to be. Similarly, Chalmers' (1996) zombie argument is supposed to show that the physical state of a person's body is logically consistent with that person having no conscious experience at all.

This sort of logical independence postulated between mental and physical states is precisely the same as that which quantum mechanics postulates between independent "degrees of freedom," either of a single object, or of two distinct objects. (For example, **position** and **velocity** are independent degrees of freedom of a baseball; similarly, **position** and **spin** are independent degrees of freedom of an electron.) Thus, if μ is a mental state and ψ is a physical state, then we will borrow from quantum mechanics the notation $\mu \otimes \psi$ to denote the conjunction "μ and ψ" whose possibility is asserted by the independence thesis.

If we use M to denote Martha's mental states, and P to denote Martha's physical states, then the logical independence thesis entails that we should use the product of M and P to denote the entire collection of Martha's states.

> *Two State Space Hypothesis:* The set of states for a person (a particular sort of mind-body composite) is $M \otimes P$, where M is the set of her mental states, P is the set of her physical states, and $M \otimes P$ is the "smallest" set of states that contains all conjunctions $\mu \otimes \psi$, where μ is a mental state (in M), and ψ is a physical state (in P).

What's not yet clear is what we should mean here by "smallest"; in particular, it's not clear whether or not states in $M \otimes P$ can be superposed.

4.2 The non-superposability of mental states

We already know the structure of the collection of Martha's physical states: quantum mechanics entails that these states can be superposed. But what shall we say about the structure of Martha's mental states? What "theory" can we draw upon to tell us the structure of these states?

I make two claims about mental states, one negative, and one positive. The negative claim is that there is no *prima facie* reason to suppose that mental states have the same structure as physical states; and in particular, there is no *prima facie* reason to suppose that mental states can be superposed. The positive claim is that our best evidence — from introspection, and from interaction with other minds — is that mental states are not superposable.[6]

Why do we think that physical states can be superposed? The answer is *not* that we "see that" one state is a superposition of two or more other states — indeed, we have no idea what that would look like. Rather, the existence of superpositions is equivalent to the postulate that there is an unobservable, theoretical relation between states; and this postulate was made because it explains phenomena (e.g. the two-slit experiment). Thus, the claim that superpositions exist is analogous to other instances where a scientific theory postulates some hidden structure, "behind the phenomena."

Do we have reason to postulate the existence of a superposition relation on *mental* states? If we are to trust the scientific experts (namely, psychologists), then the answer is No: psychologists have not yet found a need to postulate the existence of superpositions of mental states. But we have even more evidence. Indeed, each of us has special privileged access to some mental states, namely our own. I ask you, gentle reader, to tell us what state *you* are in when you are in a superposition of, say, believing that the pointer is at "not decayed" and of believing that the pointer is at "decayed". Of course, you will be at a loss — we simply have no means of identifying which mental state would correspond to $\mu_n + \mu_d$, in strong

[6]A similar idea — viz. the non-superposability of mental states solves the measurement problem — is mentioned in (Chalmers, 2003). Chalmers claims that such a view needs to be elaborated, and I do so in the following section.

contrast to quantum mechanics, where we have a very good theoretical grasp of which states are superpositions of which other states. Given that we do not have a theory of superpositions of mental states, it would be more prudent to assume — at least until further evidence comes in — that mental states are not superposable.

There is a further reason for denying that mental states can be superposed, or at the very least, those mental states (such as perceptual states) to which we seem to have direct introspective access. As Albert points out at the end of the long passage above, we seem to be aware of our own perceptual states by means of direct introspection. In other words, there is a process (called introspection) by which a person can reliably determine her own perceptual state. Suppose now — for reductio ad absurdum — that mental states can be superposed, and that the evolution of mental states obeys the dynamical laws of quantum mechanics (as Albert himself supposes). It is a mathematical fact that no process can clone quantum states (Wootters and Zurek, 1982). In other words, there is no way that we can ask "what is the quantum state of an object?", and be sure always (or even reliably) to get the correct answer. But this same no-cloning result would hold true of perceptual states, if indeed perceptual states were superposable. That is, if perceptual states were superposable, then introspection could not reliably deliver the contents of perceptual states.

Thus, against the existence of superpositions of mental states we have collected the following evidence: (1) psychological theories of the mind have not found it useful to postulate superpositions of mental states; (2) our self-awareness of our own mental states does not suggest that some mental states stand in the relation of superposition to other mental states; (3) we do not find it helpful to invoke a superposition relation in explaining the states of those other minds with which we interact; and (4) our direct introspective awareness seems to indicate that we can ascertain our own perceptual state in a direct and transparent way that would not be possible (as shown by the "no cloning theorem") if our mental states had the same structure as physical (quantum mechanical) states.

What, in contrast, is the argument in favor of the superpositions of mental states? The only possible argument I can imagine for superpositions would be an argument by analogy: all physical states can be superposed, therefore (in absence of further evidence) we should suppose that mental states can be superposed. But why should we think that what's true of *physical* states should also be true of *mental* states — unless of course, we think that mental states are somehow derivative? Perhaps one will try to claim that if there is to be appropriate correlation between physical and mental states, then there must be some structural isomorphism between the two collections of states. But in fact, I will now show that we can have *strict* correlations between physical states and mental states, even when the states have radically different structures.

The non-superposability of mental states would have profound consequences for the states of a composite mental-physical system. In particular, we saw in Section 2.3 that if the states of two objects can be superposed, then the states of their composite can be superposed, and so there are entangled states — states in which neither individual object has any determinate properties. But now suppose that states of mental physical composite $M \otimes B$ can be superposed. Then, in particular, we could superpose the two states $\mu_0 \otimes \beta$ and $\mu_1 \otimes \beta$. But by the distribution equation, this would give a state $(\mu_0 + \mu_1) \otimes \beta$, in contradiction with the fact that the states of M cannot be superposed. Therefore, the states of $M \otimes B$ also cannot be superposed, and $M \otimes B$ does not have any entangled states. Therefore, there is no danger of Martha failing to be in any determinate perceptual state.

4.3 Interactionist dynamics

To this point I have argued that mental states are not superposable; and it follows by pure mathematics that mental states cannot become entangled with physical states. In this final section I return to the measurement problem: how might physical and mental states interact during a measurement process?

Recall that Wigner (1961) claims that when a conscious observer makes a measurement on an atom in the superposition state $\psi_n + \psi_d$, then the state collapses onto one of the terms of the superposition, either ψ_n or ψ_d. But Wigner's proposal can be criticized for being ad hoc, and for requiring unexplained violations of the laws of nature. Thus, in this section I will improve on Wigner's proposal by providing a principled account of the interaction between physical and mental states. By "principled" here I mean that I will first lay down a general constraint on reasonable dynamics — i.e. a generalization of the requirement of linear dynamics. I will then show that there are dynamical laws that satisfy this constraint, and that also reproduce Wigner's account of what occurs during measurement.

Often the key to finding a natural generalization of some important fact or law is to find an appropriate formulation of that fact or law. In the case of the dynamical laws of quantum mechanics, we wish to generalize the requirement of preserving superpositions to systems that do not have superpositions. But for systems that have superpositions, the requirement of preserving superpositions is equivalent to the following principle:

> *Mixture preserving dynamics (MIX):* Dynamical evolution is given by a one-to-one and onto mapping $\psi \Rightarrow \psi'$ of states that preserves all mixtures in the following sense: if $\psi \Rightarrow \psi'$ and $\phi \Rightarrow \phi'$ then
>
> $$\lambda\psi \circ (1-\lambda)\phi \quad \Longrightarrow \quad \lambda\psi' \circ (1-\lambda)\phi'.$$

Although it is not mathematically trivial, it can be proven that if states can be superposed, then linearity is a consequence of MIX.[7]

Unfortunately, MIX is too strong a requirement. Recall that Martha is reliable with respect to the pointer's showing "not decayed" or "decayed" just in case:

$$\mu_* \otimes \beta_* \otimes \psi_n \implies \mu_n \otimes \beta_n \otimes \psi_n, \tag{6}$$

$$\mu_* \otimes \beta_* \otimes \psi_d \implies \mu_d \otimes \beta_d \otimes \psi_d, \tag{7}$$

where μ denotes Martha's mental states, β her brain states, and ψ the joint atom-pointer states. But if the states μ_d and μ_n cannot be superposed, then it is mathematically impossible to satisfy MIX as well as Equations (6) and (7).[8] In other words, the derivation of the measurement problem is based on assuming that there is a dynamical law that satisfies Equations (6) and (7); and such is true if the states μ_n and μ_d can be superposed. But given that the states μ_n and μ_d cannot be superposed, the measurement problem cannot even get off the ground, because the constraint on dynamics, namely MIX, is so strong that it rules out even formulating the idea that Martha is a reliable observer of pointer states.

Thus, if we are to formulate the idea that Martha is a reliable observer, then we need to generalize MIX. But it is obvious how to generalize MIX, namely by dropping the requirement that dynamics be given by a *one-*

[7]Sketch of proof: An affine bijection preserves all transition probabilities. Furthermore, Wigner's theorem (Molnar, 1998) shows that any transition probability preserving map is given by a "unitary" or "antiunitary" mapping on the space of states. Unitary mappings are linear, antiunitary mappings are antilinear (and hence preserve weights of superpositions).

[8]Sketch of proof: Two states ω and ρ are said to be *in the same sector* just in case there is another state σ that has nonzero transition probability to both ω and ρ. The two states on the left of Equations (6) and (7) are in the same sector, and the two states on the right are not. But the requirement MIX entails that dynamical transitions preserve the relation of being in the same sector.

to-one mapping of states. (To say that the mapping is one to one means that two distinct initial states cannot lead to the same final state. But why should we suppose that is true?) Let MIX′ denote the requirement that dynamics preserves mixtures.

We can now raise a technical question: is the reliability requirement, as formulated in Equations (6) and (7), consistent with MIX′? In this case, we have a positive result, which we state as a theorem.

Theorem. *There is a dynamical mapping that satisfies MIX′ as well as the reliability requirements given in Equations (6) and (7).*

But what happens now when the measured object (e.g. the radioactive atom) is itself initially in superposition $\psi_d + \psi_n$ of having decayed and having not decayed? Since Martha's mental states μ_d and μ_n cannot be superposed, it is *not possible* for the composite system to end up in an entangled state. Indeed, when the initial state is the composite

$$\mu_* \otimes \beta_* \otimes (\psi_d + \psi_n),$$

then the final state is a mixture of the two states $\mu_n \otimes \beta_n \otimes \psi_n$ and $\mu_d \otimes \beta_d \otimes \psi_d$.[9] It would be natural then to interpret this situation as indicating a genuine indeterminism in the dynamical laws.[10]

The preceding result might seem magical — as if the peculiar features of mental states are responsible for the good behavior of the physical world. But the result follows directly from the natural assumptions that (1) men-

[9] Sketch of proof:

[10] An expert reader will notice that this solution to the measurement problem is mathematically equivalent to the solution that would be obtained by supposing that there is a superselection rule between the two pointer states. But the advantage of the present approach is that the non-superposability of mental states in independently motivated, whereas the non-superposability of pointer states seems to be motivated only in order to solve the measurement problem.

tal states cannot be superposed, and (2) dynamical transitions preserve mixtures of states. Notice furthermore that we cannot be accused — as Wigner has been — of saying that the laws of nature run their course (in particular, preserve superpositions) until a conscious being comes along to cause a miracle (i.e. collapse a superposition onto one of its terms). No: we have claimed that there is only one, indeterministic dynamical law that applies to the entire universe consisting of physical and mental things. It is true that if God had only created mindless physical objects, then he could have created dynamical laws that are deterministic and that also satisfy MIX. But God could not, on pain of logical contradiction, have made all of the following true: (a) both physical and mental properties are instantiated, (b) there are reliable observers [as cashed out in Equations (6) and (7)], and (c) the laws of nature satisfy MIX and are deterministic. In other words, measurements do not require miracles, but they do require there to be embodied minds; and the existence of embodied minds is not trivially compatible with any choice for the laws of nature.

In conclusion, many of us operate with an outdated picture of the physical world, and of how conscious beings — such as ourselves — might interact with this world. We should turn to quantum mechanics to provide a more up to date and accurate picture. However, some of the more popular interpretations of quantum mechanics beg the question against the soul hypothesis; and it is no wonder then that their description of the world is difficult to reconcile with our experience as embodied souls. I have argued, instead, that quantum mechanics is not only consistent with the soul hypothesis, but that the soul hypothesis does work in explaining how quantum mechanics is consistent with our experience.

References

Albert, D. Z. (1992). *Quantum mechanics and experience*. Harvard University Press, Cambridge, MA.

Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press, Chicago.

Chalmers, D. J. (1996). *The conscious mind*. Oxford University Press, New York.

Chalmers, D. J. (2003). Consciousness and its place in nature. In Stich, S. P. and Warfield, T. A., editors, *Blackwell guide to the philosophy of mind*, pages 102–142. Blackwell Publishing, Boston.

Everett, H. (1957). *On the foundations of quantum mechanics*. PhD thesis, Princeton University.

Molnar, L. (1998). An algebraic approach to Wigner's unitary-antiunitary theorem. *Journal of the Australian Mathematical Society*, 65:354–369.

Wigner, E. (1961). Remarks on the mind-body question. In Good, I. J., editor, *The Scientist Speculates*, pages 284–302. Heinemann, London.

Wolterstorff, N. (1984). *Reason within the bounds of religion*. Eerdmans Publishing Co.

Wootters, W. and Zurek, W. (1982). A single quantum cannot be cloned. *Nature*, 299(5886):802–803.