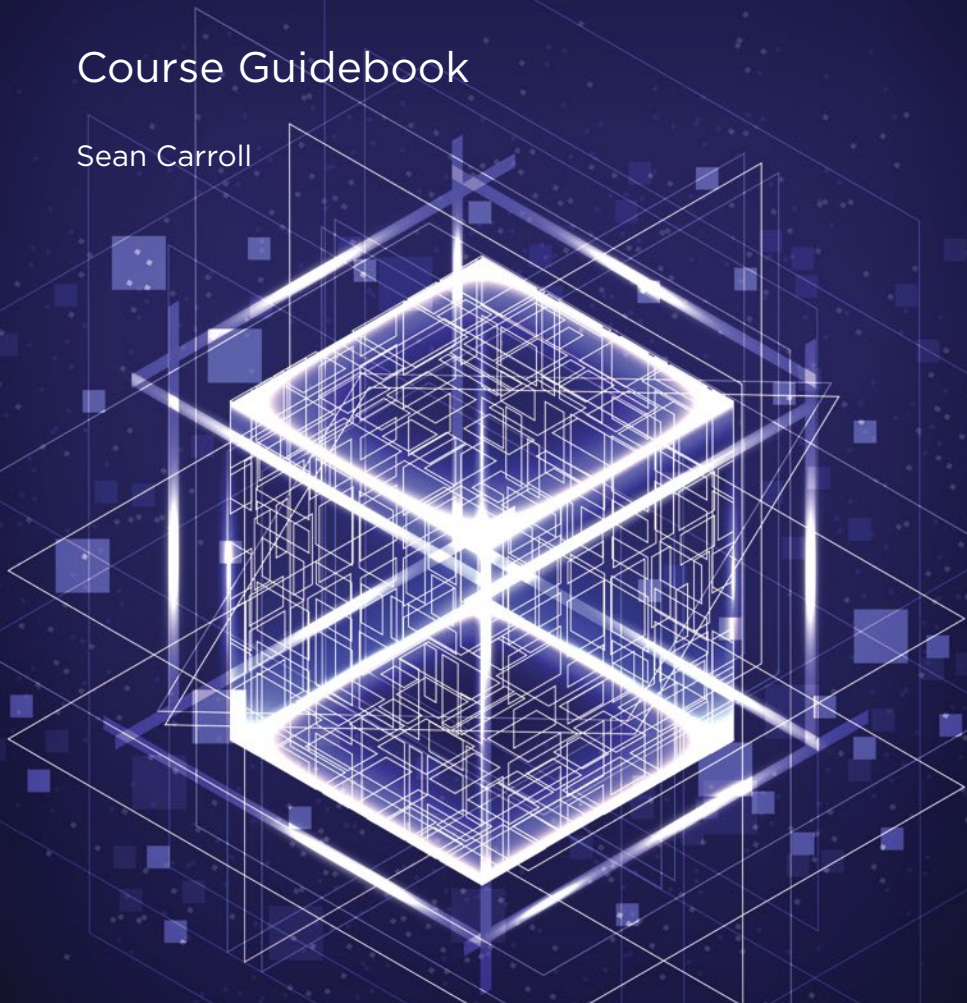


# The Many Hidden Worlds of Quantum Mechanics

Course Guidebook

Sean Carroll





**Copyright © The Teaching Company, 2023**

Printed in the United States of America

This book is in copyright. All rights reserved.

Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of The Teaching Company.

4840 Westfields Boulevard, Suite 400  
Chantilly, VA 20151-2299  
USA  
1-800-832-2412  
[www.thegreatcourses.com](http://www.thegreatcourses.com)



## Sean Carroll

Sean Carroll is the Homewood Professor of Natural Philosophy at Johns Hopkins University and both a member of the Fractal Faculty and an External Professor at the Santa Fe Institute. He received his PhD in Astrophysics from Harvard University. He is the author of several books, including *Something Deeply Hidden: Quantum Worlds and the Emergence of Spacetime*, and the host of the weekly *Mindscape* podcast. He has been awarded prizes and fellowships by the National Science Foundation, NASA, and the Guggenheim Foundation, among others.

# Table of Contents

About Sean Carroll	i
<b>1.</b> Why Suppose There's More Than One World?	1
<b>2.</b> The Classical Physics World That Never Was	9
<b>3.</b> Quantum Worlds Start with Waves and Particles	18
<b>4.</b> A Wave Function to Describe Particles	30
<b>5.</b> Copenhagen Says the Wave Function Collapses	39
<b>6.</b> Is the Wave Function Real?	47
<b>7.</b> Uncertainty in Action with Spin and Qubits	57
<b>8.</b> Quantum Entanglement and Action at a Distance	66
<b>9.</b> Entanglement Leads to Many Worlds	76
<b>10.</b> Decoherence Explains Branching Worlds	84
<b>11.</b> How Entanglement Powers Quantum Computers	92
<b>12.</b> Too Many Worlds! Five Objections Answered	101
<b>13.</b> Testing the Many-Worlds Interpretation	110
<b>14.</b> Where Does Probability Come From?	118
<b>15.</b> Quashing Worlds with Wave Function Collapse	126
<b>16.</b> Blocking Worlds with Hidden Wave Variables	134
<b>17.</b> Mind before Matter in Quantum Theory	143
<b>18.</b> The Quantum Emergence of the World We See	152
<b>19.</b> The Challenge of Quantum Gravity	160
<b>20.</b> Space Emerges from Entanglement	169
<b>21.</b> The Quantum Emergence of Time	178
<b>22.</b> Free Will, Determinism, and Many-Worlds	186
<b>23.</b> What Happens to Ethics under Many-Worlds?	195
<b>24.</b> A Future Renaissance for Quantum Mechanics	204
Bibliography	213



# 1

# Why Suppose There's More Than One World?

**T**his course explores a truly mind-bending idea: that quantum mechanics, which is the most successful framework for doing fundamental physics, predicts the existence of a large number of separate universes parallel to our own that's almost identical to the one we live in, but with small variations between them. This idea is known as the many-worlds interpretation, and it's one of the leading contenders for a rigorous underlying formulation of quantum mechanics.

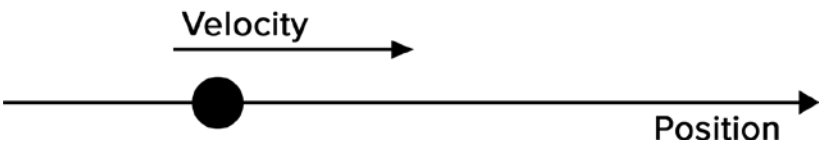
## What Is Quantum Mechanics?

Quantum mechanics is a framework for doing physics. It's a replacement for classical mechanics, which was the incredibly successful framework established in the 17th century by Isaac Newton and others. The big difference between classical mechanics and quantum mechanics is how you think about the process and notion of measuring something.

In classical mechanics, measurement is easy: You just go and measure things. In classical mechanics, a standard situation is you have an object that has some properties, such as a position and a velocity, and you measure them. If you do it clumsily, you might end up disturbing the system or changing its properties. But there's no obstacle, in principle, to measuring things carefully—without changing the system in any important way.

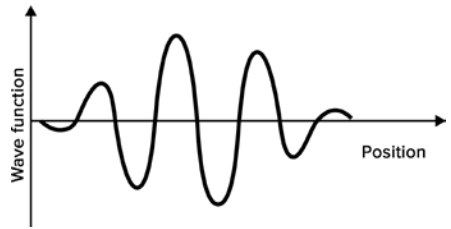
Quantum mechanics, which came along in the early 20th century, tells a radically different story. You do need special rules to handle what happens when you measure a quantum system. You can do measurements, but unlike in classical mechanics, in general you cannot precisely predict what the measurement outcome will be. All you can do in quantum mechanics is predict the probability of getting different measurement outcomes—the probability that a spinning particle will be spinning clockwise or counterclockwise, or that some other particle will be observed to have a certain position or momentum.

A fundamental idea in physics is the state of a system—meaning everything relevant there is to know about it, or everything you need to know to predict what will happen next. In classical mechanics, for example, there are states of systems. The state of a system is specified by the position and velocity of each part of which the system is made, and in principle there's no obstacle to measuring the position and velocity of everything precisely. Once you have the position and velocity of a baseball, you can use the laws of physics to say what it will do next.



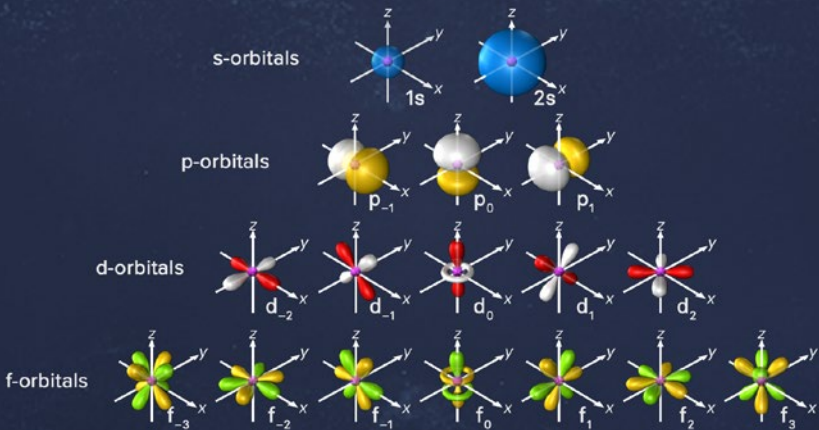
## 1. Why Suppose There's More Than One World?

The quantum system is also specified by its state, and the state is called a wave function. Wave functions can get complicated, but in the simple example of a single particle, the wave function looks like a wave. Rather than describing a particle as a point with some specific location, the wave function in quantum mechanics takes on a numerical value at every possible point in space.



The orbitals of electrons in atoms from chemistry are wave functions. But unlike in the classical story, in quantum mechanics you can never measure the wave function itself. All you can measure are certain observables, such as position or momentum or spin. And then, when you make your measurement, the wave function generally changes dramatically, no matter how careful you are. Before the observation, the wave function is generally spread out. It includes many possible measurement outcomes. But what you see when you perform a measurement is some actual, specific position, not a spread over many of them. And then, maybe most dramatically of all, once you make that measurement, the wave function changes suddenly: It collapses to be concentrated purely on the outcome you actually obtained.

## Atomic Orbitals



**In quantum mechanics, before you do a measurement, the wave function has in it a set of various different possible measurement outcomes. When you do the measurement, you see one of them and all the other possibilities disappear.**

Nobody wanted this kind of thing. Physicists were happy with the classical Newtonian paradigm, in which measurements are simple and precisely predictable.

This quantum shift in perspective was forced on physicists by trying to explain the data. In the period between about 1900 and 1927, physicists were trying to explain phenomena such as radiation from hot objects or the structure of atoms. They were forced to postulate an increasingly confusing set of new rules and principles, including carving out a special role for measurements and what you see when you look at a system. This was, and remains today, an extremely puzzling feature. Why should a fundamental theory of physics care about measurements? And after all, what counts as a measurement? When does a measurement precisely happen?

This set of questions has become known as the measurement problem of quantum mechanics, approaches to which have generally been referred to as interpretations of quantum mechanics. And this was what physicists were doing in the 1920s. But the state of the art today has become more advanced. As a consequence of this advance, practitioners in the field think of themselves as working on the foundations of quantum mechanics—discussing honest scientific theories, not just the interpretations of quantum mechanics. They're trying to figure out what is actually going on in nature—the physical mechanisms that lead to the predictions of quantum mechanics.

Quantum mechanics as a framework was put together gradually. A milestone was the fifth Solvay Conference in 1927. Albert Einstein, Niels Bohr, and many other leading physicists convened and argued over whether quantum mechanics as it existed then was a complete and correct theory or whether it needed to be revised and extended somehow. Out of this conference and other discussions among physicists emerged the Copenhagen interpretation.

That's the story you've been reading. But even in the 1920s, Einstein and other leading people thought it might not be the final answer. They appreciated the empirical successes of the quantum theory that existed at the time in fitting the data; they just didn't think the work was done yet.

In the 1950s, Hugh Everett started reconsidering the foundations of quantum mechanics and developed the theory of the universal wave function. In that picture, every possible measurement outcome actually happens. The secret is each one happens in a different world. This approach was later dubbed the many-worlds interpretation.

## Where Do the Many Worlds Come From?

In quantum mechanics, an elementary particle (something like an electron) sometimes acts like a particle and sometimes acts like a wave. To be more precise, an electron acts like a wave when it's not being measured, and it acts like a particle when it is measured.

It's referred to as a wave because this means the state of the electron is described by the wave function. A wave function in the case of a single particle like an electron is basically a smooth function of space. At every point in space, the wave function takes on a specific numerical value. This is because points in space are possible measurement outcomes. If you're measuring the location of an electron, every possible outcome is some location in space.

So, you can think of this wave function as a superposition: a weighted combination of every possible measurement outcome. This wave stretching through space is assigning a different weight to every location you could see the particle. And the same kind of thing holds true for the spin of the electron or other observable quantities.

**The wave function is a superposition: a weighted sum of every possible measurement outcome.**

Humans have a wave function, too. A person is a physical system, and physical systems are described by quantum mechanics; therefore, people have a wave function describing what's going on. It's just that human wave functions, or anything describing a large macroscopic system, are typically very localized around some particular location in space—such that to a very good approximation, people can be talked about using the language of classical mechanics: Humans appear essentially classical. But the claim is that deep down, even humans are quantum mechanical.

Wave functions obey the Schrödinger equation—named after physicist Erwin Schrödinger—which shows, for any initial choice of wave function, how that wave function will evolve over time. The Schrödinger equation indicates exactly what the wave function will do.

So, the Schrödinger equation explains how the combined system of a human plus the electron will evolve with time. You can ask questions about it, and it provides definite answers. When human beings measure the position of an electron, the overall wave function of both evolves into a very specific kind of superposition. The electron was at position  $x$ , and you observed it to be at  $x$ ; plus the electron was at  $y$ , and you observed it to be at  $y$ ; plus every other possible outcome. And then, amazingly, according to Everett, each one of those parts of the superposition evolves independently as its own world.

## Follow-Up Questions

Precisely how do these worlds come into existence? Why are they truly independent? These seem like difficult questions. The ordinary physics with which human beings are familiar doesn't have any such issues. People have no practice with thinking about them and no experience to guide them. But often there are straightforward ways to answer these questions.

It's tempting to think that branching worlds somehow correspond to different decisions that you could have made: A human being deciding to have pizza or Chinese food creates two different universes. But that's not right. You don't create new worlds just by thinking about them and making decisions. Worlds come into existence in a very specific, physical, tangible way when quantum systems are measured.

So, what is meant by *measured*? In the Copenhagen version of quantum mechanics, that was always left a little vague, but this question has a precise answer in the many-worlds interpretation.

Instead of having measurement as a primitive notion as part of the postulates of the theory, many-worlds can simply describe the conditions under which the universe branches into multiple copies. And that happens whenever a quantum system that is in a superposition becomes entangled with the outside world. Entanglement has nothing to do with consciousness or with being a living observer. If an atomic nucleus decays, for example, and then interacts with its environment and becomes entangled with the world around it, that counts as a quantum measurement—even if no living being is ever aware that it happened.

Another question that arises with many-worlds is this: Exactly how many worlds are there supposed to be? Unfortunately, this is not known. It could be infinitely many. This is not really a concern; it's just a way of saying that the set of worlds is smooth, or continuous, rather than discrete. In that case, the question of how many worlds isn't any more well defined than how many numbers there are between 0 and 1. There are infinitely many, but that's not very helpful. But it's also possible that once physicists understand the laws of physics perfectly, they will realize that there are only finitely many worlds—in which case the question does have a good, specific answer. The state of the art currently just isn't good enough to say either whether the answer is finite or infinite, or what the finite number might be.

So, many-worlds is therefore an example of a theory that predicts a multiverse—many different universes that simultaneously exist. There are other such theories, including the cosmological multiverse and the world of multiverses, but they're all different in where they start from and what they imply.

The many-worlds theory of quantum mechanics provides a compelling, simple picture of what's really going on when the properties of quantum systems are measured. But, despite its compellingness and simplicity, not everyone agrees that it's on the right track. There are other possible alternatives in the foundations of quantum mechanics. There are still a lot of working physicists who hold on to the old-fashioned Copenhagen approach or some modern, slightly souped-up version thereof. The crucial thing about those pictures is that observers aren't treated as quantum. Even though you

might be made of quantum particles, you are classical according to this way of thinking. And wave functions are also not treated as representing reality. A wave function is thought of as just a tool for predicting what observers, who are classical, will measure.

And there are other approaches. But many-worlds is the simplest of any of them. But although many-worlds tells a compelling story, at the end of the day it might not be the right answer. There are puzzles to which the answers are not yet known, and there's still remaining work to be done.



# 2

## The Classical Physics World That Never Was

**Q**uantum mechanics is a replacement for classical mechanics, which is a way of thinking about the laws of physics and how they work, put together by Isaac Newton and his contemporaries in the 1600s. This lecture explores the classical picture: what physicists thought the world was like before quantum mechanics came along.

## The History of Physics

The history of physics can be divided into three eras. The first era can be traced to Aristotle, a Greek philosopher who lived in the 4th century BC. The second era belongs to classical physics, which was codified by Newton in the 17th century. That lasted until the quantum revolution of the 20th century. So, to understand quantum mechanics, it's helpful to appreciate the classical framework it has supplanted.

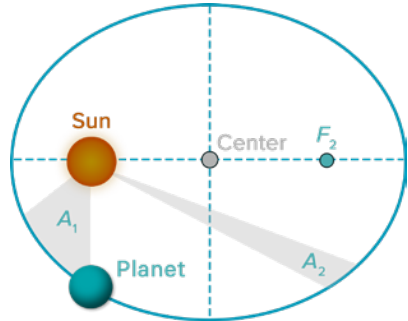
And to understand classical physics, at least a nod should be given to Aristotle. In Aristotle's view, you can think of objects as having a natural state of being, which is typically at rest wherever they want to be. To get something moving, in Aristotle's view, you have to exert a force on it. When the force ceases, the object will stop moving.

Although this is very different than the physics that came later, it's compatible with what you see in the world. A coffee cup on a table will just sit there unless you push it. And when you stop pushing it, it will stop moving. Today, that behavior is attributed to friction from the table. Aristotle's way of thinking was that it was just the natural thing the coffee cup wanted to do.

So, a big change from Aristotle to classical mechanics is that things tend to keep moving as long as you don't stop them. That's called conservation of energy and momentum in classical mechanics. Many thinkers contributed to this idea, but Newton developed it into a full-blown framework. To do that, he needed both precise definitions of important concepts and equations relating them to each other.

A major impetus for the development of classical mechanics came from astronomy, especially the motions of the planets in the solar system. Nicolaus Copernicus famously suggested that the earth and all the other planets move around the sun, rather than the sun and the planets moving around the earth. That was a good idea, but in order to fit the observations that people had collected over the years, Copernicus needed to invoke a complicated set of interlocking circles to describe the orbits.

German astronomer Johannes Kepler realized everything would be much simpler if the planets moved in ellipses rather than in circles and developed a set of simple laws of planetary motion, according to which planets move faster when they're closer to the sun. But he didn't explain why this would be the case.



Meanwhile, other scientists were developing ideas like force and acceleration.

In the late 1600s, members of the new Royal Society in England talked about how to derive the orbits of the planets rather than merely postulating them. Robert Hooke suggested that there was a force stretching from the sun to the planets that diminished with distance as 1 over the distance squared, a famous inverse square law:

$$F = 1/d^2.$$

He suggested that such behavior would naturally lead to elliptical orbits, but he didn't quite have the mathematical chops to be able to prove that.

Eventually, Hooke and Christopher Wren (a famous architect and the designer of Saint Paul's Cathedral) convinced their friend Edmond Halley (who would eventually become famous for discovering the comet) to make the journey from London to Cambridge, home of Newton, to ask him about this problem. Newton replied that he'd already calculated that an inverse square law would naturally give rise to elliptical orbits, and Halley encouraged him to publish his calculation.

This grew into Newton's masterwork, *Philosophiæ naturalis principia mathematica*, which was published in 1687. In this book, Newton not only established his law of universal gravitation, a version of the inverse square law, but he also presented the general laws of motion that formed the basis of classical mechanics as well as the first glimmers of the mathematical techniques of calculus.

## Classical Mechanics and Calculus

Classical mechanics is radically different from the Aristotelian perspective. Objects naturally remain in motion at a constant velocity unless they are acted on by an outside force. The basic idea of classical mechanics is that every system—such as a planet or anything else moving according to the classical laws of motion—has a state, and that state consists of its position and its velocity. There is also a net force acting on the system. The force could be zero, but it has a number. In fact, it's a vector, so it has a magnitude and a direction.

And then there is a crucial rule that relates the acceleration to the force—namely, Newton's second law of motion,  $F = ma$ , or force equals mass times acceleration. This says that if the force is zero, the acceleration is zero. That means that the object will keep moving at a constant velocity. And if the force is not zero, the velocity will change due to the acceleration.

So, the force on an object is simply determined by its environment—meaning all the other stuff in the universe. So, to get classical mechanics off the ground, you don't need to separately specify the force acting on an object. In principle, the forces acting on an object are fixed by the rest of the universe. That has a profound implication. If you're given the state—the position and the velocity of everything in the universe—then what happens at the next moment will be entirely fixed by the laws of physics.

**Whereas Aristotle said that things want to remain stationary once they're in their natural state, Newton said that things are going to keep moving unless you force them to stop.**

Then, you can use calculus—the mathematical technique that Newton, as well as his contemporary Gottfried Wilhelm Leibnitz, developed—to chug forward in time, determining the entire trajectory of the object based on that initial state. So, classical mechanics is deterministic: Both the future and the past are fixed by events at any one moment.

This idea took a while to sink in. Newton didn't really emphasize it. Around the year 1800, French mathematician and scientist Pierre-Simone Laplace pointed out that, if you knew the position and velocity of everything in the universe and you had infinite calculational powers, both the past and future would in principle be known to you, according to the principles of classical mechanics.

This hypothetical being that Laplace suggested—someone who knew everything and could do any calculation—is now known as Laplace's demon. It's important to recognize that it's a thought experiment: There is no prospect of ever having such information, even in the future, when computations get very good. It's meant as an illustration of the power of classical physics—that in principle everything is determined from what is going on at any one moment in time. And any system using this basic framework is called classical. That includes the theory of relativity that will come along early in the 20th century.

## Newton's Theory of Gravity

A paradigmatic example of a classical theory is Newton's theory of gravity, which says that the force of gravity between two objects is proportional to the mass of each object and to the inverse of the distance between them squared. That's the inverse square law.

### Isaac Newton's Gravity Equation

$$F = G \frac{m_1 m_2}{r^2}$$

Using this as a starting point—using  $F = ma$ , this classical mechanics equation, and his law of universal gravitation—Newton was able to rederive all of Kepler's laws of planetary motion. He also moved beyond those laws to think about how the gravitational field of each planet would perturb the motions of the other. In other words, it's not just the sun's gravitational field that matters. All the planets are constantly exerting a slight gravitational pull on the other planets, and that disturbs them from moving in perfect ellipses as Kepler would have suggested. This kind of thinking eventually led to the discovery of the planet Neptune, which was proposed to explain the anomalous motion of Uranus and the other outer planets.

But Newton's gravitational law leads to a philosophical puzzle: How does a planet know what the gravitational pull exerted by the sun is supposed to be? What is the medium that carries the force from the sun to the planets? Without knowing what that is, it would seem like what's being proposed is action at a distance. The sun is in one place, and somehow the planets know how far away the sun is—what the gravitational force is supposed to be. Action at a distance is something that physicists generally disapprove of. Newton knew this and wrote about it in the *Principia*, and he decided to leave the problem to future generations.

There is a resolution to this, at least in classical mechanics. The resolution was suggested by Laplace, who said that rather than thinking directly of one object exerting a force on another—the sun and a planet exerting a gravitational force on each other—instead imagine there is a field pervading space.

A field is a mathematical entity that exists everywhere throughout space. Rather than having a specific location like a particle, the field has a value at every point. An example of a field that you're familiar with is the air temperature on the ground at different points around the globe.

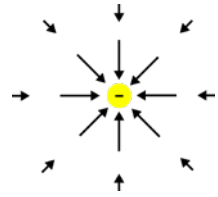
So, the idea of fields was introduced by Laplace, but it didn't matter much. The same predictions resulted from both Laplace's version of gravity and Newton's. In the 19th century, it became a much more important idea, because that saw the advent of electromagnetism.

## Maxwell's Theory of Electromagnetism

At every point in space, both the electric and the magnetic fields have a magnitude and a direction in which they point. So, the values of these fields are vectors.

The way that electromagnetism works is that charges at rest create an electric field. A charge in motion generates a magnetic field. This notion was developed by Michael Faraday and turned into a rigorous mathematical system by James Clerk Maxwell—who unified electricity and magnetism into the single theory of electromagnetism.

Maxwell's electromagnetism also explained what light is: a wave in the electromagnetic field. That's an early triumph of the idea of unification in physics: Electrical and magnetic forces are unified with light.



These theories—the theories of gravity and of electromagnetism—refer to fields rather than particles, but they're still part of classical mechanics. Rather than the position of a particle, you just use the value of the field at each point.

Electromagnetism plays a big role historically, not just because electromagnetism is intrinsically important, but with its special role for the speed of light, electromagnetism was the inspiration for the next big revolution in physics: the theory of relativity.

## Einstein's Theory of Relativity

In Maxwell's theory, electric and magnetic fields are two aspects of a single underlying phenomenon. It's just being looked at in different reference frames, or different states of motion. But what's a state of motion? Who decides what's moving and what's not?

Long ago, Galileo pointed out that there is no preferred state of rest in classical mechanics. Instead, there are relative velocities. If you have two objects, you can ask what the velocity of the objects relative to each other are, but you can't say that there's an absolute velocity for either one of them. This is called Galilean relativity, and it's part of Newtonian mechanics. When electromagnetism came along, it seemed to require something new—a medium through which these electromagnetic waves could move.

If light is supposed to be a wave of electromagnetism, what is waving? The answer was given the name the luminiferous ether. But although experimentalists looked for this ether, they found no evidence that it existed. In 1905, Albert Einstein said that there was no ether and instead postulated that the speed of light is an absolute quantity. It's measured to be the same by every observer, no matter how that observer is moving through the universe. This requires a radical rethinking of what is meant by distance in space as well as time. This rethinking is now known as special relativity.

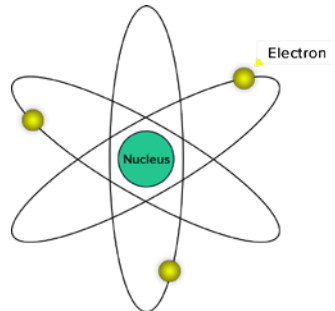
Einstein's professor Hermann Minkowski pointed out that the simplest way of thinking about Einstein's theory is to combine space and time together into a new thing called space-time—another example of unification in physics—which was adopted eventually by Einstein. And in 1915, 10 years after formulating special relativity, Einstein proposed that space-time, a field, can be warped and bent, and that bending and warping is perceived as the force of gravity. This is now called the theory of general relativity, as opposed to special relativity, which is about the speed of light being constant.

## Atoms and Particles

The forces of nature, such as the electromagnetic force and the gravitational force, are being mediated by fields. So, fields play a crucial role in the fundamental architecture of physics. Meanwhile, physicists were drawn to think about matter as consisting of atoms—or, more generally, particles.

In the early 1800s, chemist John Dalton realized that there were two possible compounds you could make out of tin and oxygen: a white powder and a black powder. He figured out that if you had a fixed amount of tin, it took exactly twice as much oxygen to make one of these compounds as the other. This fact—exactly twice as much—is hard to explain if tin and oxygen are continuous, smoothly varying quantities. But it's easy to explain if they're made of discrete atoms: One powder has a single oxygen atom, while the other has two oxygen atoms in it.

The word *atom* comes from the ancient Greek word for “indivisible.” But that turned out not to be quite accurate. In 1897, J. J. Thomson discovered particles called electrons, which are much smaller than even the lightest atoms (hydrogen atoms). Electrons have a negative electric charge, whereas a typical atom has zero charge. Eventually, Ernest Rutherford put together the well-known picture of an atom, which looks like a solar system. There is a massive, compact nucleus at the center of the atom with much later electrons orbiting around it.



Today, it is known that the nucleus is made of some combination of positively charged particles called protons and electrically neutral particles called neutrons. And protons and neutrons are both made of even smaller particles called quarks.

So, matter seems to be made of particles. Every atom in the periodic table is some combination of three kinds of particles: protons, neutrons, and electrons. And if you put yourself in the mindset of a physicist as the 19th century turned into the 20th, you might have thought that you were close to the ultimate theory of all of physics. But as it turns out, a radical rethinking was indeed in the cards, in the form of quantum mechanics.



# 3

## Quantum Worlds Start with Waves and Particles

**S**tarting in 1900, the picture of classical mechanics that had worked for centuries began to unravel. Two important discoveries—Einstein’s particles of light and de Broglie’s matter waves—ultimately led to the formulation of quantum mechanics: that waves sometimes act like particles and that particles sometimes act like waves. It turns out that waves and particles aren’t two different kinds of things; they’re different aspects of a single underlying quantum reality.

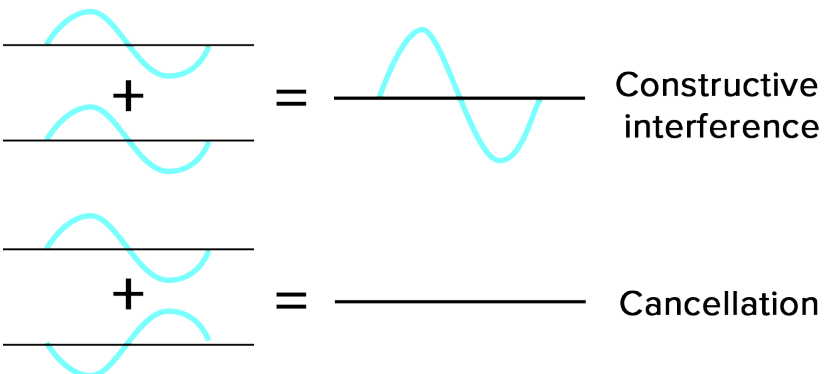
## Planck's Blackbody Radiation

Both important discoveries emerged gradually. The first breakthrough was Max Planck's work on blackbody radiation that pointed the way toward thinking of light as composed of particles rather than waves.

This was an old debate. In the 1600s, Isaac Newton had speculated that light was made of particles. One reason was that light would cast a sharp shadow. If you shine a light on an object, you can see a crisp image of the shadow behind the object. Waves don't act that way. If you put an object in water and water waves flow through the object, they go around the corners, unlike light. But Newton's idea was not universally accepted.

The Dutch physicist Christiaan Huygens thought that light was a wave. His argument was that light did flow. For example, when you pass light through a narrow slit, it spreads out. This is a phenomenon known as diffraction.

To find out whether light was a particle or a wave, physicists did experiments, and there was some evidence accumulating for the view that light was a wave. Much later, in 1803, Thomas Young demonstrated the phenomenon of interference. Since a wave can go both up and down, when you have two intersecting waves, two things can happen: They can either both go up in the same way and reinforce each other if they're moving in the same direction, or they can cancel each other out if they're moving in opposite directions. Young saw interference in light passing through two slits, and that reinforced the idea that light is like a wave.

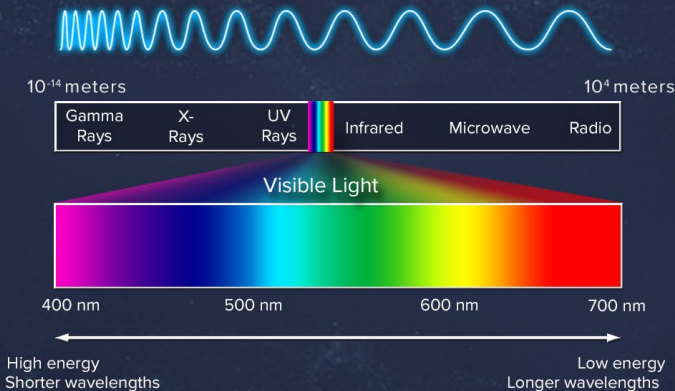


Then, in the middle of the 19th century, James Clerk Maxwell showed that electromagnetic fields naturally form waves that travel at the speed of light, and people thought it was all figured out. Light was a wave, no matter what Newton would have said. These discoveries formed the backdrop for Planck's work on blackbody radiation.

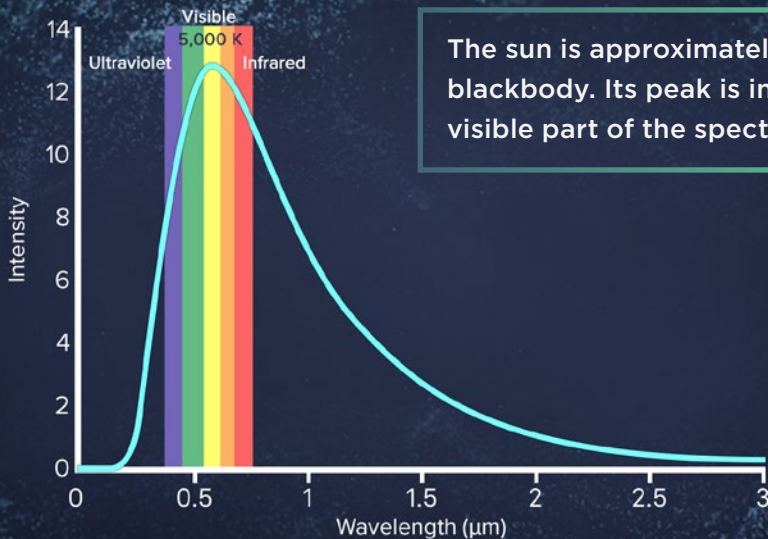
A blackbody is an object that is at a fixed, constant temperature. It absorbs all the light that falls on it; it doesn't reflect any light. So, any light that you see coming from the blackbody is being given off by that object. Note that this is an idealized, experimental object. These theoretically perfect blackbodies are not everyday objects. Put simply, blackbody radiation is electromagnetic radiation—in other words, light that is being given off, or radiated, by an idealized object.

Even though it's idealized, experiments can be done that come close to resulting in a blackbody. One of the things that was realized about blackbody radiation is that it is universal. Different objects reflect and absorb light differently, but any two blackbodies that are held at the same temperature emit the same kind of radiation. This radiation is characterized by its spectrum.

The electromagnetic spectrum is a way of describing all the different possible wavelengths of light, from x-rays to radio waves. But physicists use the word *spectrum* in a more specific sense to mean the relative amount of light emitted at each wavelength.



At a given temperature, blackbody radiation has a characteristic form. It starts at zero at very small wavelengths, rises to a peak somewhere on the spectrum—depending on the temperature of the blackbody—and then fades away again. The peak might be in the infrared for room-temperature objects, the visual part of the spectrum for slightly hotter objects, or the ultraviolet for even hotter objects.



In the late 1800s, physicists had a theory about why blackbody radiation is universal. Materials—things that blackbody objects are made of—are made of atoms. These atoms are bound together into molecules, and the temperature indicates how rapidly the atoms are vibrating back and forth. You could even use this simple picture to derive an equation saying exactly what the spectrum should be. But while the predicted spectrum fit the experimental data very well at long wavelengths, or low frequencies, it went wrong at shorter wavelengths, or higher frequencies. There was another theory under different assumptions that worked well at high frequencies but failed at low frequencies, but no one theory worked for the whole spectrum.

The solution to this problem was proposed by Planck. He assumed that the light being emitted by a blackbody is not continuous. It's emitted in discrete amount of energy, now called quanta. He suggested a new fundamental constant of nature, called Planck's constant ( $h$ ), that related the frequency ( $f$ ) of a given kind of light to the amount of energy ( $E$ ) in a single quantum being emitted:

$$E = hf.$$

In other words, the energy in one quantum is Planck's constant times the frequency. This means that higher-frequency light waves correspond to bigger bundles of energy. It's going to turn out to be more convenient for many purposes to use, rather than  $h$ , the reduced Planck's constant,  $\hbar$  ( $\hbar$ ), which is given by  $h/2\pi$ . Once you make this simple assumption, the blackbody radiation spectrum that's derived using Planck's idea matches the observations perfectly.

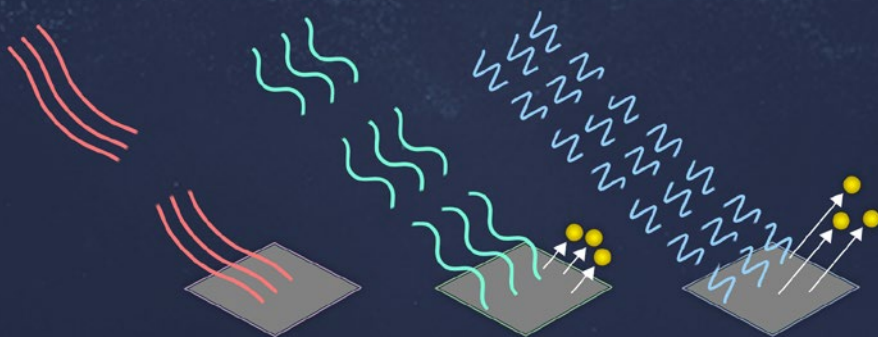
To be clear, a beam of light of a certain frequency can have as much energy in it as you want. That's about the brightness or dimness of the light. The brighter it is, the more energy it's going to have. Planck suggested that such light is emitted in certain bundles, or discrete packets, of energy rather than continuously. So, what is meant by bright light is that there are a lot of discrete quanta of energy contained in it.

There is an important nuance here: Planck didn't ever say that light is made of particles. He merely suggested that it was emitted in discrete amounts. So, you can still imagine a wave with a discrete amount of energy. Five years later, a young physicist named Albert Einstein was considering a separate problem, the photoelectric effect. He was the one who said that light is being emitted in particles.

## Einstein's Photoelectric Effect

The idea of the photoelectric effect is that if you shine light on a metal, the metal can absorb energy and will start spinning off electrons. Electrons are bound to the atoms in the metal and get energy from the light. If they get enough energy, they leave the metal. You might imagine that a brighter light with more energy in it would give more energy to the emitted electrons and that a dimmer light would result in weak electrons. But that is not what the experiments show.

The photoelectric effect experiments showed that the energy of the emitted electrons cared about the frequency of the light being shined on it—not its intensity, or its brightness. Electrons are not going to be emitted at all if the frequency of light is below a certain threshold, and above that threshold, they will be emitted, even if the light is extremely faint.



Classically, that's very hard to understand. It would make sense that the brightness of the light should matter classically. But why should its frequency? Why should the color of the light affect the energy that the electrons have when they leave the metal?

Einstein remembered Planck's idea about discrete amounts of energy in light, but he took it a step further. What if light is not just emitted in discrete bundles of energy but simply *is* discrete bundles? In other words, what if light is a particle rather than a wave?

Planck had a formula relating the frequency of light to the energy of an emitted quantum, and according to Einstein, that formula can be interpreted as the energy of each specific particle. Today, those particles of light are called photons. With Einstein's idea that light is made of particles, suddenly the photoelectric effect begins to make sense. There are atoms in the metal that are holding on to their electrons. It's going to take a certain amount of energy to dislodge an electron from an atom, but that electron is only going

to be dislodged if it's hit by a sufficiently energetic photon. Enough energy is needed per particle—per photon—to kick out that electron. The electrons don't care about the total intensity, or brightness, of the light. They only care about the energy in the specific photon that is hitting them.

High-energy photons corresponding to higher frequencies of light are going to be able to dislodge electrons and give them more energy. That is exactly what is observed experimentally. So, taken together, Planck's and Einstein's ideas seem to suggest that there is something particle-like about light. And both of them won Nobel Prizes for these ideas.

But the situation remained a bit murky. Maxwell's theory of electromagnetism had been incredibly successful, and it predicted unambiguously that light is a wave. Is it somehow both a particle and a wave?

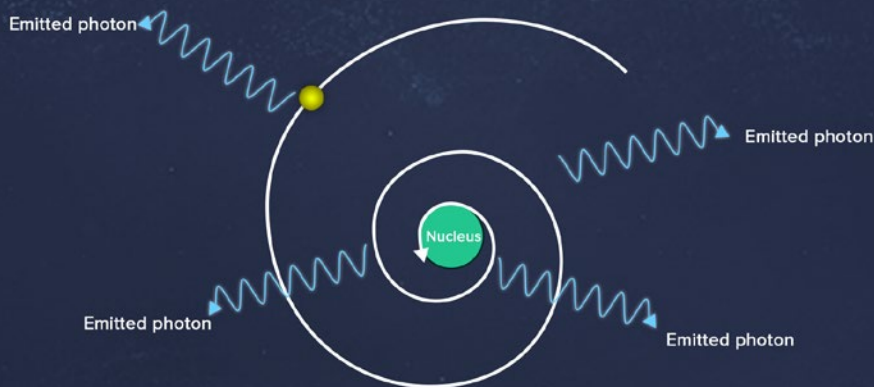
**Albert Einstein never won the Nobel Prize for relativity—special relativity or general relativity. He won it for the photoelectric effect, which was the beginning of quantum mechanics.**

## The Particle Side of the Ledger

In 1911, Ernest Rutherford put together what is known as the Rutherford model of the atom. It has a nucleus at the center, just like the sun is at the middle of the solar system, and the electrons orbiting around it are like planets orbiting the sun.

The Rutherford model was both elegant and sensible. It has become an iconic symbol of physics, or even science more generally. But it's also wrong.

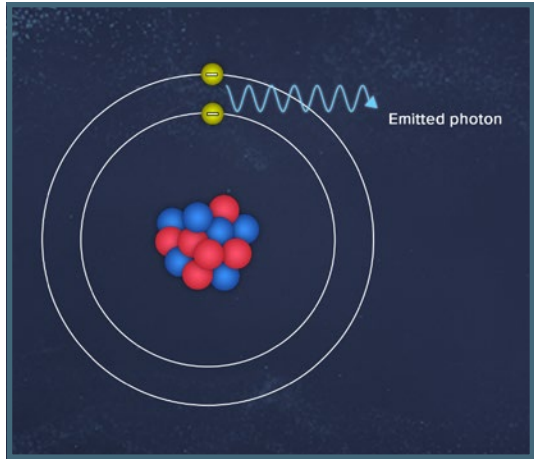
Physicists of the time knew that when an electrically charged particle moves back and forth—such as an electron rapidly orbiting the nucleus of an atom—it emits electromagnetic radiation. So, every Rutherford atom should be emitting light at a prodigious rate. And energy is conserved, so the energy that goes into that radiation should leave the electron. In other words, the electron should lose energy. And that means that rather than just orbiting the nucleus in a stable circle or ellipse, electrons should rapidly spiral downward toward the nucleus.



If that was how nature worked, all matter—including you, the chair you’re sitting on, the earth, etc.—would be completely unstable in a tiny fraction of a second. So, something is wrong with this picture.

Scientists had one more clue: Atoms could absorb or emit light, but it seemed experimentally that this happened only for certain, special wavelengths. If you have a definite type of atom, such as a gas that is made of one element or molecule, certain discrete wavelengths are absorbed when you pass light through a vial of that substance. Likewise, if you energize a gas by passing electricity through it, certain discrete, special wavelengths of light get emitted, depending on what kind of thing you’re lighting up. You might imagine that absorbed and emitted light is interacting with electrons around the atoms and changing their orbits—being absorbed and losing energy or being emitted in forms of different energy. But it wasn’t clear why there should be anything special about particular wavelengths of light or energies of electromagnetic photons.

A possible solution was presented in 1913 by Danish physicist Niels Bohr, who proposed that electrons move in orbits, just like Rutherford had imagined, but only in certain special orbits. And rather than gently spiraling inward, electrons would jump either upward or downward in energy by special, discrete amounts.



This idea of special orbits explains the general stability of atoms: Because there is a special lowest-energy orbit, the electrons can't go below that orbit and therefore don't lose all their energy. Special orbits could also explain the fact that emission and absorption only happen for certain photon energies: There are only certain jumps in energy that an electron can take and therefore only certain photons that can be absorbed.

Bohr was able to derive an equation that predicted which electron orbits were allowed, at least for the simplest atom of hydrogen. And the answer was that the angular momentum of the electron, roughly the amount of oomph that the electron has in its spin around the nucleus, had to be an integer multiple of Planck's constant  $h$ -bar—which had been invented for a completely different reason to understand blackbody radiation. Bohr saying that you could understand the orbits of electrons in atoms if you use that constant of nature established a provocative connection with the results of Planck and Einstein.

But Bohr's model was completely made up. There was no reason given for why electrons would have only those orbits. And there was no mechanism for the jumps between orbits. They appeared random, rather than obeying deterministic equations from classical physics. And the Bohr model didn't do a good job at explaining the data for atoms more complicated than hydrogen.

## De Broglie's Matter Waves

In the late 1910s, classical mechanics was still the dominant paradigm. But the microscopic atomic realm seemed to be hinting that something random and unpredictable was going on.

Radioactivity had been discovered by Henri Becquerel in 1896, and it was named during further study by Marie Curie and her husband, Pierre. Certain elements were spontaneously emitting radiation, and nobody could predict exactly when any particular atom would do so. But the success of statistical mechanics in the late 1800s had accustomed physicists to thinking about the average properties of an ensemble of particles. Even when you don't know the precise state of every particle, you can still say something quantitative and useful.

It made sense to imagine that something similar was going on with radioactivity and jumping electrons. They might be behaving in perfectly predictable ways, but you don't have the right information to make those precise predictions.

The next big step came from Louis de Broglie. Just as Einstein had suggested that there was something particle-like about light, in 1924, de Broglie had the corresponding idea: that there might be something wavelike about particles.

Just as Max Planck could put forward a relationship between the frequency of an electromagnetic wave and the corresponding quantum of energy now called a photon, de Broglie suggested that each particle that has a momentum  $p$  has a corresponding wavelength  $\lambda$  that is given by  $\lambda = h/p$ , where  $h$  is Planck's constant. This is now called the de Broglie wavelength of that particle.

It seems vague to just say that particles have wavelike properties. What is that supposed to mean in practice? One obvious guess is that there should be interference, just like with water waves or electromagnetic waves. And indeed, soon after de Broglie's suggestion, physicists tested that prediction and found evidence of it.

Experiments done by George Paget Thomson in the UK and Clinton Davisson and Lester Germer in the US showed that electrons are able to interfere with each other—something that particles shouldn't be able to do. You can shoot a beam of electrons off a crystal and then measure where they land, and the results show interference bands, just like waves of water passing around different obstacles. The individual landing spots are still points, just like particles would have.

Another important feature of de Broglie's hypothesis related to Bohr's model of quantized, or discrete, electron orbits in atoms. If electrons are ordinary particles, the requirement that they only fall into certain orbits is arbitrary. Bohr invented it. But if they're waves, that's a very different story. Maybe it makes sense why waves only fall into certain orbits: because waves spread out over a region. They're not located at a single point like a particle. So, a wave, in some sense, knows something about the region of space in which it's defined.

Imagine you have an oscillating wave with some fixed wavelength, but the wave doesn't extend over an indefinite line. It lives on a circle—for example, a circle an electron might orbit on. Then, there is automatically a relationship between the wavelength of the wave and the circumference of the circle.

If you walk around the circle and the wave goes up and down as you travel, it should return to the same value when you get back to where you started. So, for any given wavelength, there are only so many circumferences that will allow you to begin and end at the same point, or vice versa: For a fixed circumference, only certain wavelengths will be allowed. So, wavelike electron behavior places a natural limit on the number of orbits an electron can have. The wavelength has to be equal to the circumference divided by some whole number.

If the wavelength is very small, it doesn't matter very much. You can fit any number of different wavelengths in there. But when the wavelength is comparable to the distance around the circle—longer wavelengths, lower momenta, and thus lower energies, according to de Broglie's formula—there are relatively few ways to fit a wave onto a circle. So, this idea that an electron wave around the nucleus must oscillate up and down an integer number of times around the circle turns out to precisely explain Bohr's quantization condition on electron orbits and therefore why emission and absorption

only happen at certain photon energies. There are discrete, limited electron orbits at different energy levels. Interestingly, the quantum nature of the phenomenon of these discrete energy levels comes about not by positing anything intrinsically discrete—not by pixelating matter or energy. In fact, it comes from something perfectly continuous: a wave profile throughout space.

De Broglie ultimately imagined that photons and electrons consisted of both a wave and a particle. The idea is that the wave pushes around the particle, and this idea has not completely gone away.

Einstein's particles of light and de Broglie's matter waves are both great ideas, but neither rose to the standard of being a full-blown theory as physicists would understand it. These kinds of ideas, and others floating around at the time, were part of an ad hoc collection of principles that are now referred to as the old quantum theory.



# 4

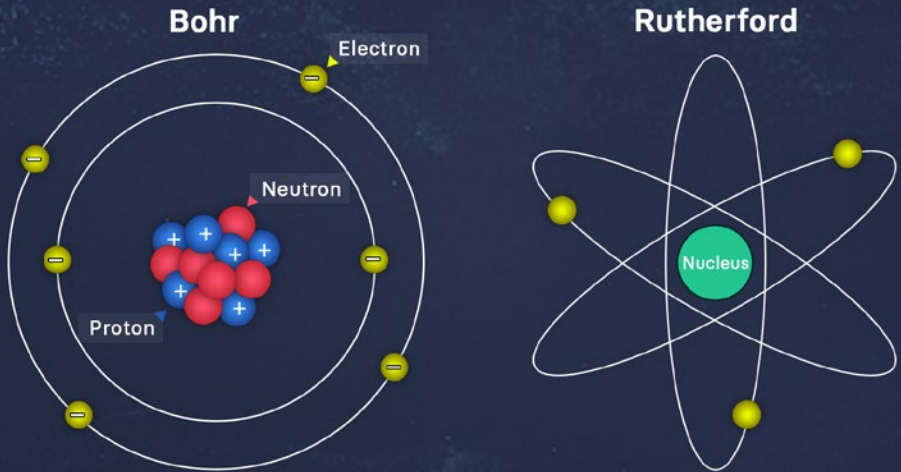
## A Wave Function to Describe Particles

**T**he old quantum theory was a mishmash of rules and principles that didn't quite hang together. The first truly modern versions of quantum mechanics were put forward in 1925, soon after de Broglie introduced the idea of matter waves. There were two distinct proposals: one called matrix mechanics and the other called wave mechanics. Soon enough, it was determined that these were actually two equivalent things, and one theory of quantum mechanics resulted.

## Matrix Mechanics

Matrix mechanics was proposed by Werner Heisenberg, a young German physicist who had been working with Niels Bohr. He had been following up on Bohr's work, trying to develop a more complete understanding of the photons that are emitted when an electron jumps from one orbit to another. He made a conceptual breakthrough in how to think about this puzzling, unpredictable behavior: He said to forget about orbits entirely.

Bohr and Ernest Rutherford had made visual models of what electrons are supposed to be doing in their orbits around atoms, but experimenters weren't able to observe orbiting electrons directly. They just saw the photons that were emitted or absorbed when the electrons purportedly jumped from one orbit to another. So, Heisenberg suggested focusing on what could actually be observed: the photons being emitted. That's a dramatic philosophical move that's very different from the usual way of doing physics. Usually, physicists think that understanding what's really happening is the whole point.



With this perspective, Heisenberg was able to formulate a mathematically rigorous way of talking about what are called observables: things you can measure about a system, such as its position and momentum, even if these quantities didn't have independent existence when you weren't observing them.

Heisenberg wrote his theory down in a paper and sent it to his former teacher, Max Born, who recognized that Heisenberg's equations could be reformulated in terms of a known set of mathematical objects called matrices. In mathematics, a matrix is an array of numbers—a series of rows and columns. Born and his student Pascual Jordan wrote a paper following up on Heisenberg's paper, and then the three of them collaborated on a joint paper that laid out the complete theory of what was called matrix mechanics. Heisenberg and Born both won Nobel Prizes for their contributions to the invention of matrix mechanics.

Matrix mechanics was important for two reasons: It was the first published full theory of quantum mechanics, and it represented Heisenberg's philosophical shift that you should worry about what you can observe and not worry about what's really happening.

Many-worlds as a version of quantum mechanics is one approach that is willing to talk about what really happens. But there's a strong strain within modern physics to claim that quantum mechanics talks about what is observable, not what really happens. And that strain can be traced back to Heisenberg and Bohr—and more generally to what is called the Copenhagen school of quantum mechanics.

## Wave Mechanics

While Heisenberg and his collaborators were thinking about matrix mechanics in 1925, Erwin Schrödinger was thinking about quantum mechanics from a completely different perspective. He and others were thinking deeply about de Broglie's idea of matter waves: the notion that particles have wavelike properties and can interfere with each other experimentally. Schrödinger knew that de Broglie's idea was missing a good equation of motion—an equation that tells you what's going to happen next.

There was an informal idea of how waves propagate, but waves are not all created equal. A light wave acts differently from a sound wave, for example. So, Schrödinger tasked himself with deriving an equation that matter waves might satisfy so that the future evolution of a wave could be predicted from some initial condition of the wave. The goal was to describe a kind of wave that would also lead to the well-known properties that particles have—to unify wavelike behavior with particle-like behavior.

Schrödinger realized that in order to satisfy the properties he wanted his equation to have, the wavelike quantity that he was dealing with would have to be made from complex numbers, not just real numbers. Mathematically, a complex number is the sum of a real number plus an imaginary number. The imaginary unit  $i$  is defined as  $\sqrt{-1}$ . Then, any real number times  $i$  is an imaginary number. A complex number is a combination of  $a + bi$ , where  $a$  and  $b$  are both ordinary real numbers.

Schrödinger's version of the matter wave is called the wave function. It's denoted by the Greek letter psi ( $\Psi$ ). Wave functions are the general way that the state of a quantum system is described, just like the position and velocity of each particle describe the state of a classical collection of particles.

For just one particle, the wave function assigns a complex number  $\Psi$  as a function of  $x$  and  $t$  to each point of space and time:

$$\Psi(x, t).$$

So, at every point in space—at every moment of time for a single particle—there's a complex number  $\Psi$ . And then there's an equation that he derived for it. It's the famous Schrödinger equation:

$$i\hbar \frac{\partial}{\partial t} \Psi = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V(x) \Psi.$$

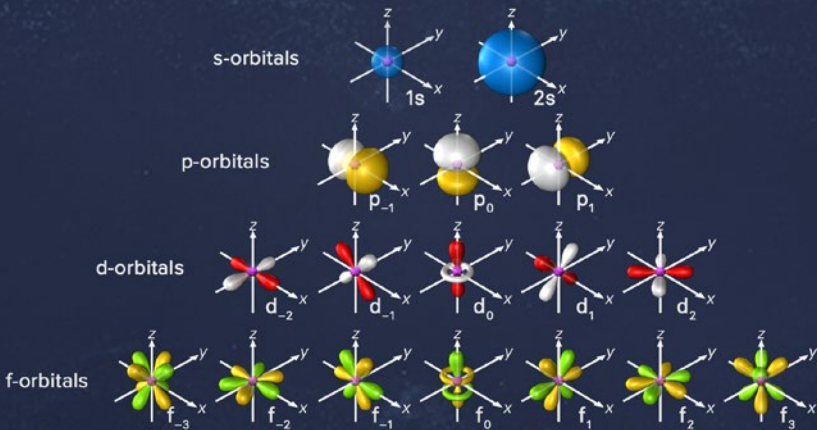
This equation describes the wave function for a single particle with mass  $m$  that's moving in a potential  $V(x)$ . The potential represents the energy that depends on the position of the particle. It's like a ball rolling on a hilly landscape—or an electron moving in the electric field created by an atomic nucleus.

Roughly, the Schrödinger equation says that an energy contained within the wave function can be calculated. Basically, it says that the more energy is being described by the system's wave function, the more rapidly it will be changing.

**What exactly is the wave function? What does it represent? Is it a physical thing, a mathematical artifice, or something new and different? Physicists don't agree on what the wave function is, despite the fact that every physicist uses wave functions. Many-worlds offers a possible answer to these questions, but there are also other interpretations.**

Schrödinger had some initial success with his wave function and its corresponding equation. He was able to recover the idea that a wave representing an electron near a nucleus would naturally fall into one of a discrete set of possible waveforms, or atomic orbitals. And the energy of those orbitals, as predicted by the Schrödinger equation, matched the experimental data. But the picture wasn't yet complete.

## Atomic Orbitals



Although Schrödinger could explain the specific atomic orbitals that people knew about, he couldn't explain why electrons seemed to unpredictably jump from one orbital to another. That was the whole point of Heisenberg's approach. To be fair, Heisenberg arguably didn't explain that either. He basically took it as a starting point. It's part of the postulates of matrix mechanics. But the advocates of matrix mechanics could at least claim that they had a theory that fit all the data. With wave mechanics, that was much less clear.

## Wave and Matrix Mechanics

Schrödinger was able to show that the basic results of wave and matrix mechanics would agree with each other. They're essentially two different languages to describe the same underlying theory. Schrödinger's equation provided a clear description of how wave functions evolve with time, which helps physicists understand the sense in which electrons are wavelike. But he also wanted to understand the particle-like aspects of electrons. That's the challenge of quantum mechanics: making sense of both.

Schrödinger hoped that solving his equation would show a ripple in the wave propagating freely throughout space, not trapped in an atomic orbital around a nucleus, for example. That wave would naturally, Schrödinger expected, focus on some particular point in space, rather than spreading out all over. The idea was that if you then looked at that wave from far away, you would see a localized ripple in the wave function that would look and act like a particle.

The problem was that when the Schrödinger equation is solved, that is not how solutions to the equation behave. Any initial vibrations in the wave function tend to spread out in space, as long as they're not trapped by an atomic nucleus or some other force. The equation essentially says that there is nothing intrinsically particle-like in the behavior of wave functions. But it's known that electrons have particle-like properties, as do photons. When an individual electron is detected on a screen as an empirical experimental event, it makes a dot. It does not make a spread-out wave pattern. A single electron is particle-like. So, what does the wave function actually represent? How does it connect to known, observed physical phenomena?

The conventional answer that's still used today was provided by Born, a coinventor with Heisenberg and Jordan of matrix mechanics, in which probability is part of the assumptions. Born suggested that the wave function is actually indicating the probability of observing a particle at each point.

Schrödinger's equation is perfectly deterministic. In many ways, it's like a classical physics equation. Born's point is that the equation is not the whole story. In his new picture, you also have to invent a new rule that will tell you what you will observe when you measure a system with a certain wave function.

Here's how this new rule works: At any point  $x$  in space, you have a wave function  $\Psi(x)$ , which is a complex number. The wave function is spread out, but  $\Psi(x)$  is a number  $a + bi$ , where  $i$  is  $\sqrt{-1}$ :

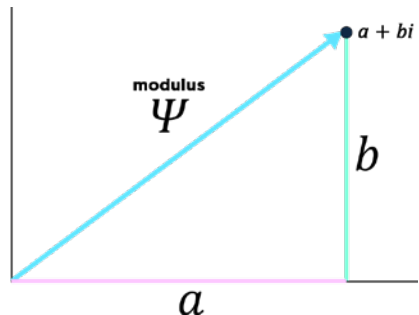
$$\Psi(x) = a + bi.$$

This number,  $\Psi(x)$ , is called the amplitude associated with the value  $x$ , where you might see the particle. If you plot the amplitude, or the value of the wave function at  $x$ , as a single complex number, you can put it on a two-dimensional graph. Let the real part,  $a$ , go horizontally and the imaginary part,  $b$ , go vertically.

So,  $a + bi$  is a point in this two-dimensional complex plane.

Then, the modulus, or the absolute value of that complex number, is the length of the line segment connecting that point,  $a + bi$ , to the origin. Then, you can use Pythagoras's theorem to tell you that the square of the modulus equals  $a^2 + b^2$ , which is written as follows:

$$|\Psi|^2 = a^2 + b^2.$$



The Born rule, named after Max Born, says that the probability of observing the particle at position  $x$  when you measure it is the modulus squared of the wave function at that position—probability equals amplitude squared.

The modulus squared is a nonnegative number (because it's some number squared). Born was able to show that if you consider the modulus squared of the wave function at every point in space and add all those values together, you get a number that remains fixed as the wave function  $\Psi$  evolves. So,  $\Psi$  changes through time, but adding up  $\Psi^2$  everywhere is a fixed, constant number. If you just add up the values of the wave function itself,  $\Psi(x)$ , that would change over time. But if you add up the modulus squared,  $|\Psi|^2$ , that result stays constant. And that fixed number is arbitrary; you can make it anything you want. All you have to do is multiply the wave function by some overall constant number—some scaling factor. That's a good thing, because you can choose numbers so that the sum of the modulus squared of the wave function equals 1.

So, the modulus squared of the wave function  $\Psi$  at each point  $x$  gives a set of nonnegative numbers that add up to 1. This is exactly what you want out of the probability of a set of exclusive events—a set of events such that one and only one of them definitely happens. The probability of any single event is a real number between 0 and 1, and the sum of all the probabilities should add up to 1—meaning that something happens.

The mathematical details are more complicated than this, but what's important was that Max Born was proposing a way of thinking about what the wave function really is, and it was different than Schrödinger had originally hoped for. Schrödinger had idealized the wave function as a physically tangible thing, like a sound wave or an electromagnetic wave—something that could become concentrated at a particular position and would look and act like a particle. Born proposed thinking about the wave function not as a physically tangible thing but as a way of calculating the probability that you will see a particle at any specific point. That probability is given by the wave function's modulus (or the amplitude) squared. Where the wave function is large, there's a high probability of seeing the electron there. Where it's small (close to 0), the probability will be small. This idea—probability is wave function squared—is now known as the Born rule, and it's the insight for which Born was eventually awarded the Nobel Prize.

Schrödinger did not approve of this idea. He originally thought that his equation would be able to restore a more classical-seeming view of physics. To interpret the wave function as just a way of calculating probabilities seemed to him like a step backward. But it did, given a certain number of interpretive moves, allow you to fit the data, which is what physicists ultimately care about.

## An Implication of the Schrödinger Equation

There's one more important aspect of the Schrödinger equation. It seemed like a bit of a footnote at first, but eventually it became the most important part of quantum mechanics. Certainly, it's the part that makes the many-worlds interpretation possible. This is the fact that there is only one wave function for everything in the universe. So, even when there's more than one particle, there's still just one wave function.

Once there are multiple particles, the wave function is not an ordinary wave moving through space. It's the set of possible positions of all the particles in the system. That can be a very large space, since there are lots of particles in the universe.

This is one way in which the mathematics of quantum mechanics becomes intimidating. That simple difference—there's only one wave function, not two different wave functions—turns out to be a world-changing distinction. It makes sense given Born's probability interpretation of the wave function.

This simple move—just one wave function—has a profound consequence. The probability of observing one particle at a particular location can depend on the observation of a completely different particle. This is the phenomenon known as entanglement, and it underlies all of the most bizarre and counterintuitive quantum phenomena, including the prediction of other worlds.



# 5

## Copenhagen Says the Wave Function Collapses

**T**he old quantum theory had been an ad hoc collection of rules. But in 1925, Werner Heisenberg proposed matrix mechanics and Erwin Schrödinger proposed wave mechanics, and both were fully formulated theories. Later, it was shown that they were in fact the same theory: different mathematical descriptions of a single thing that can be called quantum mechanics. But even then, questions of principle and philosophical questions remained—especially about the nature of probabilistic predictions and the need to talk specifically about observation and measurement. A set of answers to these questions eventually coalesced in the work of several quantum pioneers and was labeled the Copenhagen interpretation of quantum mechanics.

## The Copenhagen Interpretation

The Copenhagen interpretation of quantum mechanics wasn't called this at the time; it was simply called quantum mechanics. And not everyone believed it. Major figures in physics, such as Einstein and Schrödinger, remained skeptical. They were convinced that a better theory still needed to be formulated. But for the most part, the physics community accepted the Copenhagen view. And they set to work answering important questions about specifics—about particle physics and nuclear physics—within the framework of the Copenhagen approach.

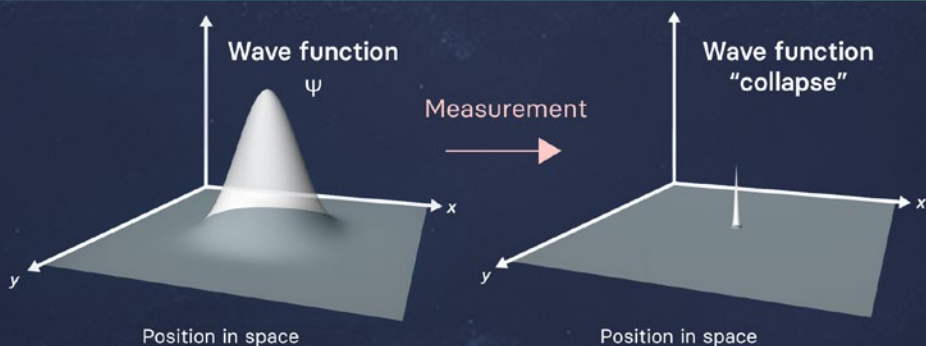
The idea of the Copenhagen interpretation grew gradually out of various writings from popular physicists of the time. But for the purposes of physics rather than history, it's not worth worrying about which precise details should be called part of the Copenhagen interpretation and which should not.

A number of ideas that eventually became part of the Copenhagen interpretation were floated even before the formulation of matrix mechanics and wave mechanics in 1925. Niels Bohr, in particular, had played a major role in influencing how people came to think about quantum theory.

Here's the basic picture of the Copenhagen approach: Consider a single electron. It can be described by a wave function  $\Psi(x)$ . When the electron is not interacting with the rest of the world, the Copenhagen interpretation says that the wave function will obey the Schrödinger equation. But when it's observed—when the position of the electron is measured—its wave function changes suddenly. It does not obey the Schrödinger equation. It collapses to a form that is highly localized around wherever you happen to observe it to be.

What value will actually be observed? You cannot predict ahead of time with any precision. What you can know is the probability of seeing different measurement outcomes using the Born rule: The probability is the wave function squared. These rules together constitute what might be thought of as the operational part of the Copenhagen interpretation.

All this suggests a way of thinking about what the wave function is really telling you. Once you do a measurement and see what a system is doing, it's inevitably tempting to think of that measurement outcome as real. But before



the measurement, the wave function will generally describe many possible measurement outcomes, all with a different amount of representation in the wave function  $\Psi$ .

So, indeed, the wave function assigns a complex number to every possible measurement outcome, and the probability of observing it is the modulus squared of that amplitude. And after the measurement, the new wave function is located only at that outcome and 0 elsewhere. Therefore, you can think of the wave function as describing beforehand a superposition of all possible measurement outcomes weighted by the associated complex number—the amplitude. The wave function is a superposition of all the different things you could potentially observe about the system. It's all the possible outcomes put together, and they're weighted in a way that will tell you the probability (amplitude squared).

## The Measurement Problem of Quantum Mechanics

But there is something undeniably unsatisfying about this set of rules, especially if they're supposed to represent a true story of how nature works at its deepest levels. And all the unsatisfying parts of the Copenhagen interpretation center on the idea of measurement.

For one thing, there isn't any other theory in physics that requires special rules for measurement. You just go and measure things in any classical version of physics. It's only quantum mechanics that says that measurement plays a special role, and you can't make your measurements as unobtrusive as you can imagine. Apparently, measurements can always lead to a rapid, dramatic change in the system—the collapse of the wave function.

But what is meant by the word *measurement*? When exactly does a measurement occur? What precisely is required? If you have a measurement, does that need to be made by a conscious observer? What if you measure something badly, slowly, or inaccurately?

Together, all these questions constitute what is called the measurement problem of quantum mechanics. Any decent approach to quantum foundations has to provide a clear answer to the measurement problem. And there are such approaches. The problem is, unfortunately, they don't agree with each other, and nobody knows which one is right. And the Copenhagen interpretation does not really provide a clear answer to the measurement problem.

## Philosophical Stances of Copenhagen

In addition to the Copenhagen interpretation's operational aspects, there are some philosophical stances that are associated with it. These center around the nature of the wave function and how it's supposed to be thought about.

When it comes to this issue, there are basically two options. One is that the wave function represents a physically real thing. In other words, it's just like other kinds of waves, the kinds of waves that are known about in the classical paradigm. The wave function of the worlds would be part of what reality actually is. But the other perspective says that the wave function is simply a tool for making predictions. It's not real. What's real are the experimental outcomes—what is observed—not the language and methods used to predict them.

The many-worlds interpretation, along with pilot-wave theories and other approaches to quantum mechanics, says that the wave function is a physically real thing—or, more precisely, that there are aspects of reality that are

represented by the wave function. The Copenhagen interpretation does not say that. Instead, it says that the wave function is just a tool for making predictions. And unlike classical statistical mechanics—where there really is some position of momentum for every atom in a gas, even if you don't know it—in classical statistical mechanics, a gas is often described in terms of a probability distribution for all the different positions and velocities. You don't know what the real state is, but you have a probability of what it might be. In the Copenhagen interpretation, in contrast, there is no preexisting objective reality that you reveal when you measure a quantum system.

There is no such thing, by the language of the Copenhagen interpretation, as the thing that is really happening before you make the measurement. According to the Copenhagen interpretation, the only things that really exist are the measurement outcomes. The wave function is objective. According to the Copenhagen interpretation, everyone agrees on what the wave function of a system is supposed to be. It has nothing to do with your personal knowledge that you have happened to accumulate over the experiments or with subjective states of knowledge in your brain. But the wave function doesn't represent reality until it gets observed.

Another crucial element that goes along with the Copenhagen way of thinking is the idea of what an observer is. When considering measurement outcomes for quantum systems, the observers who are doing the measuring can themselves be still treated classically. That might be hard to swallow because a person is made of atoms, and atoms obey the rules of quantum mechanics, so they are treated quantum mechanically. But the Copenhagen interpretation says that at some point, when you have enough atoms to make a person, it becomes appropriate to treat the collection of atoms as a classical object, even though it's made of quantum mechanical atoms.

A watershed moment in the history of quantum mechanics and how to interpret it was the fifth Solvay Conference in Brussels, Belgium. It was held in October 1927, and all the major players were there, including Einstein, Schrödinger, de Broglie, Bohr, Heisenberg, Pauli, and Rutherford. Every one of these people had read the papers and knew what the new quantum theory was, and many of them thought the time was right to codify the new knowledge into a once-and-for-all understanding of quantum mechanics.

Max Planck is usually given credit for starting off the growth of quantum theory, but it was arguably Albert Einstein who first placed quantum mechanics front and center within physics.

## Skeptics of Copenhagen

Most of the participants emerged from the Solvay Conference accepting something like the Copenhagen interpretation, but not everyone was convinced. Louis de Broglie, for example, still had his idea that the reason things sometimes behave like particles and sometimes like waves was that they were both. The wave function in his picture should be thought of as a kind of pilot wave that guides around the particles, which is what is observed in apparatuses.

De Broglie gave a talk about this idea at the Solvay Conference, and it was not received well. His ideas were ridiculed by Pauli, and de Broglie essentially abandoned his theory. Eventually, the idea of a pilot wave plus particles was revitalized by David Bohm in the 1950s, and it continues to be taken seriously today.

An even more famous skeptic of the Copenhagen interpretation was Einstein. He just didn't think that quantum mechanics, as it was understood at the time, was the complete and final answer—at least in its 1927 form.

The discussions at the Solvay Conference were a major part of what is now called the famous Bohr-Einstein debates. Although Bohr and Einstein remained admiring friends their whole lives, they disagreed about whether quantum mechanics was more or less finished.

Einstein's major objection was to the probabilistic nature of quantum mechanics. The famous quote from Einstein says, "God does not play dice with the universe." It's a true quote, and that was one of his concerns, especially early on, but it was never his biggest worry about quantum mechanics. What he really cared about was reality. Einstein worried that the Copenhagen interpretation's insistence on not talking about what is really

happening was a major step backward for physics, and part of his concern about reality was reflected in his worries about nonlocality: the “spooky action at a distance” that is still plaguing physicists’ understanding of quantum mechanics today.

Unfortunately, at the time of the Solvay Conference in 1927, Einstein had not articulated his worries about locality very clearly. It wasn’t until several years later that his objections came into sharp focus in the famous Einstein-Podolsky-Rosen paper in 1935. And there was yet another skeptic: Schrödinger.

Much like Einstein, Schrödinger favored the kind of deterministic local dynamics that were familiar from ordinary classical theories of fields, such as Maxwell’s electromagnetism. But also like Einstein, Schrödinger didn’t have any knockdown arguments to present at the Solvay Conference. His worries, like de Broglie’s and Einstein’s, went largely unheeded. Schrödinger didn’t sharpen his objections until 1935, when he proposed his famous Schrödinger’s cat thought experiment.

## Schrödinger’s Cat

The idea of Schrödinger’s cat came into being from Einstein and Schrödinger talking with each other about these issues. The idea is to seal a cat inside a box along with a certain apparatus that includes a lump of radioactive material and a Geiger counter that clicks when it detects emitted radiation. For simplicity, the initial state of the radioactive source can be described as “has not yet decayed.” Quantum mechanics, in the form of the Schrödinger equation, predicts that the quantum state of the radioactive substance will evolve into a superposition, which can be called “the source has decayed, plus it has not yet decayed.”

If you simply treat everything in the box as obeying the rules of quantum mechanics, that means the Geiger counter evolves into a superposition of “I have clicked, and I have not clicked.” It clicks when it detects some radioactive particle. In Schrödinger’s thought experiment, there is a gizmo that drops a hammer onto a glass vial filled with gas so that the hammer will fall and break the glass if and when the Geiger counter clicks. In his version of the thought experiment, Schrödinger had the vial filled with cyanide, but

sleeping gas is more ethical. Then, the prediction of quantum mechanics is the hammer and the vial evolve into a superposition of “the hammer has fallen and the vial has broken, and the hammer hasn’t fallen and the vial remains unbroken.” That’s the prediction of the Schrödinger equation.

Then, if the quantum story—the Copenhagen interpretation—is taken at face value, the cat will evolve into its own superposition of “it’s awake if the vial has not broken, and it’s asleep if the vial has broken.” This is the profound quantum leap that you have to wrap your mind around. It’s not that there is either an awake cat or an asleep cat, but you just don’t know which one. That would be the story you would tell if you thought there was a reality inside the box before the measurement happened.

According to the Copenhagen interpretation, there is just the wave function of the cat, which is in a superposition of both states. You’re assuming that the cat does not count as an observer, but the rules for what counts as an observer are murky. If you need to be a human to count as an observer, it’s clear what’s going on: When a person opens the box and peers inside, the Copenhagen interpretation says that the wave function of the cat instantly collapses onto either “awake” or “asleep,” and you can use the Born rule to calculate the appropriate probabilities.

Schrödinger’s point with this thought experiment was to take a quantum superposition, which exists in the microscopic world, and magnify it from the subatomic level to the macroscopic world. There is no clear conclusion to be drawn from this thought experiment. It’s just a vivid illustration of what quantum mechanics, at least in the Copenhagen interpretation version of it, seems to demand that people believe. And it’s meant to shine a light on the need to get a clearer understanding of how reality is supposed to behave in this theory.

As you will learn, the many-worlds version of quantum mechanics offers clear answers to the question of what happens inside the box: The wave function branches into two possibilities, and it happens long before the cat goes into any superposition. You end up with two worlds, and in each world, there’s a well-defined single cat.



# 6

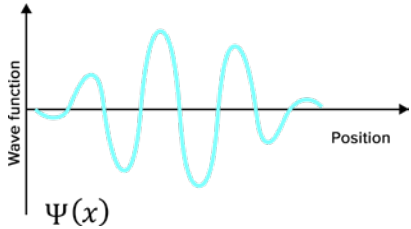
## Is the Wave Function Real?

**D**iscussions about the foundations of quantum mechanics often center on the measurement problem. But another important issue is the reality problem: Does the wave function represent reality, or is it merely a useful way of making predictions for observations? In the Copenhagen interpretation, the wave function does not represent reality. But there are other formulations of quantum mechanics—many-worlds being one of them—where the wave function does represent reality. This lecture displays some of the considerations for either side.

## The Uncertainty Principle

The uncertainty principle was formulated by Werner Heisenberg in 1927. According to common understanding, the uncertainty principle says that you can't know exactly your position and your momentum (or velocity) at the same time. But the reality is a bit more subtle.

Remember that the wave function—the central object that is studied in quantum mechanics—can be thought of as a complex-valued function of all the possible positions in space that a particle can take.  $\Psi(x)$  is the wave function, the amplitude for seeing the particle at position  $x$ .



In classical mechanics, the state of a particle is not specified by only giving its position,  $x$ . You also need to specify the momentum,  $p$ , which is mass times velocity:

$$p = mv.$$

So, if you know where a particle is in classical mechanics but don't know how fast it's moving, you can't predict anything about where it's going to be next. You don't know what direction it's moving in.

But in quantum mechanics, the wave function doesn't depend on position and momentum separately—it only depends on the position. So, in quantum mechanics—as a successor theory to classical mechanics—where is the information about momentum in the quantum wave function? How do you know what you're going to observe about the momentum,  $p$ ?

The answer is that you can also define a wave function that depends on the momentum. You can call it  $\Psi(p)$  as opposed to  $\Psi(x)$ .  $\Psi(x)$  is the position wave function, whereas  $\Psi(p)$  is the momentum wave function. And the momentum wave function will tell you the probability of measuring the momentum to have any particular value in the usual way: You take the square of the modulus (that is, the amplitude squared), and that gives you the probability. There's a Born rule for measuring the probability, just like there's a Born

rule for measuring position. So, you can talk separately about the position wave function, which tells you the probability of any position measurement, and the momentum wave function, which tells you the probability of a momentum measurement.

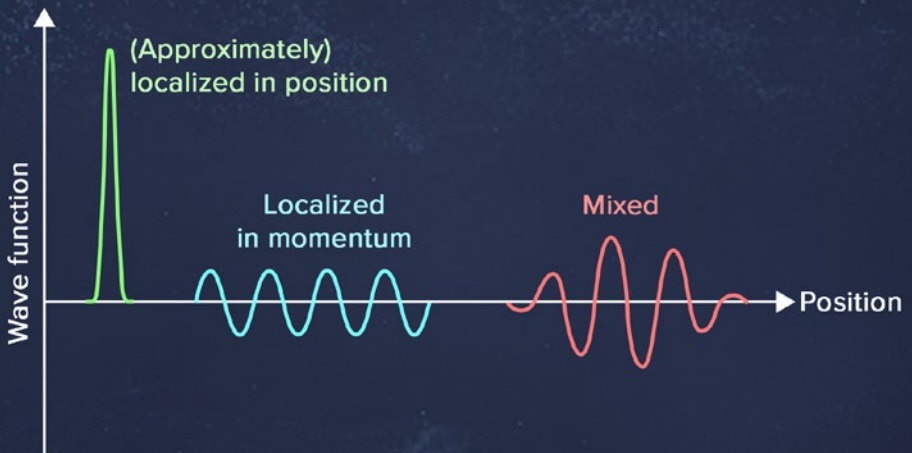
The trick is that the momentum wave function is not defined separately from the position wave function. In classical mechanics, position and momentum are independent. In quantum mechanics, the position wave function is equivalent to the momentum wave function. There is a well-defined mathematical way of going back and forth, and it's called the Fourier transform.

Neither the position wave function nor the momentum wave function by itself is fundamental. There is a single thing—call it the quantum state—that's an abstract state of description, and it's just that the quantum state can be expressed equally well as a wave function depending on position or as a wave function depending on momentum. Classically, knowing the position of a particle tells you nothing about its momentum, and vice versa, whereas quantum mechanically, there's a single quantum state that tells you everything there is to know about both position and momentum. How does that work?

The answer is there's an interesting relationship between these two ways of expressing the quantum state: the position wave function and the momentum wave function. If the position wave function is highly localized—if  $\Psi(x)$  is almost entirely concentrated around some particular point in space such that you know that if you measure, that's probably where you're going to see it—then the Fourier transform, or the map from position wave functions to momentum wave functions, tells you that the momentum wave function will be spread out all over. If you were to measure the position of a particle with a highly localized wave function, you'd have a pretty good idea ahead of time where you're going to find it. But if you take that same particle with the same wave function, and want to measure its momentum, you have no idea what answer you would get. There are many different possibilities, all of which would have substantial probability for getting that answer.

Likewise, the other way around works just as well. If you have a momentum wave function that is localized around some value—so you know essentially what you would get if you measured the velocity or momentum of the particle—then the position version of the wave function is spread out all over, so you have no idea what position you would measure for that kind of particle.

There are also wave functions that are neither exactly localized in position or momentum—they're compromises. They are somewhat compact with respect to position and somewhat compact with respect to momentum. In that case, there's uncertainty about what you would measure if you measured either the position or the momentum.



None of these options lets you say that if you measure either one, you can predict carefully what you are going to observe. This feature is the heart of the uncertainty principle: You can have a wave function that is localized in position, and you can have a wave function that is localized in momentum. You cannot have both at the same time. There is an unavoidable trade-off in how compact a wave function can be in both position and momentum at once.

Notice that it has not been stated that in quantum mechanics, measuring the position disturbs the system and leaves its momentum uncertain, or vice versa. It's true that when you measure a system, you can collapse the wave function and change it—but that's not the fundamental point of the uncertainty principle. The uncertainty principle is not a statement about measurements, though it might imply things about measurements. It's a statement about what kinds of wave functions can possibly exist.

You can measure either position or momentum. That's your choice as an experimentalist. And if you're a good experimentalist, you're going to get a well-defined answer: You're going to see the electron having a position or see it having a momentum, depending on your experiment. But there are no wave functions where both quantities are precisely defined at the same time ahead of time.

So, is the wave function all there is? That's the reality problem. Or does an electron really have a well-defined position and momentum even before you measure it, but somehow quantum mechanics doesn't know how to describe those real-world quantities?

One good way to get into these issues is to consider yet another famous thought experiment in quantum mechanics: the double-slit experiment.

## The Double-Slit Experiment

Picture a tub of water. If the water is perfectly still and you poke it, ripples will emanate outward from where you poked it. They'll move in a circular pattern, getting farther and farther away.

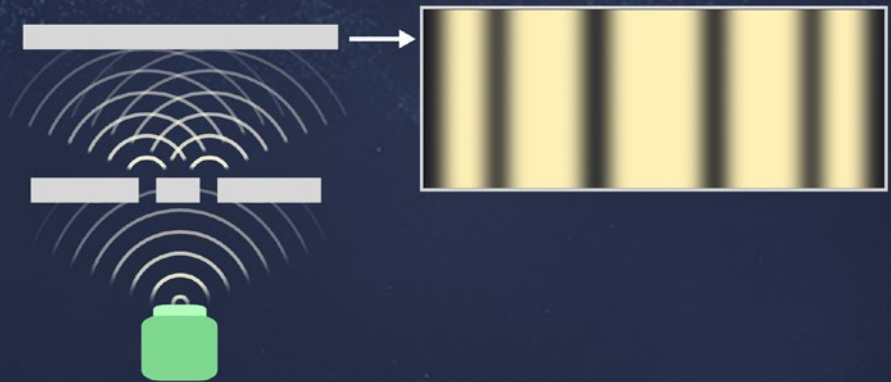
Then, imagine that you put a barrier into the water some distance away from where you start the ripples. If there is a single slit in the barrier, then some waves are going to pass through. On the other side of the barrier, the waves will again emanate from the slit in another semicircular pattern.

Now imagine that there are two slits in the barrier. Then, two circular patterns of ripples will emerge on the other side. And you've learned that when you have waves, you can have interference. So, these ripples coming out

of the two slits will interfere. At some points, they will reinforce each other—points where the two ripples are moving in the same direction—and at some points, they will cancel each other out.

Next, consider doing this with light rather than water. Take a beam of light pointed at a thin slit in a barrier and set up a detector on the other side. If you start with a single slit, then because light has wavelike properties, you will see the light spread out a bit. That's diffraction. It won't quite be a semicircle because the light has some momentum in that direction.

Then, imagine you have a barrier with two slits that are pretty close together—not much farther apart than the wavelength of the light. Then, at a detector placed on the opposite side, you see an interference pattern: bright bands with dark bands in between. Given the distances from the source of light to the slits to the detector, the light waves will interfere either constructively (build up) or destructively (cancel each other out). This version of the double-slit experiment was done by Thomas Young in the 19th century to demonstrate the wave nature of light.



If you instead send particles through the slits, some of the particles will hit the barrier and then stop, while others will pass through the slit. But if they get lucky enough to pass through the slit, they're just going to continue in a straight line. Classically, particles do not diffract; they keep moving unless they hit something. Mostly, a particle detector on the other side of the two slits will basically see the slits: It will measure a bunch of detected particles in a single location for a single slit. If there are two slits, there will be two bunches in two locations. You don't get interference patterns by sending particles through double slits. Particles don't interfere with each other. Particles have a location in space; they do not spread out everywhere.

Returning to the double-slit experiment using light, imagine you make the intensity of the light beam extremely small. When you make the light that dim, you eventually stop seeing a continuous image at the detector. Rather, you start detecting individual flashes at specific locations. This is because you've reached the realm of individual photons. What you thought for bright light was a wave passing through the slits, you now realize that when the light gets very dim, it's actually a collection of individual photons. So, a sensitive photodetector—a machine that discovers photons when they get hit by photons—can measure photons one at a time.

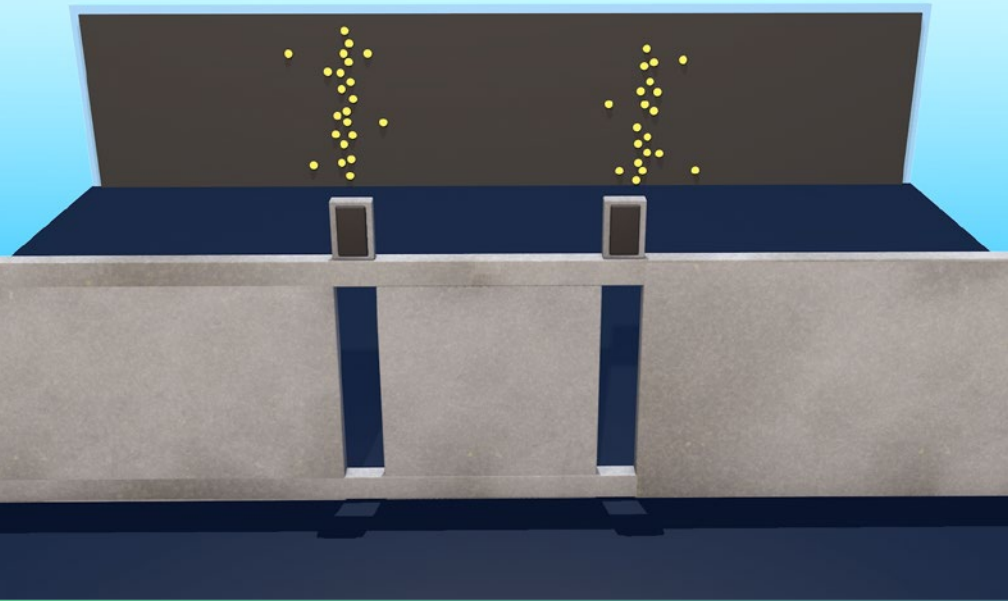
High-intensity light shows wavelike interference patterns, but low-intensity light behaves like point-like particles. How can that be reconciled?

The answer comes if you look at the individual flashes from low-intensity light. Any individual photon leaves a flash, but if you allow those flashes to accumulate over time, the flashes are detected one by one by the detector—but they do not congregate around just two points, one for each slit. Rather, the photon dots that show up in the detector cluster in high-intensity bands separated by low-intensity empty regions and then fall away if you get far away from where the slits are pointing.

In other words, photons do exhibit interference patterns—even when you're viewing them one at a time as individual particles. And the same behavior holds for electrons. This was a thought experiment for a long time, but the experiment can actually be done. Richard Feynman popularized this version of the two-slit experiment, but it wasn't until 2012 that physicists could do the experiment in a laboratory.

If you send a single electron through two slits, the detector on the other side will see a dot like a particle. But—just like light—if you keep sending electrons through one at a time, many dots will accumulate, and that accumulation of dots will look like an interference pattern.

Historically, the double-slit experiment was not one of the motivations for quantum mechanics. Its strange behavior was predicted and observed long after the basic ideas of quantum theory had been established.



If classical physics were right and electrons were just particles, this is very hard to make sense of. You're sending electrons through one at a time, yet their accumulated effect is an interference pattern that seems wavelike. This means that each electron must somehow be interfering with itself. So, it knows not to go to the places where the interference bands are dark; it knows to probably go where the interference bands are light. And if you insist on thinking of the electron as an ordinary particle—which you shouldn't—that electron either went through one slit or the other one. A particle cannot go through both. So, how did it interfere with itself?

Within quantum mechanics, the answer is that the electron is a wave function. When it's passing through two slits, there is absolutely something wavelike. Wave functions have no trouble interfering with themselves. And then it's only when the electron is finally detected on the other side that it looks like a particle. That's the short motto for quantum mechanics: It's a wave when you're not looking at it; it looks like a particle when you do.

If you wanted to stubbornly insist that the electron went through either one slit or the other, you could presumably find out by placing small detectors near the slits to measure which slit the electron goes through. For every electron you fire at the slits, you detect that it goes through one slit—not two. It looks like a particle, and then it continues on to be detected on the other side. When you detect it, it always looks like a particle when you observe it.

But if you send many electrons through and monitor which slit they go through, each electron only goes through one or the other each time. And on the other side, where you have the main detector, the interference pattern goes away. When you try to figure out which path the electron takes, you force it to take one or the other. The wave function collapses onto whichever slit you observed it to go through. So, your observation has fundamentally changed the state of the system.

The double-slit experiment seems to have implications for the question of whether the wave function represents something real or whether it should be thought of as something epistemic. Philosophers use the word *epistemic* to refer to knowledge or the ability to make predictions.

The double-slit experiment implies that the electron isn't simply described by a wave. It seems to legitimately act like a wave when you're not observing it. It's not just that individual electrons go through one slit or the other and you just don't know. The correct thing to say is that the wave function of the electron goes through both slits. You need to be able to say that because that's how the electron is able to interfere with itself—by being a wave when it goes through those two slits.

At least at face value, this kind of behavior is what makes people think that the wave function should be thought of as a real thing, not just a way of making predictions. A physical, actual wave has no problem interfering with itself. But something that you just think of as a tool for making predictions—how does that interfere with itself?

Bohr, Heisenberg, and colleagues decided that wave functions should not be thought of as directly representing reality. The fact that you don't observe wave functions directly lends that view some credence. If the wave functions were the real physical stuff of the world, why can't you just measure them? Why do they keep collapsing every time you look? Many intelligent physicists today still don't think of wave functions as representing reality.

But quantum mechanics is confusing. You should be open to different possibilities, and phenomena like the double-slit experiment suggest that you should at least take seriously the possibility that the wave function does represent reality. This is called the realist position about the wave function.



# 7

## Uncertainty in Action with Spin and Qubits

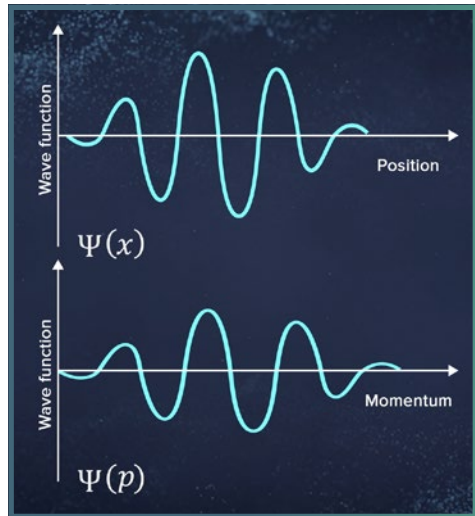
**W**ave functions are the central object of study in quantum mechanics. But they're undeniably abstract. Once you start considering more than one particle at a time, a wave function becomes very hard to visualize. As this course progresses with increasingly subtle examples of quantum phenomena in action, it will be helpful to have a simple example that can be referred to repeatedly—something that's easier to keep in mind than the wave function as a function of a bunch of particles. So, this lecture introduces the meaning of spin—of a single particle or maybe a few particles at a time—in quantum mechanics.

## Spin

The notion of spin in quantum mechanics is not much different from the classical idea. It's just rotation around an axis. The earth does this every day. In physics, systems are often characterized in terms of degrees of freedom: the various quantities that exist that specify the exact state of the system. In classical mechanics, the state of a system is its position and its momentum for a single particle or all the positions and all the momenta for many particles. Therefore, position and momentum are separate, independent degrees of freedom in classical mechanics. You have to specify them independently.

In quantum mechanics, position and momentum are not independent degrees of freedom. You can specify the wave function in terms of one, but then the wave function in terms of the other can be calculated. It's fixed by the first specification. So, you either have the wave function as a function of position or momentum. You can think of either one of those as a degree of freedom.

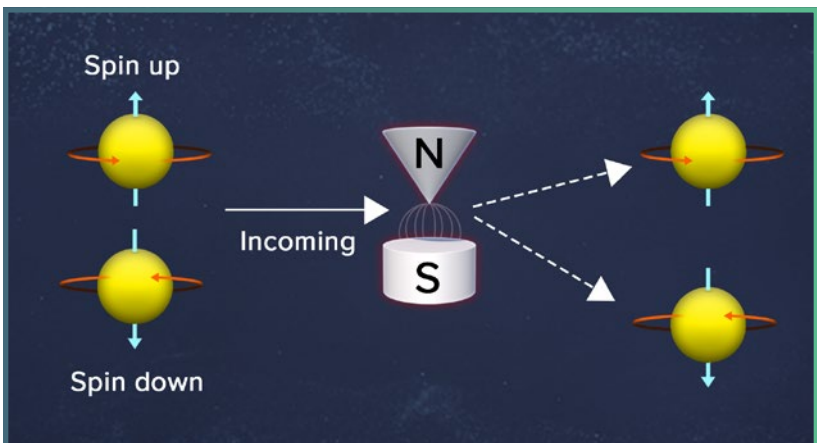
Spin is a separate degree of freedom. You need to specify how the wave function depends on spin in addition to how it depends on either position or momentum. It's an independent degree of freedom of a subatomic particle like an electron. When you learned about the energies of an electron orbiting an atomic nucleus, you discovered that in quantum mechanics, there are only certain discrete results that you can obtain when you measure that energy. Likewise, the same is true when you measure a particle's spin: There will only be certain discrete results.



For an electron, for example, there are two possible outcomes when you measure the spin of an electron. To do this, you first pick an axis—a direction in space—because you’re going to measure the spin of the electron with respect to that axis. When you make that spin observation, just like measuring the energy of an electron in an atom, there’s a discrete set of possibilities. For the spin of an electron, there are only two possible measurement outcomes: The electron is spinning either clockwise or counterclockwise.

These are conventionally referred to as either spin up or spin down. So, how do you do this kind of measurement? The good news is a spinning electron is kind of a tiny magnet. It has a north magnetic pole and a south magnetic pole, much like the earth. And the spin axis conventionally points toward the north pole. So, one way of measuring the spin of an electron is to shoot it through a focused magnetic field. According to the laws of physics, the electron gets deflected. It’s moved a little bit in its motion depending on how the spin is oriented with respect to the magnetic field that it’s passing through. So, how you orient that magnetic field chooses the axis with respect to which you measure the spin.

If you were thinking classically and were told that an electron had a certain total spin, you might make the following prediction for such an experiment: The electron would be deflected upward if its spin axis were aligned with that external magnetic field and downward if its spin were aligned in the opposite direction of the magnetic field. And if its spin were somewhere in between, then you might predict that the electron’s motion would be deflected at some intermediate angle between perfectly up and perfectly down.

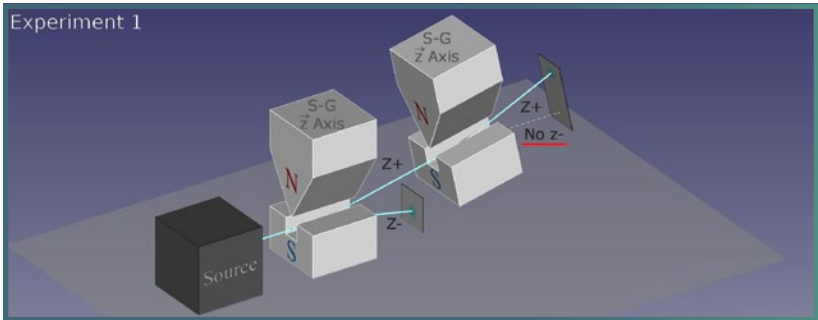


But that is not what you see when you do the experiment—you do not see a continuous set of different possible deflections. This experiment was performed in 1922 by the German physicist Otto Stern, who was an assistant to Max Born and Walther Gerlach.

In the experiment, they saw that electrons are deflected when they go through a magnetic field. But there are only two possibilities. Some electrons are deflected up at an angle; others are deflected down at the same angle. There is nothing in between. You get two dots on the screen where you pointed the electrons at. If you rotate the magnetic field so that you're measuring the spin of the electron with respect to a different axis, the electrons are still deflected in the direction of the field—either along it or against it. There are no intermediate values. You still get two dots. In other words, as a fundamental feature of quantum mechanics, spin—the value of spin that you measure, like the energy of an electron—appears to be quantized, with only two possible discrete results. It seems like the electron is either completely aligned with the magnetic field or completely anti-aligned with the magnetic field.

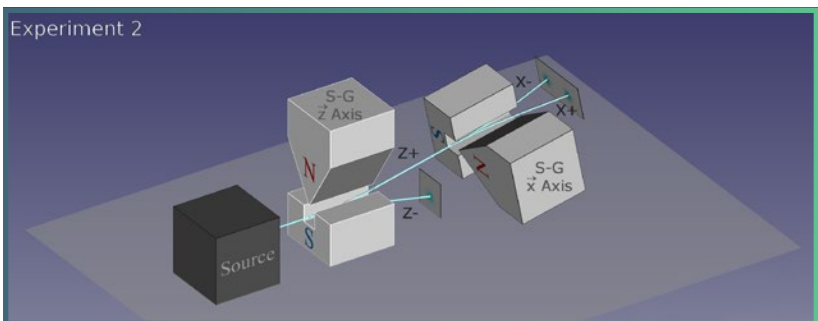
To test this, you can run the electron through two magnets in a row. The rules of textbook quantum mechanics say that if you get a certain measurement outcome and then immediately measure the same system with the same observable, you will always get the same answer. And for spin, that is indeed what happens. If you put an electron through the magnetic field and it's deflected upward by a magnet—so the electron is spin up—then it will always be deflected upward by a following magnet that is oriented in the same direction.

**The uncertainty principle expresses the fact that there are different, incompatible measurements that you can make on any particular quantum state, and horizontal and vertical spin are incompatible in the same way that position and momentum are.**

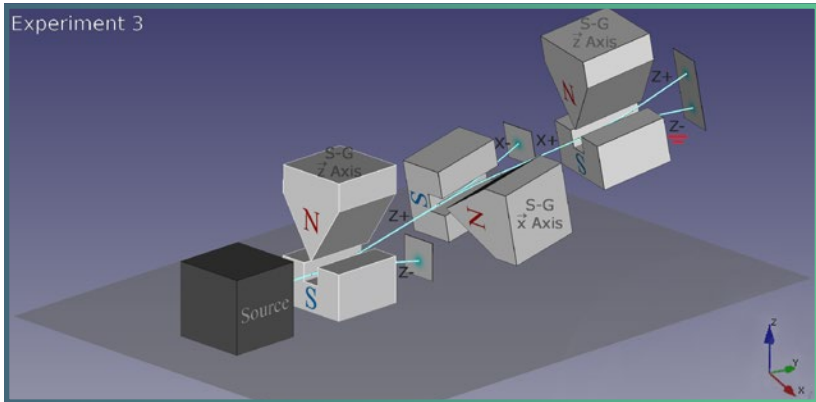


So, the wave function could have started out as a superposition of spin up and spin down. But when you measured it, according to the usual rules of quantum mechanics, it collapsed. And it collapsed, in this case, onto spin up. So, if you keep measuring it along the same axis, you're going to keep getting spin up repeatedly.

But what if you rotate one of the magnets by  $90^\circ$  and do a second measurement? You're splitting an initial beam of electrons into spin up and spin down as measured by a vertically oriented magnet and then passing only the spin-up electrons through a horizontally oriented magnet. The second magnet splits the spin-up electrons into two beams. A horizontally oriented magnet will deflect them either right or left. And if they come in spin up, half of them will be deflected to the right along the direction of the second magnet, and half of them will be deflected to the left. There is nothing in between.



Let's say you do three measurements in a row. Imagine that the first is along the vertical axis. Imagine that the electron is measured to be spin up but you keep letting it go through. So, the second measurement is along the horizontal axis. You're going to get either spin left or spin right. So, let's say you get spin left. For the third measurement, you're going to do the vertical axis again.



What you find after the third is a 50-50 chance of getting spin up or spin down. That's even though the first measurement was clearly spin up because you only looked at the answers you got that were spin up, and if you would have just done vertical measurements repeatedly, you would keep getting spin up repeatedly with 100% probability. But somehow, you did an intervening horizontal measurement and got an answer, and that somehow collapsed the wave function and reset the state. So, it's not merely a passive observation of a preexisting property. This is an example of wave functions collapsing and being reset and the fact that you cannot measure them so delicately so as to not disturb them.

This kind of measurement of a spin dramatically disturbs the system that you're looking at, and that's strange. Your classical intuition makes you visualize something like a top spinning in a certain way, and that makes you think that there's an axis around which the electron is spinning. The earth

has an axis that it spins around, and you know the orientation it has. So, perhaps it makes sense that the spin around whatever axis it's spinning around is quantized.

But that's not what the experiments show. They show that the axis around which the spin is quantized is not predetermined by the state of the particle itself. You choose any axis you want to measure with respect to. You choose it by rotating the magnet appropriately. And when you do that and measure the spin, you see that the spin is quantized with respect to your measurement axis.

This is another manifestation of the uncertainty principle: that position and momentum are not properties that an electron has once and for all, like a classical particle, but are just things you can measure about the particle. And quantum mechanics lets you predict the probability of different measurement outcomes, but those outcomes are not determined 100% in advance. In particular, there are no wave functions, or quantum states, in which a particle has a definite value of both position and momentum simultaneously.

When you specify the exact wave function for position, you can work out the probability of observing any particular momentum, and vice versa. And the uncertainty principle says that the more localized you are in position, the more spread out you are in momentum, and vice versa.

The same is true for vertical and horizontal spin, for example. These experiments are telling you that vertical and horizontal spin are not separate, real properties that the electron has before you measure it. They are just different quantities that you can measure to get a certain outcome.

And starting with this abstract underlying quantum state, you can represent it in terms of either position or momentum. It's the same information. So, vertical and horizontal spin are the same thing. They give you different but equivalent ways of expressing the spin part of the wave function. So, if you express the quantum state of a spinning particle in terms of its vertical spin, that 100% determines the probability of observing it left or right horizontal spin. The measurement outcomes you can get are determined by the underlying quantum state; it's just that that quantum state can be expressed in different but equivalent ways.

## Qubits

Systems with two possible measurement outcomes in quantum mechanics are called qubits. Whereas a classical bit has just two possible values (0 or 1), a quantum bit, or qubit, is a system that has two possible measurement outcomes, such as spin up and spin down along some specified axis. The state of a generic qubit is a superposition of both possibilities. And each possibility, spin up or spin down, is weighted by a complex number, which is the amplitude for that alternative, and the probability of getting that measurement outcome is the amplitude squared. You can write the wave function, or quantum state, of a qubit as  $\Psi$  equals  $a$  times the spin-up state plus  $b$  times the spin-down state.

The diagram shows the equation  $\Psi = a(\text{spin up}) + b(\text{spin down})$  on a dark blue background. Each 'spin up' and 'spin down' term is represented by a yellow sphere with a red horizontal line around its equator and a blue vertical arrow pointing up or down respectively.

Here,  $a$  and  $b$  are complex numbers—the amplitudes for spin up and spin down, respectively, always with respect to some determined axis. And the pieces of the wave function representing the different possible measurement outcomes—in this case, spin up or spin down—are called the components of that wave function. So, in this state,  $a$  times spin up plus  $b$  times spin down, the probability of observing the particle to be spin up would be the absolute value of  $a^2$ , and the probability for observing spin down would be the absolute value of  $b^2$ .

This diagram illustrates the components and probabilities of the qubit wave function. At the top, the equation  $\Psi = a(\text{spin up}) + b(\text{spin down})$  is shown. Labels with arrows point to the 'Wave function' ( $\Psi$ ), the 'Amplitudes' ( $a$  and  $b$ ), and the 'Components' (the spin up and spin down states). Below the equation, two separate spin states are shown. The first is labeled 'Probability' and is associated with the expression  $|a|^2$ . The second is also labeled 'Probability' and is associated with the expression  $|b|^2$ . Each probability state is represented by a yellow sphere with a red horizontal line and a blue vertical arrow pointing up or down.

And you need the condition that the absolute value of  $a^2$  plus the absolute value of  $b^2$  equals 1, because if you measure the spin, you're going to get either spin up or spin down.

$$|a|^2 + |b|^2 = 1$$

It can be hard to think of the quantum state as a wave function. It's not very wavy. It doesn't look like a smooth function of space, like the wave function of the position of a single particle. But this actually points to an important property of quantum states—one that gets obscured when you think about the wave function for just a single particle. A wave function position of one particle looks like a wave traveling in space. But more general wave functions aren't like that. They are not a function of location in space like the electric or gravitational field.

The quantum state as an abstract idea is actually a function of the space of measurement outcomes. For a qubit, the space of measurement outcomes only includes two possibilities: spin up or spin down. But when what you are observing is the location of a single particle, a quantum state assigns an amplitude to every possible location in space.

That assignment—an amplitude at every point in space—looks like a wave in ordinary space. But that's the unusual case; that's not the typical case. In general, the wave function is more abstract, and when more than one particle becomes involved, it becomes hard to visualize. Qubits are great because at least the wave function has only two components, which makes it relatively easy to visualize and think about.

**A quantum computer is just a computer that manipulates qubits in the same way that an ordinary computer manipulates classical bits.**



# 8

## Quantum Entanglement and Action at a Distance

**T**his lecture is about entanglement—a strikingly nonclassical feature of quantum mechanics. Entanglement both raises some new questions and might provide some new answers to the questions that already exist.

## Entanglement

What bugged Einstein about quantum mechanics was nonlocality: What happens at one point in space can seemingly have immediate consequences for experiments done very far away. It took him a while to codify his concerns into a well-formulated objection, but in doing so, he helped illuminate one of the most profound features of the quantum world: the phenomenon of entanglement. Entanglement arises because, in quantum theory, there is only one wave function for the entire universe. Why isn't there a wave function for each individual particle?

Consider some particle that decays into two particles. In quantum mechanics, all you can do is calculate the probability that the emitted particles will be observed moving in some direction. So, you have an initially stationary particle that decays into two particles, but all you can say is the probability that the different decayed particles will be moving in different directions. The wave function of the decayed products spreads out in a roughly spherical pattern until you observe it. When you observe the particles that are coming out, you pin down a definite direction in which the particles seem to be moving.

And when you do this experiment, you notice that the initial particle was at rest, so you can talk about the momentum of that particle. Momentum is mass times velocity:

$$p = mv.$$

The initial velocity, or momentum, of the particle is zero, and momentum is conserved in processes like this, so the total momentum of the two decayed products should also be zero. If you want to have zero total momentum out of two particles, that means that the created particles might emerge moving in different directions, but whatever direction one of them is moving in, the other has to be moving in precisely the opposite.

Classically, that's all fine. If you bounce two billiard balls off each other, they travel in opposite directions. But the quantum version is a bit funny. The first particle has a probability of moving in some direction. The second one also has a probability. You can't tell exactly what it will be ahead of time. But if they each had a separate wave function—if you had two wave functions moving in a spherical cloud—those two probabilities would be completely

unrelated. You could imagine observing one of the particles and measuring the direction in which it's moving after the decay. The other one hasn't been measured yet; it's undisturbed. Its wave function is still moving out in a sphere. How could it know that it's supposed to be moving in the opposite direction of the first particle when you do measure it?

The answer is that the two particles don't have two separate wave functions. Their behavior is described by the single wave function of the universe. In this particular case, you're ignoring most of the universe. You're focusing on just these two particles. But you can't ignore one of the particles and just focus on the other. This is the lesson of entanglement. What you observe about one particle can dramatically affect the predictions you make about the other. That's what it means that the particles are entangled.

Think about this on a mathematical level. A wave function is an assignment of a complex number—the amplitude to each possible observational outcome. And the Born rule states that the square of the amplitude equals the probability that you would observe that outcome were you to make that measurement. So, in the context of talking about more than one particle at a time, that means you assign an amplitude to every possible outcome of observing all the particles at once. If what you're observing is positions, for example, the wave function of the universe can be thought of as assigning an amplitude—a wave function value, a complex number—to every possible combination of positions for all the particles in the universe.

Although it is hard to visualize the wave function of the universe, the Schrödinger equation is definite in what it says about how the wave function behaves and tells you exactly what you need to know:

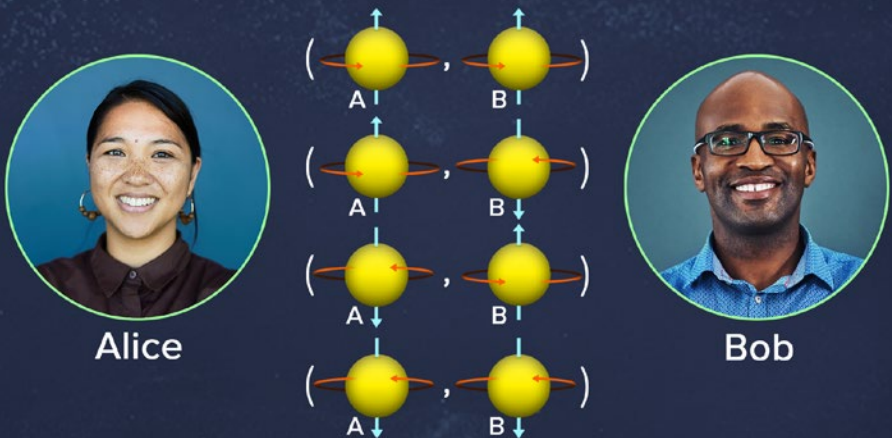
$$i\hbar \frac{\partial}{\partial t} \Psi = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V(x) \Psi.$$

## Entanglement in Action

Happily, almost everything interesting about entanglement can be cast into a much simpler context: the context of qubits. A qubit has two possible outcomes for any particular measurement you can make on it. So, quantum physicists like to consider two people named Alice and Bob who share qubits with each other. Alice has a qubit, and Bob has a qubit.

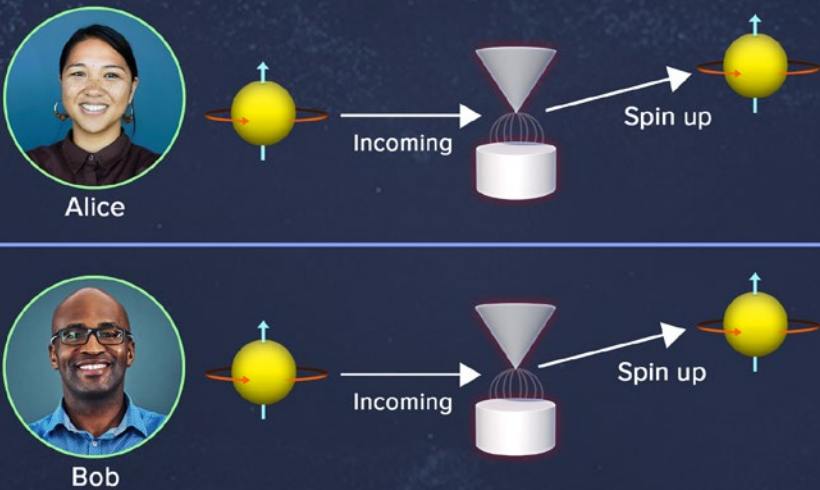
Imagine two electrons. Electron A belongs to Alice; electron B belongs to Bob. The spins of these two electrons constitute a system of two qubits, and they're described by a corresponding wave function.

The wave function assigns an amplitude to each configuration of the system as a whole with respect to something you might observe about it. Imagine measuring the spin in the vertical direction of both qubits. In that case, there are four possible measurement outcomes. For any one qubit, you can only measure spin up or spin down. But once you have two, you have four possibilities: Both spins are up, A is up and B is down, A is down and B is up, or both spins are down.



In this case, the state of the system is some superposition of the four possibilities—the basis states. You can construct any quantum state as a superposition of the basis states. And you would square the amplitudes to get the probability of measuring the state that you're actually looking at.

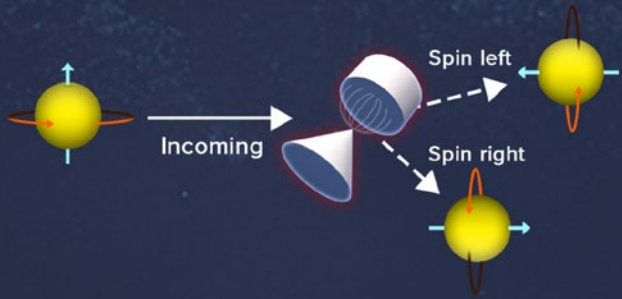
Just because there are two qubits doesn't mean that they are necessarily entangled. Consider, for example, the basis state where both of the qubits are spin up. What would happen if Alice and Bob were to measure their qubits? The hypothesis is that the state of both qubits is spin up, so if Alice measures her qubit along the vertical axis, with 100% probability she will obtain spin up. Likewise, Bob will obtain spin up with 100% probability.



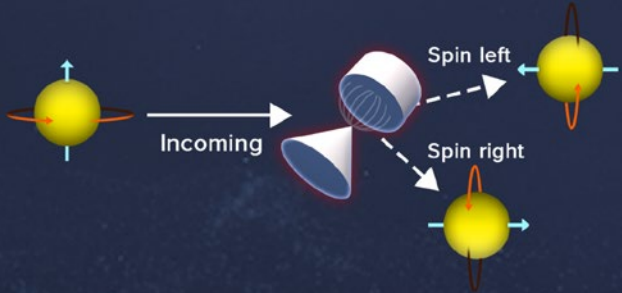
If Alice measures her spin along a different axis, then you have to do some work. If Alice measures her spin along the horizontal axis, for example, then a spin up is equivalent to a 50-50 superposition of spin right or spin left along the horizontal axis, so Alice will have a 50-50 chance of measuring spin right or spin left. And the same is true for Bob.



Alice



Bob



But there's no relationship between them. In each case, you don't learn anything about what Bob will measure just by hearing what Alice actually did measure. The measurement outcomes are completely independent. That's why the wave function of a particle can often be very casually spoken about.

Even though it's not really what the world is—because there's only one wave function for everything—when different parts of the system are unentangled with each other, it is just as if they have their own wave functions. If the two spins are both spin up, then it's just like Alice has a spin-up particle and Bob has a spin-up particle. But that's not the typical, generic case.

Instead, consider an equal superposition of two basis states, one with both spins up and the other with both spins down. In each case, the first arrow in the parentheses represents Alice's spin and the second arrow represents Bob's.

$$\Psi = \frac{1}{\sqrt{2}} (\uparrow, \uparrow) + \frac{1}{\sqrt{2}} (\downarrow, \downarrow)$$

Note that it's  $1/\sqrt{2}$  because when you square it, you need to get the probability. And to get equal probabilities— $1/2$ ,  $1/2$ —you multiply the basis states by  $1/\sqrt{2}$ .

Importantly, in this quantum state, neither spin has a wave function to call its own. The only single wave function that you have is the one referring to the combined system of the two spins. In this case, where you've added two terms together, the wave function is indeed in a superposition, both spins up plus both spins down.

Next, consider what happens when Alice and Bob start measuring their spins, starting from this particular entangled state. If Alice measures her spin along the vertical axis, in this state, she will have a 50-50 chance of getting spin up or spin down, and the same is true for Bob.



But if you learn Alice's outcome before Bob does his measurement, you know what Bob will see with 100% confidence. In that quantum state when Alice measures her spin, it collapses the wave function. And Bob will see the same thing that Alice did with 100% confidence. In the language of textbook quantum mechanics, Alice's measurement collapses the wave function onto one of the two basis states. And after that, you go from a superposition up-up plus down-down to either up-up or down-down. So, then, in either one of those two states, Bob has a known outcome. But the word *known* means that Alice knows what Bob's outcome will be—Bob doesn't necessarily know it yet. Alice has done a measurement, and Bob has consequences, but he doesn't yet have knowledge about them.

In the language of the many-worlds interpretation of quantum mechanics, Alice's measurement has branched the wave function into two possibilities, and that process has created two different Bobs. Each version of Bob will get an outcome that is known for certain once you know which Alice you have—what Alice saw on that branch. That's the power of entanglement in action.

## The EPR Puzzle

In the aftermath of the Solvay Conference in 1927, Einstein remained convinced that quantum mechanics, especially as interpreted by the Copenhagen interpretation, did a good job at making predictions for experimental outcomes. But he thought that it fell short as a complete theory of the physical world. His concerns were finally written up for publication in 1935 with his collaborators, Boris Podolsky and Nathan Rosen. This paper is universally known as EPR—Einstein, Podolsky, Rosen.

EPR thought about the position and momentum of two particles moving in opposite directions, much like the example used here that has one particle decaying into two. But it's easier to get to the point by thinking about qubits. So, consider again two spins that are in the entangled state from the example: a superposition of both spins up plus both spins down.

Imagine that Alice has a qubit, with a spin, and she keeps it in a box at home. She doesn't look at it or measure it. Bob also has his in a box, but flies in a rocket ship to a planet orbiting around the star Alpha Centauri, 4 light-years away.

The particles are entangled in the quantum state that they started with, and nothing happens to them. But entanglement is a feature of the two-particle system, and the entanglement between Alice's particle and Bob's doesn't fade away as they are moved apart from each other. As long as neither Alice nor Bob measures the spin of their qubits, the overall quantum state describing those two particles will remain the same.

Once Bob arrives safely at Alpha Centauri, then Alice finally does measure the spin of her particle. And let's say that they've agreed ahead of time that Alice will do this on the vertical axis. Before that measurement, you were completely unsure what such an observation would reveal for her spin—a 50-50 chance for spin up or spin down—and likewise for Bob. So, suppose that Alice sees spin up. Then, by the rules of quantum mechanics, you immediately know that when Bob does his measurement, he will also measure spin up, even though he's 4 light-years away.

That's in conflict with the intuition you have about how information flows through space-time. Einstein's special theory of relativity implies that nothing can travel faster than the speed of light. Yet here, according to quantum mechanics, a measurement that Alice does has an immediate effect on Bob's qubit, even though it's 4 light-years away.

So, how does Bob's qubit "know" that Alice's has been measured, and how does it know what the measurement outcome was? It would seem from the story that Alice's qubit sent information faster than the speed of light to Bob's qubit and influenced Bob's measurement outcomes. This is the "spooky action at a distance" that Einstein so memorably fretted about.

If Alice measures spin up, she instantly knows what Bob will also measure—spin up, given the quantum state at the start. Whenever he gets around to doing a measurement, you can predict what he will see. The problem is that Bob doesn't know that. In the global description you have, the state of Bob's qubit has changed. It's now either purely spin up or purely spin down, but he doesn't know that it's changed or what it's changed to. In order for him to know what the spin of the particle will be when he measures it, Alice has to tell him. She has to send him her measurement results. To do that, she'll use a means of conventional communication, such as a laser or a light beam, that's limited by the speed of light.

It turns out that well beyond this particular example, this is a general feature of quantum entanglement—the no-communication theorem—according to which an entangled pair of particles can't be used to transmit information between two particles faster than light. So, this is a weird puzzle. Quantum mechanics seems to be exploiting a subtle loophole: The spirit of relativity says that nothing travels faster than the speed of light, and seemingly quantum mechanics is violating this spirit. But it obeys the letter of the law. Actual physical particles and whatever useful information they might convey aren't traveling faster than the speed of light. This is the so-called EPR puzzle.

What Einstein aimed to show is not only that quantum mechanics was spooky, but that it couldn't possibly be a complete theory—that there had to be some underlying, better, more comprehensive model for which quantum mechanics is simply a useful approximation.

EPR believed in the principle of locality—that physical quantities describing nature are defined at specific points in space-time. They're neither spread out all over the place nor do they interact directly with other quantities at a distance. They only interact directly with other things nearby. Said in another way, given the speed-of-light restriction on special relativity, locality would seem to imply that nothing you can do to a particle at one location can have an instantaneous effect on anything, including measurements that might be performed on other particles very far away. On the face of it, the fact that two widely separated particles can be entangled, and that measuring one entangled particle instantly affects the other ones, seems to imply that locality is violated by quantum mechanics.

EPR worked hard to establish that there wasn't some clever workaround that would make everything suddenly seem local, so they suggested a principle: If you imagine you have a physical system in a specified state and there is some measurement you can do on that system such that you know with 100% certainty what the outcome will be, then you're going to say that you associate an element of reality to that measurement outcome. For example, if you have a spin that you've just measured to be spin up along some axis, you know with 100% certainty that it will be spin up again if you measure it with respect to the same axis right away.

If you have a qubit in a pure spin-up state, the EPR authors would say that there's an element of reality corresponding to the spin in the vertical direction: It's spin up. But they would also admit that there is not an element of reality corresponding to the horizontal spin—according to the conventional rules of quantum mechanics—because you don't know what you will get when you measure the horizontal spin of a qubit that is purely vertical. So, they said that a complete theory in the EPR formulation is one in which every element of reality has a direct counterpart in the theory itself. So, a complete theory is one that takes account of everything that they've called an element of reality—everything that gives you a definite prediction for what you will measure. And then they argued that quantum mechanics cannot be complete by this criterion.

As you'll discover, many-worlds offers an answer to this conundrum—but it's a different answer. It is one that Einstein would not necessarily have been happy about.



# 9

## Entanglement Leads to Many Worlds

In 1933, Einstein took a position at the new Institute for Advanced Study in Princeton, New Jersey, which became a center of activity for quantum mechanics. Around the same time, John Archibald Wheeler joined the physics faculty at Princeton University, down the road from the Institute for Advanced Study. One of the many ways in which Wheeler made an impact on physics was in the mentoring of talented graduate students. One of his students was Hugh Everett III, who would introduce a dramatically new approach to thinking about the foundations of quantum mechanics. His idea is that the wave function represents reality and evolves smoothly, according to the Schrödinger equation, and that evolution leads to multiple distinct worlds when a quantum measurement takes place.

## Conceptual Problems of Quantum Gravity

Part of Everett's inspiration was the search for a theory of quantum gravity: a marriage between general relativity, Einstein's theory of gravity, and the new principles of quantum mechanics. In the 1950s, Wheeler had become interested in this problem because the rest of physics—matter, electromagnetism, the nuclear forces—as far as anyone could tell, seemed to fit comfortably within the framework of quantum mechanics. But gravity did not fit, and it still doesn't today.

In 1915, Einstein proposed his general theory of relativity, which says that space and time are unified into space-time and that unified space-time is itself a dynamical entity. Space-time can bend and warp, and what you perceive as that bending and warping is the force of gravity. When you think of an apple falling from a tree, says Einstein, what you're seeing is the implication—the consequence—of the fact that the earth is bending the space-time around it. The theory explains a lot, and following Wheeler, it explains much more than was known when the theory was invented.

But general relativity is still a thoroughly classical theory. There are analogs in the theory of position and momentum, except that rather than the position of a particle or its velocity, there is the curvature of space-time. In classical general relativity, there's no limit, in principle, on how you might measure the curvature of space-time—that's a feature of classical physics. What you can do is try to take that theory and quantize it. In other words, you can construct a theory of wave functions of space-time rather than using a particular classical space-time. But in practice, that's proven very difficult to do.

The difficulties of quantum gravity can be classified as two different kinds: technical difficulties and conceptual difficulties. On the technical side, if you treat general relativity like any other field theory—such as Maxwell's electromagnetism—there's a recipe for trying to convert a field theory

John Archibald Wheeler became known as one of the world's experts in general relativity and popularized terms such as *black hole* and *wormhole*.

into a quantum version. But when you do that with general relativity, the calculations give infinitely big answers, which means that the theory is somehow wrong.

On the conceptual side, it's not certain what's happening when gravity is quantized. In ordinary quantum mechanics, for a typical electron wave function, you cannot predict exactly at which point in space the electron will be found when you measure it. But the notion of a point in space is perfectly well defined in ordinary quantum mechanics. You can specify some location and then ask what the probability is that you will find the electron nearby that location. That's what the Schrödinger equation and the Born rule let you do. But if reality doesn't consist of stuff distributed through space but is a much more abstract notion of a quantum wave function that describes superpositions—not just of different particles, but of different possible space-times—how do you even ask the question, Where are you going to observe a certain particle? The word *where* implies the existence of space and locations. These kinds of fundamental questions are part of what are known as the conceptual problems of quantum gravity.

Another conceptual puzzle appears when you turn to the measurement problem of orthodox quantum mechanics. By the 1950s, the Copenhagen school was established doctrine. Physicists had made their peace with the idea that wave functions will collapse when you do a measurement.

Everett, thinking about quantum gravity, asked the question, What about when the quantum system under consideration is the entire universe? It's crucial in the Copenhagen interpretation that the quantum system being measured is distinguished from the classical observer—the one who is doing the measuring. But if the system is the whole universe, all the observers that exist are part of the quantum system that's being described. There is no external observer that can be appealed to.

## The Universal Wave Function

While Wheeler and some of his colleagues thought about the technical challenges of quantum gravity, Everett became fascinated by the conceptual problems, especially how to handle measurement if the observer is part of the quantum system that you're looking at. Clearly, Everett reasoned, if you're

going to talk about the universe as a whole as a quantum mechanical system, you can't carve out a separate classical realm. Every part of the universe will have to be treated according to the rules of quantum mechanics—including the observers within it. There's no longer the possibility of pretending that observers obey the rules of classical physics. And there will only be a single quantum state: what Everett called the universal wave function. It's an implication of entanglement, the 1935 EPR paper, and Einstein's notion that there is only one wave function.

So, if everything is quantum mechanical and the universe is described by a single universal wave function, how is measurement supposed to occur if there's nothing outside the universe? When you talk about measurement, you must be talking about one subsystem of the universe—that you call the observer—measuring another subsystem of the whole system. And if you treat everything as quantum from the start, you know what it means for two different subsystems to interact with each other. You set up their wave functions and then solve the Schrödinger equation.

So, you don't need to invoke any special rules for measurement. Things bump into each other all the time—that's ordinary physics. The only difference is that you're treating the observer as part of the quantum state, not separate from it. In fact, a huge benefit of this approach to quantum mechanics is that you don't have to talk about observers at all. You can simply talk about a physical object that performs the quantum measurement, and that apparatus is also going to be part of the wave function of the universe. Eventually, you're going to want to ask what observers experience, since you are an observer yourself, but that can wait until the last step.

Everett considered the quantum state to include both the system to be observed and an apparatus doing the observing. And he asked, What would the Schrödinger equation predict for these two quantum subsystems interacting with each other?

As a measurement apparatus interacts with the quantum system, the two will become entangled with each other, since the measurement apparatus is itself quantum. There's no wave function collapse; that idea is not needed because there's no classical realm. The apparatus—which is part of the wave function—itself evolves into a superposition, and the apparatus and the system being observed become entangled with each other.

What is seen as an apparently definite measurement outcome, such as the electron is spin up, that's only relative to a particular state of the apparatus in Everett's idea. So, the apparatus's quantum state could be either that you measured the electron to be spin up or you measured the electron to be spin down. The other possible measurement outcomes still exist according to the Schrödinger equation. And Everett says to accept that—to treat these other measurement outcomes as perfectly real but as happening in separate worlds.

## An Entangled Superposition

According to Everett's theory, what happens when a measurement is made? Imagine that you have a spinning electron that could be observed to be either in a state of spin up or spin down with respect to some chosen axis—so the electron acts as a qubit. Before the measurement, the electron is typically in some superposition of up and down. You have a measuring apparatus, and you're now treating the measuring apparatus as a quantum system in its own right. You don't just magically say that the apparatus looks at the electron and sees spin up or spin down. You have a space of possible quantum states that the apparatus can be in, and it can be in a superposition of those various possible states.

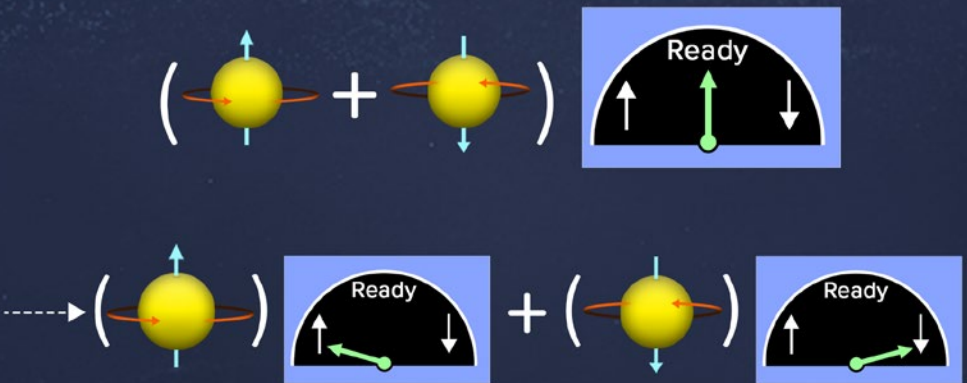
For simplicity, ignore the irrelevant details about all the different atoms that are in the apparatus—which is macroscopic—and instead imagine that it's just in a superposition of three different possibilities: The apparatus can have measured the spin to be up, the apparatus can have measured the spin to be down, or the apparatus hasn't made a measurement yet. The ready state is the state before you've made any measurement, and there are three possible states for the measurement apparatus.

If the measurement apparatus is effective, it's easy to see how the quantum state of the combined spin-plus-apparatus system must evolve according to the Schrödinger equation without actually solving it. You know that if you start with the apparatus in its ready state and with the spin in a purely spin-up state, then you're guaranteed that the apparatus evolves into a pure measured-up state. So, the initial state is that the spin is in the up state and the apparatus is ready. The final state is that the spin is up and the apparatus has measured exactly that.

Likewise, the ability to successfully measure a pure down spin implies an analogous thing: The apparatus must evolve from the ready state to measure down.

When the system is in a state that happens to give some definite measurement outcome—i.e., it's not in a superposition—the apparatus has to successfully return that outcome. It's not going to give a weird superposition or something completely different. But what will happen when the initial, unmeasured spin qubit is not in a pure up or down state but is in some superposition of both?

If you know how the system evolves starting from two different states, the evolution of a superposition of both of those states is just a superposition of the two evolutions. Starting from a spin in a superposition of up and down and from the measurement device in its ready state, this evolves into the state where the spin is up and the measurement apparatus measured it to be up, plus the spin is down and the measurement apparatus measured it to be down. In other words, the final state is an entangled superposition.



At this point, when you have that final state, it's not strictly correct to say that the spin is in a superposition or the apparatus is in a superposition. Entanglement prevents you from talking about the wave function of the spin or that of the apparatus individually, because what you observe about one can depend on what you observe about the other.

In the final state that you got, the only thing you can say is that the spin-plus-apparatus system is in a superposition. So, this final state—this entangled superposition of spin is up and it was measured to be up, plus spin is down and it was measured to be down—is the clear, unambiguous, definitive final wave function for the combined spin-plus-apparatus system, if all you do is evolve it according to the Schrödinger equation and don't allow any mysterious measurements or collapses.

Everybody agrees with this part of the story. What people disagree on is whether the wave function really does always obey the Schrödinger equation. And they also disagree about how to interpret the wave function—whether the wave function really represents everything in reality. But if you do accept those things, this final entangled state reveals the secret to Everettian quantum mechanics. It's just listening to what the Schrödinger equation says: that an accurate measuring apparatus will evolve along with the thing it's measuring into a superposition—even if the apparatus is macroscopic. This process will ultimately be described as branching into separate worlds. There's a world where the spin is up and the apparatus measured it to be up, and there's a separate world for the spin being down and the apparatus measuring it to be down.

But notice that the worlds were always there, and it's the Schrödinger equation that brings them to life. If you believe that an electron can be in a superposition of up and down, says Everett, you should believe that the universe can be in a superposition as well. That's why the Everett interpretation of quantum mechanics has come to be called the many-worlds interpretation.

## Everett's Many Worlds

Everett's many-worlds idea is so simple, obvious, and compelling—why doesn't everyone agree that this is the right way to think about quantum mechanics? The problem is that human beings never seem to come across superpositions involving macroscopic objects in their experience of the world. In particular, if you think of yourself as the apparatus—which you're allowed to do—you never seem to feel like you are in such a superposition yourself. Every physicist who has ever looked at a spin in one of these measurement apparatuses has seen a definite measurement outcome of some kind.

So, it seems that there's a mismatch between what Everett predicts and what is actually observed. And a traditional remedy to not seeing macroscopic superpositions—such as Schrödinger's cat—has been to monkey with the fundamental rules of quantum mechanics in one way or another. In some approaches, you say that the Schrödinger equation isn't always applicable; there's another way that quantum states evolve. In other approaches, you say that there's additional variables over and above the wave function; the wave function is not all there is. Such approaches involve contortions of some sort in order to not accept superpositions as the true and complete description of nature.

Everett didn't like the idea of treating measuring apparatuses as classical or the idea that you have to be silent on what really happens in the quantum realm. So, his prescription is simple: Accept the reality of what the Schrödinger equation predicts. Both parts of the final wave function are real; they simply describe separate, never-to-interact-again worlds.

Everett didn't introduce anything new into the rules of quantum mechanics. He removed some extraneous pieces from the formalism. If the multiple worlds bother you, you have to fiddle with either the nature of quantum states or their ordinary evolution in order to get rid of them.



# 10

## Decoherence Explains Branching Worlds

**W**hen Hugh Everett first developed the many-worlds picture, he argued that it's logically consistent to think of certain superpositions involving macroscopic objects as describing separate worlds. But in the 1950s, at the time he was writing, physicists hadn't yet developed the technical tools necessary to turn this idea into a complete and compelling picture. They didn't yet have the explanation for why separate worlds is the right way to think about macroscopic superpositions. That understanding came later—with the appreciation of a phenomenon called decoherence.

## Decoherence

Decoherence is a way of thinking about the entanglement between an object and the rest of the world. The idea, introduced in 1970 by the German physicist Heinz-Dieter Zeh, has become a central part of how physicists think about the dynamics of quantum systems. To the modern Everettian, decoherence is crucial to making sense of quantum mechanics. It's what explains once and for all why wave functions seem to collapse when you measure quantum systems—and indeed what a measurement really is. Words like *collapse* and *measurement* don't appear in the fundamental statement of Everettian quantum mechanics.

In quantum mechanics, there's only one wave function: the wave function of the universe. But some systems of the universe may or may not be entangled with other subsystems. An individual microscopic particle, such as a single electron, can settle into a quantum state that is completely unentangled from the rest of the world. In that case, it's sensible to talk about the wave function of the electron.

For macroscopic objects, things aren't that simple. It's easiest to think of a real-world apparatus that will measure a quantum spin. You send a spin through a magnet and then imagine a dial with a pointer that indicates whether the spin is measured to be either up or down.

An apparatus like that will not stay separate from the rest of the world, even if it looks like it's just sitting there not disturbing anything else. In reality, the air molecules in the room are constantly bumping into it. There are photons of light in the room that are bouncing off it. So, how do you deal with that? Classically, you can just ignore it. But quantum mechanically, it makes a big difference.

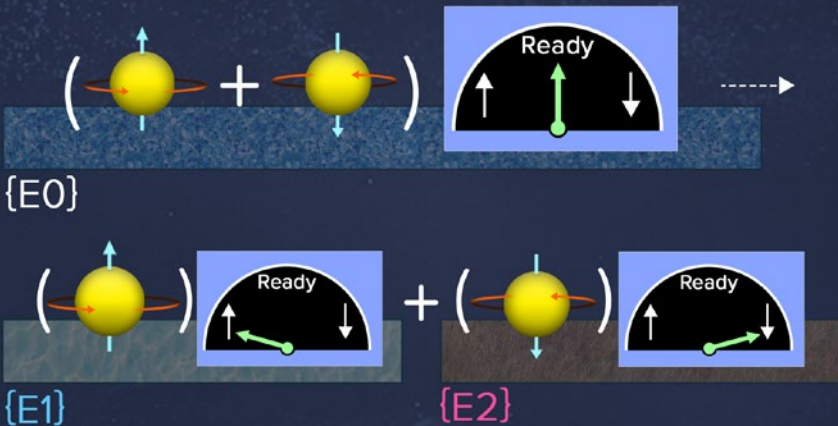
All that other stuff—the entire rest of the universe—is considered the environment. It's everything you're not paying close attention to in your quantum experiment. In ordinary situations, there's no way to stop a macroscopic object from interacting with its environment, even if the atoms of air and photons are going to hit it very gently.

Now imagine that you do somehow start with your macroscopic apparatus in a superposition of the pointer pointing toward both spin up and spin down. The atoms and photons in the room that are bumping into it are very quickly going to cause the apparatus to become entangled with the environment. Just a single photon could reflect off the dial if the pointer is pointing in one direction, but the same photon would be absorbed by the pointer if the pointer is pointing in a different direction.

So, the wave function considered previously, in which a measuring apparatus that's macroscopic became entangled with a single qubit, wasn't quite the whole story for being careful and rigorous about quantum mechanics. That story should also have included the states of the environment, which itself is quantum mechanical.



So, you have an initial state in which spin by itself is in a superposition of up and down, the apparatus is in its ready state, and the environment is in some state—call it  $\{E_0\}$ . It evolves into a state in which there's a part where the spin is up, the apparatus measured it to be up, and the environment is in a different state—call it  $\{E_1\}$ —plus the spin is down, the apparatus measured it to be down, and again the environment is in some different state—call it  $\{E_2\}$ .



The good news is it doesn't matter what the environment states actually are at the detailed level. Somehow, atoms and photons will interact differently with different states of the macroscopic measurement apparatus—and that process is decoherence. Macroscopic objects are going to become entangled with the environment. You can't keep track of everything that's going on in the environment, but you don't need to. All that matters is that the states of the environment that are attached to different states of the apparatus in the wave function are different from each other. That fact—decoherence—comes with universe-altering consequences.

## Branching

Decoherence causes the wave function to split—or branch—into multiple worlds. The reason is because individual states of the system and apparatus become entangled with different states of the environment. And there are many such states, or ways for the atoms and photons to arrange themselves in the environment. There are so many that those environment states are never going to line back up to once again become equal on the different branches. And crucially, during this process, because there's only one wave function of the universe, any observer—including yourself—branches into multiple copies of themselves, along with the rest of the universe.

After branching, each copy of the original observer finds themselves in a world with some particular measurement outcome. As far as they're concerned, the wave function seems to have collapsed. They only see what's happening on a single branch where there is some definite outcome. An omniscient physicist knows better: that the wave function is only obeying the Schrödinger equation. The collapse is just apparent. It's due to decoherence entangling the environment with the rest of the wave function and splitting it into branches.

How often does branching happen? How many worlds are there? For better or for worse, it's not known how often branching happens because it's not even certain whether that's a sensible question to ask. The real-world answer depends on whether there are a finite or infinite number of degrees of freedom in the universe. That is currently an unanswered question in fundamental physics.

What *is* known is that there's a lot of branching going on. The branching of the wave function happens every time a quantum system that is in a superposition becomes entangled with the wider world.

Understanding decoherence enables you to explain why separate branches count as different worlds. You simply derived a single quantum state describing a spin and apparatus and environment. But what makes you say that it describes two worlds rather than just a single world?

This is an interesting, knotty philosophical question as well as a scientific one. You have to think about what it means to be a world and how that maps onto what's going on in the quantum wave function. So, what does it mean to be a world? One thing you would like to have to define a world is that the different parts of it can, at least in principle, interact with each other and affect each other in some profound way.

Consider the following scenario: In Ghost World, when living beings die, they all become ghosts. These ghosts can see and talk to one another, but they cannot see or talk to living beings—nor can living beings see or talk to them. They live on a separate ghost earth, and they can build ghost houses and go to ghost jobs, but neither they nor their surroundings can interact with living beings and the stuff around them in any way. If this scenario were anything like reality, it would make sense to say that the ghosts inhabit a truly separate world—for the fundamental reason that what happens in Ghost World has no bearing on what happens in the living world.

When you apply this criterion to what you've seen in the world of quantum mechanics, you're not trying to say that different subsystems of the wave function are separate worlds. A spin and the measuring apparatus are not separate worlds. They can influence each other. Instead, you're asking whether one component of the wave function—say the apparatus wave function, specifically the component where the dial is pointing to spin up—can influence the other component of the wave function, where the apparatus dial is pointing to spin down.

## Macroscopic Objects and Irreversibility

It's important to emphasize the role that's being played by the fact that the apparatus is macroscopic. It almost sounds like the old-fashioned Copenhagen view of macroscopic being classical and microscopic being quantum. But that's not what you have.

The difference between microscopic and macroscopic in Everett's idea is simple. Microscopic objects can live in quantum superpositions without decohering. A microscopic object such as an electron is not necessarily constantly bumping into its environment. But a macroscopic object simply cannot help but interact with its environment. It cannot help but become entangled and decohering. In the Copenhagen approach, you had a vague notion of the boundary between micro quantum systems and macro classical ones. In many-worlds, there is no longer any such vague idea. Macro systems are defined as ones that quickly decohere.

It's also important to emphasize that true decoherence is an irreversible process. Once it happens, it's not going to unhappen—that would be two separate worlds coming back together. You don't expect that to ever occur. The reason is basically the same reason that you can mix cream into coffee, but you cannot simply unmix them by randomly stirring your spoon. And the reason for that is that there are many more ways for cream and coffee to be mixed together than for them to be unmixed. By just playing around, you're never going to end up in the very unlikely unmixed state.

Likewise, when a macro system in quantum mechanics becomes entangled with its environment, there are many different environment states that it could become entangled with. Think of the position and velocity of every photon in the room. The chances that two different environment states entangled with two different states of a macroscopic object will randomly line up with each other and erase the decoherence is unimaginably small. It's not something you need to worry about occurring.

So, the many-worlds formulation of quantum mechanics removes any mystery about the measurement process and the collapse of the wave function. You don't need special rules for what happens when you make an observation.

The only thing that ever happens is that the wave function keeps chugging along in accordance with the Schrödinger equation. And there's nothing special about what constitutes a measurement or an observer. A measurement is simply any interaction that causes a quantum system to become entangled with its environment, leading to decoherence and branching into separate worlds. And an observer is any system that brings such an interaction about—any macroscopic system that keeps bumping into its environment. Consciousness and awareness have nothing to do with it. The observer could be an earthworm, a microscope, or a rock. There's nothing special about macroscopic systems. The only thing that is useful about them is that they can't help but interact and become entangled with their environments. The price you pay for such a powerful, simple, compelling unification of quantum dynamics is that you get a large number of separate, distinct worlds.

## The Multiverse versus Many Worlds

As similar conceptually as they might sound, the set of all branches of the wave function is very different from what cosmologists often call the multiverse. The cosmological multiverse is a collection of different regions of space—generally very far away from each other—where local conditions look very different and maybe even have different local laws of physics. That's an entirely different physical picture than many worlds.

The worlds that Everett says exist are truly parallel to each other. They simultaneously exist, but they're not embedded in any bigger space. If you want to know how far away the worlds are from each other, there's no such thing as the distance between the worlds. There's no such thing as where the other worlds are located.

The Everettian worlds simply exist simultaneously and parallel to each other. One way of thinking about it is the worlds are not located in space; space exists separately within each world.

Everett himself, writing his PhD thesis in the 1950s, wasn't familiar with the modern picture of decoherence, so his own story wasn't quite as robust and complete as the one that's been painted here. But his way of rethinking the measurement problem and offering a unified picture of quantum dynamics was compelling from the start.



# 11

## How Entanglement Powers Quantum Computers

**T**his lecture explores the basics of quantum computers and how they operate. There is a natural connection between the ideas of the many-worlds interpretation and quantum computers. And quantum computers are in the forefront of modern thinking about quantum mechanics and how to put it to work.

## Classical Computation Theory

In principle, a computer theoretically is a simple thing. It takes some input, processes it through the operation of an algorithm, and provides corresponding output. The main point of an algorithm is that it is precise and rigorous. Algorithms are spelled out, every step, in brutal detail.

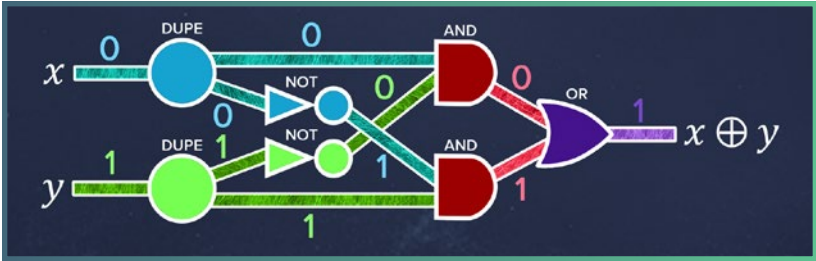
It turns out to be easiest to consider this chain (input, then algorithm, then output) if the information of interest is broken down to its fundamental unit: the bit. A bit is a single piece of information; it's a variable that can take on only two values. You can label the values in different ways, but 0 and 1 are common, as are true and false.

Any number can be expressed as a string of bits. All you have to do is write it in binary, or base-2, notation. In binary notation, the rightmost digit is the number of 1s in the number (either 0 or 1), and the next digit is the number of 2s, then the number of 4s, the number of 8s, and so on, through all the powers of 2 ( $2^N$ ).



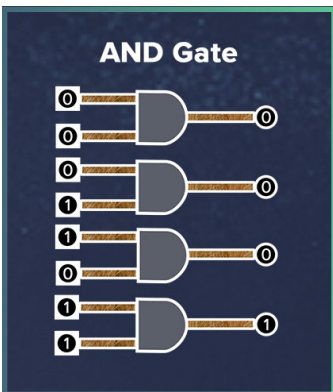
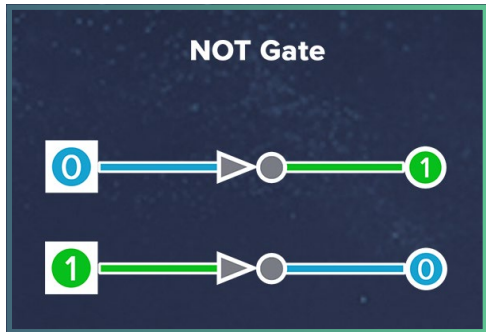
For example, the number 13 in binary becomes 1 1 0 1. That's a single 8, a single 4, zero 2s, and a single 1:  $8 + 4 + 1 = 13$ . Then, you can use some kind of code to also assign such numbers to letters or other symbols. In this way, any information can be expressed in terms of binary numbers—and therefore in terms of bits. Words, pictures, and sounds can all be turned into bits. And this makes things easy because you don't need to think about how a computer should manipulate an arbitrary set of symbols or pictures; you only need to think about how computers manipulate bits or strings of bits.

When you're dealing with single bits, 1 is often thought of as representing true and 0 as false. Now what you need is an algorithm, which will process input bits and return some output. What really matters is that the algorithm itself doesn't depend on a specific input—that you can feed it different kinds of bits, and whatever bits you feed it, it will do its job and provide some output. An algorithm might add or multiply two or more numbers, for example.



You can think of an algorithm in terms of a set of certain basic operations called logic gates. The logic gates are connected by wires, and imagine that the information flows through the wires. The individual bits travel down the wires, and then the gate takes in a certain set of bits and outputs another set of bits.

Consider a simple example called the NOT gate. A NOT gate takes in one bit and outputs one bit. And the operation of the NOT gate is to flip the bit, meaning that if you get an input of 0, then the output of the NOT gate is 1, and vice versa.



There are other famous gates that take in two bits and output just one. For example, the AND gate is true only if both input bits are true. So, if both inputs are 1, you get a 1. But in any other combination, the AND gate will output 0.

Similarly, the OR gate asks if either one of the input bits is true and returns 1 as long as either one of the inputs is 1 and returns 0 if they're both 0.

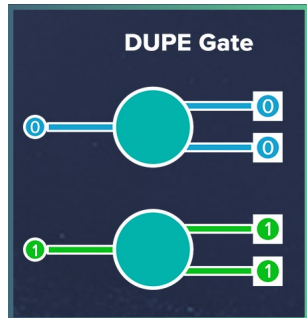
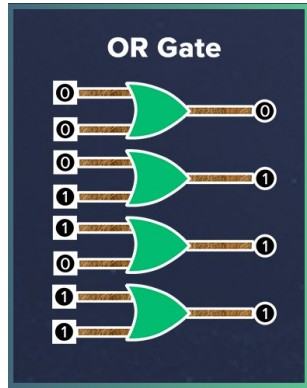
A DUPE gate (for *duplicate*) takes a single bit and outputs two copies of it. In quantum computers, no such gate exists; there's no way in quantum computers to simply duplicate a qubit.

In classical computing, this simple set of gates—NOT, AND, OR, and DUPE—is universal. Remarkably, any deterministic algorithm (meaning whenever you put the same input into the algorithm, you will always get the same output) that operates on a set of bits can be constructed using just these basic gates. All you have to do is string them together in an appropriate combination and order.

## The Conservation of Information

When you ordinarily think about the operations that are allowed under the rules of classical mechanics, many times you don't keep track of absolutely everything happening in the universe—all the little bits that tend to be ignored. When you flip a switch by hand, for example, from a 0 to a 1, there are noise, friction, and other kinds of interactions with the environment.

So, you can easily imagine irreversible operations, even though you know that the underlying rules of classical mechanics are perfectly reversible—they work equally well forward or backward in time. You forget that, because you allow for noise, friction, and entropy. When you're imagining classical algorithms, you can forget about that kind of environmental interaction. You can just invoke whatever manipulations of the bits you want to imagine.



In quantum mechanics, however, you do need to keep track of absolutely everything that's going on. And that's ultimately because you want to take advantage of the fact that the quantum version of bits—qubits—can be entangled with each other. That's going to give power to quantum computers; otherwise, you're just pushing around classical information. If any of your qubits starts interacting with the outside world, then there's the possibility that they will decohere. If decoherence sneaks in, it can ruin the entanglement between different qubits in the quantum computer. And that is ultimately because there is only one wave function of the universe.

Both the equations of classical mechanics and the Schrödinger equation of quantum mechanics have an important property—namely, that information is conserved. This is the basic idea behind Laplace's demon (which was introduced in lecture 2): Specifying the state of the system at one time is enough to determine what it will be, in principle, at any other time.

In the context of a logic gate, conservation of information is simply that you can reverse the gate. Given the output, if you knew which gate you ran it through, you can reconstruct the input. So, a NOT gate, for example, is perfectly reversible. In fact, the reverse of a NOT gate is just a NOT gate—it reverses itself. Two NOT gates in a row return the classical bit you started with.

But a classical gate like the AND gate is not reversible. An AND gate takes in two bits and produces just one. If it's 1 and 1 coming in, you get 1; anything else gives you a 0. So, there are multiple possible inputs that get you a 0 output. There are four possible input states overall: two for the first bit and two for the second bit. And there are only two possible output states; there's just one bit. So, reversal can't be possible. The AND gate will take multiple inputs and map them to the same output, so you don't know where the output came from.

The reason you are nevertheless allowed to think about AND gates in classical computation theory is precisely because you're ignoring friction, dissipation, and lots of other things. The system is not isolated by itself. You could indeed reverse the operation of any gate if you were Laplace's demon—if you could track the motion of every atom and every photon in the universe. But

you're not. So instead, thinking as a real-world, finite human being, you just allow yourself to think of certain observable processes as if they were truly irreversible. In the macroscopic world, that's what appears to be true. And that includes the idea of passing classical bits through certain gates. You could also do that for a quantum computer, but it would undermine the whole point.

In a quantum computer, you're going to want to keep qubits isolated from the environment to preserve their entanglement with each other. And as a result, you're only going to be interested in reversible gates—gates that do not involve any friction or dissipation. And one immediate result of that is that all of your gates are going to have to have the same number of input qubits and output qubits. It's only in that case that the amount of information of input and output matches. So, if you only care about some subset of the information—if we have two input qubits but are only going to pay attention to one output qubit—that's fine. You can always throw the rest away at the end of the process. But that means that you cannot just borrow the AND gate and the OR gate from classical computation theories, since those gates are not reversible.

There is one exception to the principle of conservation of information. In quantum mechanics, there are special rules for measurement. Measurement is an irreversible process, and measurement destroys information. If you have a state before the measurement in a superposition and then you measure it—so you have an unknown superposition of spin up and spin down, measure it, and see that it's spin up—you cannot reconstruct the starting state after you've measured it. You don't know what the superposition was.

So, the general form of a quantum computer is to take some input qubits; act on them with perfectly reversible gates, preserving all the information; produce some output qubits; and then, only at the very end of the process, measure some of those output qubits. Any peeking along the way would destroy information and entanglement and potentially ruin the computation.

It's because of this principle of conservation of information that there is no DUPE gate for quantum computers. You can't take an arbitrary qubit and create a duplicate of it. That would bring new, more information into the world. In fact, this is a famous theorem in quantum mechanics: the no-cloning theorem. It is impossible to clone any quantum state.

So, what can you do if you can't have a DUPE gate? There's a quantum version of the NOT gate called the CNOT, or controlled NOT, gate, which has two inputs and two outputs. It's the same operation as before, but it's thought of in a slightly different way.

So, just as in classical computers, it turns out that for quantum computers, you can specify a small universal set of gates from which any deterministic algorithm you want can be constructed. In fact, you can do this with the CNOT gate plus a special series of one-qubit gates.

So, a complete quantum algorithm will consist of a set of input wires, an appropriate series of connected gates, and a set of output wires, plus instructions about how you are to measure the qubits to get the answer to the computation. Because it's quantum mechanics, there is a new feature. It is possible that you might only get the right answer with a certain probability rather than with certainty. The Born rule still matters. If you start with a superposition, you might not get 100% predictability in your measurement outcome. But that's OK. Even for a very important computation, all you have to do is repeat the calculation over and over again. As long as you're more than 50% likely to get the right answer, repeating it over and over again builds your confidence that you have the answer correct.

**British physicist David Deutsch, one of the strongest proponents of the many-worlds interpretation, pioneered one of the most exciting areas of physics: quantum computing.**

## Quantum versus Classical Computers

It turns out to be difficult to get a reliable handle on the important question of what problems might be solved much more efficiently on a quantum computer than a classical computer. There's an ongoing race to solve difficult problems faster on quantum computers.

But it's hard to answer that question. One reason is because it's difficult to determine precisely how good a classical computer can be. You want to show that a quantum computer is faster than the classical one, and that's hard to do unless you are sure how fast the classical computer can, in principle, be. More than once, people have announced a faster quantum algorithm for some problem, only for someone else to later invent a classical algorithm that is just as fast.

So, the state of knowledge is very much uncertain, but there are some things that are known. There's a famous example called Shor's algorithm, named after physicist Peter Shor, that provides a way to factor integers (meaning breaking a number into its prime factors, such as the number 15 is 3 times 5, with both 3 and 5 being prime numbers).

Shor's algorithm offers a way to factor integers very quickly. It is believed that there is no classical way to factor numbers quickly—that it would take an exponentially large number of steps for a classical computer to factor a large number. But as of 2023, that hasn't been proven rigorously. This is an important practical application because factoring large numbers plays a crucial role in cryptography and computer security systems.

One problem that is tailor-made for quantum computers is simulating quantum mechanical systems. A good-sized quantum computer will provide novel insights into the behavior of systems in chemistry, physics, and molecular biology. But a good-sized quantum computer will be hard to build even though the technology is progressing rapidly. And the biggest obstacle is decoherence. If just a single photon bumps into your qubit and becomes entangled, that can ruin the whole computation.

And blackbody radiation exists. Everything in the world with a nonzero temperature emits some photons—even the machine you're making your quantum computer out of. Their number can at least be attempted to be limited by keeping the apparatus extremely cold, so typical quantum computers are kept close to absolute zero in temperature. And there are also techniques to correct quantum errors—events that can fix a stray qubit or two that has gone awry.

And there are quantum computers currently in operation. There are as many as 400 or more qubits involved, and there are plans to soon go to thousands of qubits. A typical CPU in a home computer is 32 or 64 bits. The difference is that in a classical computer, there's no problem linking together many processors. You can go from one tiny processor to a giant processor; the biggest supercomputers will have millions of core processors.

That idea of linking things together doesn't work in a quantum computer. It's too hard to maintain the entanglement between two qubits that aren't even in the same processor. So, the future of quantum computing looks very promising—but there are technological challenges remaining to be addressed and crucial questions to be answered.



# 12

## Too Many Worlds! Five Objections Answered

**T**he underlying formalism of the many-worlds interpretation of quantum mechanics—the basic rules that give rise to the theory—is efficient and successful. Nevertheless, not everyone agrees that the theory is on the right track. People worry generally that the theory stretches the imagination to the breaking point. This lecture confronts some of the objections to the many-worlds interpretation.

## Occam's Razor

The simplest and most direct objection to the many-worlds interpretation is that it violates Occam's razor: the principle that you should look for the simplest, effective explanation for any set of phenomena. Or, in the words of William of Occam, the medieval scholastic philosopher, "Entities should not be multiplied beyond necessity." In other words, you should favor the theory that involves the fewest entities.

Superficially, that does not sound like the many-worlds interpretation of quantum mechanics. In many-worlds, there's maybe an infinite number of universes. That sounds like a lot of entities. But it depends on what is meant by an entity. Think about, for example, the mathematical idea of the whole numbers—the set of numbers 0, 1, 2, 3, ... . This set is infinitely big. But nobody thinks that the idea of the whole numbers is complicated.

**Many-worlds is extremely simple in the number of ideas underlying the formalism, with just two: the quantum state and the Schrödinger equation.**

As a concept, the whole numbers are very simple. You start with a number labeled 0, and you construct the next number by adding 1. And then you construct the next number after that by adding 1 again. And you repeat. You get all of the whole numbers. The set of the whole numbers is considered simple because it's not the number of elements in the set that matters. What matters is the number of ideas. The whole numbers are simple as a concept because there's a simple rule that generates them: You start with 0, you add 1, and then you repeat. There's a simple algorithm for making the thing, and that's how simplicity is measured.

The same thing goes for many-worlds. While it's true that there are many worlds, the theory isn't constructed by postulating the existence of many-worlds. All that's posited are wave functions and the Schrödinger equation. And those ingredients are part of anybody's version of quantum mechanics.

Once you've made those assumptions, the worlds are there. If you believe that a single electron or a set of qubits can be in a superposition—which you have to believe in quantum mechanics—then you should believe that the whole universe can be in a superposition. It's intuitively challenging, but the equations are unambiguous.

If you really don't like the other worlds, you have a responsibility to do the work of developing an alternative theory in which the dynamics of the wave function doesn't bring the other worlds to life, as the ordinary Schrödinger equation does.

## Time Asymmetry

With quantum mechanics according to Everett, there seems to be a difference between going forward in time and going backward. It's a technical worry that wave functions branch toward the future, but they don't unbranch—they don't come together from the past to the future. Two separate worlds never combine back into a single one. That is an asymmetry. Branching increases the number of worlds as you move into the future. That number, according to the theory, is lower in the past. So, this is a temporal asymmetry, or an asymmetry in time. In many-worlds, you can distinguish the past from the future by counting how many worlds there are, if that's something you could do.

This is somewhat puzzling. The Schrödinger equation itself—which is in the basic postulates—doesn't pick out any directionality to time. It's perfectly time symmetric. You can run the Schrödinger equation backward or forward equally well. It's reversible.

But this is actually a familiar problem with a familiar solution. In classical mechanics, the fundamental equations—Newton's laws of motion—are also time symmetric. But in the real world, there are many processes that show a pronounced arrow of time pointing from the past to the future. You can mix cream into coffee, but you can't unmix them. You're born young and grow older as time passes. You can remember the past but not the future. How is all this possible if the fundamental laws of physics are time symmetric?

The answer starts with the realization that human beings, as finite creatures, don't observe the world at a fundamental level. Laplace's demon would do that, but nobody is Laplace's demon. What a human actually sees are macroscopic features of systems. You don't see the microscopic, individual atoms and molecules that macrosystems are made of. What you see, in other words, is a coarse-grained universe: one in which many possible arrangements of the atoms can look indistinguishable to you.

When you have a cup full of coffee and cream, you might be able to see the basic outline of where the cream is and where the coffee is, but you don't see the position of each individual molecule of coffee or cream. So, for any given macroscopic configuration of matter, you can ask how many arrangements of the microscopic molecules might look that way, and the answer measures the entropy, or the disorderliness, of the system.

This was understood by Austrian physicist Ludwig Boltzmann in the 1800s. High entropy corresponds to situations where many arrangements would look the same to you. A low-entropy arrangement corresponds to relatively few ways that you can rearrange the atoms and molecules.

Entropy tends to increase because there are more ways to be high entropy than to be low entropy. That's the second law of thermodynamics, a famous rule in physics that underlies the arrow of time. Entropy increases as you go to the future, and it was lower in the past. But this rule that entropy is increasing—which sounds like the rule that the number of branches of the wave function of the universe is increasing—only makes sense if you imagine that you started in a low-entropy state.

And, indeed, you did. About 14 billion years ago, near the big bang, the entropy of the universe was exceedingly low. The entropy has been increasing ever since, and that process of entropy increase is what's responsible for all the different ways in which the past is different from the future. And the same basic story holds for many-worlds quantum mechanics. The wave function branches toward the future not because of the fundamental Schrödinger equation, but because in the distant past, all of the branches were combined into just one branch or a relatively small number of branches. And ever since then, the universe has just been doing its thing—which naturally involves differentiating into multiple branches.

Increasing entropy is irreversible. Technically, it's actually only very, very unlikely for entropy to go down, but it's so fantastically unlikely that you need not ever worry about encountering it in the entire history of the observable universe if the systems you're thinking about are macroscopically big in their extent.

The single assumption that you start with low entropy—with just one branch—is all you need to explain both the second law of thermodynamics and the branching structure of many-worlds. So, the fact that the branching happens in one direction of time but not the other is not a separate objection to many-worlds.

## Infinity

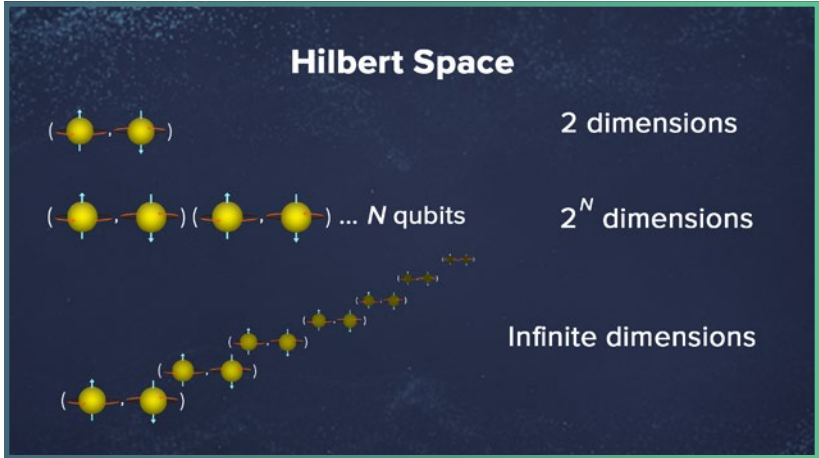
Infinity by itself is not a worry, but it raises questions: Are there truly an infinite number of worlds? And, if so, does that mean that every conceivable world exists somewhere?

The answer to the first question is that it's unknown whether there is a finite number or infinite number of worlds. The answer to the second question is no.

The space of all possible quantum states is called Hilbert space. It is a vector space, or a collection of a certain number of vectors, which are things you can add and multiply by numbers. Every vector space has a certain number of dimensions. A line is a one-dimensional vector space. A plane is a two-dimensional vector space. Ordinary space is a three-dimensional vector space. The number of possible worlds that there can be in quantum mechanics is described by the number of dimensions of Hilbert space. When you ask how many worlds there could be, you're asking how big the Hilbert space of the real world is—how many dimensions it has.

You can think up certain model systems that can be described exactly and say how big their corresponding Hilbert spaces are. But when it comes to the real world, it's unknown how big Hilbert space is. A single qubit has a two-dimensional vector, or Hilbert, space. If you have a collection of  $N$  qubits,

then the corresponding Hilbert space has  $2^N$  dimensions. That's pretty big. If you have a lot of particles, each one has a spin, and you have a pretty big Hilbert space to describe all that at once.



The real world is much worse than that. When you have just a single particle that can move through space, its Hilbert space is infinite-dimensional. You can think of that by thinking of every different location in space as a place you could observe the particle, and Hilbert space will have one dimension for every possible observational outcome. That's an infinite number of dimensions for a single particle moving through points in space.

It's reasonable to think that if just a single particle is described by an infinite-dimensional Hilbert space, then the whole universe, which has a lot of particles in it, should also have an infinite-dimensional Hilbert space. But interestingly, this is maybe not the case.

In the real world, there is a complicating factor. The real world is not a single particle moving through space. The real world has gravity in it. And in a theory with gravity, like the real world, when you try to pack too many particles into the same region of space, the whole thing will collapse under the force of gravity and make a black hole. And black holes only have a finite-dimensional Hilbert space associated with them.

The entropy of a black hole is finite, and the entropy is related to the size of Hilbert space. So, a black hole is as much stuff as can be put into a finite region of space, and it's thought that black holes only have finite-dimensional Hilbert spaces. So, it's possible that the universe itself—or at least maybe the visible part of the universe—is also described by a finite-dimensional Hilbert space. Maybe the thought that a single particle has an infinite-dimensional Hilbert space is too simple and working harder results in a finite-dimensional answer.

Once again, modern physics is not quite up to the task of providing a definitive answer to this one. It's not known whether the Hilbert space of the real world is finite- or infinite-dimensional. But if it is finite-dimensional, then the number of worlds in many-worlds is a large number—but not infinite.

## Immortality

Imagine a deadly device that is triggered by a quantum measurement. There's a 50% chance of triggering a gun that shoots a bullet into your head at close range and a 50% chance of doing nothing, according to some quantum measurement device. Then, by the rules of many-worlds, this implies the existence of two branches of the wave function: one of which contains a live version of you and the other of which contains a dead version of you.

The philosophical twist is that the branch on which the gun is fired isn't one that any version of you ever gets to experience. There is your descendant in that world, but your descendant in that world is dead—not experiencing anything—whereas your descendant continues unharmed on the branch where the gun didn't fire.

So, in some sense, you will live forever, even if you repeat this crazy procedure over and over again. One might go so far as to argue that you shouldn't object to going through this experiment, as long as you don't care that other people will be sad if you go away. But in the branches where the gun fired, you don't really exist because you're dead, while in the single branch where it failed to fire time after time, you're perfectly healthy and know that you're around.

This is called quantum immortality as a thought experiment. It's been popularized by physicist Max Tegmark. Of course, it's not recommended that you try such an experiment or even contemplate it. In fact, the logic behind not caring about the branches in which you've been killed in this experiment is more than a little wonky.

To see that, think about life in a classical universe—so there's just one world. If you thought you lived in such a universe, ask yourself whether you would mind if someone sneaked up behind you and shot you in the head so that you died instantly. Again setting aside other people getting upset, most people would say that they do mind and would not be in favor of that happening. But by the previous logic, even in the classical universe, you really shouldn't mind. After all, once you're dead, there's no "you" to be upset about what happened. That's a little misguided as far as intuition goes. So, what's going on?

The point being missed by this analysis is that you are upset now—while you are still alive and feeling—by the prospect of being dead in the future, especially if that future comes sooner rather than later. And that's a perfectly valid perspective. Much of how you think about your current life depends on a projection into the rest of your existence. Cutting that existence off is something you are perfectly allowed to object to, even if you won't be around to be bothered by it after it happens.

And if you accept that reasoning, then the quantum immortality experiment turns out to be just as bleak and unpalatable as your immediate intuition might suggest. It's acceptable for you to yearn for a happy and long life for all the future versions of you that will end up in various branches of the wave function, as much as it would be valid for you to hope for one long life if you just thought there was a single world.

## Energy Conservation

It's traditionally taught that the universe contains a certain amount of energy and that the total energy is conserved. So, if you're constantly making new worlds in the many-worlds interpretation of quantum mechanics, where does the energy to make those new worlds come from?

Unlike some of the more philosophical worries, where opinions may differ, this is a physics question, and for this one, the equations provide a definite answer. Energy, overall, is perfectly conserved in the many-worlds interpretation of quantum mechanics. What's difficult is translating the equations that say that into words.

The equation in question is the Schrödinger equation, which is unambiguous: However you want to define the energy of a quantum state, that energy quantity is conserved under the Schrödinger evolution of the wave function. And that's the only kind of evolution that exists in many-worlds.

Many-worlds is the one theory of quantum mechanics where the wave function of the universe just obeys the Schrödinger equation, conserving energy all the time.

**Many-worlds is special in that energy is at least conserved in the overall wave function.**



# 13

## Testing the Many-Worlds Interpretation

**T**he previous lecture offered a survey of some of the possible objections people might have to the many-worlds formulation of quantum mechanics. While those particular objections were not convincing, some objections are worth worrying about more than others. This lecture and the next will discuss two of the more worrisome aspects of the theory: One is its testability, and the other is its ability to recover the probabilistic predictions of ordinary quantum mechanics.

## Falsifiability and the Demarcation Problem

How would you ever know if there really were all those other worlds out there? This worry is often phrased in terms of an idea called falsifiability: What is the experiment you could do to show that the theory you're thinking about is false, if it were false?

The idea of falsifiability arises as an answer to what is called the demarcation problem: What counts as science and what does not? This is a general question in the philosophy of science, not just physics. And like many issues in philosophy, the final answer is not certain.

Falsifiability was popularized by British intellectual Karl Popper, one of the most influential philosophers of science in the 20th century. Before Popper, plenty of people thought about the demarcation problem, and philosophers typically tended to favor the idea that scientific theories get confirmed by collecting evidence that favors them.

But Popper worried about that way of thinking. He worried that you could seem to confirm a theory just because your theories are so vague that they predict everything. If your theory is that flexible, you can always come up with an explanation no matter what you observe. As Popper correctly emphasized, that's a bad quality for a theory to have.

What you want in a scientific theory is to be compatible with some hypothetical future observations but incompatible with others. In other words, the falsifiability criterion states that truly scientific theories are the ones that offer specific predictions that you might imagine testing by possible or conceivable observations.

In Popper's view, if a theory could, in principle, be proven false by the appropriate experimental results, then it's a scientific theory. If it can't be proven false by any conceivable experimental results, then it doesn't qualify as scientific.

Popper had a picture of how he thought science should work and make progress: Namely, you imagine all possible falsifiable theories and then set about falsifying all the ones that you can. What is left standing are the theories that are true. So, you've discovered the truth through this process of falsification.

The argument is that many-worlds should not be rejected on the grounds that it's not falsifiable: first, because it is falsifiable, and second, because falsifiability is a trickier concept than is generally understood. Let's address the second issue first to try to understand what made falsifiability such an attractive idea in the first place and how it might apply or not apply to many-worlds.

The intuition tends to be that you gather support for ideas by verifying or confirming them, and Popper set out to offer an alternative view. He was inspired by his worries about theories that claimed to be scientific, and he wanted to argue that these theories were not as scientific as they claimed. Falsifiability, in his mind, was meant as a corrective to the claims of such theories to be scientific.

In the opinion of most modern contemporary philosophers of science, falsifiability doesn't actually work as a solution to the demarcation problem. Scientists, as opposed to philosophers, like to bring up falsifiability, partially because it's easy to understand as a concept. Scientists don't like to get dragged into convoluted philosophical arguments, and falsifiability is an attractive motto as an answer to the demarcation problem. But the reality of scientific practice and the deep philosophical implications thereof are messier than that.

You might imagine that there exists a popular scientific theory, but then along comes some killer experiment, and the results definitively rule out the theory. In a Hollywood movie, that would be a dramatic moment. Even as the oversimplified version of history that scientists often tell themselves, it's useful to imagine these key experiments suddenly changing everyone's minds. But it's not how things actually work in the real world.

One big reason why falsifiability is not seen as the final answer to the demarcation problem is that, in contrast with the romantic ideal of a single experiment that shuts the door on a theory's prospects, most scientific discoveries have a more ambiguous impact, especially at first.

In the real world of scientific practice, when data seemed to disagree with your theory, your work is not over; instead, it's just beginning. You don't know ahead of time what that means—not until you collect more data and think about the possible alternative theories.

Individual experiments are crucially important, but they are very rarely the final word. In any real-world case, it's simple to explain why an overly naive view of falsifiability would fail to capture the messy reality of the scientific process. Experimental results can be wrong, or they can be right but misinterpreted. Theories can be slightly altered to fit the data without changing their fundamental assumptions. And, importantly, this kind of messiness is the point; it's not a bug but a feature of how science works. Science proceeds via an ongoing dialogue between theory and experiment, always searching for the best-possible understanding. It's not that falsified theories are cleanly lopped off one by one.

A more modern view about the demarcation problem is that all kinds of evidence can bear on how a theory is thought about—both confirming evidence and falsifying evidence. The idea behind what is called Bayesian reasoning is that different probabilities, or credences, are assigned to all sorts of propositions, and then those credences are adjusted when the data comes in. If the data is likely or unlikely to be predicted by the theory, the credence in the theory increases or decreases accordingly.

Popper's concern that theories are too vague is dealt with by Bayesian reasoning because the credence in the theory doesn't increase just because it's consistent with the data. It only increases when it would have been less consistent with different data.

## Definiteness and Empiricism

The idea of falsifiability as the ultimate correct way to demarcate between science and nonscience is not actually very popular among contemporary philosophers of science. The real practice of science is far messier than that would have you believe.

Regardless, Popper's idea of falsifiability gets at two of the most important aspects of what makes science special: definiteness and empiricism.

- ▼ A good scientific theory says something specific and inflexible (definite) about how nature works. It shouldn't be possible to just take any conceivable set of facts and claim that they're compatible with your theory. There need to be ideas and facts about the world that the theory does not allow you to accommodate.
- ▼ Empiricism happens because the ultimate purpose of a theory is to account for what is observed. No theory in science should be judged to be correct solely based on the fact that it's beautiful or it seems reasonable—or other qualities that can be established without going out and looking at the world.

So, the falsifiability idea as a criterion for science versus nonscience attempts to operationalize these two characteristics by insisting that a theory that says something definite and can be experimentally tested counts as science. The idea is explicitly to exclude purported scientific theories that really don't say anything at all because they can be stretched to cover absolutely anything.

Whatever a person might think of the many-worlds interpretation of quantum mechanics, it does not say that anything goes. It says something definite: that the wave function is an exact description of reality and that it always obeys the Schrödinger equation. The extra worlds—thought of as branches of the wave function—pop out of that formalism.

The worlds are not put in as postulates; they are a prediction of the theory, not the theory itself. And that specific prediction—that there are other branches describing separate worlds—admittedly cannot be tested. It can't be tested now or anytime in the future because people cannot physically interact with those other worlds. But that's OK because every theory makes some predictions that can't be tested. You don't judge a theory on its untestable predictions; you judge it on its testable ones.

So, the real question is this: Do the predictions of many-worlds fit the data that is collected? The theory is thoroughly empirical because it aims to provide an explanation for the experimental data. All the experimental data that has been collected in favor of quantum mechanics is predicted by many-worlds.

Put another way, there is a crucial difference between the many-worlds interpretation as a version of quantum mechanics and the theories that Popper was worried about. Many-worlds makes definite predictions. It predicts both the usual phenomena associated with quantum mechanics and the existence and properties of other branches of the wave function. It's not that predictions aren't made; it's just that the laws of physics and the situation of the world prevent scientists from directly seeing some of their predicted implications.

## Many-Worlds Is Falsifiable

Many-worlds is, without any doubt, completely falsifiable—even in the most naive sense. Many-worlds doesn't say there are a lot of worlds; it says that the quantum state always obeys the Schrödinger equation. That is a very-easy-to-falsify statement. All you have to do is observe the quantum state not obeying the Schrödinger equation.

But there is some subtlety here. It's true that quantum states seem to violate the Schrödinger equation whenever you observe them. The wave function appears to collapse. But if the many-worlds interpretation is right, that collapse is only apparent because of decoherence and the fact that you are also quantum systems yourself becoming entangled with the systems that you're looking at. But in order to falsify many-worlds interpretation—it's still possible to do that—all you need is for a quantum system to be violating the Schrödinger equation all by itself, even when it's not decohering or undergoing a measurement process. And that is a phenomenon that you can go look for, and experimental physicists are looking for it.

There are a set of alternatives to many-worlds interpretation called objective collapse models in which wave functions really do collapse—it's not merely apparent—and there is an active experimental program looking for these collapses. If a good experimenter finds that quantum wave functions are collapsing without decoherence, measurement, etc., then many-worlds will be instantly falsified.

So, arguing that the many-worlds interpretation is not scientific because it's not falsifiable is straightforwardly mistaken. Admittedly, people don't always give up, even when you point out the mistake they're making. They will start

saying that many-worlds as a comprehensive theory may be falsifiable, but the particular implication that there are other worlds is not. And therefore, they will say, that implication should not be taken seriously.

This is an interesting question for the philosophy of science. If you have a theory that does very well at fitting the situations that you have looked into, how far should you trust it when it gets extended into regimes that you haven't looked into—or even regimes where you cannot ever look into? It's an interesting question because the answer is not simple or obvious. It depends on other considerations.

Of course, extending predictions of the theory into regimes that you haven't yet observed, by its nature, happens all the time. In fact, it's inevitable. The very notion of a prediction involves saying something about the future. And the future has not yet been observed. It would surely be over-limiting to only use your theories to explain things that you have already observed and not allow yourself to make predictions about the future.

But there is also a difference between different kinds of predictions you might make. Sometimes, you extend a theory into a regime where the basic situation is extremely similar to ones you've already experienced, even if they're not identical. That's what is the case for making predictions about tomorrow based on what happened yesterday, for example. It's also the case when you imagine that the basic rules of classical mechanics or quantum mechanics will work perfectly well on the surface of Pluto or on planets around a different star. Those experiments haven't yet been directly done, but the situations are so similar in nature that there's no reason to expect any substantial deviation. You should always be prepared for such a deviation, and you'll update your beliefs if it happens, but you can put a relatively small probability on it happening ahead of time.

So, is the prediction from the many-worlds interpretation—that there are other branches of the wave function—more like a familiar situation or an unfamiliar one? Arguments can be made for either side. You can say that these other worlds are very familiar because you've experienced your world and know how quantum mechanics works. You've seen superposition, entanglement, and decoherence in action, and the prediction of other worlds is just a matter of applying those concepts to the universe as a whole. You haven't added anything new conceptually to get those worlds to pop into

existence. But you can also say that the other worlds are unfamiliar because you've never seen a new world being created and then witnessed multiple worlds afterward.

So, in this situation, it pays to be explicit about what the alternatives are. It seems clear that the unaltered rules of quantum mechanics, if they are taken to apply to everything in the universe, predict the appearance of multiple worlds. And if you want to doubt the existence of those other worlds, you have to invent a disappearing worlds theory. This can be done, and it has been done. The way it happens is by suggesting new dynamical equations and then by taking the predictions of those equations seriously. And when this theory building is attempted, it's found that it is possible. But what's also found—as often happens in science—is that it's harder than you might expect to get everything to work out without coming into conflict with other experimental constraints.

Given all these considerations, you might reasonably take the stance that it just doesn't matter whether the other worlds actually exist. Why should you care about things that you can never observe, even in principle? It matters because, among other things, the ultimate laws of physics aren't known, so the theories that currently exist are the point from which scientists start trying to construct better ones. Every kind of theory naturally suggests a certain set of open questions, possible variations on the theory, and basic principles you might expect to keep as your understanding evolves. Many-worlds is one such theory. And it suggests new avenues to one of physics's most difficult problems—that of quantum gravity. It's unknown whether those suggested avenues are going to pan out, but the possibility is undeniably important. That's why theories should be taken seriously, even when they predict things that will never be observed: because taking them seriously helps point scientists toward even better theories.



# 14

## Where Does Probability Come From?

**T**his lecture confronts a true puzzle for the many-worlds interpretation: the origin and nature of probability. While there are several things that are not yet completely understood about many-worlds, probability is likely the single biggest challenge.

## Probability in Many-Worlds

The probability puzzle in many-worlds is sometimes referred to as deriving the Born rule. Remember, the Born rule, from Max Born, is a way of thinking about how you use the wave function in the Schrödinger equation. You square it, and the probability of seeing a measurement outcome equals the wave function squared. So, the probability question is sometimes phrased as, How do you derive the Born rule? But that gives you the wrong impression.

The real question is not why the wave function is squared rather than the wave function itself, or the wave function cubed, or whatever it might be. It's clear from the mathematical setup that the wave function squared is the unique set of numbers that have the properties you want out of something called probability. You get a bunch of nonnegative numbers, and they add up to 1. The real question, though, is, How do probabilities enter the many-worlds interpretation at all? Why is there any need to talk about probability in this point of view? After all, the Schrödinger equation is perfectly deterministic. There's nothing unpredictable about the evolution of the wave function of the universe according to Everettian quantum mechanics. Does it make sense, then, to speak of the probability of ending up on some particular branch where there will be a future version of yourself on every branch?

In the textbook, or Copenhagen, version of quantum mechanics, this is not an issue. There is no need to derive the Born rule for probabilities because it's asserted as one of the postulates of the theory. In other words, the Born rule is one of the foundational assumptions on which the Copenhagen version depends. So, why can't you do the same thing—assume that probability equals the wave function squared—in many-worlds?

The answer is that even though the rule would sound the same in both cases—it would be probabilities are given by the wave function squared—their meanings would be very different. The textbook version of the Born rule really is a statement about the frequency with which things happen—how often something occurs in the past and future. But many-worlds has no room for an extra postulate like that. You know exactly what will happen just from the basic rule that the wave function always obeys the Schrödinger equation.

So, probability in many-worlds, if it exists at all, is going to have to be a statement about how you should believe things are going to happen and how you should act, not about how often actually things do happen. And what you should believe isn't something that has a place in the postulates of a physical theory; it should be implied by them.

So, this question in the many-worlds context forces you to think about the nature of probability. And again, this is a fundamentally philosophical question, not a scientific one. But philosophy is needed to make sense of physical theories.

One of the fun but challenging aspects of many-worlds is that you need to think carefully about the philosophy you're following.

## The Philosophy of Probability

Consider tossing a fair coin, which should come up heads 50% of the time and tails 50% of the time. Of course, that's only true in the long run, and nobody is surprised when you toss a fair coin twice and it comes up tails both times. For just a few coin tosses, you wouldn't be surprised at almost any outcome. But as you do more and more, you expect—in conventional probability theory—the total number of heads to come closer to 50%. So, an obvious move to make philosophically is to define the probability of getting heads as the fraction of times you actually would get heads if the coin were to be tossed an infinite number of times—the mathematical limit as you toss the coin more and more times.

This notion of what is meant by probability is sometimes called frequentism, as it defines probability as the relative frequency of an occurrence in the limit of a very large number of trials. And that matches well with the intuitive notions of how probability functions in many familiar circumstances, such as when you toss coins, roll dice, or play cards.

To a frequentist, probability is an objective notion about reality. It only depends on features of the coin. It doesn't depend on you personally or your state of knowledge. The coin flip is a 50% chance, no matter how much you

know about the universe. But note that it's only going to be objective if you can talk about the limit of an infinite number of events. And that doesn't actually happen in the real world.

Frequentism fits comfortably with the textbook picture of quantum mechanics and the Born rule. Maybe you don't actually send an infinite number of electrons through a magnetic field to measure their spins, but you can easily imagine sending a very large number of them. In other words, you can gather enough statistics to convince yourself that the probability in quantum mechanics really is given by the wave function squared.

Many-worlds is a different story. If Everett is right, there is a 100% probability that each possibility will be realized in some particular world. Fortunately, there's an alternative way of thinking about probability that long predates quantum mechanics. This is the notion of epistemic probability, which is that the probability has to do with what you know, rather than some hypothetical infinite number of trials.

In your everyday life, you're constantly judging the likelihood of different events that are only going to happen once. In such cases, it makes no sense to talk about the limit of an infinite number of occurrences. For example, you might talk about the probability that a certain team is going to win the Super Bowl. But you can also talk about the probability of past events. Maybe you don't remember what time you left work last Thursday, but you bet it was probably between 5 and 6 pm since that's usually when you head home. That's a probability assignment to an event that's already happened. It doesn't make sense to think about that as the limit of an infinite number of trials. Nevertheless, you apply the idea of probability in these cases—not because you're going to imagine doing them an infinite number of times, but because you don't know exactly what it is.

In these cases, you're assigning credences to the various propositions under consideration. A credence is your personal degree of belief that some proposition is true. Like any probability, credences must range between 0 and 1—or between 0% and 100%—and your total set of credences for the possible outcomes of some specified event must add up to 100%. You're saying that something's going to happen with 100% probability. Your credence in a particular proposition or event happening can change as you gather new information.

Statisticians have formalized how you change your credences in the presence of new information under the label of Bayesian inference, named after the Reverend Thomas Bayes, an 18th-century Presbyterian minister and amateur mathematician. He derived an equation showing how you should update your credences when you obtain new information. Actually, Pierre-Simon Laplace derived the most general version of this equation, but it's called Bayes's rule.

So, there is a venerable notion of probability that applies even when something is only going to happen once, not an infinite number of times. But it's a subjective notion rather than an objective one: It depends on your particular state of mind. Different people in different states of knowledge might very well assign different credences to the same outcomes for some event.

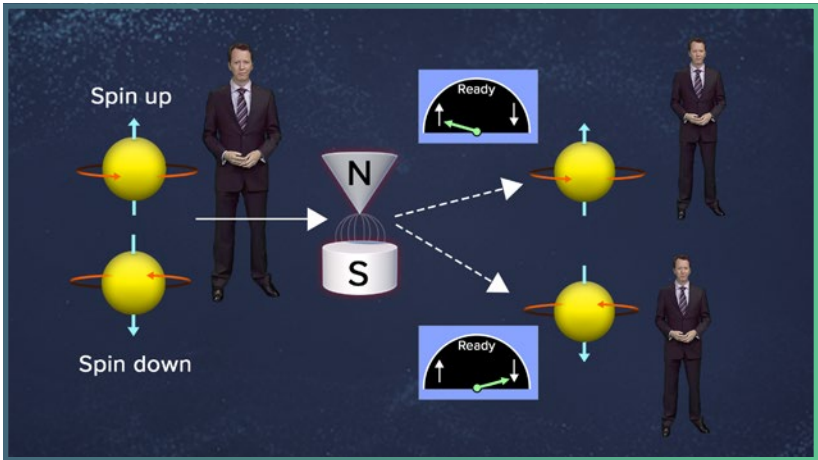
If someone is flipping a coin and you know that the coin is not fair—that it's weighted to usually come up heads—your credences are going to be different than someone who thinks it is a fair coin. The way that subjectivity has sneaked into people's way of talking about probability makes some uncomfortable. But it's OK as long as everyone agrees to follow the rules about how to update your credences when you learn something new. Those are the rules of Bayesian inference.

It's still not obvious that basing probability on your knowledge rather than on frequencies is really a step forward in the context of quantum mechanics that you care about right now, because in order for that to apply—in order to say that you have a subject of credence—there has to be a sense in which you don't know something. And then if you don't know exactly what's going to happen, you can assign a credence and update it via Bayes's rule. But how does that work with the Born rule in many-worlds?

## The Born Rule in Many-Worlds

Let's say that you have an electron and you pass it through a magnetic field, measuring the spin. That particle will be measured by an apparatus that is macroscopic, so that apparatus will become entangled with its environment. That is decoherence. And then the wave function branches. That is the standard many-worlds story.

In that story, there is a single world that splits into two. There are now two people where you used to be just one. And it makes no sense in this situation to ask which one of those two people—the person on the spin-up branch or the person on the spin-down branch—is really you. They're both descendants of the previous version of you. Likewise, before the branching happens, it makes no sense to wonder which branch you will end up on. Both of the people that have a result to that spin measurement have every right to think of themselves as you.



This is a situation you only face in many-worlds, not in the previous history of science or philosophy. And neither one of those two people—the spin-up person and the spin-down person—is wrong. Each of them is a separate person, both of whom trace their beginnings back to the same, previous person. In the many-worlds interpretation, the life span of a person should be thought of as a branching tree with multiple individuals at any one time, rather than as a single trajectory.

But this is saying, finally, how some lack of perfect knowledge sneaks into the many-worlds story. You're not assigning a credence to whether you will end up in one world or another. The resolution is to switch from asking, Which branch will you end up on? to, Which branch are you on immediately after the branching has happened?

Think about the moment right after decoherence has occurred. The world has branched; there are now two separate copies of the world. As a matter of empirical reality, decoherence happens extraordinarily rapidly—generally in a tiny fraction of a second. The branching happens first, and you branch along with the rest of the universe but only find out about it slightly later, for example, by looking to see whether the electron went up or down when it passed through the magnetic field. So, for a brief while, there are two people who have evolved from you who are precisely identical. Each of them lives on a separate branch of the wave function, but neither one knows which branch they're on.

There's nothing unknown from the God's-eye view about the wave function of the universe. You know—from solving the Schrödinger equation, knowing that you're going to send a spin through a magnetic field—that the universe contains two branches, and you know the amplitude associated with each of those two branches, if you set up your experiment appropriately. But the actual people on those branches don't know which branch they're on before they see the measurement outcome.

This state of affairs is called self-locating uncertainty, or indexical uncertainty. It's a situation where you know everything there is to know about the whole universe—the whole wave function of the universe—except where you are within the universe. There are multiple copies of you that look the same. Which one of them is you?

In that moment after branching, both copies of you are subject to self-locating uncertainties since they don't know which branch they're on. But they can assign a credence to being on one branch or the other. What credence should they assign? What should their credence be for being on the spin-up branch or the spin-down branch?

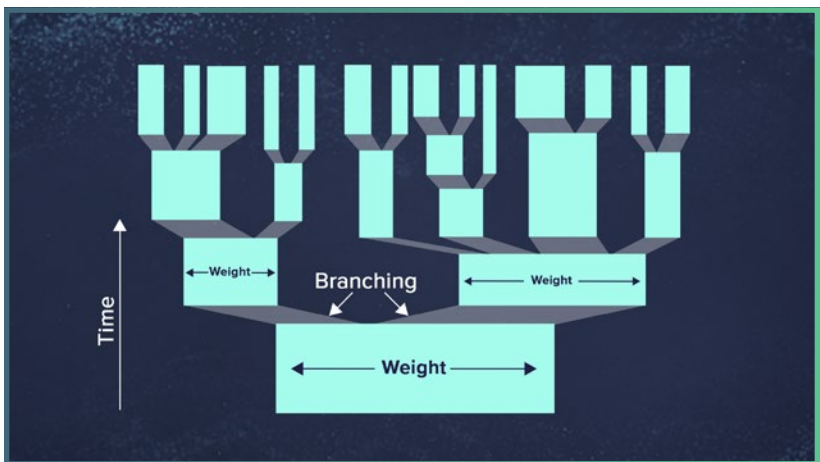
You know that you want to get the Born rule, which says that the probability is given by the amplitude of the wave function squared. But if you didn't already know about the Born rule, you might think that the amplitudes in the wave function were irrelevant to your self-locating uncertainty credences—to the probabilities of measuring a certain outcome.

When you go from one branch to two, for example, another alternative would be to simply assign equal probability to each one of those. After all, they are two separate universes—two separate branches of the wave function. This

idea, called branch counting, simply says that you give each branch that you make equal probability because each one is a separate universe by itself. This is intuitively appealing, but it can't work, even if you didn't know the Born rule.

Instead of being simplistic and assigning equal probability to each branch no matter what, you could assign equal probability to branches when they have equal amplitudes. Analysis of this alternative suggestion indicates that the amplitudes in a quantum wave function lead to different branches having different weights. And the weight of a branch is equal to the amplitude of that branch squared. When there are two branches with unequal amplitudes, there are only two worlds. The different amplitudes don't change the number of worlds. But crucially importantly for thinking like an Everettian, the two branches don't have equal weight. The one with a larger amplitude should be given more probability and contributes more to the overall wave function.

The weights of all the branches of any particular wave function always add up to 1, and when one branch splits into two, you don't simply make more universe by duplicating the existing one. The total weight of the two new worlds is equal to that of the single world you started with. The worlds get thinner as the branching proceeds. Therefore, there's a uniquely sensible way to assign credences in the many-worlds interpretation, and it's precisely the Born rule.





# 15

## Quashing Worlds with Wave Function Collapse

**T**he many-worlds formulation of quantum mechanics features a very simple underlying structure, but its consequences are mind boggling, so it's sensible that people are going to keep looking for alternatives with equally simple physics but easier-to-interpret philosophy. If you think about many-worlds as just “wave functions plus Schrödinger equation,” there are some plausible ways to change that overall situation. One is that you could alter the Schrödinger equation so that multiple worlds never develop. This lecture considers this possibility.

## The GRW Theory

Erwin Schrödinger had an original hope that his equation would describe waves—the wave function—but that they would naturally localize themselves over time into blobs that behaved like particles when viewed from far away because he knew that there were particle-like things in the universe. It's possible that some modification of his equation could actually achieve that ambition and maybe even provide a natural resolution to the measurement problem without introducing the idea that there need to be multiple worlds. As an ambition, that's easy to say, and it makes sense, but it's harder than it sounds in practice.

Consider an electron in a pure spin-up state. That can equally well be expressed as an equal superposition of spin left and spin right. So, if you observe that spin along a horizontal magnetic field, you'll have a 50% chance of observing either outcome. But precisely because of the exact equality between the two options, it's hard to imagine how a deterministic equation could predict that you see one or the other rather than both or some split between them. And the original Schrödinger equation is deterministic. You need something to break the balance between spin left and spin right.

You might imagine that you could introduce new variables—hidden variables—but that possibility will be examined in the next lecture. Within the realm of trying to change the Schrödinger equation, you can start thinking more dramatically than just adding another term to it. Rather than gently tinkering with the Schrödinger equation, you can introduce a separate way for wave functions to evolve—one way that squelches the appearance of the multiple branches that are causing philosophical puzzles.

There is plenty of experimental evidence that ensures that wave functions usually obey the Schrödinger equation—at least when you're not observing them and collapsing the wave function. But you can consider the possibility that rarely, but importantly, wave functions do something very different than that. So, what might that different thing be? The goal here is to avoid the existential horror of multiple copies of the macroscopic world being described

within a single wave function. The Copenhagen interpretation says that wave functions collapse, and that's the whole story. So, what if you imagined that wave functions undergo occasional spontaneous collapse? What if a wave function just randomly converts suddenly from being spread out over different possibilities—such as the positions in space of an electron—to being relatively well localized around just one point?

This is the key new feature of what are called dynamical, or objective, collapse models—the most famous of which is called the GRW theory after its inventors, Giancarlo Ghirardi, Alberto Rimini, and Tullio Weber.

Think about an electron in free space, not bound to an atomic nucleus. According to the conventional Schrödinger equation, the natural evolution of such a free particle is for its wave function to spread out and become increasingly diffuse, at least until it gets measured. To this picture, GRW adds a new postulate: that at every moment of time, there is some probability that the wave function will change radically and instantaneously by suddenly becoming localized around some point in space. In essence, the wave function in GRW really does collapse objectively, physically. It's not just an apparent collapse from the point of view of an observer, like many-worlds would say.

Who tells you what point the new wave function is going to localize around? There's a truly random element, and the probability distribution from which you choose where the particle localized is given by the original wave function squared—in other words, the Born rule. But you're not applying it to measurement. It happens unprompted, suddenly, and unpredictably. The new wave function is strongly concentrated around a central point so that the particle is now essentially in one location as far as macroscopic observers are concerned.

So, GRW is not some nebulous interpretation of quantum mechanics; it's a brand-new physical theory. The dynamics of the theory are different. In fact, the theory postulates two new constants of nature: There is a width to the newly localized wave function, and there's a rate, a probability per second, at which the dynamical objective collapse occurs. Realistic values for these parameters are  $10^{-5}$  centimeters for the width and  $10^{-16}$  for the probability of collapse per second.

Where do these numbers come from? If these were any other numbers, you would rule out the theory experimentally. In other words, there's no a priori principle that indicates these numbers, but those are the ones that fit the data. And with these numbers, a typical electron will evolve for about  $10^{16}$  seconds before its wave function spontaneously collapses—that's about 300 million years. So, in the lifetime of the observable universe, 14 billion years, most electrons, for example, only localize a handful of times. This is considered a feature of the GRW theory, not a bug.

This is the whole point of the GRW theory. If you go messing around with the Schrödinger equation, you'd better do it in such a way as to not ruin all the successes of conventional quantum mechanics in fitting the data. Quantum experiments are done all the time with single particles or collections of a small number of particles; it would be disastrous—or at least very different than what is actually observed—if the wave functions of those particles kept spontaneously collapsing. So, if there is a truly random element in the evolution of quantum systems, it had better be incredibly rare for individual particles.

How does such a mild, rare alteration of the theory manage to get rid of macroscopic superpositions, such as with Schrödinger's cat? The answer is entanglement, which comes to the rescue here much as it does with decoherence in many-worlds.

Consider the familiar example of measuring the spin of an electron. As you pass an electron through a magnet, the wave function of the electron evolves into a superposition of deflected upward and deflected downward, according to the Schrödinger equation. You're going to measure which way it goes, for example, by detecting the deflected electron on a screen. Since it's just one electron, GRW doesn't change its dynamics. But the screen is hooked up to a dial that has a pointer, which indicates up or down. The spin-up part of the electron's wave function becomes entangled with the pointer indicating up, and the spin-down part becomes entangled with the pointer indicating down. That's the same, so far, as in the Everettian story. But an Everettian will say that the pointer is a big macroscopic object that quickly becomes entangled with the environment, leading to decoherence and branching of the wave function.

A GRW advocate doesn't want to have the wave function branching. But something related happens. It's not that the original electron wave function spontaneously collapses; you'd have to wait for millions of years for that to become likely. But once that electron becomes entangled with the apparatus, the pointer in the apparatus might contain something like  $10^{24}$  electrons, protons, and neutrons. All these particles are entangled in an obvious way: They're in different positions depending on whether the pointer indicates up or down.

Even though it's very unlikely that the wave function for any one particle will undergo a spontaneous collapse before you look to see what the answer is, the chances are very good that at least one of the many particles will collapse. That should happen roughly  $10^8$ , or 100 million, times per second.

You might not be impressed because you're thinking that that's just a tiny subset of the particles that are becoming localized in the macroscopic pointer. But here's where the magic of entanglement comes in. Because of entanglement, if the wave function of just one particle is spontaneously localized—and remember, the pointer is in a superposition of spin up and spin down—the rest of the particles that one is entangled with will come along with it. So, within the pointer, all the particles are entangled with each other in a superposition of all pointing toward up and all pointing toward down. When a macroscopic object like a pointer is in that kind of superposition, all the particles in that object are entangled with all the other particles. In one part of the wave function, all the particles are located in the configuration of “pointer indicating up,” while in the other part, they're all located in the configuration “pointer indicating down.”

Let's say that one electron spontaneously localizes so it's in the position that it has if the pointer is indicating up. Then, the entire part of the wave function that represents “pointer indicating down” disappears—not just for that one particle, but for all the other particles it was entangled with. This is just like when Alice and Bob measured their spins and the rest of the wave function disappeared. So, the overall wave function goes very rapidly from describing an apparatus pointing in a superposition of two answers to one that is definitively one or the other. In other words, the GRW theory manages to

operationalize and make perfectly objective that split between the classical world and the quantum world that the partisans of the Copenhagen approach are forced to invoke.

Classical behavior is seen in objects that contain so many particles that it becomes likely that the overall wave function will undergo a series of rapid collapses and localizes.

The GRW theory has advantages and disadvantages. The primary advantage is that it is a well-posed, specific theory. It addresses the measurement problem in a very straightforward way. The multiple worlds that exist in the Everett approach are eliminated by a series of unpredictable wave function collapses. What's left is a single world, to human eyes, that maintains the successes of quantum theory in the microscopic world because there, particles rarely collapse while exhibiting classical behavior macroscopically—which is exactly what people experience in their everyday lives. GRW provides a perfectly realist account; it does not invoke any fuzzy notions about consciousness in its explanation of experimental outcomes.

GRW can be thought of as Everettian quantum mechanics plus a random process that cuts off new branches of the wave function as they appear. It is the quintessential disappearing-worlds theory. Moreover, GRW is experimentally testable, and experiments are ongoing, with the ultimate goal of either confirming GRW or ruling it out.

Among GRW's disadvantages are the fact that this new rule for spontaneous collapse is completely ad hoc. It's also out of step with everything else that is known about physics. It seems suspicious that nature would not only choose to violate its usual law of motion at random intervals but do so in just such a way that you wouldn't have yet been able to experimentally detect it. But then again, to be fair, every version of quantum mechanics violates classical intuition in some way, so maybe you just have to learn to live with stuff like this.

Another disadvantage has prevented GRW and related theories from gaining much traction among most theoretical physicists. It's unclear how to construct a version of the GRW theory that works not only for particles but also for fields. GRW says that particles spontaneously localize, but in modern physics, the fundamental building blocks of nature are fields, not particles.

This isn't necessarily an insurmountable obstacle. Maybe the difficulty of reconciling objective collapse models with quantum field theory is just a limitation of the imaginations of modern theoretical physicists rather than an insuperable obstacle. Therefore, it's worthwhile looking at slight variations of this basic idea, searching for one that perhaps seems less arbitrary. And there are a few ways that the original idea of sudden random collapses can be modified.

## The CSL Theories

Remember, Schrödinger—when he first thought about the Schrödinger equation—had the hope that wave functions for free particles would naturally evolve to look like localized particles. That didn't happen. Actual solutions to the Schrödinger equation tend to spread out all over the place. That's what led Max Born to suggest the probability interpretation of the wave function in the early days of quantum theory.

But recovering a version of Schrödinger's original hope can be attempted by slightly changing his equation. There's a modern idea along these lines called continuous spontaneous localization (CSL) theory. The idea is similar in spirit to GRW in that wave functions localize all by themselves—they violate the original Schrödinger equation. In the CSL theories, the localization happens gradually and all the time rather than suddenly at random intervals.

There needs to still be some randomness in the equations in order to recover the unpredictable nature of observed, ordinary quantum evolution, but it's almost as if the wave function knows what a classical trajectory for a particle would look like and tries to stick as close to that as possible. These CSL theories have similar features to the original GRW proposal. They seem ad hoc, and they don't conserve energy. But the good news is they're experimentally testable, and those tests are happening.

## Penrose's Theory

Remember, GRW delineates the quantum-classical boundary by making spontaneous collapses very rare for individual particles and then rapid for large collections of particles. An alternative approach would be to make collapses occur whenever the system reaches a certain threshold—like a rubber band breaking when it is stretched too far. When you get a superposition of macroscopic things that look too different from each other, maybe the wave function just collapses.

A well-known example of an attempt along these lines has been put forward by mathematical physicist Roger Penrose, who is famous for his work in general relativity but also, like Einstein, cares about quantum mechanics. Penrose's theory uses gravity to define when these wave functions are going to spontaneously collapse.

Specifically, he says that a wave function will collapse when it describes a superposition of macroscopic objects that have appreciably different gravitational fields. The trick here is that the criterion of the gravitational fields being appreciably different turns out to be difficult to specify precisely. A single electron will not collapse in Penrose's theory no matter how spread out its wave function is, while a macroscopic object like a pointer is large enough to cause collapse as soon as it starts evolving into different states.

Most experts in quantum mechanics have not warmed to Penrose's idea about these spontaneous collapses—in part because they are skeptical that gravity should have anything to do with the fundamental formulation of quantum mechanics. But it's possible that a precise version of Penrose's criterion, thought of as decoherence in disguise, could be developed. The gravitational field of an object can be conceived of as part of its environment. And if two different components of the wave function have different gravitational fields, maybe they become effectively decohered. That might provide a sensible way of thinking about when these objective collapses are supposed to occur.



# 16

## Blocking Worlds with Hidden Wave Variables

If you're going to try to make sense of quantum mechanics one way or another, the cozy classical picture of reality will be modified in some substantial way. In many-worlds, there are other worlds that you can't directly see or interact with. In objective collapse models, there are random violations of the Schrödinger equation. This lecture turns to a different alternative to the many-worlds interpretation: the existence of hidden variables.

## Hidden-Variable Theories

Ever since the beginning of quantum mechanics, it was always lurking in the back of minds of physicists that the wave function didn't tell the whole story. What if, for example, there was a wave function, but there were also other physical variables in addition to it? After all, physicists were used to thinking in terms of probability distributions from their experience with statistical mechanics, in which you don't specify the exact position and velocity of every atom in a box of gas. But it turns out that that's OK because all you need to know are their overall statistical properties, such as the average number of particles at any given location with some specified velocity. That's enough to let you make predictions.

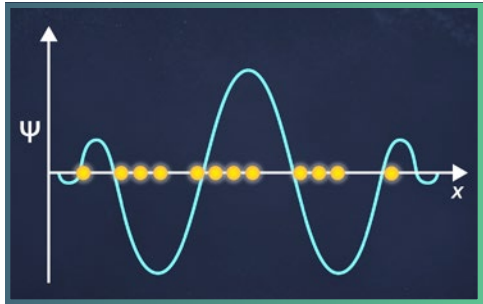
But in the classical view, it is taken for granted that there is something called the exact position and velocity of every particle—even if you don't know it, it's there. So maybe quantum mechanics is like that: Even though there are random chances in the predictions, there are definite quantities associated with observational outcomes you will eventually get, but you don't know what they are. In this view, the wave function somehow captures part of the statistical reality, but it's not telling the whole story. This possibility is appealing to people who want a realist, deterministic version of quantum mechanics but are bothered by the philosophical problems that seem to come along with the many-worlds approach.

But there's a roadblock. The wave function can't be exactly like a classical probability distribution for a straightforward mathematical reason. A true probability distribution assigns probabilities directly to outcomes. The probability of any given event has to be a real number between 0 and 1. But the wave function is a complex number. It assigns a complex number called the amplitude to every possible outcome. And amplitudes have both a real and an imaginary part, and either one of them can be either positive or negative. You get a probability distribution by squaring the absolute values of such amplitudes. But if you want to explain what is experimentally observed, you can't directly work with that probability distribution.

You have to keep track of the wave function in order to match experiments. In particular, the fact that amplitudes can be negative is what allows for interference. Probability distributions are always nonnegative numbers; they don't show interference.

So, you want to have physical quantities that lead deterministically to measurement outcomes but also allow the explanation of interference effects. To address this problem, think of the wave function not just as a state of knowledge, but as a real, physically existing thing. But then also imagine that there are additional variables that could represent the positions of real particles in the theory. These extra quantities are conventionally called hidden variables.

In hidden-variable theories, there's still a wave function, but the role of the wave function is different than the role it plays in Everettian quantum mechanics or even objective collapse models. In those models, the wave function directly represents the real physical world, whereas in hidden-variable models, the wave function takes on a new role—what's called a pilot wave. The wave function pilot wave guides the real particles as they move around. It's as if the particles are small floating barrels and the wave function describes waves and currents in the water that push the barrels around.



Interestingly, to make it all work, the wave function still needs to obey the Schrödinger equation. But a new equation, called the guidance equation, that governs how the wave function pushes around the particles is needed. The particles are guided toward where the wave function is large and away from where the wave function is nearly 0.

In hidden-variable theories, the answer to the fundamental question of why electrons behave sometimes like particles and sometimes like waves is because there are both particles and waves.

## Bohmian Mechanics

The first hidden-variable theory was presented by Louis de Broglie in 1927 at the Solvay Conference, but his ideas were harshly criticized, and he basically abandoned them afterward. Then, in a famous book from 1932 called *The Mathematical Foundations of Quantum Mechanics*, the Hungarian mathematician and physicist John von Neumann proved a theorem that said that you could not develop a hidden-variable theory that works. But it turned out that he had not proven what many physicists at the time thought he had proven.

Then, in 1951, David Bohm published an influential textbook on quantum mechanics, in which he basically adhered to the received Copenhagen approach. But in thinking through the issues of the theory as he wrote the textbook, he developed a new theory that was very similar to that of de Broglie but with more-clearly-fleshed-out details. In Bohm's theory, particles are guided by a quantum potential that is constructed out of the wave function. Today, this approach of a quantum wave function pushing around particles is called the de Broglie-Bohm theory, or simply Bohmian mechanics.

Bohmian mechanics posits both a wave function and a separate set of particles. The particles have definite positions—just like classical particles do—but these positions are unknown to you until you measure them. What you can prepare in a laboratory experiment is the wave function, not the particle positions. So, this wave function will evolve exactly according to the Schrödinger equation.

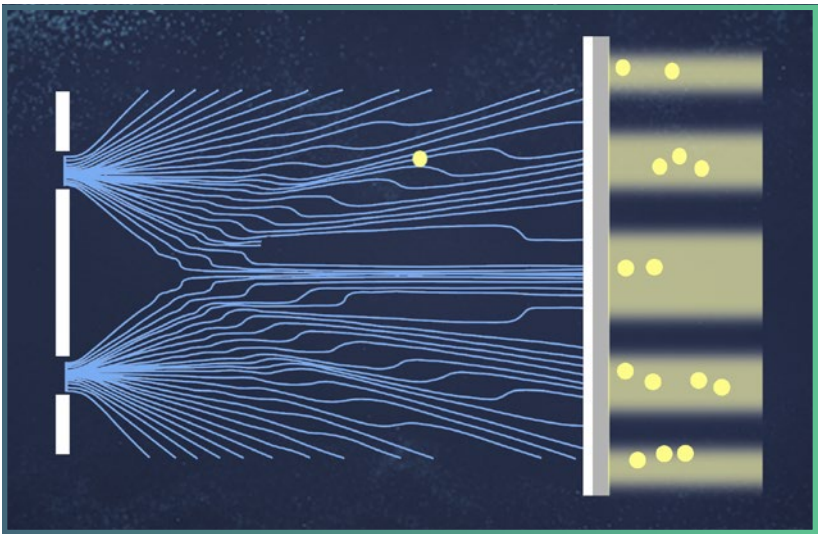
The wave function doesn't seem to recognize that the particles are there; it's unaffected by what the particles do. But the particles are pushed around by the wave equation under the guidance equation. So, roughly, the role of the guidance equation is to push the particles toward places where the wave function is big and away from where it is small. That's how you're going to match things like the interference bands that are seen in the double-slit experiment.

However, there is a very important property here. The way in which any one particle is guided in Bohmian mechanics depends not only on the wave function but also on the positions of all the other particles in the system. It is a feature, in other words, of the Bohmian guidance equation that particles far

away can directly influence the motion of a particle nearby. That is explicit nonlocality. The motion of a particle right in front of you can depend, in principle, on the positions of other particles arbitrarily far away. Nonlocality plays a crucial role in understanding how Bohmian mechanics can reproduce the predictions of ordinary quantum mechanics.

Think about the double-slit experiment, which illustrates vividly how quantum phenomena are simultaneously wavelike (you see interference patterns) and particle-like (when the particles hit the screen, you see a dot). Furthermore, the interference pattern goes away when you look toward which slit the particles go through. When you put a detector at the slits, there's no more interference.

In Bohmian mechanics, this ambiguity is not mysterious. There are both particles and waves. The particles are what you observe; the wave function affects their motion, but you have no way of seeing the wave function directly. The dots on the screen are caused by the particles.



So, according to Bohm, the wave function evolves through both slits, just as it would according to the Schrödinger equation or in Everettian quantum mechanics. In particular, there will be interference effects where the wave function adds or cancels once it reaches the screen. But it's not the wave function itself that leaves the individual dots on the detector screen. What you see are the individual particles hitting it. These particles are being pushed around by the wave function, so they're more likely to hit the screen where the wave function is large—that's where you see bands—and less likely to do so where the wave function is small.

## Bell's Theorem

When Bohm first published his theory, most physicists ignored it; they were already entrenched in the Copenhagen consensus. But one physicist was impressed with Bohm's suggestion: John Bell, a theorist at CERN, the great particle physics laboratory. Bell was inspired to understand how to reconcile the fact that Bohm had apparently written down a perfectly good theory with the apparent fact that there was a theorem from von Neumann that said it wasn't possible. And Bell ended up proving a new result now known as Bell's theorem.

According to that theorem, it is impossible to reproduce the predictions of conventional quantum mechanics unless you allow for some kind of nonlocal interaction. Bohmian mechanics is perfectly deterministic, but it's also resolutely nonlocal. Separated particles can affect each other instantaneously.

This might be why Einstein, who was very sympathetic to the idea of hidden variables, never constructed an implementation of the idea that he was satisfied enough with to publish. He didn't like the idea of nonlocality. And Bell proved that the nonlocality is not optional. If you want to have hidden variables and recover ordinary quantum mechanics, you need nonlocality.

So, what does Bell's theorem say about many-worlds (because there's no apparent nonlocality)? One of the reasonable assumptions behind Bell's theorem is that measurements have definite outcomes. In many-worlds, that's just not true in the way that it needs to be true for Bell's theorem. There are

different outcomes that are observed in different worlds. So, many-worlds is consistent with Bell's theorem simply because the assumptions that Bell makes don't apply to many-worlds.

And like many-worlds, Bohmian mechanics needs to account for the Born rule—one of the things that has to be recovered if you're proposing a new theory of the foundations of quantum mechanics. The Born rule says that the probability of observing a particle at a given location is given by the wave function squared.

On the surface, this seems hard to reconcile with the idea that the particle positions are completely independent variables. You could put the particles wherever you want and Bohmian mechanics are perfectly deterministic. So, why do you get the Born rule? Why is the probability of seeing a particle in different positions so dependent on the wave function? The answer is that while in principle particle positions can be anywhere you want, in practice—in Bohmian mechanics—there is a natural distribution for them to have.

Imagine you have a wave function and some fixed number of particles. To recover the Born rule, all you have to do is start with a distribution of those particles that is consistent with the Born rule. In other words, you have to distribute the positions of the particles such that the distribution looks like it was chosen randomly from a probability distribution given by the wave function squared. There will be more particles where the amplitude squared is large and fewer particles where it is small. Such a distribution in Bohmian mechanics is called an equilibrium distribution. And then you can prove that if you start your particles in an equilibrium distribution, the Born rule will remain valid as time passes. In other words, if you start your particles in a probability distribution that matches what you expect from ordinary quantum mechanics, you will continue to recover that expectation going forward.

It is believed by many Bohmians that if you start in a nonequilibrium initial distribution—a distribution of particles other than what you would expect from the Born rule—you will evolve toward equilibrium just as a gas of classical particles in a box evolves toward an equilibrium thermal state. But the status of this idea is not yet settled. Bohmian mechanics is still a minority pursuit among physicists.

The key difference here from other formulations of quantum mechanics is that the particles that you're observing have a definite position in reality. It's just that you haven't measured those positions yet.

## The Disappearance of Interference

A basic picture emerges in the hidden-variable paradigm. The wave function is real. It guides particles to certain positions—more particles where it's large and fewer where it's close to 0. Particles passing through a double slit will be more likely to hit the detector where the wave function is highest and less likely to hit the detector where the wave function interferes and cancels out.

To recover all of ordinary quantum mechanics, you still have to confront the question of what happens when you look to see which slit the particle goes through in the double-slit experiment, because in Bohmian mechanics, you can't say that the wave function collapses when you make that observation. Wave functions don't collapse in Bohmian mechanics; as with Everettian quantum mechanics, they simply always obey the Schrödinger equation. So, how are you supposed to explain the disappearance of the interference pattern in the double-slit experiment?

You explain it in the same way you do in many-worlds. While the wave function doesn't collapse, it does evolve. And, in particular, you should consider the wave function for the detector, the apparatus, and the electrons going through the slits. The Bohmian world, just like the Everettian world, is completely quantum. There is no stooping to an artificial split between classical and quantum realms.

**Bohmian mechanics is an explicit construction that does what many physicists thought was impossible: It constructs a precise deterministic theory that reproduces all of the predictions of textbook quantum mechanics without requiring any mysterious incantations about the measurement process or distinction between quantum and classical realms.**

The wave function for the detector will become entangled with that of an electron that passes through the slit, so a kind of branching will occur. You have both different parts of the wave function entangled with different things in their environment.

But there's a difference between Bohmian mechanics and many-worlds. The difference is that the variables describing the apparatus—the hidden variables that aren't there in many-worlds—will be at locations corresponding to one of these branches, not to all of them. In other words, in the mind of a Bohmian, there aren't two equally real branches like there are in the many-worlds interpretation.

Rather, there is a real branch—that's the branch where the particles are. And measurement just makes it apparent which branch that is. For all intents and purposes, it is just like the wave function has collapsed, or just like decoherence has branched the wave function. But instead of assigning equal reality to all the branches, the particles you're made of are only located on one particular branch.

So, rather than the objective collapse models from the previous lecture, where you have branches and then you erase them, here you have all the branches and a pointer that says that one branch is the real one. That's what happens in hidden-variable theories.

Many Everettians are dubious about this kind of explanation, and some have claimed that Bohmian mechanics isn't really any different from Everettian quantum mechanics. It just includes some superfluous variables that serve no purpose but to theoretically assuage some anxieties that you might have about splitting into multiple copies of yourself.

Apart from the presence of explicit nonlocality and the accusation that the theory is basically just the many-worlds interpretation, there are other problems inherent in Bohmian mechanics. And the way the dynamics of Bohmian mechanics works is far from elegant.



# 17

## Mind before Matter in Quantum Theory

**T**his lecture does two different but related things. One is to address the question of consciousness. The other is to describe yet another alternative to many-worlds: the so-called epistemic approaches to quantum mechanics, in which the wave function doesn't represent reality but our knowledge of the world.

## The Question of Consciousness

There's a long history of people suggesting either that human consciousness is necessary to understand quantum mechanics or that quantum mechanics is necessary to understand consciousness. But according to everything that is currently known, there is no good evidence that there is actually a connection between the two.

Defining what is meant by consciousness is notoriously difficult. Let's start more simply by contemplating the more general idea of thinking. In the modern world, it is believed that thinking depends on various physical processes in your brain. In particular, a human brain contains about 86 billion neuron cells, and those neurons are linked together by various connections. There are electrochemical signals that jump between the neuronal connections when you think. The strength of different neuronal connections has to do with memory and knowledge.

It is conceivable that the rates of different neurons firing—or the ways in which different neurons string together—depend on quantum entanglement in an interesting way. But there's no hard-and-fast evidence.



In general, decoherence—becoming entangled with the environment and branching the wave function—is expected to be ubiquitous and rapid in a warm, wet environment like the human brain. It would be surprising if the quantum entanglement between two neurons or two chemicals in your neurons lasted long enough to have much of an impact on your thinking. It's possible—the world is quantum mechanical at heart, and that includes brains—but it doesn't seem especially likely that the quantum mechanics of your brain matters over and above chemistry and classical mechanics.

But accounting for consciousness, as it is traditionally thought about, is not a straightforward matter of the rates of neural processes. So, you can think about whether quantum entanglement changes the rates of neurons firing, but there are deeper, more philosophical problems.

The philosopher David Chalmers proposed to distinguish between the “easy problems” of consciousness—How do you sense things? How do you react to them? How do you think and talk about them?—and the “hard problem” of consciousness, which has to do with your subjective, first-person experience of the world. The hard problem is about what it is like to be you rather than someone else.

Typically, even philosophers imagine that eventually neuroscience will figure out the easy problems—which are not actually easy. Neuroscience is quite difficult, but significant progress has been made in mapping out the connections between neural processes and thoughts/awareness. Progress will likely continue, and more and more will likely be understood about how you relate what happens in the brain to what you say that you're thinking.

Meanwhile, when people worry about the truly puzzling aspects of consciousness, they're concerned with the hard problem: How does a collection of physical particles in the brain attain the ability to feel what it is like to be something or experience something? Whatever the answer might be to the hard problem, quantum mechanics doesn't seem to have anything to do with it.

Roger Penrose teamed up with anesthesiologist Stuart Hameroff to develop a theory in which the objective collapse of the wave functions—in particular, of certain microtubules in the brain—could help explain why people experience consciousness. However, not many neuroscientists think that objective

collapse of the wave function plays an important role inside the brain, so this doesn't seem to be a leading contender for helping explain the nature of consciousness.

More importantly, it's unclear why this kind of quantum messiness should be important for the hard problem of consciousness. It's conceivable that there are some subtle quantum processes in the brain involving microtubules (or something different), but it's very unlikely. It's possible that those processes affect the rate at which neurons fire, but even if they do, this is of no help in bridging the gap between the firing of your neurons and the idea of your subjective, self-aware experience. The reason the hard problem is supposed to be hard is because it's unknown how to bridge that gap, and the Penrose-Hameroff theory doesn't help do that.

Many philosophers and scientists have no trouble believing that the gap between physical things going on in the brain and subjective experience is actually very bridgeable. But a tiny change in the rate of some particular neurochemical process doesn't seem to be relevant to understanding how. And if it were, there's no reason that whatever quantum mechanical effect that's being evoked in a brain couldn't also be repeated in a computer, so it wouldn't have anything particularly unique to say about human consciousness.

Everettian quantum mechanics—the many-worlds approach—has nothing specific to say about the hard problem of consciousness that wouldn't be shared by any other view in which the world is entirely physical. In a view where the world is just physical, the relevant facts about consciousness include just two: Consciousness arises from brains, and brains are coherent (made of mutually interacting parts) physical systems. Two collections of neurons on two noninteracting branches of the wave function are two distinct brains, not one collective brain.

This is not making extra assumptions about consciousness or personal identity to discuss quantum mechanics in the many-worlds picture. It's just a quintessentially mechanistic theory. There's no special role in many-worlds for observers or experiences; they're just yet other macroscopic systems that will decohere and branch the wave function. Conscious observers branch along with the rest of the wave function when decoherence happens, just like rocks

and other nonconscious things. The challenge of understanding consciousness is precisely as difficult in many-worlds than it would have been without quantum mechanics at all.

Admittedly, there are many important aspects of consciousness that scientists don't currently understand. But that is unsurprising. In fact, it is precisely what you should expect. The human mind generally and consciousness in particular are extremely complex phenomena—arguably the most complex in the known universe. The fact that they are not fully understood shouldn't tempt people into proposing entirely new laws of fundamental physics to help explain them. The laws of physics, as they are currently known, are enormously better understood and experimentally verified than the functioning of brains and the relationship between brains and minds.

Someday, scientists might have to contemplate modifying the laws of physics in order to successfully account for consciousness, but in terms of scientific progress, that should be the last resort. You don't change the part of your scientific understanding that's working well to explain the part that you don't have any right to think you should understand yet.

The question can also be flipped on its head: If quantum mechanics doesn't help account for consciousness, is it nevertheless possible that consciousness plays a central role in accounting for what happens according to quantum mechanics?

In the earliest days of quantum theory, when the Copenhagen interpretation came on the scene, the idea of measurement was directly implicated in the rules of the theory. And if measurement plays an explicit role in the axioms of quantum mechanics—unlike in classical mechanics—it's perfectly natural to wonder whether there isn't something special about the interaction between a conscious, observing mind and a quantum system. At the most down-to-earth level, could the collapse of the wave function be caused by the conscious perception of certain aspects of physical objects?

According to the textbook view, wave functions collapse when they are measured. But precisely what constitutes a measurement is left vague. The Copenhagen interpretation posits that there is a distinction between quantum

and classical realms, and it treats measurement as an interaction between a classical observer and a quantum system. But where the line is drawn is hard to specify.

If you have a Geiger counter observing emission from a radioactive source, for example, it'd be natural to treat that Geiger counter as part of the classical world. It's a macroscopic device. But we don't have to do that. Even in Copenhagen, you could imagine treating Geiger counters as quantum systems that obey the Schrödinger equation. It's only when the outcome of a measurement is actually being perceived by a human being that—in this way of thinking—you absolutely have to have the wave function collapse. The reason is simply an empirical one: No human being has ever reported being in a superposition of seeing and experiencing different measurement outcomes.

A final possible place that you could draw the cut between the classical and quantum realms is between observers who can testify as to whether they are in a superposition and everything else. And since the perception of not being in a superposition is part of consciousness, it's not completely crazy to ask whether it's that consciousness that is causing the collapse of the wave function.

This idea was put forward as early as 1939 by Fritz London and Edmond Bauer. It was later popularized, and today it is often associated with physicist Eugene Wigner, who actually later changed his mind about the role of consciousness in quantum theory. It's not generally a view that's spoken of approvingly at physics conferences, but there are some scientists who continue to take it seriously.

Eugene Wigner was a well-known physicist who won the Nobel Prize for his work on symmetry.

## The Epistemic Approach

There are alternative theories to many-worlds where consciousness is not needed. A modern descendant of this idea is known as the epistemic approach to quantum mechanics. It doesn't care about consciousness, but it does change what is meant by quantum mechanics.

An epistemic approach, from the Greek word *episteme* for “knowledge,” naturally concerns itself with what human agents know. So, an epistemic approach to quantum mechanics—in contrast with the many-worlds interpretation, objective collapse models, and hidden-variable theories—denies the reality of the wave function entirely. This doesn’t mean to deny the fact that wave functions are important in doing quantum mechanics. Rather, this is saying that you can use wave functions (they can solve the Schrödinger equation, etc.), but you don’t claim that they represent part of reality. Instead, in the epistemic approach, wave functions simply characterize our knowledge—in particular, the incomplete knowledge about the outcome of future measurements.

Such an approach is called epistemic because it thinks of wave functions as capturing something about what is known, as opposed to ontological approaches. Ontology is the study of what exists, so an ontological approach to quantum mechanics treats the wave function as describing objective reality, and an epistemic one is about our knowledge.

A simple-minded epistemic strategy cannot work in the most naive and straightforward way. You can’t just take the wave function, square it, and say that’s the probability distribution for knowing different things. The reason is because the wave function interferes with itself. That is seen in the double-slit experiment. It is therefore crucially important that you keep track of the distinction between the positive and negative parts of the wave function—the real and imaginary parts—rather than just squaring it and treating that as our knowledge.

Rather than giving up, however, you can try to be more sophisticated in how you think about the relationship between the wave function and the real world. Imagine a formalism that allows you to use wave functions to calculate the probabilities of seeing certain measurement outcomes while not attaching any underlying reality to those outcomes before you measure them. That is the task taken up by epistemic approaches. And there have been many attempts to interpret the wave function epistemically, just as there are competing collapse models or hidden-variable theories, etc.

The most prominent is called quantum Bayesianism, sometimes shortened to QBism. Bayesian inference suggests that you carry around a set of credences for various propositions to be true or false. And then you update those

credences following Bayes's rule when new information comes in. All versions of quantum mechanics use Bayes's theorem in some version or another, and in many approaches to understanding quantum probability, Bayesian reasoning plays a crucial role.

But QBism is distinguished from other approaches by making quantum credences personal rather than universal. In other words, according to QBism, the wave function of an electron is not a once-and-for-all thing that everyone should in principle agree on. Rather, everyone has their own personal idea of what the electron's wave function is, and they use that idea to make predictions about observational outcomes.

It's not arbitrary. In QBism, there are rules about how rational people will assign wave functions to different quantum systems. And they're basically the same rules that a non-QBist would use, given what they know about the system. If many experiments are done and the experimentalists share their knowledge about what they observe, then the QBists claim that they will come to a degree of consensus about what the various wave functions are. But at the end of the day, they're still fundamentally measures of personal beliefs. They are not objective features of the world.

Thinking about quantum mechanics in QBist terms has led to various interesting developments in the mathematics of probability, and it offers insight into quantum information theory. But most physicists are still going to want to know what reality is supposed to be in this view. If reality is not represented by a wave function, what is it? And the answer is not clear.

Modern QBists are following in the footsteps of the original pioneers of the Copenhagen interpretation—the people following Bohr and Heisenberg who didn't want to talk about what the world really is but wanted to focus on observational outcomes. Or, at least as an ongoing research program, QBists have chosen not to dwell much on questions concerning the nature of reality—questions about which other theorists care very much.

The fundamental ingredients of the QBist theory are agents—people with thoughts and beliefs—and those agents accumulate experiences. Quantum mechanics, in this view, is a way for agents to organize their beliefs and update them in the light of new experiences. The idea of an agent is central

to the theory. This is in stark contrast to other formulations of quantum mechanics, according to which observers are just physical systems like anything else.

Sometimes, QBists will talk about reality as something that comes into existence over time as observations are made. So, the idea in QBism is not that there is no such thing as reality but that reality cannot possibly be captured by a God's-eye view—by any seemingly objective, third-person perspective. Reality is constructed collectively from the experiences of many different observers.

QBism is a relatively young approach to quantum foundations. There's much development yet to be done. It is possible that eventually QBism will run into insurmountable roadblocks and that interest in it will fizzle out. But it's also possible that the insights gained from QBism will ultimately be reinterpreted as a useful way of talking about the experiences of observers within some other straightforwardly realist version of quantum mechanics. Or it might be the case that QBism, or something close to it, represents a true revolutionary way of thinking about the fundamental nature of reality—one that puts agents like you at the center of your best description of the world.



# 18

## The Quantum Emergence of the World We See

**Y**ou know by now that, according to many-worlds, what's called reality is completely and exactly described by a quantum state, which evolves according to the Schrödinger equation. And one way of thinking about that quantum state is as a wave function, which is mathematically described as a vector in an abstract, super-high-dimensional vector space called Hilbert space. The problem is that a vector moving in Hilbert space seems far removed from the world you see around you, which is full of stuff distributed through three-dimensional space. This raises a whole new class of questions: Why is there space? Why is this particular stuff distributed through space? And how does it all emerge out of the vector in Hilbert space? These difficult questions are uniquely suggested by the many-worlds approach; in the other approaches, the answers are assumed from the start.

## The Nature and Emergence of Space

When you choose hidden variables, or when you specify the way in which wave functions collapse in objective collapse models, you generally use the idea of locations in space. It is a primitive concept that's built into part of the assumptions of the theory. In many-worlds, it is not. In many-worlds, the very idea of space and locations within space must be derived because all you start with is a vector in Hilbert space.

Why is the concept of space so central to physics in the first place? The answer resides in the concept of locality. In ordinary, everyday physics, locality just means that things interact with each other when they're nearby in space. But today, there's more than one notion of locality. The word *locality* is used in at least two somewhat different senses: measurement locality and dynamical locality. And measurement locality is false, but dynamical locality is still true, as far as is known.

The EPR thought experiment seemingly showed that there is something nonlocal about quantum measurement. And later, Bell's theorem would support this. Bell's theorem implies that any theory in which measurements have definite

measurement outcomes—which basically means every approach to quantum mechanics other than many-worlds—will feature measurement nonlocality.

Dynamical locality, on the other hand, refers to the smooth evolution of the quantum state when no measurement is being performed—in other words, what the Schrödinger equation says happens to the quantum state. That's a context in which, unlike measurements, physicists expect everything to be perfectly local. Disturbances in one location are only going to immediately affect things nearby. If something is far away from a disturbance (ignoring quantum measurement), that disturbance will have to move at the speed of light or more slowly. This kind of locality is enforced by a rule in special relativity that nothing can travel faster than light.

**Measurement nonlocality seems to be part of most versions of quantum mechanics.**

Bell says that quantum mechanics measurements will be nonlocal. But think about the ways in which ordinary non-measurement physics is local. With that in mind, you can dig into the question of how the structure of your observed reality emerges from the abstract quantum wave function. You live in a world that looks like a collection of objects located in space. Furthermore, these objects behave approximately classically. There are occasional quantum jumps—Geiger counters can click, and spins can go up and down—but mostly the world around you is pretty well approximated by classical physics. In Everettian quantum mechanics, there's a story that gets told about many such worlds, but the postulates of the theory—wave functions plus smooth evolution—don't even mention worlds.

Where do the worlds come from? Why are there worlds at all? And why do the worlds look approximately classical? The first thing to emphasize is that the worlds aren't fundamental; they are emergent. This means that the idea of worlds—copies of reality that emerge when you do a dynamical measurement—is not built into the theory at the deepest level. Rather, these worlds are an approximate, but very useful, way of describing an incredibly complex situation: what the quantum state is all about.

If you were a superintelligence with exact knowledge of the quantum state of the universe, you would never have to talk about dividing the wave function of the universe into branches or about a collection of worlds because you would know the whole wave function exactly. You could just follow its evolution to whatever accuracy you liked. But you are a finite, bounded human being with only limited knowledge and limited computational resources. In the real world, it's enormously convenient and helpful to divide the wave function into branches representing worlds—you only see one world at a time, after all. And you're perfectly allowed to take advantage of this convenience because the individual worlds don't interact with each other.

So, it's not just that you don't have complete information and therefore don't talk a language of complete information. It's that you're allowed to talk in a particular language that only uses a certain subset of the information. That doesn't mean that the worlds you talk about aren't real. A distinction is drawn between fundamental and emergent. A separate distinction is drawn between real and not real.

You choose to invoke Everettian worlds when carving up the wave function because it's convenient for you, but it's crucially important that you don't do that carving of the wave function randomly. There are right and wrong ways to divide the wave function into branches—and the right ways leave you with an independent set of worlds obeying approximately classical laws of physics.

Emergence isn't automatic or ubiquitous. It happens in a very specific set of circumstances. Emergence happens when there exists a special way of describing a system that involves much less information than a complete, comprehensive description would, but it nevertheless gives you a useful handle on what's going on.

With Everettian worlds, you don't need to keep track of the entire wave function of the universe to make useful predictions. All you have to keep track of is what happens in a single, individual world where you live. To a good approximation, you can treat what happens in each world using classical mechanics, with just the occasional quantum intervention when microscopic systems that are in a superposition are entangled.

That's why Newton's laws of gravitation and motion are sufficient to fly a rocket to the moon. You don't need to know the complete wave function of the universe. Your individual branch of the wave function describes an emergent, almost classical world. That's a convenient feature of the world you live in, but you're still allowed to ask why the wave function of the universe takes the form of a superposition of many nearly classical worlds. Why is this particular way of chunking up reality so useful?

**Emergence is a way of capturing important features of a system using far less information than you might think you need.**

## The Preferred Basis Problem

This issue is sometimes called the preferred basis problem. Why is the basis that you use in Hilbert space taking the form of classical configurations? You can think about it just by returning to the example of a spin being measured. You know that you can express the wave function of a spinning particle in terms of the possible outcomes if were you to measure it along, for example, the vertical axis. Generally, the state will be a superposition of spin up and spin down.

But alternatively, you know that you can express the state as a superposition of spin right and spin left. Those are the possible outcomes if you were to measure it along the horizontal axis. It doesn't matter which option you choose; they are two individually good choices. But you do have to pick one because the two descriptions are different ways of talking about the same thing. A pure spin-up state can be written as a combination of spin left and spin right, or vice versa.

When you carry out an actual measurement along the vertical axis, you can ask why the results you see are only ever spin up or spin down. If an electron that is purely spin up can be written as a combination of spin left and spin right, why don't you occasionally see spin left and spin right as the outcome, even if the magnetic field that is doing the measuring is oriented vertically? How does the electron know what you're measuring?

That's the preferred basis problem: because the two choices of basis are either spin up and spin down or spin left and spin right. They are two equally good bases, but by doing measurements in a certain way, you're only measuring one and not the other. What's going on?

The answer to this question was essentially figured out in the 1980s using decoherence. To get there, it's useful to turn to the Schrödinger's cat thought experiment, which starts with a sealed box containing a cat and a vial of sleeping gas. The experimenter rigged a hammer to fall, break the vial, releasing the gas, and putting the cat to sleep—but only when a detector, such as a Geiger counter, clicks upon detecting a particle of radiation. And you put next to the detector a radioactive source for which you know the rate that particles are emitted from it. Therefore, you can calculate the probability that the counter is going to click and release the hammer after any given period.

This radioactive emission is a fundamentally quantum process. Informally, it's talked about as an occasional, random emission of a particle. But really what's going on is the smooth evolution of a wave function of the atomic nuclei within the source. Each nucleus evolves from a state of purely undecayed (100% not decayed) to a superposition of undecayed plus decayed. And over time, the decayed part gradually grows. The emission appears random because, as always, the detector doesn't measure the full wave function—it only sees undecayed or decayed.

The point of the Schrödinger's cat thought experiment is to take a microscopic quantum superposition and magnify it to a manifestly macroscopic situation. When that happens is as soon as the detector clicks. All the business with the sleeping gas and the cat is just to make the amplification of the quantum superposition to the macroscopic world seem more vivid and counterintuitive.

You know how to account for this in many-worlds. In the many-worlds theory, the wave function smoothly evolves into an equal superposition of two possibilities: The cat is asleep, and the cat is awake. But decoherence says that the cat is also entangled with its environment, which consists of all the air molecules and photons and anything else in the box.

So, as far as many-worlds is concerned, the branching into separate worlds happens essentially right away, after the detector clicks. By the time the experimenter gets around to opening the box, there are already two branches of the wave function, each of which has a single cat and a single experimenter—not a superposition. So, that helps explain Schrödinger's original worry. It's not about the observer opening the box; it's about decoherence inside the box.

But there's another worry that gets raised within many-worlds: Why is it that when you open the box, the particular decohered quantum states you see are either an awake cat (100%) or an asleep cat (100%)? There could be other states that are superpositions of awake and asleep. But you don't see that when you open the box. Awake and asleep together represent just two particular states for the cat system—just like spin up and spin down do for the electron.

That's why this is called the preferred basis problem. Why are those basis states—the one in which the cat's awake and the one in which the cat's asleep—preferred over any others, in the sense that when you open the box,

those are the ones you see? The physical process that matters to you here is the stuff in the environment, such as the gas molecules and the photons, interacting with the physical system under consideration. Whether a particular particle actually does interact with the cat depends on where the cat is. This is locality coming back into the game: the position, in space, of the cat.

So, a given photon might very well be absorbed by the cat if the cat is awake—the cat's up, it's prowling around the box, a certain photon hits it and gets absorbed. But that same photon would miss the cat if the cat were asleep on the floor. So, it interacts differently with the cat, depending on those two different possibilities in the superposition.

What is special about the basis states 100% awake and 100% asleep is that those individual states describe well-defined configurations in space. And space is the quantity with respect to which physical interactions are local. A particle can bump into a cat if the particle and the cat come into physical contact. Therefore, what's special about the two parts of the cat wave function—awake and asleep—is that they come into contact with different particles in their environments.

Therefore, the state of the cat and the state of the environment will become entangled if the cat starts out in a superposition. But that's precisely decoherence. That's when you get branching into different worlds, whereas if the cat is in a pure basis state—if it's just awake, for example—then no extra entanglement and decoherence happens. Every photon interacts with an awake cat in the same way. This is a crucial distinction: There's an interaction—the cat can absorb or reflect (or whatever) the photon—but there's no entanglement because the environment doesn't interact differently with different parts of the cat's wave function if it is purely awake or purely asleep.

And this is the basic answer to the question of why you see the particular worlds that you do. The preferred basis states—the ones that are robust to being monitored by the environment—are those that describe coherent objects in space because such objects interact consistently with their environments. The technical term here is *pointer states* because these are the states in which the pointer of a macroscopic measuring device indicates a definite value. The pointer on the measuring apparatus is not in a superposition in a pointer state.

In the case of measuring the spin of an electron going through a magnet, the initial superposition of spin states is amplified to a macroscopic superposition in which some pointer on an apparatus is in a superposition of pointing up and pointing down. The environment will interact with those two states differently—that leads to decoherence and branching. So, it's the pointer basis in which a well-behaved classical approximation makes sense. And therefore, it is that pointer basis that is the kind of basis defining emergent worlds.

So, unlike other models, many-worlds provides an explanation to why only certain kinds of states form worlds rather than other ones.

**Decoherence and locality play a crucial role in accounting for why a world that is ultimately governed by the rules of quantum mechanics appears the way it does. These are the phenomena that ultimately link the austere simplicity of Everettian quantum mechanics to the messy particularity of the world you see.**



# 19

## The Challenge of Quantum Gravity

**D**espite decades of effort, physicists have not yet constructed a completely successful theory of quantum gravity. This lecture explains why constructing a successful theory of quantum gravity is so difficult.

## Special versus General Relativity

The current best understanding of classical gravity is the theory of general relativity, which was presented by Albert Einstein in 1915. Like the phrase *quantum mechanics*, the word *relativity* by itself does not refer to a specific physical theory. Relativity is a framework within which theories can be constructed. Theories that qualify as relativistic are ones that share a common picture of the nature of space and time. In that picture, the physical world is described by events happening in a single, unified space-time.

Even before relativity, it was perfectly possible to talk about space-time. Even in Newtonian physics, there was three-dimensional space and one-dimensional time, and to locate an event in the universe, you had to specify both. The difference between Newton and Einstein is, before Einstein, there wasn't that much motivation to combine space and time to make a single, four-dimensional space-time. But once relativity came along, it became perfectly natural to combine them.

There are two big ideas that fall under the theory of relativity: special relativity and general relativity. Special relativity, which came together in 1905, is based on the idea that everyone measures light to travel at the same speed in empty space. It doesn't matter if you're taking measurements from a rocket ship speeding through space or from your couch. No matter how you're moving, you will always measure light to travel at the same speed. In other words, the measured speed of light will be the same to everyone no matter how they're moving relative to each other.

**Special relativity is a framework. Different kinds of theories are constructed within the framework of special relativity, and they are called relativistic theories. But general relativity is a specific theory. It's general because it includes gravity, whereas special relativity does not.**

Think about what happens when you combine this idea that light moves at a fixed, universal speed with the principle of relativity, which is an insistence that there is no absolute frame of reference for motion. That's an idea that goes back to Galileo; Galilean relativity was part of Newtonian mechanics. In this picture, every observer has their own frame of reference in which they observe the movement of other objects through space and time. So, combining these two ideas creates a puzzling question: How would two observers who are moving at different speeds possibly observe that light travels at the same constant speed? Because of relativity, you know it's not that there is a correct frame of reference in which you should do that measurement—everyone gets the same answer. How does everyone get the same speed even if they're moving with respect to each other?

Einstein's solution was to say that, indeed, the idea that the speed of light is the same to everyone is correct. What you have to do is alter your idea of how space and time work individually. After all, what is a speed? A speed is an amount of distance—an amount of space that you travel—in a given amount of time. So, Einstein's idea was that if you modify both space and time, you can leave the speed of light invariant. Space, time, and relativity are not individually absolute and honestly real as they were for Newton. It's only the combination called space-time that is universal and agreed upon by everyone.

So, everyone's going to live in space-time, and they will naturally divide it into space and time. But how you do that dividing into distances in space and intervals in time will, in general, be different for different observers. In fact, it will be relative to their motion through the universe.

A key point of special relativity is that different observers will measure different amounts of elapsed time, depending on how they move through the universe. This is a radical change from what Newton would have said. For Newton, time was universal and absolute. In Newton's world, two observers who start at the same point, zip through the universe, and come back will have measured the same amount of time. In Einstein's world, they will not have measured the same amount of time. Time for Einstein is not universal; it's personal. It's relative to how you move through the universe.

Quantum mechanics and special relativity are fully compatible with each other. The problem comes with general relativity—the other big idea in the history of relativity besides special relativity. The difference between them

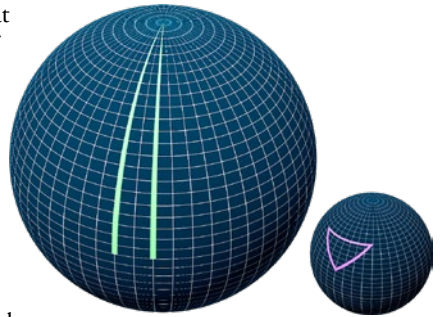
is that you have gravity in general relativity. In general relativity, gravity is thought of as a feature of space-time itself—in particular, the curvature of space-time. And this is where the problem comes in reconciling with quantum mechanics.

The crucial insight of general relativity was that four-dimensional space-time isn't simply a static background on which the interesting dynamics of physics take place. Space-time can bend and warp, and it will do so in response to the presence of matter and energy.

It's a crucial feature of Euclidean geometry—the flat geometry originally described by Euclid—that initially parallel lines remain parallel forever. From that, other features can be derived, such as the angles inside a triangle always add up to  $180^\circ$ . But space-time, Einstein eventually realized, can be thought of as having a non-Euclidean geometry, in which these important, venerable facts about geometry are no longer true. Initially parallel rays of light, for example, can be focused together even while just moving through empty space because that space is not Euclidean. And the effects of this warping of geometry of space and time are what is recognized as gravity in modern physics.

General relativity came with numerous mind-stretching consequences as soon as Einstein wrote it down. Implicit in the equation are things that Einstein hadn't been thinking of at the moment, such as the expansion of the universe and the existence of black holes. It took physicists—Einstein and others—a long time to appreciate the implications of general relativity. But by now, it's well tested and accepted.

But just like Newton's laws govern the evolution of a classical system, Einstein derived an equation that governs the curvature of space-time in response to matter and energy. The basic idea was pithily summarized by John Wheeler: "Matter tells space-time how to curve, and space-time tells matter how to move." That's general relativity in a nutshell. It's this give-and-take between matter and space-time that is called gravity—from apples falling from trees to black holes to the big bang to gravitational waves.



## Quantizing General Relativity

General relativity is a classical theory. It talks about the geometry of space-time, but that geometry is unique. There's not a wave function for it. It evolves deterministically, and, in principle, you can measure the geometry of space-time to arbitrary precision without disturbing it.

Once quantum mechanics arrived not long after general relativity, it was perfectly natural to try to quantize general relativity—to start with the classical theory and use it to invent a theory of quantum gravity. But it turns out that this is easier said than done. What makes relativity unique, whether it's special or general relativity, is that it is a theory of space-time itself. It is not a theory of stuff within space-time. So, other quantum theories that work well describe wave functions that assign, for example, the probability of observing something at a definite, well-defined location in space and moment in time. But quantum gravity, by contrast—whatever it is—has to be a quantum theory of space-time itself, just like general relativity is a classical theory of space-time itself. And this feature raises some issues.

Think about what it might mean to eventually have a quantum theory of space-time. You'd have to start by imagining that space-time could take on a number of different configurations or geometries. Then, you'd imagine that space-time could exist in a superposition of various different geometries. And then, following what is known about quantum mechanics, presumably you would need a wave function that would assign an amplitude to different possible geometries of space-time—just like the wave function assigns an amplitude to different possible locations or spins of an electron.

But that's a funny thing to think about. If you imagine, for example, two branches of such a wave function describing two different space-time geometries, it is hard to compare them. There is no unique way of specifying that two events in two different branches correspond to the same point in space-time because the two branches don't share a space-time in common.

So, apparently, space-time can't play the same central role in quantum gravity that it does in the rest of physics. There isn't a single thing called space-time; there is a superposition of many different space-time geometries. You're not

allowed to ask anymore what the probability might be to find an electron at a certain point in space in quantum gravity because there's no objective way of specifying which point you're talking about.

So, quantum gravity comes with a set of conceptual questions that distinguish it from other quantum mechanical theories. These questions can have important ramifications for the nature of the universe, including questions like what happened at the beginning or whether there was a beginning to time at all. You can even ask whether space and time are themselves fundamental or emerge out of something deeper.

Just like the foundations of quantum mechanics, the field of quantum gravity was relatively ignored for decades as physicists concentrated on other questions, and obstacles got in the way of making serious progress on quantizing gravity. A major roadblock is the difficulty of getting direct experimental data.

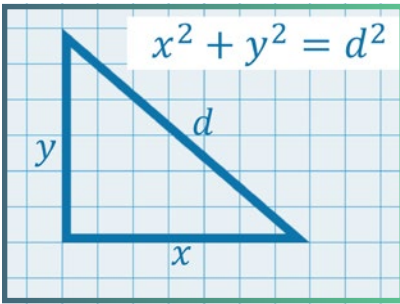
Gravity is ultimately a very weak force. Think about two electrons, which are both negatively charged. Like charges repel, so electromagnetically, they push each other apart. But gravitationally, they both have mass, so they pull each other together. Both of these are inverse square laws, but that electric repulsion is about  $10^{43}$  times stronger than their gravitational attraction. This means that in any realistic experiment involving a few particles where you might expect quantum effects to become relevant, the force of gravity is utterly negligible.

You could imagine building a particle accelerator powerful enough to smash particles together at the Planck energy, where quantum gravity becomes important. But unfortunately, the Planck energy, as discovered by Max Planck, is about  $10^{15}$ —one quintillion times higher than what can be achieved currently at the biggest particle accelerator, the Large Hadron Collider in Switzerland. If the technology that exists in current particle accelerators were scaled up, the resulting accelerator would have to be light-years in diameter to reach the Planck energy. It's not a feasible construction project for now.

## Technical Problems with Quantum Gravity

There are also technical problems with the theory itself, in addition to the conceptual ones. General relativity is, at heart, a classical field theory. The field involved is called the metric tensor field.

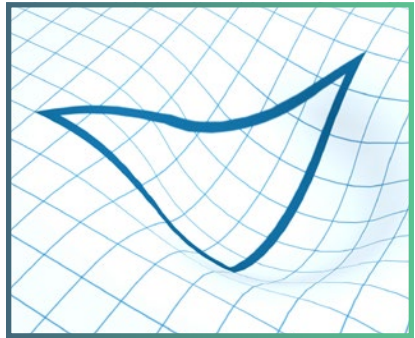
Think about flat Euclidean tabletop geometry. You might put  $x$  and  $y$  coordinates on the tabletop to help you specify points. Then, if you have some



line segment between two points, it will describe some increment  $\Delta x$  in the  $x$  direction and some increment  $\Delta y$  in the  $y$  direction. And then, you can calculate the length ( $d$ ) of that line segment by using the Pythagorean theorem, which states that, for a right triangle, the length of the hypotenuse squared is the sum of the squares of the shorter sides:

$$d^2 = x^2 + y^2.$$

In a general space-time, the metric replaces the Pythagorean theorem, which works in flat Euclidean geometry but has to be generalized and modified when the space-time is curved. In fact, the metric lets you calculate the length of any possible trajectory—even ones that are wildly curvy. And knowing the length of every possible trajectory determines the overall geometry of space-time itself.



Space-time has a metric even in special relativity and Newtonian physics. But that just means that there's a well-defined way to measure distances in space and intervals in time. But in both special relativity and Newtonian physics, the metric that's used is rigid and unchanging, and it's flat geometrically. The curvature of the metric is 0 at every point. The big insight of general relativity was to make the metric field into something that is dynamical. It is affected by matter and energy. And once you do that, you can attempt to quantize that field just as you would any other—that is, you can take the classical theory of curved space-time invented by Einstein and convert it into a theory of wave functions with an appropriate version of the Schrödinger equation. It's just that now the wave function is a function of the metric—the curvature of space-time—rather than the position of a particle.

When you apply this quantization procedure to something like electromagnetism, you find that small ripples in the electromagnetic field become discrete, particle-like lumps of energy. That's the origin of photons—quantizing the classical electromagnetic field. The same thing happens when you think about small ripples in a quantized gravitational field: They look like particles. And they're called gravitons. Nobody's ever detected an individual graviton, and it's possible that nobody ever will because the gravitational field is so incredibly weak. But if you accept the basic principles of general relativity in quantum mechanics, the existence of gravitons is inevitable.

You can then ask questions like, What happens when gravitons scatter off each other? Or what happens when they scatter off other particles? Sadly, what you find after you quantize classical general relativity is that the theory predicts nonsense, if it predicts anything at all. You need an infinite number of input parameters to calculate any particular quantity of interest. In other words, this theory—quantized general relativity in the most straightforward fashion—has zero predictive power.

But that doesn't mean that all is lost. You can restrict your attention to an effective field theory of gravity, in which only long wavelength, low-energy vibrations in the fields are considered. And that actually makes perfect sense. Even in quantum mechanics, as long as you stick to this limited regime, you have a theory that makes sense.

Happily, on Earth and in most of the Milky Way Galaxy, gravity is weak, so the effective field theory approach works—as long as you're not close to the vicinity of a black hole. So, there's no trouble in calculating the gravitational field in the solar system, for example, even in terms of quantum gravity.

But that's not a theory of everything—or even a theory of gravity—that is valid in all possible circumstances at all possible energies. If that's what you want, something dramatic and different is called for.

The most popular contemporary approach to quantum gravity is string theory, in which particles are replaced by small segments of one-dimensional strings. Today, string theory is the most pursued approach to exploring quantum gravity. But there are other ideas that maintain their adherents. The second most popular approach is loop quantum gravity, which began as a way of directly quantizing classical general relativity but doing so using vectors in space-time moving around closed loops.

String theory aspires to be a theory of all the forces and all the different kinds of matter all at once, while loop quantum gravity only aims to understand gravity. But unfortunately, the obstacles to getting experimental data relevant to quantum gravity are equally formidable for any of the known alternatives, so it's unknown which approach, if any, is on the right track.



# 20

## Space Emerges from Entanglement

**W**here can you find space within a wave function? And how can you even do that? You want to identify features of the quantum wave function that resemble space as it's known—and, in particular, since you know about curved space-time in general relativity, something that would correspond to the metric field. In classical gravity and relativity, the metric field allows you to tell the distance between two points in space-time. This lecture considers how distances show up in physics as it's currently understood. And for simplicity, this lecture addresses only distance in space. Combining space and time to make space-time is a separate question, and time will be addressed in the following lecture.

## The Story of Empty Space

The best current theories of nature are quantum field theories. As a starting point, the idea of quantum field theory is simple. There's a classical theory of fields stretching throughout all of space. A field is simply something that takes on a value at every point in space. The fields aren't localized. The field might be zero in some places and nonzero elsewhere, but it exists everywhere. And then that classical field theory is promoted to a quantum field theory. So, there is a wave function of the fields—just like in ordinary quantum mechanics when you have a wave function of the position of an electron or another particle. Remarkably, when you take a classical field and quantize it, the tiny vibrations in those fields appear to you as particles. So, in the modern view, all the things that you usually think of as particles—such as electrons and photons—can and should, at the deepest level, be thought of as vibrations in quantum fields.

But forget about the actual particles that you are made of for a moment. Think about empty space. In classical physics, there's not a lot to say about empty space. But in quantum field theory, empty space corresponds to the vacuum. Out of all the quantum states describing the theory, whichever has the lowest energy is the vacuum state. Intuitively, the vacuum state has no particles in it because if you add a particle to the vacuum state, you have more energy there now. You have an amount of energy equal to the mass of the particle times the speed of light squared:  $e = mc^2$ . So, to you, the vacuum state of the quantum field theory looks like empty space. But it's not really completely empty. In quantum field theory, the fields are always there, even if they're in their minimum-energy vacuum configuration.

Classically, a field's minimum energy state might be that the field is zero everywhere. But the uncertainty principle says that you can't pin down quantum systems to that definite configuration. So, in a quantum field theory, the minimum energy vacuum configuration will have some interesting structure to it. You can therefore interrogate various properties of the quantum vacuum state and get interesting answers.

Think about a particular thought experiment you can do. Imagine dividing the universe into two regions. You draw an imaginary plane somewhere in space and have two regions: left and right. The fields live everywhere, including in empty space. So, the fields continue uninterrupted across the imaginary plane because it's the vacuum state everywhere.

Classically, to construct any particular field configuration, you would have to specify both what the field is doing on the left and what the field is doing on the right. And there is energy when the field changes its value, depending on how fast it's changing its value. So, if there is a mismatch in the value of the fields across the boundary, that's a sharp discontinuity—a rapid change in the value of the field. But it costs energy for the field to change from point to point. In particular, a discontinuous jump would imply a large amount of energy at that location—at your imaginary plane. This is why, usually in field theory, you consider field configurations that vary smoothly from place to place, rather than suddenly and discontinuously.

There is a quantum version of this way of thinking. At the quantum level, the classical statement that says that field values tend to match across a boundary turns into the statement that the fields in the left region and the right region are highly entangled with each other. Just as in the classical case, the reason comes down to energy. Keeping the fields completely unentangled, or separate from each other, would cost a lot of energy. So, the lowest energy state to the vacuum state of the quantum field theory has a lot of entanglement crossing that boundary.

In fact, if you go to two different regions of space, you can ask about the entanglement between all sorts of different things. Imagine dividing all of space into equal-sized boxes. Classically, you would say that in each box, the field has a value. And to avoid infinite energy densities, the values must match at the boundaries between the boxes. In quantum field theory, therefore, what's happening in one box must be highly entangled with what's happening in the neighboring boxes.

And if a box of quantum fields inside is entangled with its neighbors, and those neighbors are entangled with their neighbors, it stands to reason that the fields in one particular box should reasonably be entangled not only with their nearest neighbors but also with the fields one box away. And if you study quantum field theory, you can prove that that's actually what happens. The way

it works is there is less entanglement with the fields one box away than for the direct neighbors, but there's still some. And this pattern does indeed continue throughout space: There are fields throughout the universe. You divide the universe into boxes, and the fields in any one box are entangled with fields in every other box of the universe—although the amount of entanglement becomes less and less as you consider boxes that are farther and farther apart.

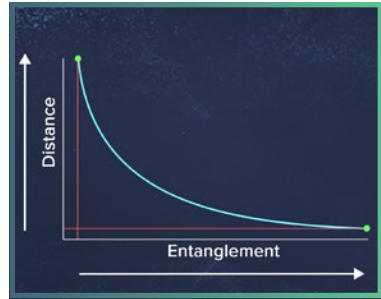
This is the standard story of empty space in quantum field theory. So, you're not yet starting with an abstract quantum state and deriving space and distance. You're taking hints from what you already know is the case in ordinary quantum field theory. But you want to do the latter: start with an abstract quantum state and derive distances, not postulate them from the beginning.

Happily, you can invoke a strategy of reverse engineering. Given some abstract quantum state, you can imagine many different ways of dividing that quantum state—or the Hilbert space in which that quantum state lives—into subsystems. You know that if you start with two spins (two subsystems), for example, you can multiply them to get a bigger Hilbert space. Now you want to go the other way around. Is there a particular way to chop up Hilbert space into subsystems so that those subsystems represent regions of space, in the sense that their entanglement structure matches that of empty space of the vacuum in quantum field theory?

There is—if you believe that locality is an important feature of the world. And indeed it is important. So, given the dynamics of the wave function as described by the Schrödinger equation, there will typically be a unique way of dividing the wave function into subsystems so that the subsystems are entangled locally. In other words, there will be exactly one right way to divide the quantum state so that things look local and interact with each other locally.

Therefore, you can place that as a demand. And it turns out that this requirement offers a unique way to decompose the quantum system. Locality tells you the right way to divide Hilbert space. In fact, if you take a quantum state and divide Hilbert space into subsystems, then any two subsystems will generally have some amount of entanglement between them. If they were vibrating quantum fields in the vacuum state, then you know exactly how that entanglement behaves: It's high if they're nearby and low if they're far away. Reverse engineering this, if you find subsystems that are highly entangled in

the vacuum state of the theory, then you can define them to be nearby. You define a distance so that things that are highly entangled have a short distance between them and things that are farther away are less entangled. In other words, a metric—a way of measuring the distance between things—on space has emerged from the entanglement structure of the quantum state.



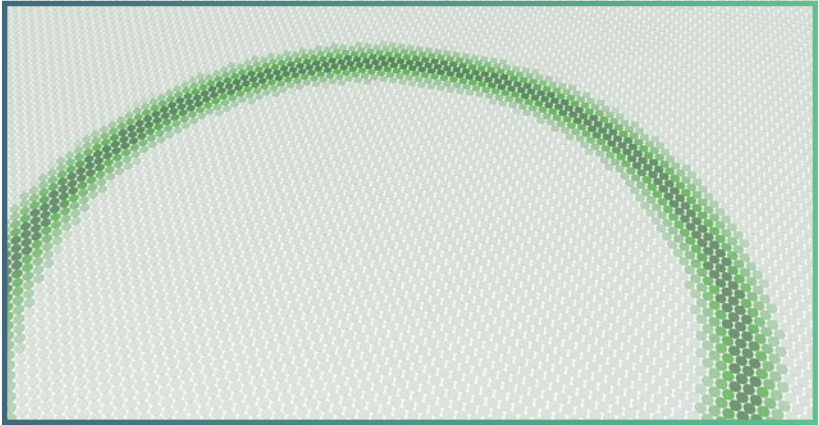
## When Space Isn't Empty

The story so far applies to empty space. You can recover a unique spatial structure and metric by looking at the lowest energy state of the theory and asking how to subdivide it in the right way. What about when you don't have empty space? What if you add some energy to the vacuum state? Ideally, you'd like for the geometry of space to adapt to the energy—that's what Einstein says happens in classical general relativity. So, you want to find a way to get an emergent metric on space that obeys Einstein's equation.

The final answer here is unknown. Research is in progress. In particular, science can't say that the emergent space from the quantum wave function really does curve and bend exactly like Einstein's equation says it should. But science can say that if certain simple assumptions are made, then the emergent space-time will curve in just that way.

To get there, let's return to thinking of the amount of entanglement between two regions in ordinary quantum field theory. In flat space-time, the amount of entanglement between two regions—not just two regions separated by a plane, but let's say a blob inside and outside—is proportional to the area of the boundary that separates them. And the reason for this is because the fields, with their quantum vibrations in one part of space, will be entangled with regions all over. But they're much more entangled with nearby regions, so the overall amount of entanglement is concentrated near the regions that are very close to each other. As a result, the total amount of entanglement between the inside and the outside depends mostly on the amount of entanglement at

the boundary. And that, in turn, is proportional to the area of the boundary. The greater the area, the more nearby regions there are—one side versus the other—and, therefore, the more total entanglement.



Maybe this feature of empty space—namely, that the entanglement between two regions is proportional to the area of the boundary that separates them—continues to be true even when you're not in empty space. Perhaps even when you add some energy so that you're no longer in the vacuum state, it is still true that the amount of entanglement between the inside and outside is proportional to the area of the boundary. This is simply a conjecture; you know it wouldn't be true in a theory without gravity. Einstein said that gravity is just the curvature of space-time. If there were no such thing as gravity, space would be rigid and fixed—it would be like special relativity. When you have that, it's easy to mess with your quantum state. If you're doing quantum electrodynamics, for example, you can easily increase or decrease the amount of entanglement between any two regions you want. The entanglement traces, and is proportional to, the boundary in the vacuum state, but in non-vacuum states, it can be anything.

The new conjecture here is that with gravity, when you try to increase or decrease the amount of entanglement between the inside of a region and the outside, the geometry of space itself adjusts to make sure that the area dividing the two regions is always proportional to the entanglement between

them. In other words, when space-time isn't empty, it acquires just the right amount of curvature to ensure that the entanglement between any two regions stays proportional to the area of the boundary between them by bending or stretching those regions and their boundary appropriately.

**The geometry of space is fixed by quantum entanglement, and that geometry seems to obey the dynamical rules of general relativity.**

## The Entropy of Black Holes

This conjecture has some support from an interesting source: the entropy of black holes. In the 1970s, Stephen Hawking and Jacob Bekenstein made the argument that black holes have a finite, nonzero entropy. And they derived a formula for that entropy in which the entropy is proportional to the area of the boundary of the black hole. That is called the event horizon of the black hole. In classical statistical mechanics, entropy is a way of counting how many ways there are to rearrange the constituents of a system, such as atoms or molecules, so as to keep it looking the same macroscopically. The same basic idea is true in quantum mechanics, but there's an important new wrinkle.

Remember that when you have two subsystems that are entangled with each other, you can't talk about the wave function for either subsystem separately. You can only talk about the wave function for the combination. But it turns out, as shown by John von Neumann, that you can think of each subsystem of this entangled collection as being described not by a single wave function but as a probability distribution over wave functions. Very much like in classical mechanics, you have a probability distribution over possible microscopic states. And in the same way, these entangled subsystems, therefore, have a nonzero entropy—even if you know the overall wave function exactly. So, entropy for a quantum subsystem is a reflection of the degree of entanglement it has with the outside world. This is a provocative connection between entropy and entanglement that says that the amount of entanglement between a black hole and the rest of the world is proportional to the area of the event horizon of the black hole.

While that's reasonable for black holes, now consider a more sweeping conjecture: that the entropy of any region of space is proportional to its boundary area. A version of this conjecture was first put forward in a provocative paper as recently as 1995 by physicist Ted Jacobson, who postulated that space curves just so that the area of a surface—that is, the boundary of some region—remains proportional to the entanglement of the region it encloses. You can't calculate the area of the surface without knowing something about the geometry of the larger space that it's part of. Meanwhile, entanglement tells you something about matter broadly construed—about the quantum fields living within space-time.

Jacobson also argued that entanglement could be directly related to the energy source that appears in Einstein's equation of general relativity. Through these maneuvers, Jacobson was able to derive Einstein's equation for general relativity rather than simply postulating it like Einstein did.

This is a radical change of perspective. Einstein thought in terms of energy—a definite quantity associated with particular configurations of stuff in the universe. Jacobson and others have argued that you can reach the same conclusions by thinking about entropy and entanglement. These are collective phenomena; they emerge from the mutual interactions of many small constituents of a system. And this simple shift in focus might offer a crucial way forward in the quest to discover a fundamentally quantum theory of gravity.

Jacobson's work, though, was not in the context of emergent space-time. He took space-time as an ingredient and talked about entanglement of quantum fields living within that space-time. But this course's approach is to take this idea and put it into the context of many-worlds quantum mechanics, where space itself is emergent. It's the entanglement structure, after all, that lets you define what you mean by distances and, therefore, by geometry. So, it's very natural to relate changes in the geometry to changes in entanglement. And in this way, the space-time that emerges from entanglement of a quantum state is very naturally dynamical—it very naturally responds to energy, just the way Einstein predicted.

Physicists aren't quite at the point yet where they can simply say that this emergent geometry on space evolves with time in exactly the right way to describe an emergent space-time that obeys Einstein's equation—even though

that's the ultimate goal. What they can do is specify a list of requirements under which that's exactly what does happen. And these individual requirements seem reasonable; for example, at long distances, physics looks like quantum field theory. Many of them, though, remain yet unproven. So far, the only rigorous results are available in situations where the gravitational field is relatively weak. So, there is not yet a way of describing black holes or the big bang or any of the other big puzzles that quantum gravity will eventually help solve, although there are some promising ideas.



# 21

## The Quantum Emergence of Time

**T**he question at hand is how space-time can emerge from entanglement in the context of quantum gravity within the many-worlds framework. But the previous lecture only considered how space emerges, not space-time. It is tempting to wonder whether time, like space, might be emergent rather than fundamental and whether entanglement might have something to do with it. The answer—tentatively and hypothetically—is yes on both counts, although the details remain sketchy.

## Time and the Schrödinger Equation

The many-worlds interpretation says that there is a quantum state, and it evolves according to the Schrödinger equation. But if you take the Schrödinger equation at face value, time is there in a fundamental way. On one side of Schrödinger's equation is something measuring the amount of energy being captured and characterized by the particular wave function you're looking at. On the other side of Schrödinger's equation is the rate at which the wave function is changing with respect to time. So, without time, you are not starting with the ordinary Schrödinger equation.

And that Schrödinger equation itself has an interesting consequence for the nature of time. If the Schrödinger equation is correct in its ordinary way of being presented, it follows that the universe lasts eternally toward both the past and the future for almost all quantum states that can be considered. You can evolve the equation forward and backward in time equally—at any moment and for almost any wave function. At no point do you hit a singularity or some kind of stopping point.

You might think that this feature of the Schrödinger equation conflicts with the oft-repeated fact that the big bang is the beginning of the universe—that there is no time before the big bang. But it's not actually known whether that's true. The big bang is supposed to represent a singularity—a moment in time where the energy density and the curvature of space-time become infinitely big. The equations break down. But singularities are a prediction of classical general relativity. They're not a prediction of quantum gravity or the Schrödinger equation. It's not known what quantum gravity predicts, so this is a place to keep an open mind. If quantum gravity operates according to some version of the Schrödinger equation, then for almost all quantum states, time runs from minus infinity in the past to plus infinity in the future.

There are lots of caveats there because the full picture is not known. What is thought of as the big bang might simply be a transitional phase. There might be an infinitely old universe before the big bang. It's also possible that what an observer like you who lives in space-time thinks of and measures as time might not be the same quantity as the time parameter in the Schrödinger

equation. One of Einstein's lessons is that there's more than one notion of time. So, maybe the  $t$  parameter that appears in Schrödinger's equation does indeed go forever, but the time you experience doesn't.

$$i\hbar \frac{\partial}{\partial t} \Psi = -\frac{\hbar^2}{2m} \nabla^2 \Psi + V(x) \Psi$$

## The Wheeler-DeWitt Equation

When it comes to quantum gravity and emergent space-time, there's not much that's certain, but there is one particular kind of uncertainty that is worth focusing on. The Schrödinger equation says that the rate of change of the wave function is driven by how much energy the quantum state represents. So, what if you consider quantum states whose energy is precisely zero? In that case, the equation says that the system doesn't evolve at all. Time disappears from the story. If the rate of change of the state is proportional to the energy in the state and the energy is zero, then there's no rate of change—there's no change at all. There's no change with respect to time, in particular.

You might think that it's implausible that the real universe has exactly zero energy. The universe is full of planets and stars, and all of those things have energy in them. But general relativity suggests that this isn't certain. Even at the classical level, you can imagine that the universe as a whole has zero energy. Of course, the individual things within the universe, such as stars and planets, seem to contain energy. But there's a subtlety: You have to consider that the gravitational field itself in general relativity contributes to the energy of the universe. Surprisingly, the energy of the gravitational field is generally a negative number.

The universe could be open and infinite in space, or it could be closed, and space could wrap around itself. It could be like a three-dimensional sphere or torus—some compact geometry—rather than stretching to infinity. In a closed spatial universe, the gravitational energy, which is negative, precisely cancels the positive energy from everything else. A closed universe has exactly zero energy, regardless of what's inside.

As a matter of empirical fact, it's not known whether the real universe is closed or whether space extends infinitely far in all directions. After all, humans can only see so far out into space since light travels at a finite speed—the speed of light—and there was only a finite amount of time since the big bang. Scientists can't see past the big bang. Past what they can see, maybe things extend forever, or maybe space closes back in on itself to form a closed three-dimensional version of the surface. It's just not known. But the point is that a closed universe is a viable possibility. And if space is closed, the mathematics of general relativity says that there is a sense in which the universe has precisely zero energy.

That's a purely classical statement in the context of classical general relativity, but there is a quantum mechanical analog. When you take general relativity and try to quantize it with wave functions and so forth, you have a version of the statement that says that a closed universe has exactly zero energy. This statement was developed by John Wheeler and Bryce DeWitt in the 1960s. The Wheeler-DeWitt equation says that the state of the universe described by quantum gravity does not evolve as a function of time. So, when the energy is zero, there is no evolution over time.

That idea might seem crazy, or at least ruled out by the data, in flagrant contradiction to observational experience. The universe certainly seems to evolve over time; planets orbit stars, apples fall from trees, and clocks tick. So, how can the manifest evidence of people's senses and experiences—all of which indicate that the universe evolves with time—be reconciled with the abstruse theoretical idea that somehow time doesn't exist?

The problem of time in quantum gravity is a known puzzle. And this problem is where the possibility of emergent time might come to the rescue. If the quantum state of the universe obeys the Wheeler-DeWitt equation—which is plausible but far from certain—then time has to be emergent rather than fundamental.

## The Page-Wootters Approach

There are ways that this might actually work. One way it might work was suggested by physicists Don Page and William Wootters in 1983. A simple analog of what they proposed starts with an ordinary quantum system that

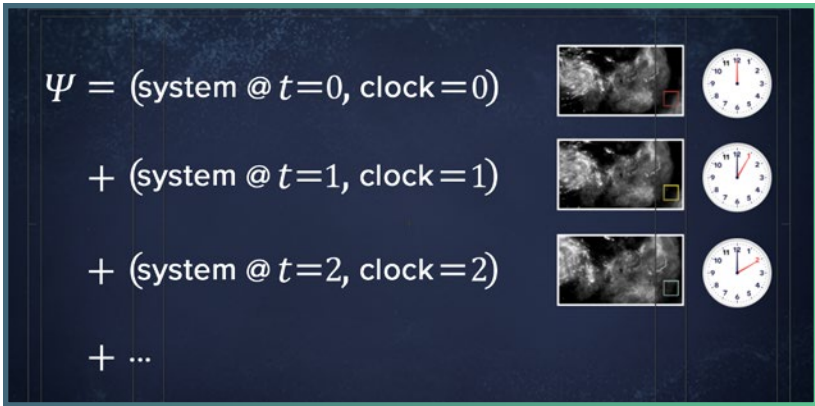
does evolve with time and shows how you could equally well think of such a system as a system without time. In some sense, time isn't necessary in this point of view; it's a choice to make to think about the system as evolving with time. And hopefully, understanding this will allow you to go backward: Not only can you start with the system with time evolution and show how time wasn't necessary, but you can also start from a system that is timeless, with no time evolution, and show how time can emerge.

Imagine an ordinary quantum system that consists of two parts: a clock (it doesn't need to be a literal clock, but it's something that keeps time) and everything else in the universe. Imagine that both the clock and the rest of the universe evolve in time as usual. Then, you take snapshots of the quantum state of both the clock and the rest of the universe every hour.

In any particular snapshot, the quantum state describes the clock reading some time and the rest of the system in whatever configuration it was in at that time. All the snapshots together results in a collection of instantaneous quantum states of the system at different times. The great thing about quantum states is that you can simply add them together. You can make a superposition of all the individual quantum states to make a new quantum state. You wouldn't be able to do that in classical mechanics. You can't take the state of one particle at one location and add it to a different state where the particle is at a different location. But quantum states are vectors. You live in Hilbert space, which is a vector space, and that means you can add them together.

So, let's make a new quantum state. You started with a quantum system evolving over time and took snapshots at different moments; now, you just add together all the snapshots. You're constructing a single unchanging quantum state that does not evolve with time by adding all the snapshots together.

And there is no one specific, once-and-for-all, time reading on the clock. You made a superposition, so the clock subsystem is in a superposition of all the times at which you took snapshots. That doesn't sound much like the real world, but it is something that you can write down as a quantum state:  $\Psi$  equals whatever the system was doing at  $t = 0$ , and the clock reads 0; plus whatever the system was doing at  $t = 1$ , and the clock reads 1; plus whatever the system was doing at  $t = 2$ , and the clock reads 2, etc.



But within that superposition of all the snapshots, the state of the clock is entangled with the rest of the state of the system. If you imagine measuring the state of the clock and see that it reads some particular time, then the rest of the universe will also be in a particular corresponding state—just like Bob’s spin is correlated with Alice’s measurement outcome in the EPR scenario. So, at any given reading on the clock, the universe is in whatever state the original evolving system was caught in at precisely that time.

In other words, there’s not really time in this superposition state. It’s not changing over time; it’s static. But because there is entanglement, there’s a relationship between what the clock might be reading and what the rest of the universe is doing. And the state of the rest of the universe is precisely what it would be if it were evolving as the original state did as a function of time.

This is what it means for time to be emergent in quantum mechanics. Time here is not a fundamental notion in this weird static state that you have constructed. Time means what the clock reads in this part of the overall quantum superposition. In that way, there is something that you can think about as time: It has emerged from a completely static state, thanks to the magic of entanglement.

This scenario—the possibility that time is simply an emergent phenomenon rather than something fundamental—has intriguing and potentially important consequences. To see that, let’s consider the opposite idea. Imagine

that time really is fundamental—that the fundamental equation of physics is the Schrödinger equation with the time parameter  $t$ . Time is a fundamental part of this version of physics.

Then, it follows that time is eternal. Time has neither a beginning nor an end from any one state. The Schrödinger equation indicates how the thing evolves both forward and backward in time; it indicates what the state will be a moment after and what it was a moment before. There's no point at which that evolution breaks down.

But if time is emergent, it's a very different story. One of the unanswered questions in quantum mechanics is whether Hilbert space is finite-dimensional or infinite-dimensional. In an ordinary quantum theory of particles or fields in smooth space-time, it can be shown that the Hilbert space must be infinite-dimensional if you don't have gravity. So, roughly, there are an infinite number of places a particle could be or of possible field configurations. But with the combination of quantum gravity and emergent space-time, there are indications that Hilbert space is actually finite-dimensional. Again, roughly, you might think there are an infinite number of things you could do with the quantum field. But above a certain energy, they just collapse to make a black hole. And it is known that the entropy of a black hole is finite, so there's only a finite number of possibilities.

If Hilbert space is finite-dimensional and time is emergent, then an interesting new possibility opens up. In contrast with the ordinary Schrödinger equation, for which time is eternal, a finite-dimensional Hilbert space offers only a finite number of distinct quantum states. That limits the possible ticks on the clock to a finite number. You have a clock subsystem that you're using to measure emergent time, and you have different states that it can be in, but if there's a finite dimensional Hilbert space, there's only a finite number of possible clock states. This means that time will have both a beginning and an end. You could consider the possibility that time cycles around forever—that it moves on a circle somehow. But there will only be a finite number of moments in each cycle. Maybe that possibility has something to do with how scientists should think about the big bang and the various puzzles associated with the beginning of the universe. As of now, it's just not known.

Every physicist who considers the questions of quantum gravity and emergent space-time would love to make a testable experimental prediction that would help them know whether they were on to something. Unfortunately, current ideas are too tentative to know precisely what those predictions might be and so ambitious that the technological capacity might not be up to the task of testing the predictions that do exist. So, at the moment, this small corner of physics remains stubbornly theoretical.



# 22

## Free Will, Determinism, and Many-Worlds

**T**his lecture and the next will explore the implications of the many-worlds interpretation for ordinary human lives. If you think that you will evolve into many future selves rather than one unique future self, how does that change how you should think about yourself—and even how you should behave in such a branching universe?

## Decisions and Choices

One misapprehension people have about many-worlds is that different worlds are created when you make a decision. You really shouldn't think about it that way. The idea of making a decision—which is central to human lives—is not part of the fundamental laws of physics. Decisions are one of those useful approximate emergent notions that is convenient to invoke when describing human-scale phenomena. What's labeled “making a decision” is a way of talking about a set of neurochemical processes happening in your brain.

So, here's one way of thinking about this: Do the physical processes happening in your brain when you make a decision cause the wave function of the universe to branch, with different decisions being made in each one of the different branches?

No, you do not cause the wave function to branch by making a decision. How branches happen has nothing to do with human brains making decisions. In large part, it's due to what is meant by something causing something else. This is a philosophy question, but it's a necessary one to confront here. Branching is the result of a microscopic process amplified to macroscopic scales. A system that is in a quantum superposition—typically a microscopic system—becomes entangled with a larger system, and that becomes entangled with its environment. That's decoherence, and it's what causes branching.

But a decision, on the other hand, is a macroscopic phenomenon from the start. There are no decisions being made by the electrons and atoms in your brain; they're just obeying the laws of physics.

What is meant by words like *decisions* or *choices*? Decisions and choices and their consequences are useful concepts when considering things at the macroscopic, human-sized level. It's perfectly OK to think of choices as really existing and making a difference—as long as you confine such talk to the regime in which they apply. Choices in your life really exist, but they

**When you make a decision, are there realities corresponding to every series of alternative choices that you could have made—universes that actualize all the possibilities of your life?**

exist in the same way that tables and chairs really exist. They're not part of fundamental physics, but they're part of a vocabulary—a language used for discussing the macroscopic world that you can wrap your brain around. This is a world where you don't have complete information about what every single atom is doing but, nevertheless, have coarse-grained information that is enough for you to make reasonable predictions about how things behave and what will happen next. That's the higher level—emergent level—at which decisions and choices can be discussed.

In other words, you could choose to talk about a person as a bunch of particles obeying Schrödinger's equation, or you could equally well talk about them as an agent—a person who has the ability to make decisions and whose decisions can affect the world around them. The important point is that you can't use both descriptions at once. You can't start a sentence in the language of microscopic particles in the Schrödinger equation and end the sentence in the language of human beings making decisions. These vocabularies are applicable at different scales. So, your decisions don't cause the wave function to branch, because your decisions and the wave function branching exist as concepts at two very different levels of reality.

The wave function branching is a relevant concept at the level of fundamental quantum physics. Your decisions are a relevant concept at the everyday macroscopic level of people, not at the fundamental physics level.

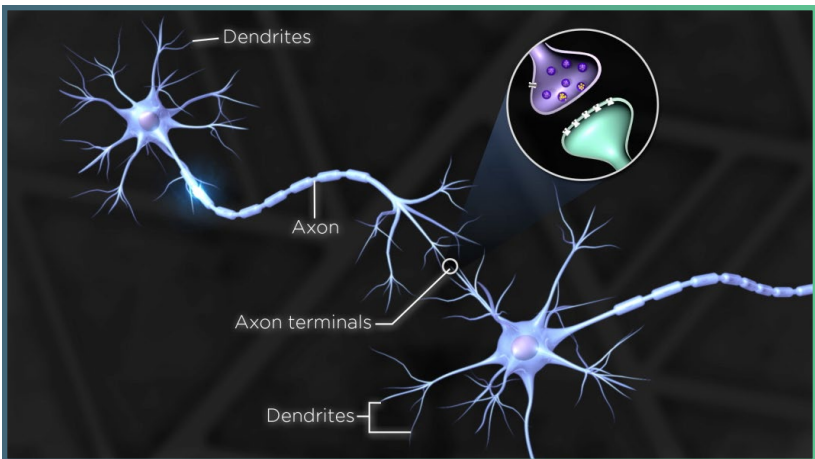
So, there is no sense in which your decisions cause branching. What causes branching is a tiny quantum system in a superposition becoming entangled with the rest of the universe. But even if you think that way, which is the philosophically respectable way to think, you can still ask whether there exist other branches where you made different decisions.

And, indeed, there can be. The subtle point that reconciles all this is how you should think about the notion of causality. Generally, it does make sense to imagine causal influences flowing from the fundamental microscopic level to the emergent macroscopic level. So, it makes sense to say something like this: "The table is solid because of the structure of the molecules in the wood, the properties of electrons in the atoms, the Pauli exclusion principle, and various microscopic ideas." But the other way doesn't work. You can't say, "The molecules in the wood have a certain structure because the table is solid."

Therefore, it is better to say, “Some microscopic process happened that caused branching, and it ended up that on different branches, I made different decisions.” That’s very different from saying, “I made a decision, and that caused the wave function of the universe to branch.” Your decisions don’t cause branching, but branching in the right circumstances can cause decisions.

But for the most part, when you make a decision that seems to you like a close call at the time you make it, almost all the probability ends up being concentrated on a single branch, not spread equally over many alternatives. Why is that true?

Think about what physically happens in the neurons in the brain. They’re cells that consist of a central body and a number of appendages. Most of the appendages on a neuron are called dendrites, and they take in signals from the surrounding neurons. But one of the appendages is an axon: a longer fiber down which outgoing signals are sent. So, you get both input and output from each neuron. Charged molecules, or ions, build up in the neuron until they reach a point where some kind of electrochemical pulse is triggered, and that pulse travels down the axon and across synapses to the dendrites of other neurons. And when you have many such events happening, you have the makings of a “thought.”



For the most part, these processes can be thought of as being purely classical, or at least deterministic. A certain combination of neural firings will lead to a certain outcome. It's chemistry.

Quantum mechanics does play a role at some level whenever there is a chemical reaction—because, after all, quantum mechanics sets the rules for how electrons want to jump from one atom to another, bind two atoms together, and so on. But when you get enough atoms together in one place, their overall net behavior can be described without any reference to quantum concepts. Otherwise, you wouldn't have been able to learn chemistry without first learning about the Schrödinger equation and worrying about the measurement problem. Most of chemistry is classical because there lots of atoms and molecules, and any small bits of quantum uncertainty will wash out.

So, decisions in the brain are best thought of as classical events, not quantum ones. While you might be unsure which decision you will eventually make, the outcome is most likely already encoded in your brain. If you were Laplace's demon—if you knew the state of all the neurons and all the chemicals in your brain—you would know with extremely good probability what you were going to do next.

To be fair, the quantitative extent to which this is true is not certain, simply because there's still a lot that's unknown about the physical processes behind thinking. It is conceivable that chemical reactions in your brain may play out differently depending on the entanglement between the different atoms involved. If that turns out to be true, then there would be a sense in which your brain is a quantum computer. It would be a limited sense but a very real one. When you think about decisions as the inevitable outcome of microscopic processes, though, the overall picture tends to look classical and deterministic.

At the same time, an honest Everettian admits that there will always be branches of the wave function on which quantum systems are appearing to do very unlikely things. If there's only a 0.001% chance of something happening, according to many-worlds, it happens on some branch of the wave function. So, even if the classical approximation to your brain implies that you're going to make a particular decision, there is some tiny amplitude for a bunch of neurons to do unlikely things in the wave function of the universe and

cause you to do something else. But it's not your decision that is causing the branching even then; it's the branching that you interpret as affecting your decision. In other words, by the time you make a choice—by the time you recognize that your cognition in your brain has come to a decision about what to do—the chemical reactions in your neurons have already branched the wave function. You just happened to have ended up on a branch in which the neurons have aligned in such a way so as to make a certain decision.

So, under the most straightforward understanding of the chemistry going on in your brain, most of your thinking has nothing to do with entanglement, branching of the wave function, or other worlds. You shouldn't imagine that making a difficult decision splits the world into multiple copies, each containing a version of you that chose differently.

## Free Will versus Determinism

Similarly, quantum mechanics has nothing to do with the question of free will as a human ability. It's tempting to think that it might because free will is often contrasted with determinism, which is the idea that the future is completely determined by the present state of the universe. If the future is determined, what room is there for you to make choices? In the textbook presentation of quantum mechanics, measurement outcomes are truly random.

It was discovered at the beginning of the 20th century that physics is not deterministic after all. Maybe that new idea from the beginning of quantum mechanics opens the door a crack for free will to sneak back in after it was banished by the Newtonian clockwork paradigm of classical mechanics.

There's a lot wrong with all this. First, the distinction between free will and determinism is not the right contrast to examine. Determinism is contrasted with indeterminism. Free will should be opposed to no free will. And you can live in any one of the four possibilities: determinism and free will, indeterminism and free will, determinism and no free will, or indeterminism and no free will.

*Determinism* is straightforward to define. Given the exact current state of a system, the laws of physics determine precisely what the state will be at later times. It is much trickier to define *free will*. It's usually defined as something like "the ability to have chosen otherwise." That sounds like it's well defined and compact, but there's an ambiguity there. The usual idea of free will imagines a comparison between what really happened in the actual world versus a different hypothetical scenario. And you know what really happened—you were in a situation, made a decision, and acted accordingly.

Hypothetically, if you wound the clock backward to that original situation, could you have decided differently? That's supposed to be the defining question of free will. The subtlety is that when you're playing this game of winding the clock backward, it's crucial to specify exactly what is being kept fixed between the real and hypothetical situations. What remains the same? Is it supposed to be absolutely everything, down to the last microscopic detail? Or are you just imagining fixing the available macroscopic information but allowing for variation within invisible microscopic details?

Let's compare what actually happened to a hypothetical rerunning of the universe. Imagine that the alternative universe starts from exactly the same starting condition and then remains exactly the same, down to the precise state of every elementary particle. In a classical, deterministic universe, the outcome—your decision—would be precisely the same. The choice would have been determined from the beginning. So, classically, it's a pretty simple dichotomy: There's no possibility that you could have made a different decision if you started from exactly the same initial state.

By contrast with this, according to textbook quantum mechanics, an element of true randomness is introduced, so you can't confidently predict exactly the same future outcome from the same initial conditions. But whatever the underlying theory of quantum mechanics turns out to be—whether it's many-worlds, objective collapse, hidden-variables, or something else—as far as the experience of observers is concerned, quantum mechanics is not deterministic. The underlying equations of something like many-worlds or hidden-variable theories are deterministic, but that doesn't matter for this kind of question because people don't experience those underlying equations. The best predictions they can make are truly not deterministic.

But, yet again, the fact that different outcomes are possible has nothing to do with what is usually meant by *free will*. Just because there can be a different outcome doesn't mean that it's because some kind of personal, superphysical, volitional influence was manifested over the laws of nature. It just means that there are some unpredictable quantum random numbers, and they came up differently. What really matters for the traditional notion of free will is not whether you are subject to deterministic laws of nature but whether you're subject to laws of any sort—laws that are impersonal and that you cannot overrule. The fact that you can't predict the future isn't the same as the idea that you are free to bring it about.

Even in textbook quantum mechanics, human beings are still collections of particles and fields obeying the laws of physics. There's no more free will there than there was in the classical, deterministic clockwork universe. And for that matter, deep down, quantum mechanics is not necessarily indeterministic—and this is where many-worlds comes into the picture. Many-worlds is a counterexample to the idea that the laws of nature truly are indeterministic. In many-worlds, you evolve perfectly deterministically from a single person now into multiple persons at some future time. No choices come into the matter anywhere. And most hidden-variable theories are also deterministic. Things just seem random because the values of the hidden variables are not known.

So, the universe might very well be deterministic and yet appear indeterministic. Presumably what matters to free will are the appearances of determinism, not the underlying reality. But the reason why this consideration of determinism versus indeterminism is not the final answer is that you also are free to contemplate weaker notions of free will. If you thought that free will meant the ability to override the laws of physics when you make decisions, it doesn't matter whether those laws are deterministic. But that's not the only thing it might mean.

Let's dispense with the idea that you could have some kind of perfect microscopic knowledge of the world (because you can't) and instead refer to the macroscopic knowledge that is actually available to you. In that case, a different form of unpredictability arises—not from the fundamental laws of physics but from the lack of your perfect knowledge.

Let's say that you know something about your current mental state. At the microscopic level, there will typically be many different specific arrangements of atoms and molecules in your body and brain that are 100% compatible with that knowledge of what you're calling your mental state. Some of those arrangements may lead to very different neural processes—sufficiently different that you would end up acting very differently if those arrangements had been true. So, how do you conceptualize this lack of perfect microscopic knowledge when describing real human beings?

The best you can realistically do to describe the way human beings—or other conscious agents in the world—act in reality is to attribute volition, or the ability to choose differently, to them. You attribute volition, or choices, to people when each person goes through the day. Attributing volition is what you do when you talk about yourself and others in real life because in the real world, it doesn't matter whether you could predict the future from perfect knowledge of the present because you don't have such knowledge, nor will you ever.

So, you conclude that whatever the fundamental laws of physics are, the right way of talking about people is as agents making choices. And this has led philosophers going back as far as Thomas Hobbes to propose what is called compatibilism between underlying deterministic laws and the reality of human decision-making. Most modern philosophers are compatibilists about free will. That doesn't mean that it's right, but the point is that there's no necessary conflict between impersonal underlying laws of physics and talking about human beings as if they have free will. In this picture, free will is real—just like tables and temperature and branches of the wave function. It is a higher-level emergent phenomenon. The microscopic laws of physics have nothing to do with it.

As far as quantum mechanics is concerned, it doesn't matter whether you are a compatibilist or an incompatibilist concerning free will. In neither case should quantum uncertainty affect your stance one way or the other. Even if you can't predict the outcome of a quantum measurement, that outcome stems from the laws of physics, not any personal choices made by you. You don't create the world by your actions; your actions are part of your world.



# 23

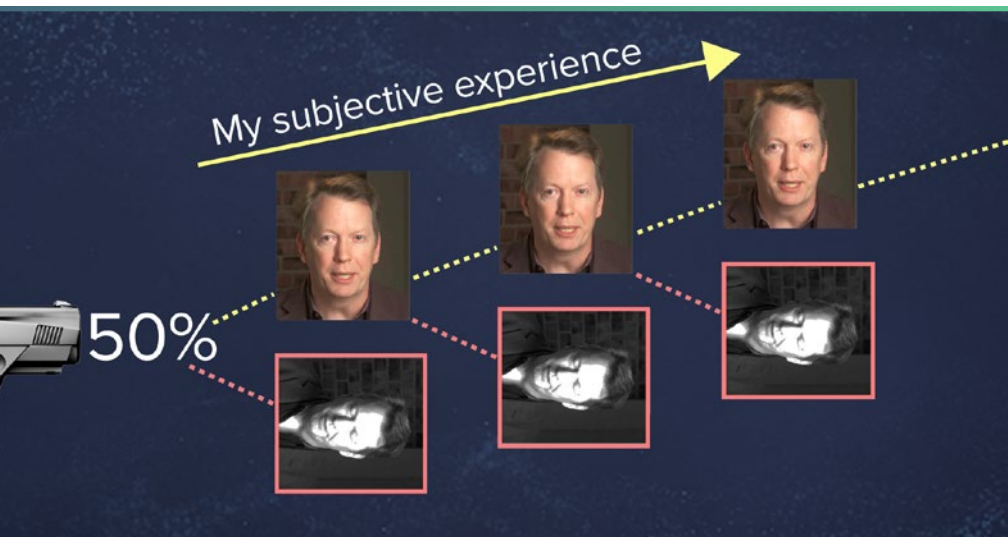
## What Happens to Ethics under Many-Worlds?

**T**he many-worlds interpretation suggests a radical revision to your view of reality. In addition to the world you experience, this scenario suggests that there are countless other worlds. It would be natural to think that if you change your fundamental view of reality in this dramatic way, that should imply correspondingly significant changes in how you should live your life. If you care about the well-being of future people in the universe, now you have to care about a whole multiverse of them. But despite the fact that many-worlds is a very different view of the universe, the implications for people's lives and how they should live them turn out to be relatively minimal, at least in the most straightforward way of thinking about this. To each individual on some branch of the wave function, life goes on just as if they had lived in a single world with stochastic quantum events. But the stakes are high, so these issues are worth exploring carefully.

## The Quantum Immortality Thought Experiment

Think about the quantum immortality thought experiment you learned about in lecture 12. Imagine a deadly device that is triggered by a quantum measurement that is supposed to have a 50% chance of triggering a gun that shoots a bullet into your head at close range and a 50% chance of doing nothing. According to many-worlds, that implies the existence of two branches of the wave function, one of which contains a living version of you and the other of which contains a dead version.

From your perspective, the branch on which the gun fired isn't one that any version of you ever gets to experience. Your descendant—or who would be your descendant on that world—is dead. But your descendant does live on the branch where the gun didn't fire. So, is there a sense in which you live forever? In fact, some people have claimed that you shouldn't be reluctant to do this experiment because there's a version of you who was still alive in the multiverse.



But that's not the right way to think about it. Even in a single-universe theory, you don't want to be suddenly killed—even if you wouldn't be around afterward to be upset. The point is that you are upset now by the prospect that you could be killed at some future time. Just saying that once you're dead, you don't have any regrets afterward doesn't mean that you don't mind being dead.

This goes back to that idea that if you accept many-worlds, it's important to treat individuals on different branches of the wave function as distinct persons, even if they descended from the same individual in the past.

There is an important asymmetry between how you think about your future versus your past in the many-worlds scenario. Ultimately, that's because of special conditions when the universe started at or near the big bang, when there was not that much entanglement and, correspondingly, just one branch of the wave function. Ever since, the number of branches has been growing as the universe evolves.

It's not known why the universe had this special initial condition. But for whatever reason, the quantum past and future are very different in the real universe. As a result, any one individual can trace their life backward in a unique line of continuity through a single person—a single set of lives. But if you go forward in time, you're going to branch into multiple people, according to the many-worlds view. There is not one future self that is picked out as really you; all the future people are legitimately future versions of you in some sense.

But it's equally true that there is no one person constituted by all the future individuals. They're not one person living on different branches. They are separate—as separate as identical twins or distinct people—despite descending from a single cell. And these people on separate branches cannot influence each other. They can't talk to each other in any way, and they can't know what happens to each other. They happen to share a common past but have completely distinct futures. As a result, you might care about what happens to the other versions of yourself that live on other branches, but that doesn't mean you should think of them as you—they're different people.

## A Philosophical Thought Experiment

Let's consider why this is a sensible attitude using a type of philosophical thought experiment that involves handing out money. Philosophers like to imagine that they can operationalize what you care about by asking how you would respond to different offers of being given money or having money taken away. So, imagine that you're just about to perform a measurement of the vertical spin on an electron that you've prepared in an equal superposition of spin up and spin down. As you know, you predict a 50-50 chance of seeing either spin up or spin down. In many-worlds, there would be a world where it's spin up and one on which it's spin down. But before you press the button to start the experiment, a random philanthropist enters your lab and offers you the following bargain: If you measure the spin up, they will give you a million dollars. But if you measure the spin down, you will give them a dollar. Under ordinary ways of thinking about rational decision-making, you would be wise to take the deal.

For all intents and purposes, it's just as if you're being offered a bet with equal chances of winning a million dollars or losing a dollar. Even if you know in many-worlds that there will certainly be some future selves that will be out a dollar, the expected reward to such a bet is quite large, if you think about it as a probability. So, you're going to take that offer.

But now imagine that you pushed the button before the philanthropist came in, and what you observed is spin down—the result that would have lost you a dollar in the previous scenario. The philanthropist comes in and sees that you've already done the experiment, but they want to make the deal and give you the following sales pitch: There is a version of you on another branch of the wave function that is right now being given a million dollars. But since you were unlucky enough to measure spin down, it's only fair, they say, that you should give them a dollar in this branch.

There's no reason for you to be happy about this just because a different branch has a version of you getting rich. You do not benefit. You, the person who is on the spin-down branch, doesn't have any material consequence from knowing or thinking that there's a different version of you that is a million dollars richer, nor is there any reason to feel any moral compulsion to give up your dollar. You are not the person on the other branch, and they are not you.

You do share a past—you used to be the same person, in a very real sense. But now you are separate. Once you've measured the spin and the branching has occurred, there are two different people. Neither your experiences nor your rewards should be thought of as being shared by various copies of you on different branches. The philanthropist missed a chance to make that bet, and even if you refuse to hand over the dollar, the people on the other branch have no way of knowing that. It can't have any direct effect on whether the version of you on the other branch gets paid the million or not.

That might seem like a reasonable policy to have when it comes to your own personal well-being. But what about that of others? How does knowing about the existence of other worlds affect the notions of moral or ethical behavior? Is the way that people should behave in the world affected by the fact that there could be different worlds in which different people feel different consequences?

Let's contrast two different scenarios. One is many-worlds. It's the usual story of decoherence, branching, and so forth, and you're going to ask how you should behave in that universe. The other scenario is some version of single-world quantum mechanics. So, there's still quantum mechanics—what happens in a quantum event is supposed to be truly random, stochastic, and unpredictable, and it will follow the statistics given by the Born rule—but there's only one stochastic world.

And imagine that you accept the arguments that derive the Born rule in many-worlds. If you don't accept the arguments for the Born rule in many-worlds, you don't need to go this far in your arguments. You don't need to worry about morality or anything like that. If you don't think that many-worlds recovers the Born rule, then many-worlds would just be nonviable from the start.

So, what you're really comparing is a single world with random happenings along the way to a set of many-worlds where everything allowed happens in some world—but, crucially, to which you attach different weights to different happenings, always in accordance with the Born rule. Does the actual existence of these other worlds have any effect on your moral decisions, even if you think about probability in the same way?

## Deontology versus Consequentialism

The difficulty is that people don't agree on how to think about morality. The right way to define morality is itself a controversial subject, even if everyone were to agree that there's only a single world. But it's instructive to consider two broad categories of moral theory: deontology and consequentialism. For deontologists, moral behavior is a matter of obeying the right rules. The idea is that an action is inherently right or wrong; it's not a matter of what the consequences of that action might turn out to be. Consequentialists have the alternative view—that you should work to maximize the beneficial consequences of your actions. Utilitarians who advocate maximizing some measure of overall well-being are paradigmatic consequentialists.

How would these different schools of moral thought be affected by agreeing that multiple worlds exist with slightly different conditions? At first glance, deontology would seem to be unaffected by the possible presence of other worlds, at least in the straightforward cases. If the whole point of your theory is that actions are intrinsically right or wrong, regardless of what outcomes they lead to, the existence of more worlds in which those outcomes can occur doesn't seem like it should really matter.

A typical deontological rule would be Immanuel Kant's categorical imperative, which says to act only according to that maxim whereby you can at the same time will that it should become a universal law. In other words, you should choose actions that you think other people in the same circumstances should also choose.

If this is your preferred moral approach, it would seem like it will be safe here to replace a universal law with a multiversal law. In other words, any law worth having should be a law that should hold in all branches of the wave function. That shouldn't alter any substantive judgment about what kind of actions might qualify. You're not saying what is a moral action or an immoral action. You're just saying that if that's the way you reason about morality—actions are intrinsically good or bad—the fact that you live in one universe versus many probably doesn't change many things.

So, in the most straightforward versions of deontology, the fact that a multiverse might exist—that many-worlds might be the right version of quantum mechanics—doesn't change a lot about how you should act.

But in philosophy, much as in physics, it's very difficult to make a claim that something is truly impossible. And indeed, it's possible to come up with a version of deontology that would lead you to behave differently in a single stochastic world versus many worlds.

Consequentialism is a slightly trickier question. A typical no-nonsense utilitarian, for example, will believe that there is a quantity called utility that measures the amount of well-being associated with all the conscious creatures in the world. And the utilitarian will think that this quantity can be added together among all the creatures to obtain a total utility for society and that the morally right course of action is the one that maximizes this total utility.

Imagine further that you judge the total utility in the universe to be some positive number. You think that, overall, the existence of people in the world is a good thing—they're more happy than unhappy. If you didn't think that, then you'd rather that the people within the universe just didn't exist because the maximum utility is a negative number. Some people do think that way, but it's generally considered to be an extreme position.

So, imagine that you think that, overall, utility is positive. It might seem to follow that if the universe has positive utility and your goal is to maximize utility, creating a new copy of the whole universe would be one of the most morally valorous actions you could possibly take. The right thing to do, then, would be to branch the wave function of the universe as often as possible. Somehow, you're making a huge number of quantum measurements. And by making new universes, you're making new people and, therefore, increasing the utility.

But something about this doesn't seem right. If you're just sitting in your basement and measuring quantum spins, the outcome of those measurements has no impact on the lives of actual people in this universe or any other. Does it really make sense to think that such a morally praiseworthy result should be given to that sitting-in-your-basement activity?

Happily, there are a few ways out of this puzzle. One is simply to deny the assumptions. Maybe this kind of no-nonsense utilitarianism just isn't the best moral theory after all.

**As counterintuitive as many-worlds might seem, at the end of the day, it doesn't really change how you should live your life.**

Alternatively, the philosopher Robert Nozick imagined what he called the utility monster, a hypothetical being that is so good at experiencing pleasure and happiness that the most moral thing anyone could do—the way that you could most effectively increase the utility of the world—would be to keep the utility monster as happy as possible, even if all the actual human beings in the world would thereby suffer.

To most people, this means that this kind of utilitarianism is too simplistic. And the idea of just adding up utilities among different people doesn't always lead to the results you might initially have hoped for, usually because you can come up with some scenarios where some people have to suffer a lot just so that the greater good can benefit. Depending on the situation, that will often strike people as completely unfair. This is a standard worry about utilitarianism—nothing specifically related to many-worlds.

But there's another solution to the dilemma posed about sitting in your basement measuring spins, and this one comports more directly with the many-worlds philosophy. Remember, when you learned about deriving the Born rule, you learned how to apportion credences and conditions of self-locating uncertainty. You know the wave function of the universe, but you don't know which exact branch you're on. The answer was that your credence should be proportional to the weight of the branch—the absolute value of the amplitude squared—multiplying that branch. This weight is a crucially important aspect of how worlds are thought about in an Everettian picture. It doesn't just frame how probability is thought about. Conservation of energy also only makes sense if you multiply the energy of each branch by its associated weight so that you aren't creating energy out of nowhere every time a world branches.

It makes sense, then, that you should do the same with utility. If you have a universe with some given total utility and you measure a spin to branch it in two, the post-branching utility should be the sum of the utilities times the weights of their branches so that you don't get some fake increase in utility just because you have two branches where you used to have one.

This is what was really meant when it was said that you should think of the branches as counting by an amount proportional to their weights. It goes for utility just as much as probability or energy. And then, in the likely event that your spin measurement didn't affect anyone's utility in the actual universes, the total utility will be unchanged by your measurements and the subsequent branching. You have made more universes, but they're correspondingly thinner—and that's just what your intuition should expect. It's also what you would directly conclude from the decision-theoretic approach to probability in many-worlds. So, if you think it through carefully, the tentative conclusion is that the existence of the many worlds shouldn't change your ideas about moral action under consequentialism in any noticeable way.

But as with deontology, it is also possible to cook up a system in which the difference between many-worlds and random collapse theories really would be morally relevant.

Excluding such artificial constructions, though, many-worlds doesn't seem to have many moral implications. The picture of branching as creating an entirely new copy of the universe is a vivid one, but it's not quite right. It's better to think of it as dividing the existing universe into almost identical slices—each one of which has a smaller weight than the original. If you follow that picture carefully, you conclude that it's correct to think about your future exactly as if you lived in a single stochastic universe that obeys the Born rule. You might not live in such a universe, but practically, it's safe and consistent to navigate moral choices as if you do.



# 24

## A Future Renaissance for Quantum Mechanics

**Q**uantum mechanics is about a century old, but it remains an active area of research among physicists, not to mention philosophers and others. And some of this research investigates the very foundations of the theory. But the large part of research in quantum theory considers the applications of quantum mechanics. And amid all the fretful contemplating about multiple universes and philosophical conundrums, quantum mechanics has been extraordinarily successful as a predictive theory.

### Birthday for Quantum Mechanics



## 20th-Century Quantum Mechanics

Quantum mechanics is needed to do many things within physics—for example, to predict the decay or scattering of elementary particles at the Large Hadron Collider in Geneva. And the rules of quantum mechanics, even if the Copenhagen interpretation is not perfectly rigorous, are still clear enough that such predictions can be made—and verified and tested to exquisite accuracy.

So, even though it might not be known, or agreed upon, what is really happening under the hood with quantum mechanics, physicists have been extremely good at putting it to use. And for a long time, that was a perfectly adequate state of affairs as far as the actual research of working physicists was concerned. There just wasn't much appetite for delving into foundational issues. People mostly accepted the Copenhagen interpretation—and, even more importantly, it's not that they accepted all the details and philosophical claims of Copenhagen but that they didn't worry too much about it.

But for the most part, over the years, the ways that quantum mechanics was used by professional physicists didn't need detailed analysis of the foundation because they were using it in fairly straightforward—maybe even simplistic—ways.

Once the basic ideas of quantum mechanics were put into place, there were countless interesting questions to be addressed using those basic ideas—from radioactivity and particle physics to chemistry, superconductivity, and the early universe. Foundational questions still lingered in the background, but it was hard to see how to address them.

Starting in the 1950s, some progress was made. There was the invention of Bohmian mechanics and Everett's many-worlds theory. These are interesting and viable models that are legitimate scientific theories; they're not just interpretive frameworks to think about Copenhagen or some other version of quantum mechanics. But even though these are legitimate and certainly different scientific theories, it remained difficult—or perhaps impossible—to distinguish between them and other foundational approaches by doing experiments. So, to the average physicist, it seemed like a sensible attitude to sweep these issues under the rug—to just not spend time fretting about them. They could spend their valuable research time on more practical matters with direct experimental implications.

These days, however, things are beginning to change. As is often the case in science, much of the change is being driven by technological innovation—not just what physicists want to do but what they have the capacity to pull off. Starting around the 1980s, as computers became ubiquitous in many fields, physicists began to realize that quantum computers are special: They might have capabilities that an ordinary classical computer would not. And those special capabilities require that entangled qubits are shielded from the threat of decoherence. Protecting qubits from decoherence is difficult, but physicists are rapidly getting better at it.

To even have this conversation, people need to know what decoherence is. They need to understand how it works well enough to try to shield their qubits from it. And this quantum computing success has opened the door to other applications of the special features of entangled quantum systems—from fields like quantum cryptography, which is about encoding messages, to quantum metrology, which is about measuring tiny quantities to high precision.

And underlying this technological success is a need to better understand the ideas that make them possible, such as decoherence—and, for that matter, the whole new field of quantum information. And that has motivated a renewed attention to the foundations of quantum mechanics. This revived interest in

foundational questions goes beyond the standard understanding of quantum mechanics that begins and ends with the Schrödinger equation and the Born rule.

But it's not just a matter of technology spurring understanding. That's a hopeful, optimistic story to tell, but all along there was a plucky minority of physicists and philosophers who have continued to think about these foundational issues. In some ways, what's happening is that their long-standing work is finally paying off.

John Bell was inspired by David Bohm's hidden-variables theory and kept his work secret from his colleagues at CERN. It wasn't his respectable, everyday research; it was a side project. Now, however, the significance of Bell's work is widely recognized. For evidence, look no further than the Nobel Prize in Physics in 2022, which was given to three physicists—Alain Aspect, John Clauser, and Anton Zeilinger—for doing experimental tests of Bell's theorem.

This 2022 prize for testing Bell's theorem was the first time that a Nobel Prize had been given for work in the foundations of quantum mechanics since Max Born won it in 1954 for the probability rule. Many other Nobel Prizes have relied on quantum mechanics—whether it's superconductivity or particle physics—but they weren't given specifically for advancing the foundations of quantum theory itself rather than putting it to work in some particular context. There has been other important foundational work that's been done—for example, understanding the role of decoherence and pointer states—and it's very possible that this work will be the subject of future Nobel Prizes.

These days, physics as a whole is an incredibly vibrant field in which people are studying everything from atoms to metals to DNA molecules to stars and planets. Some tiny part of this work is called fundamental physics because the subject matter is at the deepest level.

Fundamental physics probes the basic laws that undergird observed phenomena. It includes fields such as particle physics, gravity, and the most basic parts of cosmology—not to mention the deep frameworks of relativity and quantum mechanics. And progress within fundamental physics, as opposed to physics as a whole, has slowed down in recent decades. The first half of the 20th century witnessed an extraordinary explosion of progress in fundamental physics, and it's just hard to live up to that standard. The first

half of the 20th century is the period in which special relativity and general relativity were developed and the fact that the universe was expanding was discovered observationally after being theorized. Quantum mechanics and quantum field theory were invented, and the road toward modern particle physics was started down.

The second half of the 20th century was also very exciting, but to some extent, it was a matter of following up on the breakthroughs of the first half: understanding black holes and cosmology through general relativity, identifying the cosmic microwave background, and using techniques from quantum mechanics and quantum field theory to discover new particles and assemble the standard model of particle physics. By the 1960s and 1970s, there were a large number of surprising new discoveries in particle physics, and those discoveries helped physicists assemble the standard model as it's known today. But since then, all the new discoveries confirm the predictions that the standard model was already making. When the Higgs boson was discovered in 2012, it was a huge deal, but it had been predicted in the 1960s. The only arguable exception that the standard model keeps fitting the data is the fact that neutrinos have mass, as opposed to being the massless particles that are predicted in the simplest versions of the standard model. But many people suspected from the start that they would predict that neutrinos really did have mass. That idea was finally verified in the 1990s. When it was verified, it didn't require a major rethinking of the fundamental foundations to accommodate neutrino masses in the standard model.

The only truly surprising experimental result in fundamental physics since the 1970s has been the discovery of the acceleration of the universe. That's attributed to dark energy in empty space, and it was found in 1998. And even that was anticipated. Physicists knew that it was a possibility, although most working physicists would have bet against it. And even today, there is not a 100% consensus on what the dark energy is that is making the universe accelerate—but most cosmologists have a favorite candidate: the cosmological constant that was suggested by Einstein in 1917. The cosmological constant is equivalent to a tiny but constant energy density inherent in the fabric of space itself, so it is also sometimes called vacuum energy. In some sense, the mystery is not what the vacuum energy is; it's why the vacuum energy is so tiny, rather than why it exists at all.

In some ways of thinking about it, given all the quantum fields buzzing around in empty space, you might have expected the energy density of empty space to be much larger than it is—there’s a lot going on in empty space. Unfortunately, even if that 1998 discovery manages to help quantify how much the amount of vacuum energy is, that’s just one number at the end of the day—how much energy is in empty space. And knowing that one number doesn’t provide much useful guidance in pushing understanding into new regimes.

## 21st-Century Quantum Mechanics

The 21st century, by contrast, has seen essentially no surprising experimental results in fundamental physics. And that’s not because there aren’t open questions. Some of the things that aren’t known include

- ▼ whether the forces of nature are somehow unified together,
- ▼ what the ultimate nature of dark matter or dark energy is,
- ▼ what happens in black holes or the big bang, and
- ▼ why there’s more matter than antimatter in the universe.

There’s a long list of similar questions. The complete understanding doesn’t exist yet. It’s just that physicists aren’t yet able to collect data that helps them answer the questions they have. And there’s very little way of knowing which of these ideas are the most promising.

How do you make progress in such a situation? A sensible strategy is to take a step backward to reexamine the foundations of the field—especially look at those questions that have long been swept under the rug. Maybe the reason why some of the questions are so difficult is because physicists never took the time to truly understand quantum mechanics in the first place.

Ongoing research, especially in quantum gravity—things like black holes and string theory—has provided hints that might be needed to understand quantum mechanics better in order to address these other problems. Several lines of research and string theory and other approaches to quantum gravity have pointed to the idea that space-time is emergent rather than fundamental. If that’s true, then what is fundamental? If the world is quantum mechanical

to someone who appreciates many-worlds, the obvious answer would be the quantum state of the universe. But how the everyday world emerges from such an abstract idea is still far from clear—a better understanding is called for.

This course has explored the idea that such a better understanding will involve the Everettian, or many-worlds, approach to quantum theory. That's the unique approach where you don't have to assume anything about the nature of space-time from the beginning. You can just carefully examine the behavior of a quantum wave function and ask when and how a classical-looking space-time (and all the other stuff you're familiar with) emerges from it. Maybe this approach—stepping back and thinking about the foundations of quantum theory and many-worlds in particular—will ultimately turn out to be a dead end. That's life as a research scientist. But given what is now known, it does have a simple and compelling perspective to offer. There are no true collapses of the wave function, nor are there any hidden variables. There's just the quantum state. It evolves smoothly, according to the Schrödinger equation. Everything else emerges from that simple setup. To show precisely how it emerges is a challenge, but it's an exciting one.

The simple setup of many-worlds leads to dramatic and challenging consequences. As Hugh Everett himself realized, taking the wave function in the Schrödinger equation at face value leads physicists to predict the existence of other worlds—ones that are almost identical to the world you live in, except for the outcomes of certain quantum measurements. But nobody wanted the universes predicted by many-worlds. They weren't being used to solve any puzzles that already existed. But the theory predicts them, so it behooves science to figure out how best to deal with them.

Not everyone agree. Indeed, there's an interesting correlation between a person's academic specialty and their happiness with many-worlds. Physicists—especially those working in gravitation, cosmology, and fundamental physics—are often enthusiastic Everettians. But many philosophers of physics worry about the challenges of dealing with an enormous number of new universes, so they tend to lean toward hidden-variables or objective collapse models.

<p><b>1901–1950</b></p> <ul style="list-style-type: none"> <li>Special relativity</li> <li>General relativity</li> <li>Expanding universe</li> <li>Quantum mechanics</li> <li>Quantum field theory</li> <li>Modern particle physics</li> </ul>	<p><b>1951–2000</b></p> <ul style="list-style-type: none"> <li>Black holes and cosmology</li> <li>Cosmic microwave background</li> <li>Standard model of particle physics</li> <li>Neutrinos have mass</li> <li>Acceleration of the universe from dark energy</li> </ul>	<p><b>2001–present</b></p> <ul style="list-style-type: none"> <li>Higgs boson</li> <li>LIGO discovers gravitational waves</li> </ul>
--	--	--

## The Future of Quantum Mechanics

The foundations of quantum mechanics generally and many-worlds in particular are on the cusp of a renaissance. A combination of factors has led a growing number of intelligent people to start thinking carefully about these ideas. That is leading to new progress in the field—and that progress is needed.

It's easy to state what many-worlds is as a theory. What's difficult is connecting it to the world that's observed.

Physicists are trying to understand things like how space-time emerges, why the world looks classical at all, and why locality is so important to fundamental physics. This point of view might help better explain longstanding puzzles of how to quantize gravity, why the energy scales of particle physics are so different from those of gravity, what happens inside black holes, what happens when black holes evaporate, and what is so special about the big bang—not to mention even grander questions, such as why there is something rather than nothing and why these specific laws of physics exist rather than other ones.

This project of developing the many-worlds approach could also help in much more down-to-earth ways by finding new technological applications for quantum theory. The other worlds don't help provide technological applications, but they could inspire people. David Deutsch, for example—one of the pioneers of quantum computing—was directly motivated by Everett's ideas. You don't have to accept that many-worlds is true to think about quantum computers or other applications. But if you do, it might lead you to think in new and helpful ways.

As impressive as the 20th century was in terms of breakthrough ideas in physics, there's still a lot remaining to be done. Facing up to the foundational puzzles of quantum mechanics might help inaugurate a new era of deep discoveries about the fundamental nature of reality—and, perhaps, the many worlds hidden inside it.

# Bibliography

There are a number of books dealing with quantum mechanics in general and many-worlds in particular as well as other approaches. These are some of the most useful, ranging from fairly basic to relatively advanced.

## Overall Reference:

Carroll, Sean. *Something Deeply Hidden: Quantum Worlds and the Emergence of Spacetime*. Dutton, 2019. This book is similar in level to this course and follows a similar basic outline. The topics of most of the lectures are also discussed in the book.

## Introductions to Quantum Mechanics:

Orzel, Chad. *How to Teach Quantum Physics to Your Dog*. Scribner, 2010. This is a gentler introduction to quantum theory with a humorous bent, but it's still accurate on the physics.

Bub, Tanya, and Jeffrey Bub. *Totally Random: Why Nobody Understands Quantum Mechanics*. Princeton University Press, 2018. In graphic-novel form, an introduction focusing on entanglement and philosophical issues.

Rovelli, Carlo. *Helgoland: Making Sense of the Quantum Revolution*. Riverhead Books, 2021. A short and elegantly written introduction to the basic ideas of quantum theory with a different (non-Everettian) perspective from this course.

Deutsch, David. *The Fabric of Reality*. Penguin, 1997. A book that includes quantum mechanics, evolution, and computation in a unique and thrilling way.

**Historical Perspective:**

Becker, Adam. *What Is Real? The Unfinished Quest for the Meaning of Quantum Physics*. Basic Books, 2018. A history of quantum mechanics, including both scientific and philosophical issues. Unusual in that it pays attention to the foundations of quantum mechanics and follows their development up to the present day, rather than stopping in the 1930s.

Byrne, Peter. *The Many Worlds of Hugh Everett III: Multiple Universes, Mutual Assured Destruction, and the Meltdown of a Nuclear Family*. Oxford University Press, 2010. An in-depth biography of Hugh Everett III.

**More Depth on Many-Worlds:**

Wallace, David. *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford University Press, 2012. Written by a well-respected philosopher of physics, this is considered the definitive modern comprehensive introduction to Everettian quantum mechanics.

Everett, Hugh, III. *The Everett Interpretation of Quantum Mechanics: Collected Works 1955–1980 with Commentary*. Edited by Jeffrey A. Barrett and Peter Byrne. Princeton University Press, 2012. A reprint of Everett's original work, including various versions of his thesis as well as substantial later correspondence and commentary.

Saunders, Simon, Jonathan Barrett, Adrian Kent, and David Wallace. *Many Worlds? Everett, Quantum Theory, and Reality*. Oxford University Press, 2010. A collection of essays discussing issues in many-worlds, both pro and anti.

### **Other Approaches to Quantum Foundations:**

Albert, David. *Quantum Mechanics and Experience*. Harvard University Press, 1992. An elegant philosophical discussion of quantum foundations.

Maudlin, Tim. *Philosophy of Physics: Quantum Theory*. Princeton University Press, 2019. A somewhat more technical introduction with details on hidden variables and Bohmian mechanics.

von Baeyer, Hans Cristian. *QBism: The Future of Quantum Physics*. Harvard University Press, 2016. An enthusiastic introduction to epistemic approaches to quantum mechanics.









