Quantum Time

Chapter 11 of *From Eternity to Here*, Sean Carroll.

> Sweet is by convention, bitter by convention, hot by convention, cold by convention, color by convention; in truth there are but atoms and the void.
>
> — Democritus[i]

Many people who have sat through introductory physics courses in high school or college might disagree with the claim, "Newtonian mechanics makes intuitive sense to us." They may remember the subject as a bewildering merry-go-round of pulleys and vectors and inclined planes, and think that "intuitive sense" is the last thing that Newtonian mechanics should be accused of making.

But while the process of actually calculating something within the framework of Newtonian mechanics—doing a homework problem, or getting astronauts to the moon—can be ferociously complicated, the underlying concepts are pretty straightforward. The world is made of tangible things that we can observe and recognize: billiard balls, planets, pulleys. These things exert forces, or bump into one another, and their motions change in response to those influences. If Laplace's Demon knew all of the positions and momenta of every particle in the universe, it could predict the future and the past with perfect fidelity; we know that this is outside of our capabilities, but we can imagine knowing the positions and momenta of a few billiard balls on a frictionless table, and at least in principle we can imagine doing the math. After that it's just a matter of extrapolation and courage to encompass the entire universe.

Newtonian mechanics is usually referred to as "classical" mechanics by physicists, who want to emphasize that it's not just a set of particular rules laid down by Newton. Classical mechanics is a way of thinking about the deep structure of the world. Different types of things—baseballs, gas molecules, electromagnetic waves—will follow different specific rules, but those rules will share the same pattern. The essence of that pattern is that everything has some kind of "position," and some kind of "momentum," and that information can be used to predict what will happen next.

This structure is repeated in a variety of contexts: Newton's own theory of gravitation, Maxwell's nineteenth-century theory of electricity and magnetism, and Einstein's general relativity all fit into the classical framework. Classical mechanics isn't a particular theory; it's a paradigm, a way of conceptualizing what a physical theory is, and one that has demonstrated an astonishing range of empirical success. After Newton published his 1687 masterwork, *Philosophiæ Naturalis Principia Mathematica*, it became almost impossible to imagine doing physics any other way. The world is made of things, characterized by positions and momenta, pushed about by certain sets of forces; the job of physics was to classify the kinds of things and figure out what the forces were, and we'd be done.

But now we know better: Classical mechanics isn't correct. In the early decades of the twentieth century, physicists trying to understand the behavior of matter on microscopic scales were gradually forced to the conclusion that the rules would have to be overturned, and replaced with something else. That something else is *quantum mechanics*, arguably the greatest triumph of human intelligence and imagination in all of history. Quantum mechanics offers an image of the world that is radically different from that of classical mechanics, one that scientists never would have seriously contemplated if the experimental data had left them with any other choice. Today, quantum mechanics enjoys the status that classical mechanics held at the dawn of the twentieth century: It has passed a variety of empirical tests, and most researchers are convinced that the ultimate laws of physics are quantum-mechanical in nature.

But despite its triumphs, quantum mechanics remains somewhat mysterious. Physicists are completely confident in how they *use* quantum mechanics—they can build theories, make predictions, test against experiments, and there is never any ambiguity along the way. And nevertheless, we're not completely sure we know what quantum mechanics really *is*. There is a respectable field of intellectual endeavor, occupying the time of a substantial number of talented scientists and philosophers, that goes under the name of "interpretations of quantum mechanics." A century ago, there was no such field as "interpretations of classical mechanics"—classical mechanics is perfectly straightforward to interpret. We're still not sure what is the best way to think and talk about quantum mechanics.

This interpretational anxiety stems from the single basic difference between quantum mechanics and classical mechanics, which is both simple and world-shattering in its implications:

> According to quantum mechanics, what we can *observe* about the world is only a tiny subset of what actually *exists*.

Attempts at explaining this principle often water it down beyond recognition. "It's like that friend of yours who has such a nice smile, except when you try to take his picture it always disappears." Quantum mechanics is a lot more profound than that. In the classical world, it might be difficult to obtain a precise measurement of some quantity; we need to be very careful not to disturb the system we're looking at. But there is nothing in classical physics that prevents us from being careful. In quantum mechanics, on the other hand, there is an unavoidable obstacle to making complete and nondisruptive observations of a physical system. It simply can't be done, in general. What exactly happens when you try to observe something, and what actually counts as a "measurement"—those are the locus of the mystery. This is what is helpfully known as the "measurement problem," much as having an automobile roll off a cliff and smash into pieces on the rocks hundreds of feet below might be known as "car trouble." Successful physical theories aren't supposed to have ambiguities like this; the very first thing we ask about them is that they be clearly defined. Quantum mechanics, despite all its undeniable successes, isn't there yet.

None of which should be taken to mean that all Hell has broken loose, or that the mysteries of quantum mechanics offer an excuse to believe whatever you want. In particular, quantum mechanics doesn't mean you can change reality just by thinking about it, or that modern physics has rediscovered ancient Buddhist wisdom.[ii] There are still rules, and we know how the rules operate in the regimes of interest to our everyday lives. But we'd like to understand how the rules operate in every conceivable situation.

Most modern physicists deal with the problems of interpreting quantum mechanics through the age-old strategy of "denial." They know how the rules operate in cases of interest, they can put quantum mechanics to work in specific circumstances and achieve amazing agreement with experiment, and they don't want to be bothered with pesky questions about what it all means or whether the theory is perfectly well-defined. For our purposes in this book, that is often a pretty good strategy. The problem of the arrow of time was there for Boltzmann and his collaborators, before quantum mechanics was ever invented; we can go very far talking about entropy and cosmology without worrying about the details of quantum mechanics.

At some point, however, we need to face the music. The arrow of time is, after all, a fundamental puzzle, and it's possible that quantum mechanics will play a crucial role in resolving that puzzle. And there's something else of more direct interest: That process of measurement, where all of the interpretational tangles of quantum mechanics are to be found, has the remarkable property that it is *irreversible*. Alone among all of the well-accepted laws of physics, quantum measurement is a process that defines an arrow of time: Once you do it, you can't undo it. And that's a mystery.

It's very possible that this mysterious irreversibility is of precisely the same character as the mysterious irreversibility in thermodynamics, as codified in the Second Law: It's a consequence of making approximations and throwing away information, even though the deep underlying processes are all individually reversible. I'll be advocating that point of view in this chapter. But the subject remains controversial among the experts. The one sure thing is that we have to confront the measurement problem head-on if we're interested in the arrow of time.

The quantum cat

Thanks to the thought-experiment stylings of Erwin Schrödinger, it has become traditional in discussions of quantum mechanics to use cats as examples.[iii] Schrödinger's Cat was proposed to help illustrate the difficulties involved in the measurement problem, but we're going start with the basic features of the theory before diving into the subtleties. And no animals will be harmed in our thought experiments.

Imagine your cat, Miss Kitty, has two favorite places in your house: on the sofa and under the dining room table. In the real world, there are an infinite number of positions in space that could specify the location of a physical object such as a cat; likewise, there are an infinite number of momenta, even if your cat tends not to move very fast. We're going to be simplifying things dramatically, in order to get at the heart of quantum mechanics. So let's imagine that we can completely specify the state of Miss Kitty—as it would be described in classical mechanics—by saying whether she is on the sofa or under the table. We're throwing out any information about her speed, or any knowledge of exactly what part of the sofa she's on, and we're disregarding any possible positions that are not "sofa" or "table." From the classical point of view, we are simplifying Miss Kitty down to a two-state system. (Two-state systems actually exist in the real world; for example, the spin of an electron or photon can either point up or point down. The quantum state of a two-state system is described by a "qubit.")

Here is the first major difference between quantum mechanics and classical mechanics: In quantum mechanics, there is *no such thing* as "the location of the cat." In classical mechanics, it may happen that we don't know where Miss Kitty is, so we may end up saying things like "I think there's a 75 percent chance that she's under the table." But that's a statement about our ignorance, not about the world; there really is a fact of the matter about where the cat is, whether we know it or not.

In quantum mechanics, there is no fact of the matter about where Miss Kitty (or anything else) is located. The space of states in quantum mechanics just doesn't work that way. Instead, the states are specified by something called a *wave function*. And the wave function doesn't say things like "the cat is on the sofa" or "the cat is under the table." Rather, it says things like "if we were to look, there would be a 75 percent chance that we would find the cat under the table, and a 25 percent chance that we would find the cat on the sofa."

This distinction between "incomplete knowledge" and "intrinsic quantum indeterminacy" is worth dwelling on. If the wave function tells us there is a 75 percent chance of observing the cat under the table and a 25 percent chance of observing her on the sofa, that does not mean there is a 75 percent chance that the cat *is* under the table and a 25 percent chance that she *is* on the sofa. There is no such thing as "where the cat is." Her quantum state is described by a *superposition* of the two distinct possibilities we would have in classical mechanics. It's not even that "they are both true at once"; it's that there is no “true” place where the cat is. The wave function is the best description we have of the reality of the cat.

It’s clear why this is hard to accept at first blush. To put it bluntly, the world doesn’t look anything like that. We see cats and planets and even electrons in particular positions when we look at them, not in superpositions of different possibilities described by wave functions. But that’s the true magic of quantum mechanics: What we see is not what there is. The wave function really exists, but we don’t see it when we look; we see things as if they were in particular ordinary classical configurations.

None of which stops classical physics from being more than good enough to play basketball or put satellites in orbit. Quantum mechanics features a “classical limit” in which objects behave just as they would had Newton been right all along, and that limit includes all of our everyday experiences. For objects such as cats that are macroscopic in size, we never find them in superpositions of the form “75 percent here, 25 percent there”; it’s always “99.9999999 percent (or much more) here, 0.0000001 percent (or much less) there.” Classical mechanics is an approximation to how the macroscopic world operates, but a very good one. The real world runs by the rules of quantum mechanics, but classical mechanics is more than good enough to get us through everyday life. It’s only when we start to consider atoms and elementary particles that the full consequences of quantum mechanics simply can’t be avoided.

How wave functions work

You might wonder how we know any of this is true. What is the difference, after all, between “there is a 75 percent chance of observing the cat under the table” and “there is a 75 percent chance that the cat is under the table”? It seems hard to imagine an experiment that could distinguish between those possibilities—the only way we would know where it is would be to look for it, after all. But there is a crucially important phenomenon that drives home the difference, known as *quantum interference*. To understand what that means, we have to bite the bullet and dig a little more deeply into how wave functions really work.

In classical mechanics, where the state of a particle is a specification of its position and its momentum, we can think of that state as specified by a collection of numbers. For one particle in ordinary three-dimensional space, there are six numbers: the position in each of the three directions, and the momentum in each of the three directions. In quantum mechanics the state is specified by a wave function, which can also be thought of as a collection of numbers. The job of these numbers is to tell us, for any observation or measurement we could imagine doing, what the probability is that we will get a certain result. So you might naturally think that the numbers we need are just the probabilities themselves: the chance Miss Kitty will be observed on the sofa, the chance she will be observed under the table, and so on.

As it turns out, that's not how reality operates. Wave functions really are wavelike: A typical wave function oscillates through space and time, much like a wave on the surface of a pond. This isn't so obvious in our simple example where there are only two possible observational outcomes—"on sofa" or "under table"—but becomes more clear when we consider observations with continuous possible outcomes, like the position of a real cat in a real room. The wave function is like a wave on a pond, except it's a wave on the space of all possible outcomes of an observation—for example, all possible positions in a room.

When we see a wave on a pond, the level of the water isn't uniformly higher than what it would be if the pond were undisturbed; sometimes the water goes up, and sometimes it goes down. If we were to describe the wave mathematically, to every point on the pond we would associate an *amplitude*—the height by which the water was displaced—and that amplitude would sometimes be positive, sometimes be negative. Wave functions in quantum mechanics work the same way. To every possible outcome of an observation, the wave function assigns a number, which we call the amplitude, and which can be positive or negative. The complete wave function consists of a particular amplitude for every possible observational outcome; those are the numbers that specify the state in quantum mechanics, just as the positions and momenta specify the state in classical mechanics. There is an amplitude that Miss Kitty is under the table, and another one that she is on the sofa.

There's only one problem with that setup: What we care about are probabilities, and the probability of something happening is never a negative number. So it can’t be true that the amplitude associated with a certain observational outcome is equal to the probability of obtaining that outcome—instead, there must be a way of calculating the probability if we know what the amplitude is. Happily, the calculation is very easy! To get the probability, you take the amplitude and you square it.

(probability of observing *X*) = (amplitude assigned to *X*)^2.

So if Miss Kitty’s wave function assigns an amplitude of 0.5 to the possibility that we observe her on the sofa, the probability that we see her there is $(0.5)^2 = 0.25$, or 25 percent. But, crucially, the amplitude could also be -0.5, and we would get exactly the same answer: $(-0.5)^2 = 0.25$. This might seem like a pointless piece of redundancy—two different amplitudes corresponding to the same physical situation—but it turns out to play a key role in how states evolve in quantum mechanics.[iv]

Interference

Now that we know that wave functions can assign negative amplitudes to possible outcomes of observations, we can return to the question of why we ever need to talk about wave functions and superpositions in the first place, rather than just assigning probabilities to different outcomes directly. The reason is interference, and those negative numbers are crucial in understanding the way interference comes about—we can add two (nonvanishing) amplitudes together and get zero, which we couldn’t do if amplitudes were never negative.

To see how this works, let’s complicate our model of feline dynamics just a bit. Imagine that we see Miss Kitty leave the upstairs bedroom. From our previous observations of her wanderings through the house, we know quite a bit about how this quantum cat operates. We know that, once she settles in downstairs, she will inevitably end up either on the sofa or under the table, nowhere else. (That is, her final state is a wave function describing a superposition of being on the sofa and being under the table.) But let’s say we also know that she has two possible routes to take from the upstairs bed to whatever downstairs resting place she chooses: She will either stop by the food dish to eat, or stop by the scratching post to sharpen her claws. In the real world all of these possibilities are adequately described by classical mechanics, but in our idealized thought-experiment world we imagine that quantum effects play an important role.

Now let's see what we actually observe. We'll do the experiment two separate ways. First, when we see Miss Kitty start downstairs, we very quietly sneak behind her to see which route she takes, either via the food bowl or the scratching post. She actually has a wave function describing a superposition of both possibilities, but when we make an observation we always find a definite result. We're as quiet as possible, so we don't disturb her; if you like, we can imagine that we've placed spy cameras or laser sensors. The technology used to figure out whether she goes to the bowl or the scratching post is completely irrelevant; what matters is that we've observed it.

We find that we observe her visiting the bowl exactly half the time, and the scratching post exactly half the time. (We assume that she visits one or the other, but never both, just to keep things as simple as we can.) Any one particular observation doesn't reveal the wave function, of course; it can only tell us that we saw her stop at the post or at the bowl that particular time. But imagine that we do this experiment a very large number of times, so that we can get a reliable idea of what the probabilities are.

But we don't stop at that. We next let her continue on to either the sofa or the table, and after she's had time to settle down we look again to see which place she ends up. Again, we do the experiment enough times that we can figure out the probabilities. What we now find is that it didn't matter whether she stopped at the scratching post, or at her food bowl; in both cases, we observe her ending up on the sofa exactly half the time, and under the table exactly half the time, completely independently of whether she first visited the bowl or the scratching post. Apparently the intermediate step along the way didn't matter very much; no matter which alternative we observed along the way, the final wave function assigns equal probability to the sofa and the table.

Next comes the fun part. This time, we simply choose not to observe Miss Kitty's intermediate step along her journey; we don't keep track of whether she stops at the scratching post or the food bowl. We just wait until she's settled on the sofa or under the table, and we look at where she is, reconstructing the final probabilities assigned by the wave functions. What do we expect to find?

In a world governed by classical mechanics, we know what we should see. When we did our spying on her, we were careful that our observations shouldn't have affected how Miss Kitty behaved, and half the time we found her on the sofa and half the time on the table no matter what route she took. Clearly, even if we don't observe what she does along the way, it shouldn't matter—in either case we end up with equal probabilities for the final step, so even if we don't observe the intermediate stage we should still end up with equal probabilities.

But we don't. That's not what we see, in this idealized thought-experiment world where our cat is a truly quantum object. What we see when we choose not to observe whether she goes via the food bowl or the scratching post is that she ends up on the sofa 100 percent of the time! We never find her under the table—the final wave function assigns an amplitude of zero to that possible outcome. Apparently, if all this is to be believed, the very presence of our spy cameras changed her wave function in some dramatic way. The possibilities are summarized in the table.

What route we see Ms. Kitty take	**Final probabilities**
Scratching post	50% sofa, 50% table
Food bowl	50% sofa, 50% table
We don't look	100% sofa, 0% table

This isn't just a thought experiment; it's been done. Not with real cats, who are unmistakably macroscopic and well described by the classical limit; but with individual photons, in what is known as the "double slit experiment." A photon passes through two possible slits, and if we don't watch which slit it goes through, we get one final wave function; but if we do, we get a completely different one, no matter how unobtrusive our measurements were.

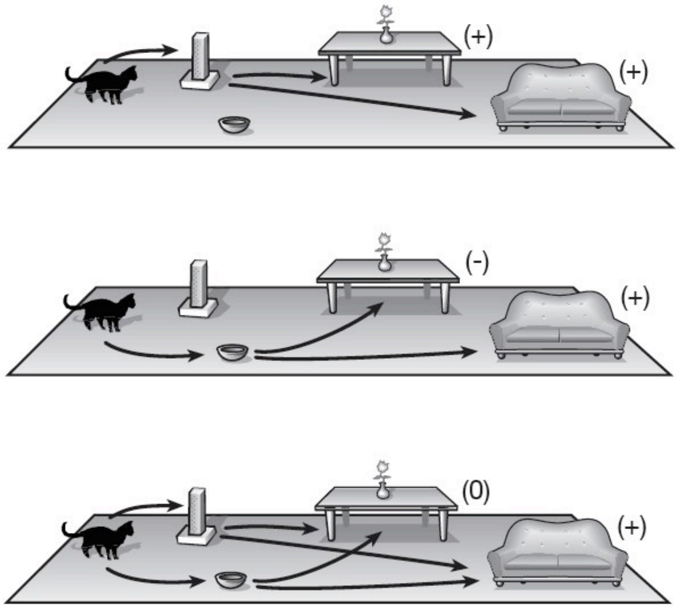

Figure 57: Alternative evolutions for Miss Kitty's wave function. At the top, we observe her stop at the scratching post, after which she could go to either the table or the sofa, both with positive amplitude. Middle, we observe her go to the food bowl, after which she could go to either the table or the sofa, but this time the table has a negative amplitude (still a positive probability). Bottom, we don't observe her intermediate journey, so we add the amplitudes from the two possibilities. We end up with zero amplitude for the table (since the positive and negative contributions cancel), and positive amplitude for the sofa.

Here's how to explain what is going on. Let's imagine that we do observe whether Miss Kitty stops by the bowl or the scratching post, and we see her stop by the scratching post. After she does so, she evolves into a superposition of being on the sofa and being under the table, with equal probability. In particular, due to details in Miss Kitty's initial condition and certain aspects of quantum feline dynamics, the final wave function assigns equal *positive* amplitudes to the sofa possibility and the table possibility. Now let's consider the other intermediate step, that we see her stop by the food bowl. In that case, the final wave function assigns a negative amplitude to the table, and a positive one to the sofa—equal but opposite numbers, so that the probabilities turn out precisely the same.[v]

But if we don't observe her at the scratching-post/food-bowl juncture, then (by the lights of our thought experiment) she is in a superposition of the two possibilities at this intermediate step. In that case, the rules of quantum mechanics instruct us to add the two possible contributions to the final wave function—one from the route where she stopped by the scratching post, and one from the food bowl. In either case, the amplitudes for ending up on the sofa were positive numbers, so they reinforce each other. But the amplitudes for ending up under the table were opposite for the two intermediate cases—so when we add them together, they precisely cancel. Individually, Miss Kitty's two possible intermediate paths left us with a nonzero probability that she would end up under the table; but when both paths were allowed (because we didn't observe which one she took), the two amplitudes interfered.

That's why the wave function needs to include negative numbers, and that's how we know that the wave function is "real," not just a bookkeeping device to keep track of probabilities. We have an explicit case where the individual probabilities would have been positive, but the final wave function received contributions from two intermediate steps, which ended up canceling each other.

Let's catch our breath to appreciate how profound this is, from our jaundiced classically-trained perspective. For any particular instantiation of the experiment, we are tempted to ask: Did Miss Kitty stop by the food bowl, or the scratching post? The only acceptable answer is: no. She didn't do either one. She was in a superposition of both possibilities, which we know because both possibilities ended up giving crucial contributions to the amplitude of the final answer.

Real cats are messy macroscopic objects consisting of very large numbers of molecules, and their wave functions tend to be very sharply concentrated around something resembling our classical notion of a "position in space." But at the microscopic level, all this talk of wave functions and superpositions and interference becomes brazenly demonstrable. Quantum mechanics strikes us as weird, but it's the way nature works.

Collapse of the wave function

The thing about this discussion that tends to rub people the wrong way—and with good reason—is the crucial role played by observation. When we observed what the cat was doing at the scratching-post/food-bowl juncture, we got one answer for the final state; when we didn't make any such observation, we got a very different answer. That's not how physics is supposed to work; the world is supposed to evolve according to the laws of nature, whether we are observing it or not. What counts as an "observation," anyway? What if we set up surveillance cameras along the way, but then never looked at the tapes? Would that count as an observation? (Yes, it would.) And what precisely happened when we did make our observation?

This is an important collection of questions, and the answers are not completely clear. There is no consensus within the physics community about what really constitutes an observation (or "measurement") in quantum mechanics, nor on what happens when an observation occurs. This is the "measurement problem," and is the primary focus of people who spend their time thinking about interpretations of quantum mechanics. There are many such interpretations on the market, and we're going to discuss two: the more or less standard picture, known as the "Copenhagen interpretation," and a view that seems (to me) to be a bit more respectable and likely to conform to reality, which goes under the forbidding name of the "many-worlds interpretation." Let's look at Copenhagen first.[vi]

The Copenhagen interpretation is so named because Niels Bohr, who in many ways was the godfather of quantum mechanics, helped to develop it from his institute in Copenhagen in the 1920s. The actual history of this perspective is complicated, and certainly involves a great deal of input from Werner Heisenberg, another quantum pioneer. But the history is less crucial to our present purposes than the status of the Copenhagen view as what is enshrined in textbooks as the standard picture. Every physicist learns this first, and then gets to contemplate other alternatives (or choose not to, as the case may be).

The Copenhagen interpretation of quantum mechanics is as easy to state as it is hard to swallow: when a quantum system is subjected to a measurement, its wave function *collapses*. That is, the wave function goes instantaneously from describing a superposition of various possible observational outcomes to a completely different wave function, one that assigns 100 percent probability to the outcome that was actually measured, and 0 percent to anything else. That kind of wave function, concentrated entirely on a single possible observational outcome, is known as an “eigenstate.” Once the system is in that eigenstate, you can keep making the same kind of observation, and you’ll keep getting the same answer (unless something kicks the system out of the eigenstate into another superposition). We can’t say with certainty which eigenstate the system will fall into when an observation is made; it’s an inherently stochastic process, and the best we can do is assign a probability to different outcomes.

We can apply this idea to the story of Miss Kitty. According to the Copenhagen interpretation, our choice to observe whether she stopped by the food bowl or the scratching post had a dramatic effect on her wave function, no matter how sneaky we were about it. When we didn’t look, she was in a superposition of the two possibilities, with equal amplitude; when she then moved on to the sofa or the table, we added up the contributions from each of the intermediate steps, and found there was interference. But if we chose to observe her along the way, we collapsed her wave function. If we saw her stop at the scratching post, once that observation was made she was in a state that was no longer a superposition—she was 100 percent scratching post, 0 percent food bowl. Likewise if we saw her stop at the food bowl, with the amplitudes reversed. In either case, there was nothing left to interfere with, and her wave function evolved into a state that gave her equal probabilities to end up on the sofa or under the table.[vii]

There is good news and bad news about this story. The good news is that it fits the data. If we imagine that wave functions collapse every time we make an observation—no matter how unobtrusive our observational strategy may be—and that they end up in eigenstates that assign 100 percent probability to the outcome we observed, we successfully account for all of the various quantum phenomena known to physicists.

The bad news is that this story barely makes sense. What counts as an “observation”? Can the cat herself make an observation, or could a nonliving being? Surely we don’t want to suggest that the phenomenon of *consciousness* is somehow playing a crucial role in the fundamental laws of physics? (No, we don’t.) And does the purported collapse really happen instantaneously, or is it gradual but just very fast?

Irreversibility

At heart, the thing that bugs us about the Copenhagen interpretation of quantum mechanics is that it treats "observing" as a completely distinct kind of natural phenomenon, one that requires a separate law of nature. In classical mechanics, everything that happens can be accounted for by systems evolving according to Newton's laws. But if we take the collapse of the wave function at face value, a system described by quantum mechanics evolves according to two completely separate kinds of rules:

1. When we're not looking at it, a wave function evolves smoothly and predictably. The role that Newton's laws play in classical mechanics is replaced by the *Schrödinger equation* in quantum mechanics, which operates in a precisely analogous way. Given the state of the system at any one time, we can use the Schrödinger equation to evolve it reliably into the future and into the past. The evolution conserves information, and is completely reversible.
2. When we observe it, a wave function collapses. The collapse is not smooth, nor perfectly predictable, and information is not conserved. The amplitude (squared) associated with any particular outcome tells us the probability that the wave function will collapse to a state that is concentrated entirely on that outcome. Two different wave functions can very easily collapse to exactly the same state after an observation is made; therefore, wave function collapse is not reversible.

Madness! But it works. The Copenhagen interpretation takes concepts that would seem to be nothing more than useful approximations to some deeper underlying truth—distinguishing between a "system" that is truly quantum-mechanical and an "observer" who is essentially classical—and imagines that these categories play a crucial role in the fundamental architecture of reality. Most physicists, even those who use quantum mechanics every day in their research, get along perfectly well speaking the language of the Copenhagen interpretation, and choosing not to worry about the puzzles it presents. Others, especially those who think carefully about the foundations of quantum mechanics, are convinced that we need to do better. Unfortunately there is no strong consensus at present about what that better understanding might be.

For many people, the breakdown of perfect predictability is a troubling feature of quantum mechanics. (Einstein among them; that's the origin of his complaint that "God does not play dice with the universe.") If the Copenhagen interpretation is right, there could be no such thing as Laplace's Demon in a quantum world; at least, not if that world contained observers. The act of observing introduces a truly random element into the evolution of the world. Not *completely* random—a wave function may give a very high probability to observing one thing, and a very low probability to observing something else. But *irreducibly* random, in the sense that there is no piece of missing information that would allow us to predict outcomes with certainty, if only we could get our hands on it.[viii] Part of the glory of classical mechanics had been its clockwork reliability—even if Laplace's Demon didn't really exist, we knew he could exist in principle. Quantum mechanics destroys that hope. It took a long while for people to get used to the idea that probability enters the laws of physics in some fundamental way, and many are still discomforted by the concept.

One of our questions about the arrow of time is how we can reconcile the irreversibility of macroscopic systems described by statistical mechanics with the apparent reversibility of the microscopic laws of physics. But now, according to quantum mechanics, it seems that the microscopic laws of physics aren't necessarily reversible. The collapse of the wave function is a process that introduces an intrinsic arrow of time into the laws of physics: Wave functions collapse, but they don't un-collapse. If we observe Miss Kitty and see that she is on the sofa, we know that she is an eigenstate (100 percent on the sofa) right after we've done the measurement. But we don't know what state she was in *before* we did the measurement. That information, apparently, has been destroyed. All we know is that the wave function must have had some nonzero amplitude for the cat to be on the sofa—but we don't know how much, or what the amplitude for any other possibilities might have been.

So the collapse of the wave function—if, indeed, that's the right way to think about quantum mechanics—defines an intrinsic arrow of time. Can that be used to somehow explain "the" arrow of time, the thermodynamic arrow that appears in the Second Law and on which we've blamed all the various macroscopic differences between past and future?

Probably not. Although irreversibility is a key feature of the arrow of time, not all irreversibilities are created equal. It's very hard to see how the fact that wave functions collapse could, by itself, account for the Past Hypothesis. Remember, it's not hard to understand why entropy increases; what's hard to understand is why it was ever low to begin with. The collapse of the wave function doesn't seem to offer any direct help with that problem.

On the other hand, quantum mechanics is very likely to play some sort of role in the ultimate explanation, even if the intrinsic irreversibility of wave function collapse doesn't directly solve the problem all by itself. After all, we believe that the laws of physics are fundamentally quantum-mechanical at heart. It's quantum mechanics that sets the rules, and tells us what is and is not allowed in the world. It's perfectly natural to expect that these rules will come into play when we finally do begin to understand why our universe had such a low entropy near the Big Bang. We don't know exactly where this journey is taking us, but we're savvy enough to anticipate that certain tools will prove useful along the way.

Uncertainty

Our discussion of wave functions has glossed over one crucial property. We've said that wave functions assign an amplitude to any possible outcome of an observation we could imagine doing. In our thought experiment, we restricted ourselves to only one kind of observation—the location of the cat—and only two possible outcomes at a time. A real cat, or an elementary particle or an egg or any other object, has an infinite number of possible positions, and the relevant wave function in each case assigns an amplitude to every single possibility.

More important, however, there are things we can observe other than positions. Remembering our experience with classical mechanics, we might imagine observing the momentum rather than the position of our cat. And that's perfectly possible; the state of the cat is described by a wave function that assigns an amplitude to every possible momentum we could imagine measuring. When we do such a measurement and get an answer, the wave function collapses into an “eigenstate of momentum,” where the new state assigns nonzero amplitude only to the particular momentum we actually observed.

But if that’s true, you might think, what’s to stop us from putting the cat into a state where both the position and momentum are determined exactly, so it becomes just like a classical state? In other words, why can’t we take a cat with an arbitrary wave function, observe its position so that it collapses to one definite value, and then observe its momentum so that it also collapses to a definite value? We should be left with something that is completely determined, no uncertainty at all.

The answer is that there are no wave functions that are simultaneously concentrated on a single value of position and also on a single value of momentum. Indeed, the hope for such a state turns out to be maximally frustrated: If the wave function is concentrated on a single value of position, the amplitudes for different momenta will be spread out as widely as possible over all the possibilities. And vice versa: If the wave function is concentrated on a single momentum, it is spread out widely over all possible positions. So if we observe the position of an object, we lose any knowledge of what its momentum is, and vice versa.[ix] (If we only measure the position approximately, rather than exactly, we can retain some knowledge of the momentum; this is what actually happens in real-world macroscopic measurements.)

That’s the true meaning of the Heisenberg Uncertainty Principle. In quantum mechanics, it is possible to “know exactly” what the position of a particle is—more precisely, it’s possible for the particle to be in a position eigenstate, where there is a 100 percent probability of finding it in a certain position. Likewise, it is possible to “know exactly” what the momentum is. But we can never know precisely the position and momentum at the same time. So when we go to measure the properties that classical mechanics would attribute to a system—both position and momentum—we can never say for certain what the outcomes will be. That’s the uncertainty principle.

The uncertainty principle implies that there must be some spread of the wave function over different possible values of either position or momentum, or (usually) both. No matter what kind of system we look at, there is an unavoidable quantum unpredictability when we try to measure its properties. The two observables are complementary: When the wave function is concentrated in position, it’s spread out in momentum, and vice versa. Real macroscopic systems that are well-described by the classical limit of quantum mechanics find themselves in compromise states, where there is a small amount of uncertainty in both position and momentum. For large enough systems, the uncertainty is relatively small enough that we don't notice at all.

Keep in mind that there really is no such thing as "the position of the object" or "the momentum of the object"—there is only a wave function assigning amplitudes to the possible outcomes of observations. Nevertheless, we often can't resist falling into the language of *quantum fluctuations*—we say that we can't pin the object down to a single position, because the uncertainty principle forces it to fluctuate around just a bit. That's an irresistible linguistic formulation, and we won't be so uptight that we completely refrain from using it, but it doesn't accurately reflect what is really going on. It's not that there is a position and a momentum, each of which keeps fluctuating around; it's that there is a wave function, which can't simultaneously be localized in position and momentum.

In later chapters we will explore applications of quantum mechanics to much grander systems than single particles, or even single cats—quantum field theory, and also quantum gravity. But the basic framework of quantum mechanics remains the same in each case. Quantum field theory is the marriage of quantum mechanics with special relativity, and explains the particles we see around us as the observable features of the deeper underlying structure—quantum fields—that make up the world. The uncertainty principle will forbid us from precisely determining the position and momentum of every particle, or even the exact number of particles. That's the origin of "virtual particles," which pop in and out of existence even in empty space, and ultimately it will lead to the phenomenon of Hawking radiation from black holes.

One thing we don't understand is quantum gravity. General relativity provides an extremely successful description of gravity as we see it operate in the world, but the theory is built on a thoroughly classical foundation. Gravity is the curvature of spacetime, and in principle we can measure the spacetime curvature as precisely as we like. Almost everyone believes that this is just an approximation to a more complete theory of quantum gravity, where spacetime itself is described by a wave function that assigns different amplitudes to different amounts of curvature. It might even be the case that entire universes pop in and out of existence, just like virtual particles. But the quest to construct a complete theory of quantum gravity faces formidable hurdles, both technical and philosophical. Overcoming those obstacles is the full-time occupation of a large number of working physicists.

The wave function of the universe

There is one fairly direct way of addressing the conceptual issues associated with wave function collapse: Simply deny that it ever happens, and insist that ordinary smooth evolution of the wave function suffices to explain everything we know about the world. This approach—brutal in its simplicity, and far-reaching in its consequences—goes under the name of the "many-worlds interpretation" of quantum mechanics, and is the leading competitor to the Copenhagen interpretation. To understand how it works, we need to take a detour into perhaps the most profound feature of quantum mechanics of all: entanglement.

When we introduced the idea of a wave function we considered a very minimalist physical system, consisting of a single object (a cat). We would obviously like to be able to move beyond that, to consider systems with multiple parts—perhaps a cat and also a dog. In classical mechanics, that's no problem; if the state of one object is described by its position and its momentum, the state of two objects is just the state of both objects individually—two positions and two momenta. So it would be the most natural thing in the world to guess that the correct quantum-mechanical description of a cat and a dog would simply be two wave functions, one for the cat and one for the dog.

That's not how it works. In quantum mechanics, no matter how many individual pieces make up the system you are thinking about, there is *only one wave function*. Even if we consider the entire universe and everything inside it, there is still only one wave function, sometimes redundantly known as the "wave function of the universe." People don't always like to talk that way, for fear of sounding excessively grandiose, but at bottom thats simply the way quantum mechanics works. (Other people enjoy the grandiosity for its own sake.)

Let's see how this plays out when our system consists of a cat and a dog, Miss Kitty and Mr. Dog. As before, we imagine that when we look for Miss Kitty, there are only two places we can find her: on the sofa or under the table. Let's also imagine that there are only two places we can ever observe Mr. Dog: in the living room or out in the yard. According to the initial (but wrong) guess that each object has its own wave function, we would describe Miss Kitty's location as a superposition of under the table and on the sofa, and separately describe Mr. Dog's location as a superposition of in the living room or in the yard.

But instead, quantum mechanics instructs us to consider every possible alternative for the entire system—cat plus dog—and assign an amplitude to every distinct possibility. For the combined system, there are four possible answers to the question "What do we see when we look for the cat and the dog?" They can be summarized as follows:

(table, living room)

(table, yard)

(sofa, living room)

(sofa, yard)

Here, the first entry tells us where we see Miss Kitty, and the second where we see Mr. Dog. According to quantum mechanics, the wave function of the universe assigns every one of these four possibilities a distinct amplitude, which we would square to get the probability of observing that alternative.

You may wonder what the difference is between assigning amplitudes to the locations of the cat and dog separately, and assigning them to the combined locations. The answer is *entanglement*—properties of any one subset of the whole can be strongly correlated with properties of other subsets.

Entanglement

Let's imagine that the wave function of the cat/dog system assigns zero amplitude to the possibility (table, yard), and also zero amplitude to (sofa, living room). Schematically, that means the state of the system must be of the form

(table, living room) + (sofa, yard).

This means there is a nonzero amplitude that the cat is under the table and the dog is in the living room, and also a nonzero amplitude that the cat is on the sofa and the dog is in the yard. Those are the only two possibilities allowed by this particular state, and let's imagine that they have equal amplitude.

Now let's ask: What do we expect to see if we only look for Miss Kitty? An observation collapses the wave function onto one of the two possibilities, (table, living room) or (sofa, yard), with equal probability, 50 percent each. If we simply don't care about what Mr. Dog is doing, we would say that there is an equal probability for observing Miss Kitty under the table or on the sofa. In that sense, it's fair to say that we have no idea where Miss Kitty is going to be before we look.

Now let's imagine that we instead look for Mr. Dog. Again, there is a 50 percent chance each for the possibilities (table, living room) or (sofa, yard), so if we don't care what Miss Kitty is doing, it's fair to say that we have no idea where Mr. Dog is going to be before we look.

Here is the kicker: Even though we have no idea where Mr. Dog is going to be before we look, if we first choose to look for Miss Kitty, once that observation is complete we know exactly where Mr. Dog is going to be, even without ever looking for him! That's the magic of entanglement. Let's say that we saw Miss Kitty on the sofa. That means that, given the form of the wave function we started with, it must have collapsed onto the possibility (sofa, yard). We therefore know with certainty (assuming we were right about the initial wave function) that Mr. Dog will be in the yard if we look for him. We have collapsed Mr. Dog's wave function without ever observing him. Or, more correctly, we have collapsed the wave function of the universe, which has important consequences for Mr. Dog's whereabouts, without ever interacting with Mr. Dog directly.

This may or may not seem surprising to you. Hopefully, we've been so clear and persuasive in explaining what wave functions are all about, that the phenomenon of entanglement seems relatively natural. And it should; it's part and parcel of the machinery of quantum mechanics, and a number of clever experiments have demonstrated its validity in the real world. Nevertheless, entanglement can lead to consequences that—taken at face value—seem inconsistent with the spirit of relativity, if not precisely with the letter of the law. Let's stress: There is no real incompatibility between quantum mechanics and special relativity (general relativity, where gravity comes into the game, is a different story). But there is a tension between them that makes people nervous. In particular, things seem to happen faster than the speed of light. When you dig deeply into what those "things" are, and what it means to "happen," you find that nothing really bad is going on—nothing has actually moved faster than light, and no real information can be conveyed outside anyone's light cone. Still, it rubs people the wrong way.

The EPR paradox

Let's go back to our cat and dog, and imagine that they are in the quantum state described above, a superposition of (table, living room) and (sofa, yard). But now let's imagine that if Mr. Dog is out in the yard, he doesn't just sit there, he runs away. Also, he is very adventurous, and lives in the future, when we have regular rocket flights to a space colony on Mars. Mr. Dog—in the alternative where he starts in the yard, not in the living room—runs away to the spaceport, stows away on a rocket, and flies to Mars, completely unobserved the entire time. It's only when he clambers out of the rocket into the arms of his old friend Billy, who had graduated from high school and joined the Space Corps and been sent on a mission to the Red Planet, that the state of Mr. Dog is actually observed, collapsing the wave function.

What we're imagining, in other words, is that the wave function describing the cat/dog system has evolved smoothly according to the Schrödinger equation from

(table, living room) + (sofa, yard)

to

(table, living room) + (sofa, Mars).

There's nothing impossible about that—implausible, maybe, but as long as nobody made any observations during the time it took the evolution to happen, we'll end up with the wave function in this superposition.

But the implications are somewhat surprising. When Billy unexpectedly sees Mr. Dog bounding out of the spaceship on Mars, he makes an observation and collapses the wave function. If he knew what the wave function was to begin with, featuring an entangled state of cat and dog, Billy *immediately* knows that Miss Kitty is on the sofa, not under the table. The wave function has collapsed to the possibility (sofa, Mars). Not only is Miss Kitty's state now known even without anyone interacting with her, it seems to have been determined instantaneously, despite the fact that it takes at least several minutes to travel between Mars and Earth even if you were moving at the speed of light.

This feature of entanglement—the fact that the state of the universe, as described by its quantum wave function, seems to change "instantaneously" throughout space, even though the lesson of special relativity was supposed to be that there's no unique definition of what "instantaneously" means—bugs the heck out of people. It certainly bugged Albert Einstein, who teamed up with Boris Podolsky and Nathan Rosen in 1935 to write a paper pointing out this weird possibility, now known as the "EPR paradox."[x] But it's not really a "paradox" at all; it might fly in the face of our intuition, but not of any experimental or theoretical requirements.

The important feature of the apparently instantaneous collapse of a wave function that is spread across immense distances is that it cannot be used to actually transmit any information faster than light. The thing that bothers us is that, before Billy observed the dog, Miss Kitty back here on Earth was not in any definite location—we had a 50/50 chance to observe her on the sofa or under the table. Once Billy observes Mr. Dog, we now have a 100 percent chance of observing her to be on the sofa. But so what? We don't actually know that Billy did any such observation—for all we know, if we looked for Mr. Dog we would find him in the living room. For Billy's discovery to make any difference to us, he would have to come tell us about it, or send us a radio transmission—one way or another, he would have to communicate with us by conventional slower-than-light means.

Entanglement between two far-apart subsystems seems mysterious to us, because it violates our intuitive notions of "locality"—things should only be able to directly affect nearby things, not things arbitrarily far away. Wave functions just don't work like that; there is one wave function that describes the entire universe all at once, and that's the end of it. The world we observe, meanwhile, still respects a kind of locality—even if wave functions collapse instantaneously all over space, we can't actually take advantage of that feature to send signals faster than light. In other words: As far as things actually bumping into you and affecting your life, it's still true that they have to be right next to you, not arbitrarily far away.

On the other hand, we shouldn't expect that even this weaker notion of locality is truly a sacred principle. In the next chapter we'll talk a little bit about quantum gravity, where the wave function applies to different configurations of spacetime itself. In that context, an idea like "objects can only affect each other when they are nearby" ceases to have any absolute meaning. Spacetime itself is not absolute, but only has different amplitudes for being in different configurations—so the notion of "the distance between two objects" becomes a little fuzzy. These are ideas that have yet to be fully understood, but the final theory of everything is likely to exhibit non-locality in some very dramatic ways.

Many worlds, many minds

The leading contender for an alternative to the Copenhagen view of quantum mechanics is the so-called *many-worlds interpretation*. "Many worlds" is a scary and misleading name for what is really a very straightforward idea. That idea is this: There is no such thing as "collapse of the wave function." The evolution of states in quantum mechanics works just like it does in classical mechanics; it obeys a deterministic rule—the Schrödinger equation—that allows us to predict the future and past of any specific state with perfect fidelity. And that's all there is to it.

The problem with this claim is that we appear to *see* wave functions collapsing all the time, or at least to observe the effects of the collapse. We can imagine arranging Miss Kitty in a quantum state that has equal amplitudes for finding her on the sofa or under the table; then we look for her, and see her under the table. If we look again immediately thereafter, we're going to see her under the table 100 percent of the time; the original observation (in the usual way of talking about these things) collapsed the wave function to a table-eigenstate. And that way of thinking has empirical consequences, all of which have been successfully tested in real experiments.

The response of the many-worlds advocate is simply that you are thinking about it wrong. In particular, you have misidentified *yourself* in the wave function of the universe. After all, you are part of the physical world, and therefore you are also subject to the rules of quantum mechanics. It's not right to set yourself off as some objective classical observing apparatus; we need to take your own state into account in the wave function.

So, this new story goes, we shouldn't just start with a wave function describing Miss Kitty as a superposition of (sofa) and (table); we should include your own configuration in the description. In particular, the relevant feature of your description is what you have observed about Miss Kitty's position. There are three possible states you could be in: You could have seen her on the sofa, you could have seen her under the table, and you might not have looked yet. To start with, the wave function of the universe (or at least the bit of it we're describing here) gives Miss Kitty equal amplitude to be on the sofa or under the table, while you are uniquely in the state of not having looked yet. This can be schematically portrayed like this:

(sofa, you haven't yet looked) + (table, you haven't yet looked).

Now you observe where she is. In the Copenhagen interpretation, we would say that the wave function collapses. But in the many-worlds interpretation, we say that your own state becomes entangled with that of Miss Kitty, and the combined system evolves into a superposition:

(sofa, you see her on the sofa) + (table, you see her under the table).

There is no collapse; the wave function evolves smoothly, and there is nothing special about the process of "observation." What is more, the entire procedure is reversible—given the final state, we could use the Schrödinger equation to uniquely recover the original state. There is no intrinsically quantum-mechanical arrow of time in this interpretation. For many reasons, this is an altogether more elegant and satisfying picture of the world than that provided by the Copenhagen picture.

The problem, meanwhile, should be obvious: The final state has you in a superposition of two different outcomes! The difficulty with that, of course, is that you never *feel* like you're in such a superposition. If you actually did make an observation of a system that was in a quantum superposition, after the observation you would always *believe* that you had observed some specific outcome. The problem with the many-worlds interpretation, in other words, is that it doesn't seem to accord with our experience of the real world.

But let's not be too hasty. Who is this "you" of which we are speaking? It's true: The many-worlds interpretation says that the wave function of the universe evolves into the superposition shown above, with an amplitude for you seeing the cat on the sofa, and another amplitude for you seeing her under the table. Here is the crucial step: The "you" that does the seeing and perceiving and believing is not that superposition. Rather, "you" are either one of those alternatives, or the other. That is, there are now two different "yous," one who saw Ms. Kitty on the sofa and another who saw her under the table, and they both honestly *exist* there in the wave function. They share the same prior memories and experiences—before they observed the cat's location, they were in all respects the same person—but now they have split off into two different "branches of the wave function," never to interact with each other again.

These are the "many worlds" in question, although it should be clear that the label is somewhat misleading. People sometimes raise the objection to the many-worlds interpretation that it's simply too extravagant to be taken seriously—all those different "parallel realities," infinite in number, just so that we don't have to believe in wave function collapse. That's silly. Before we made an observation, the universe was described by a single wave function, which assigned a particular amplitude to every possible observational outcome; after the observation, the universe is described by a single wave function, which assigns a particular amplitude to every possible observational outcome. Before and after, the wave function of the universe is just a particular point in the space of states describing the universe, and that space of states didn't get any bigger or smaller. No new "worlds" have really been created; the wave function still contains the same amount of information (after all, in this interpretation its evolution is reversible). It has simply evolved in such a way that there are now a greater number of distinct subsets of the wave function describing individual conscious beings such as ourselves. The many-worlds interpretation of quantum mechanics may or may not be right; but to object to it on the grounds that “Gee, that’s a lot of worlds,” is wrong-headed.

The many-worlds interpretation was not originally formulated by Bohr, Heisenberg, Schrödinger, or any of the other towering figures of the early days of quantum mechanics. It was proposed in 1957 by Hugh Everett III, who was a graduate student working with John Wheeler at Princeton.[xi] At the time (and for decades thereafter), the dominant view was the Copenhagen interpretation, so Wheeler did the obvious thing: He sent Everett on a trip to Copenhagen, to discuss his novel perspective with Niels Bohr and others. But the trip was not a success—Bohr was utterly unconvinced, and the rest of the physics community exhibited little interest in Everett’s ideas. He left academic physics to work for the Defense Department, and eventually founded his own computer firm. In 1970, theoretical physicist Bryce DeWitt (who, along with Wheeler, was a pioneer in applying quantum mechanics to gravity) took up the cause of the many-worlds interpretation, and helped popularize it among physicists. Everett lived to see a resurgence of interest in his ideas within the physics community, but he never returned to active research; he passed away suddenly of a heart attack in 1982, at the age of fifty-one.

Decoherence

Despite its advantages, the many-worlds interpretation of quantum mechanics isn’t really a finished product. There remain unanswered questions, from the deep and conceptual—why are conscious observers identified with discrete branches of the wave function, rather than superpositions?—to the dryly technical—how do we justify the rule that “probabilities are equal to amplitudes squared” in this formalism? These are real questions, to which the answers aren’t perfectly clear, which is (one reason) why the many-worlds interpretation doesn’t enjoy universal acceptance. But a great deal of progress has been made over the last few decades, especially involving an intrinsically quantum-mechanical phenomenon known as *decoherence*. There are great hopes—

although little consensus—that decoherence can help us understand why wave functions *appear* to collapse, even if the many-worlds interpretation holds that such collapse is only apparent.

Decoherence occurs when the state of some small piece of the universe—your brain, for example—becomes so entangled with parts in the wider environment that it is no longer subject to interference, the phenomenon that truly makes something "quantum." To get a feeling for how this works, let's go back to the example of the entangled state of Miss Kitty and Mr. Dog. There are two alternatives, with equal amplitudes: the cat is under the table and the dog is in the living room, or the cat is on the sofa and the dog is in the yard:

(table, living room) + (sofa, yard).

We saw how, if someone observed the state of Mr. Dog, the wave function would (in the Copenhagen language) collapse, leaving Miss Kitty in some definite state.

But now let's do something different: Imagine that nobody observes the state of Mr. Dog, but we simply ignore him. Effectively, we throw away any information about the entanglement between Miss Kitty and Mr. Dog, and simply ask ourselves: What is the state of Miss Kitty all by herself?

We might think that the answer is a superposition of the form (table)+(sofa), like we had before we had ever introduced the canine complication into the picture. But that's not quite right. The problem is that interference—the phenomenon that convinced us we needed to take quantum amplitudes seriously in the first place—can no longer happen.

In our original example of interference, there were two contributions to the amplitude for Miss Kitty to be under the table: one from the alternative where she passed by her food bowl, and one from where she stopped at her scratching post. But it was crucially important that the two contributions that ultimately canceled were contributions to *exactly the same final alternative* ("Miss Kitty is under the table"). Two contributions to the final wave function are only going to interfere if they involve truly the same alternative for everything in the universe; if they are contributing to different alternatives, they can't possibly interfere, even if the differences involve the rest of the universe, and not Miss Kitty herself.

So when the state of Miss Kitty is entangled with the state of Mr. Dog, interference between alternatives that alter Miss Kitty's state without a corresponding change in Mr. Dog's becomes impossible. Some contribution to the wave function can't interfere with the alternative "Miss Kitty is under the table," because that alternative isn't a complete specification of what can be observed; it could only interfere with the alternatives "Miss Kitty is under the table and Mr. Dog is in the living room" that is actually represented in the wave function.[xii]

Therefore, if Miss Kitty is entangled with the outside world but we don't know the details of that entanglement, it's not right to think of her state as a quantum superposition. Rather, we should just think of it as an ordinary *classical* distribution of different alternatives. Once we throw away any information about what she is entangled with, Miss Kitty is no longer in a true superposition; as far as any conceivable experiment is concerned, she is in either one state or the other, even if we don't know which. Interference is no longer possible.

That's decoherence. In classical mechanics, every object has a definite position, even if we don't know what the position is and can only ascribe probabilities to the various alternatives. The miracle of quantum mechanics was that there is no longer any such thing as "where the object is"; it's in a true simultaneous superposition of the possible alternatives, which we know must be true via experiments that demonstrate the reality of interference. But if the quantum state describing the object is entangled with something in the outside world, interference becomes impossible, and we're back to the traditional classical way of looking at things. As far as we are concerned, the object is in one state or another, even if the best we can do is assign a probability to the different alternatives—the probabilities are expressing our ignorance, not the underlying reality. If the quantum state of some particular subset of the universe represents a true superposition that is un-entangled with the rest of the world, we say it is "coherent"; if the superposition has been ruined by becoming entangled with something outside, we say that it has become "decoherent." (That's why, in the many-worlds view, setting up surveillance cameras counts as making an observation; the state of the cat became entangled with the state of the cameras.)

Wave function collapse and the arrow of time

In the many-worlds interpretation, decoherence clearly plays a crucial role in the apparent process of wave function collapse. The point is not that there is something special or unique about “consciousness” or “observers,” other than the fact that they are complicated macroscopic objects. The point is that any complicated macroscopic object is *inevitably* going to be interacting (and therefore entangled) with the outside world, and it’s hopeless to imagine keeping track of the precise form of that entanglement. For a tiny microscopic system such as an individual electron, we can isolate it and put it into a true quantum superposition that is not entangled with the state of any other particles; but for a messy system such as a human being (or a secret surveillance camera, for that matter) that’s just not possible.

In that case, our simple picture in which the state of our perceptions becomes entangled with the state of Miss Kitty’s location is an oversimplification. A crucial part of the story is played by the entanglement of us with the external world. Let’s imagine that Miss Kitty starts out in a true quantum superposition, un-entangled with the rest of the world; but we, complicated creatures that we are, are deeply entangled with the outside world in ways we can’t possibly specify. The wave function of the universe assigns distinct amplitudes to all the alternative configurations of the combined system of Miss Kitty, us, and the outside world. After we observe Miss Kitty’s location, the wave function evolves into something of the form

> (sofa, you see her on the sofa, world$_1$) + (table, you see her under the table, world$_2$),

where the last piece describes the (unknown) configuration of the external world, which will be different in the two cases.

Because we don’t know anything about that state, we simply ignore the entanglement with the outside world, and keep the knowledge of Miss Kitty’s location and our own mental perceptions. Those are clearly correlated: If she is on the sofa, we believe we have seen her on the sofa, etc. But after throwing away the configuration of the outside world, we’re no longer in a real quantum superposition. Rather, there are two alternatives that seem for all intents and purposes classical: Miss Kitty is on the sofa and we saw her on the sofa, or she's under the table and we saw her under the table.

That's what we mean when we talk about the branching of the wave function into different "worlds." Some small system in a true quantum superposition is observed by a macroscopic measuring apparatus, but the apparatus is entangled with the outside world; we ignore the state of the outside world, and are left with two classical alternative worlds. From the point of view of either classical alternative, the wave function has "collapsed"; but from a hypothetical larger point of view where we kept all of the information in the wave function of the universe, there were no sudden changes in the state, just a smooth evolution according to the Schrödinger equation.

This business about throwing away information may make you a little uneasy, but it should also sound somewhat familiar. All we're really doing is coarse-graining, just as we did in (classical) statistical mechanics to define macrostates corresponding to various microstates. The information about our entanglement with the messy external environment is analogous to the information about the position and momentum of every molecule in a box of gas—we don't need it, and in practice can't keep track of it, so we create a phenomenological description based solely on macroscopic variables.

In that sense, the irreversibility that crops up when wave functions collapse appears to be directly analogous to the irreversibility of ordinary thermodynamics. The underlying laws are perfectly reversible, but in the messy real world we throw away a lot of information, and as a result we find apparently irreversible behavior on macroscopic scales. When we observe our cat's location, and our own state becomes entangled with hers, in order to reverse the process we would need to know the precise state of the outside world with which we are also entangled, but we've thrown that information away. It's exactly analogous to what happens when a spoonful of milk mixes into a cup of coffee; in principle we could reverse the process if we had kept track of the position and momentum of every single molecule in the mixture, but in practice we only keep track of the macroscopic variables, so reversibility is lost.

In this discussion of decoherence, a crucial role was played by our ability to take the system to be observed (Miss Kitty, or some elementary particle) and isolate it from the rest of the world in a true quantum superposition. But that's clearly a very special kind of state, much like the low-entropy states we start with by hypothesis when discussing the origin of the Second Law of Thermodynamics. A completely generic state would feature all kinds of entanglements between our small system and the external environment, right from the start.

None of which is intended to give the impression that the application of decoherence to the many-worlds interpretation manages to swiftly solve all of the interpretive problems of quantum mechanics. But it seems like a step in the right direction, and highlights an important relationship between the macroscopic arrow of time familiar from statistical mechanics and the macroscopic arrow of time exhibited when wave functions collapse. Perhaps best of all, it helps remove ill-defined notions such as "conscious observers" from the vocabulary with which we describe the natural world.

With that in mind, we're going to go back to speaking as if the fundamental laws of physics are all completely reversible on microscopic scales. This isn't an unassailable conclusion, but it has some good arguments behind it—we can keep an open mind, while continuing to explore the consequences of this particular point of view. Which leaves us, of course, right where we started: with the task of explaining the apparent lack of reversibility on macroscopic scales in terms of special conditions near the Big Bang. To take that problem seriously, it's time that we start thinking about gravity and the evolution of the universe.

---

[i] Quoted in von Bayer (2003), pp. 12–-13.

[ii] This is not to say that the ancient Buddhists weren't wise; but their wisdom was not based on the failure of classical determinism at atomic scales, nor did they anticipate modern physics in any meaningful way, other than the inevitable random similarities of word choice when talking about grand cosmic concepts. (I once heard a lecture claming that the basic ideas of primordial nucleosynthesis were prefigured in the Torah; if you stretch your definitions enough, eerie similarities are everywhere.) It is disrespectful to both ancient philosophers and modern physicists to ignore the real differences in their goals and methods in an attempt to create tangible connections out of superficial resemblances.

[iii] More recently, dogs have also been recruited for the cause. See Orzel (2009).

[iv] We're still glossing over one technicality—the truth is actually one step more complex (as it were) than this description would have you believe, but it's not a complication that is necessary

---

for our present purposes. Quantum amplitudes are really *complex numbers*, which means they are combinations of two numbers: a real number, plus an imaginary number. (Imaginary numbers are what you get when you take the square root of a negative real number; so “imaginary two” is the square root of minus four, and so on.) A complex number looks like $a + bi$, where a and b are real numbers and “i” is the square root of minus one. If the amplitude associated with a certain option is $a + bi$, the probability it corresponds to is simply $a^2 + b^2$, which is guaranteed to be greater than or equal to zero. You will have to trust me that this extra apparatus is extremely important to the workings of quantum mechanics—either that, or start learning some of the mathematical details of the theory. (I can think of less rewarding ways of spending your time, actually.)

[v] The fact that any particular sequence of events assigns positive or negative amplitudes to the two final possibilities is an assumption we are making for the purposes of our thought experiment, not a deep feature of the rules of quantum mechanics. In any real-world problem, details of the system being considered will determine what precisely the amplitudes are, but we’re not getting our hands quite that dirty at the moment. Note also that the particular amplitudes in these examples take on the numerical values of plus or minus 0.7071—that’s the number which, when squared, gives you 0.5.

[vi] At a workshop attended by expert researchers in quantum mechanics in 1997, Max Tegmark took an admittedly highly-unscientific poll of the participants’ favored interpretation of quantum mechanics (Tegmark, 1998). The Copenhagen interpretation came in first with thirteen votes, while the many-worlds interpretation came in second with eight. Another nine votes were scattered among other alternatives. Most interesting, eighteen votes were cast for “none of the above/undecided.” And these are the experts.

[vii] So what does happen if we hook up a surveillance camera, but then don’t examine the tapes? It doesn’t matter whether we look at the tapes or not, the camera still counts as an observation, so there will be a chance to observe Ms. Kitty under the table. In the Copenhagen interpretation, we would say, “The camera is a classical measuring device whose influence collapses the wave function.” In the many-worlds interpretation, as we’ll see, the explanation is “the wave function of the camera becomes entangled with the wave function of the cat, so the alternative histories decohere.”

[viii] Many people have thought about changing the rules of quantum mechanics so that this is no longer the case; they have proposed what are called “hidden variable theories” that go beyond the standard quantum-mechanical framework. In 1964, theoretical physicist John Bell proved a remarkable theorem: No local theory of hidden variables can possibly reproduce the predictions

---

of quantum mechanics. This hasn't stopped people from investigating nonlocal theories—ones where distant events can affect each other instantaneously. But they haven't really caught on; the vast majority of modern physicists believe that quantum mechanics is simply correct, even if we don't yet know how to interpret it.

[ix] There is a slightly more powerful statement we can actually make. In classical mechanics, the state is specified by both position and velocity, so you might guess that the quantum wave function assigns probabilities to every possible combination of position and velocity. But that's not how it works. If you specify the amplitude for every possible position, you are done—you've completely determined the entire quantum state. So what happened to the velocity? It turns out that you can write the same wave function in terms of an amplitude for every possible velocity, completely leaving position out of the description. These are not two different states; they are just two different ways of writing exactly the same state. Indeed, there is a cookbook recipe for translating between the two choices, known in the trade as a "Fourier transform." Given the amplitude for every possible position, you can do a Fourier transform to determine the amplitude for any possible velocity, and vice versa. In particular, if the wave function is an eigenstate, concentrated on one precise value of position (or velocity), its Fourier transform will be completely spread out over all possible velocities (or positions).

[x] Einstein, Podolsky, and Rosen (1935).

[xi] Everett (1957). For discussion from various viewpoints, see Deutsch (1997), Albert (1992), or Ouellette (2007).

[xii] Note how crucial entanglement is to this story. If there were no entanglement, the outside world would still exist, but the alternatives available to Miss Kitty would be completely independent of what was going on out there. In that case, it would be perfectly okay to attribute a wave function to Miss Kitty all by herself. And thank goodness; that's the only reason we are able to apply the formalism of quantum mechanics to individual atoms and other simple isolated systems. Not everything is entangled with everything else, or it would be impossible to say much about any particular subsystem of the world.