Lecture 12, Thurs Feb 23: Interpretation of QM (Copenhagen, Dynamical Collapse, MWI, Decoherence)

At this point in the course, we're finally in a position to step back and ask,"What is quantum mechanics telling us about reality?" It should be no surprise that there isn't a consensus on this question (to put it mildly)! But regardless of your own views, it's important to know something about the various positions people have defended over the years, as the development of these positions has sometimes gone hand in hand with breakthroughs in quantum mechanics. (We'll see an example of this later, with the Bell inequality, and arguably quantum computing itself is another example.)

Most discussions about the implications of quantum mechanics for our understanding of reality center around the so-called Measurement Problem.

In most physics texts (and in this class, for that matter...), measurement is introduced as just a primitive operation that we don't try to understand more deeply. However, there's a fundamental weirdness about measurement in QM, which stems from the fact that the theory seems to demand both:

1. Unitary Evolution
   when no one is watching $|\Psi\rangle \to U|\Psi\rangle$
2. Measurement
   which collapses a state to a single possibility $|\Psi\rangle \to i$ with probability $|\langle\Psi|i\rangle|^2 = |\alpha_i|^2$

In other words, quantum mechanics seems to work in a way that's deterministic, reversible, and continuous most of the time (1), *except* during measurement (2), which is the only time we see it work in a way that's probabilistic, irreversible, and sudden. So we can phrase the question as:

"How does the universe know when to apply unitary evolution and when to apply measurement?"

People have argued about this for about 100 years. The discussion is sometimes compared to the discussion about the nature of consciousness (which has gone on for millennia) in that they both tend to devolve into people talking in circles around each other.

But it's worth understanding the main schools of thought, starting with…

The Copenhagen Interpretation

This was the prefered interpretation of most of the founders of quantum mechanics. It's closely associated with Niels Bohr and his institute in Copenhagen (hence the name) and with Werner Heisenberg. Note that the different founders said different and sometimes contradictory things, sometimes in very abstruse language, so it's notoriously hard to pin down what the "Copenhagen interpretation" actually is!

Basically, though, the Copenhagen viewpoint is that there are two different worlds (or parts of reality): the quantum world and the classical world. We live in the classical world, where objects have definite locations, the objects can be measured without disturbing them, etc. But in doing experiments we've discovered that there also exists the quantum world "beneath" ours, which obeys very different rules.

<u>Measurement</u>, in this view, is the operation that bridges the two worlds.

It lets us "peek under the hood" into the quantum world and see what's going on.

Bohr wrote long tracts saying that even to make statements about the quantum world and the classical world is to presuppose that *is* a classical world in which those statements can be made. So there's some "boundary" or "cut" between the quantum and classical worlds. The exact location of this boundary might be fuzzy, and might vary depending on what sort of question we're asking. But in any case, we should never make the error of insisting that our commonsense, classical concepts remain valid on the quantum side of the boundary.

Believers in the Copenhagen interpretation love to say things like: "if this doesn't make sense to you, then you're just stuck in the old way of thinking, and you need to change. The problem is not with quantum mechanics, it's with you."

The next interpretation, which is closely related is…

S.U.A.C. : "Shut Up And Calculate!"

This is probably the preferred "interpretation" of most physicists, chemists, and others who work with quantum mechanics.

It says that at the end of the day, quantum mechanics works: it correctly predicts the results of experiments. And that's all we can reasonably ask of a scientific theory, or all that it's fruitful to ask.

Prof. Aaronson likes to say that the Copenhagen interpretation is basically just S.U.A.C. without the S.U. part! Copenhagen starts from the intuition that "it's pointless to philosophize about what this means," but then elevates *that* to a philosophy, which of course is a little ironic.

In any case, while the S.U.A.C. view has some obvious practical advantages, it seems clear that it can't satisfy people's curiosity forever. This is not only because science has always aspired to *understand what the world is like*, with experiments and predictions just a means to that end. A second reason is that, as experimenters become able to create ever larger and more complicated quantum superpositions—in effect, "breaching" the Bohr/Heisenberg boundary between the quantum and classical worlds—it becomes less and less viable to "quarantine" quantum mechanics as just a weird mathematical formalism that happens to work for predicting the behavior of electrons and photons. The more QM impinges on the world of everyday experience, the more it seems necessary to come to terms with whatever it says about that world.

This seems like a good time for a digression about two celebrated thought experiments, which were invented to probe exactly this "breaching"...

Schrödinger's Cat

There were physicists in the 20s and 30s who never accepted the Copenhagen interpretation, of whom the most famous were Einstein and Schrödinger. They came up with many examples to try to show just how untenable it is to have a rigid boundary between the quantum and classical worlds, if you think hard about it.

By far the most famous example is Schrödinger's Cat, which first appears with Einstein remarking in a letter that if you think of a pile of gunpowder as being inherently unstable, you could model it as a quantum state which looks like $|$  $\rangle + |$  $\rangle$.

Then Schrödinger adds some flair by asking, "What happens if we create a quantum state that corresponds to a superposition of a state in which a cat is alive and one where the cat is dead?" He isolates the state from its external environment by putting it in a box $|$  $\rangle + |$  $\rangle$.

The point of the thought experiment is that the formal rules of quantum mechanics apply *whenever* you have distinguishable states, regardless of their size. In particular, they say that you can create arbitrary linear combinations of such states. But by the time we're talking about something as big as a cat, it seems patently obvious that we should have to say something about the nature of what's going on before measurement. Otherwise we'd devolve into extreme solipsism—saying, for example, that the cat only exists once we've opened the box to observe it.

Wigner's Friend

is a similar thought experiment. It says that Eugene Wigner—the physicist who proposed the experiment—could be put into an equal superposition of thinking one thought and thinking another one, which we model as

$$\frac{1}{\sqrt{2}}(|\text{Wigner}_0\rangle + |\text{Wigner}_1\rangle).$$

Now consider the joint state of Wigner and a friend who hasn't yet measured his state:

$$|\text{Friend}\rangle \otimes \frac{1}{\sqrt{2}}(|\text{Wigner}_0\rangle + |\text{Wigner}_1\rangle)$$

From Wigner's point of view, he's thinking one thought or the other one. But from his friend's point of view, he isn't thinking *either* of them until a measurement gets made. At that point we'll have an entangled state like

$$\frac{1}{\sqrt{2}}(|\text{Friend}_0\rangle|\text{Wigner}_0\rangle + |\text{Friend}_1\rangle|\text{Wigner}_1\rangle)$$

But then what happens if another friend comes along, and then another?

The point is to highlight an apparent incompatibility between the perspectives of different observers. It seems like *either* we need to retreat into a sort of solipsism—holding that an event that happened for Wigner might not have happened for his friend—or else we need some way of regarding measurement as fictitious.

OK, now let's discuss a few more interpretations of quantum mechanics. Our next one isn't really an "interpretation," but rather a demand for a new physical theory.

Dynamical Collapse

If quantum mechanics doesn't make sense to us, it's worth at least considering the possibility that it's not a complete theory. I.e., that maybe it does a good job of describing microscopic systems, but we're not looking at all of the rules that govern reality.

In more detail: maybe there exist some physics rules that we haven't discovered which say that qubits *normally* evolve via unitary transformations, but that the bigger the system is (or something), the more likely it is to collapse. In that case, we could view collapse as a straightforward *physical* process that turns pure states into mixed states.

$$\sum_i \alpha_i |i\rangle \to |i\rangle \text{ with probability } |\langle \Psi | i \rangle|^2 = |\alpha_i|^2$$

So in the Schrödinger's cat example, dynamical collapse would say that it doesn't matter how isolated the box is. There's some yet-unknown physical law that says that a system that big would quickly evolve into a mixed state.

Note that in principle, there's a measurement that can distinguish the two states above.

Such a measurement would, admittedly, be absurdly hard to implement. In fact, a recent result by Prof. Aaronson says, informally, that if you have the technological capability to distinguish the two states above, then you also have the technological capability to rotate between the cat's "alive" and "dead" states. For this reason, the Schrödinger's cat experiment involves *far* less animal cruelty than most people say! If you can do the experiment at all, and prove that you did it, then you can also bring a dead cat back to life.

But setting aside technological difficulties, for us the relevant point is this: in saying that the first state evolves to the second, we're proposing new physics, *different* from standard quantum mechanics, that in principle has testable implications.

In other words, this isn't *really* interpreting quantum mechanics, it's just proposing new laws of physics! Physicists have a high bar for such proposals; the burden of proof is on the person proposing the new law to explain in quantitative detail how it works. In this case, that would mean giving a criterion for *exactly* which systems are "big" enough, or whatever, to trigger a collapse like the above—and ideally, deriving that criterion from more fundamental laws. Some suggestions include:

- Collapse happens when some number of atoms get involved
- Collapse happens after a certain mass is reached
- Collapse happens when a system reaches a certain level of "complexity" (defined how?)
- etc.
    - On their face, all these views seem contradictory to our understanding of physics, which relies on *reductionism*: each atom just keeps obeying the same simple equations, regardless of how big or complicated a system the atom might be part of.

To escape that problem, one famous proposal is the…

Ghirardi-Rimini-Weber (GRW) Theory

which says that each atom has some tiny probability of collapsing (or if you like, "being measured by God") at each point in time. And if even one atom of Schrödinger's cat was "measured by God," that would cause the entire cat to collapse to the Alive or Dead states: this part is just the usual partial measurement rule of quantum mechanics. By analogy, measuring just one qubit of $\frac{1}{\sqrt{2}}(|00\ldots0\rangle + |11\ldots1\rangle)$ will resolve all of the qubits to 0 or 1. So in the GRW theory, Schrödinger cats are inherently unstable—and the bigger the system, the shorter the time it can be maintained in a Schrödinger-cat-like state.

another option is the…

Penrose Theory

which says that superpositions spontaneously collapse when enough mass gets involved, and the mass is separated by a big enough distance across different branches of the superposition.

Why mass and distance? mass here ▼ or ▼ mass there

Say we have the superposition of $|\quad\rangle + |\quad\rangle$. General relativity tells us that mass curves the nearby space-time: indeed, bends it like a mattress. That means that a mass in one location would make spacetime curve differently than the same mass somewhere else.

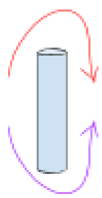
The thing is, no one really knows how to combine general relativity and quantum mechanics; it's one of the biggest unsolved problems in physics. In particular, ordinary quantum mechanics presupposes space and time, or at least a *causal structure*: in terms of quantum circuits, if you like, a definite collection of qubits and gates acting on those qubits in some order. No one quite knows what it means to have a quantum superposition of different causal structures—yet, that *seems* to be what we'd be talking about in the situation with the widely-separated masses. So, Penrose's proposal is basically that this could be the place where quantum mechanics breaks down, to be superseded by a more complete theory that includes "spontaneous collapses." And then he has further ideas about how all of this might be related to consciousness, which we won't go into.

One difficulty with this sort of theory, in general, is that the experimenters keep producing examples of bigger and bigger states in superposition. And as long as that continues, it seems like the believers in these theories will always need to be on the defensive—adjusting their answers to questions like "so, how much mass *is* enough to collapse a state?" to avoid contradicting the latest experiments.

Early on, we discussed the significance of the double-slit experiment as performed with photons. Later on, though, people managed to do the same experiment with protons, then molecules, and in 1999 the Zeilinger group in Vienna performed it with buckyballs: molecules with 60 atoms and hundreds of electrons. Since then the double-slit experiment has been done with larger molecules still.

To go even further...

Superconducting Qubits



If you take a coil, like maybe a micrometer across, and cool it to almost absolute zero, you can get a current that's in a superposition of the electrons flowing clockwise or counterclockwise around the coil.

This constitutes a quantum superposition involving billions of particles!

We'll come back to these superconducting coils at the end of the course, as they're an important technology for quantum computers.

Penrose has a specific prediction for the scale at which collapse happens, which might be testable in our lifetime. But with GRW, the prediction is basically just made to avoid contradicting any existing experiments.

One position, popular among people who want nature to be efficiently simulatable by a classical computer (and thus don't want quantum computers to work) says that:

> A frog can be in a superposition of two states. However, a complex quantum computer wouldn't work, because quantum systems spontaneously collapse after they achieve "sufficient complexity" (whatever that means).

This position is interesting because it could be falsified by building a scalable quantum computer, and reaching falsifiable theories is what moves these discussions from philosophy to science.

<u>What happens if we keep doing experiments and quantum mechanics keeps perfectly describing everything we see?</u>

In particular, suppose we *don't* want to add any new physical laws, but we also insist on being *scientific realists*—holding that there exists a real state of the universe, and that the job of physics is to describe that state, not just to predict the results of measurements made by apes like ourselves.

Well, that combination of choices basically gets you to…

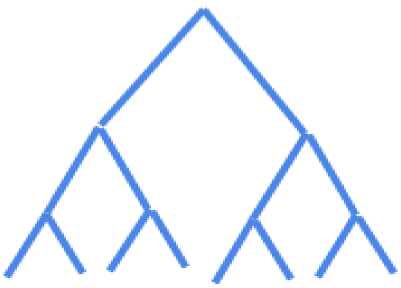
Everett's Many Worlds Interpretation (1957)

This famous view holds that the entire universe has a single quantum state $|\Psi\rangle$, and the entire history of the universe is just the vector $|\Psi\rangle$ going through unitary evolution.

On the Everett view, what we call "measurement," or "collapse," is just a special case of quantum systems becoming entangled with each other when they interact. In particular, *your brain*—not to mention, your measuring apparatus, the air molecules in the room, etc. etc.—all become entangled with the quantum system that you're measuring. You can think of it as a giant Controlled-NOT gate, with the system you're observing as the control qubit and you as the target qubit.

$$\frac{|0\rangle + |1\rangle}{\sqrt{2}}|\text{You}\rangle \to \frac{|0\rangle|\text{You}_0\rangle + |1\rangle|\text{You}_1\rangle}{\sqrt{2}}$$

The thing is, if you take this view seriously, it implies that *you yourself* have now "branched" into two possibilities, one where you observed the qubit in the $|0\rangle$ state, and another where you observed the qubit in the $|1\rangle$ state. Because unitary evolution is linear, these two branches are unlikely ever to interfere with each other again, or at least not for many quadrillions of years (more about that later). So your experience is *as if* only one of the branches was realized—but in truth, neither branch is more real than the other.



More generally, according to MWI, the universe "branches" each and every time a microscopic quantum state gets amplified to have macroscopic effect—even if there's no one around to observe the amplification (e.g., if it's in the interior of the Sun or something). So there's a staggering amount of branching happening! And because of the well-known phenomenon of *chaos*, we can expect the branching to influence pretty much everything we care about. So for example, there are some branches of the quantum state of the universe where Austin is sunny at a specific moment one month from now, others where it's rainy, others where it's been destroyed in a nuclear war, etc., and no one of these branches is more "real" than the rest.

Some variants of the Many Worlds interpretation choose words carefully to avoid sounding like there's *literal branching into different, equally-real worlds of experience*, but that's basically what they all imply. When Everett came up with MWI as a grad student at Princeton, his advisor—the famous John Wheeler—told him to remove from his paper all references to the physical reality of parallel worlds, because it wouldn't chime with the physics establishment at the time. So Everett did so, and partly as a result, it took almost 20 years for the rest of the physics community to rediscover Everett's proposal and understand what it meant.

Soon after publishing his thesis about MWI—which he called the "relative state interpretation"—Everett left theoretical physics to become a nuclear war strategist for the Pentagon. The only public lecture Everett ever gave on MWI was here at UT Austin, decades later, when some people were finally coming around to the idea. David Deutsch, the biggest current advocate of the Many Worlds Interpretation and one of the founders of quantum computing, was there.

One issue that we should return to is interference between branches.

If the different branches could interfere with each other, it would be as if not just the future but *the past* was constantly shifting, with no definite sequence of things that happened and were recorded. To avoid that, we need the "$|0\rangle|\text{You}_0\rangle$ branch" to *not* affect the "$|1\rangle|\text{You}_1\rangle$ branch," and vice versa. Both branches might be equally real, but once you're *in* one of the branches, you ought to be able to continue doing physics *as if* your branch was the only real one.

Fortunately, the usual rules of quantum mechanics give us this property: we don't need to add anything extra to them. Recall: to calculate the amplitude of a given basis state $|x\rangle$ in a larger

$|x\rangle$

superposition, you add up a contribution from every possible "path" that ends at . Interference would only happen if two "macroscopically different" paths led to *exactly* the same outcome—meaning, every single atom in the universe in the same position. But, while that's not impossible, it's massively "thermodynamically disfavored," which basically means that it's less likely to happen than seeing an egg unscramble itself. In fact, the constant proliferation of branches—the way the universe's state, $|\Psi\rangle$, constantly sprouts new branches, but they almost never recombine—can be seen as literally an *instance* of the Second Law of Thermodynamics in action, as a process that constantly increases the universe's entropy from the low value it had at the Big Bang.

> And *why* did the universe have such a low entropy at the Big Bang? Well, to this day no one knows the answer to *that*, except that if it didn't, we probably wouldn't be here to wonder about it...

Interestingly, if we *also* believed that the universe was only finitely large—and in particular, that it could be fully described by the unitary evolution of a finite number of qubits (say 10^{122} of them)—then *eventually* we'd run out of room, and the branches would necessarily start colliding with each other. But even under that assumption, there doesn't seem to be any reason for this happen in (say) the next 10^{100} years.

We said before that measurement is the one random and irreversible part of quantum mechanics. But Many Worlds denies that even that part is random or irreversible. After applying a unitary transformation U that describes a measurement process, in principle we could always apply U^{-1} to make the measurement "unhappen." But just like with unscrambling an egg, thermodynamics isn't going to make it easy.

Let's now discuss some of the most common other questions people have about Many Worlds.

(1) Even if we accept that "measurements" have no fundamental physical status—still, where do the *apparent* probabilities come from?

That is, why does measuring a qubit $\alpha|0\rangle + \beta|1\rangle$, in the $\{|0\rangle, |1\rangle\}$ basis, yield the outcomes $|0\rangle$ and $|1\rangle$ with probabilities $|\alpha|^2$ and $|\beta|^2$ respectively?

It's not enough to say that sometimes we see 0 and sometimes we see 1. Quantum mechanics gives very specific probabilities that each will occur. But if the world is just branching once for each observation, then how can we justify these probabilities as corresponding to anything meaningful? Does an "$|\alpha|^2$ fraction of souls" go down one branch while a "$|\beta|^2$ fraction of souls" goes down the other? Or: does the "splitting of the worlds" happen in such a way that amplitudes of $\frac{3}{5}$ and $\frac{4}{5}$ would correspond to $\frac{9}{25}$th "volume of worldness" going one way, and $\frac{16}{25}$th going the other?

Some philosophers don't like this because if all the worlds are equally real, then why wouldn't they just occur with equal probabilities? Why bother with amplitudes at all?

Everett's response was to argue that if the universe branched many times in succession, then in "almost all branches" (where "almost all" is measured by amplitude), it would *look like* the Born probability rule was obeyed. But many people in the past half-century have been unsatisfied with that argument, seeing it as circular—indeed, as smuggling the Born rule into the definition of "almost all branches"! So they've continued to look for something better.

There are many arguments, which we won't go into here, that try to formalize the intuition that the Born probabilities are just "baked into" how quantum mechanics works. After all, unitary evolution already singles out the 2-norm as special by preserving it, so then why shouldn't the probabilities *also* be governed by the 2-norm? More pointedly, one can argue that, if the probabilities were governed by something other than the 2-norm, then we'd get bizarre effects like faster-than-light communication. But, while these arguments help explain why the Born rule is perhaps the only choice of probability rule that makes internal mathematical sense, they still leave slightly mysterious how probability enters *at all* into Everett's vision of a deterministically evolving wavefunction.

In Everett's defense, one could ask the same questions—"where do these probabilities come from? why should they follow the Born rule, rather than some other rule?"—in *any* interpretation, not just in MWI.

(2) "If there's no experiment that could differentiate the Copenhagen Interpretation from Many Worlds, why bother arguing about it?"

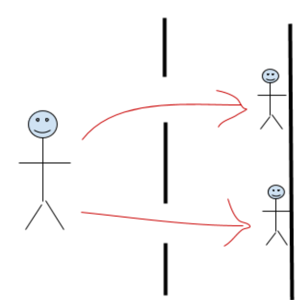
Many Worlders say that the opponents of Galileo and Copernicus could also claim the same about the Copernican versus Ptolemaic theories, since Copernican heliocentrism made no difference to the predictions of celestial movement.

Today, we might say that the Copernican view is better because you could fly outside of the solar system and see all the planets (including Earth) rotating around the far more massive sun; it's only our parochial situation of living on Earth that ever motivated geocentrism in the first place.

But if we push this analogy further, it might be harder to think of anything similar for the Many Worlds interpretation, since quantum mechanics itself explains why we can't really get outside of the universe to see the branching—or even get outside our own branch to interact in any way with the other decoherent branches.

There is one neat way you could imagine differentiating the two, though...

Before we talked about doing the double-slit experiment with larger and larger systems. Bringing that thread to its logical conclusion, <u>what if we could run the double-slit experiment with a person going through the slits?</u>



It seems like it would then be necessary to say that "observers" can, indeed, exist in superpositions of having one experience and having a different one. This is what Many Worlds said all along, but seems to put a lot of rhetorical strain on the Copenhagen interpretation.

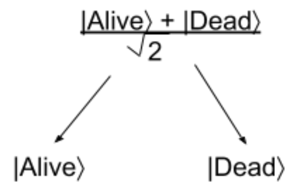
If you talk to modern Copenhagenists about this they'll take a quasi-solipsistic view, saying that if this experiment were run, "the person being behaving quantumly doesn't count as an observer, only I, the experimenter do."

Of course, the Wigner's Friend experiment was trying to get at this same difficulty.

The third question we want to tackle is the **Prefered Basis Problem**. It says:

"Let's say I buy into the argument that the universe keeps branching, well then…"

3) In what basis is this branching occurring?



We talked about Schrödinger's cat as branching into the $|\mathrm{Alive}\rangle$ state and the $|\mathrm{Dead}\rangle$ state.

But mathematically, we could equally well have decomposed the cat's state in a basis like

$$\frac{(|\mathrm{Alive}\rangle + |\mathrm{Dead}\rangle)}{\sqrt{2}}, \frac{(|\mathrm{Alive}\rangle - |\mathrm{Dead}\rangle)}{\sqrt{2}}$$

So is there anything besides our intuition to "prefer" the first decomposition over the second one?

There's a whole field of physics that tries to answer questions like these, called...
Decoherence Theory

which says that there are certain bases whose states tend to be robust to interactions with the environment, but most bases aren't like that.

So in the example above, decoherence theory would explain that an alive cat doesn't easily decohere if you poke it, but that a cat in the $\frac{(|\mathrm{Alive}\rangle + |\mathrm{Dead}\rangle)}{\sqrt{2}}$ state does, because the $|\mathrm{Alive}\rangle$ and $|\mathrm{Dead}\rangle$ branches interact differently with the environment. This, according to decoherence theory, is more-or-less how the laws of physics pick out certain bases as being special.

From the standpoint of decoherence theory, we can say that an event has "definitely happened" if and only if there exist many records of the event scattered all over the place, so that it's no longer feasible to erase them all.

This is perhaps best compared to putting an embarrassing picture on Facebook. If only a few friends share it, you can still take it down. On the other hand, if the picture goes viral, then the cat is out of the bag, and deleting all the copies becomes next to impossible.