CS191 – Fall 2014
Lecture 11: Foundations, EPR and Bell's theorem

Mohan Sarovar
(Dated: October 3, 2014)

Now that you have learned the basic quantum formalism from an axiomatic point of view and are skilled in manipulating the mathematical objects we need, I want to loop back and ask what exactly the quantities you have been manipulating mean. You've probably heard that there are some "weird" aspects to quantum mechanics and its historically been somewhat of a nonintuitive theory. Maybe you have already run into some of this weirdness in the first few weeks of this course. In this lecture I want to dig into this further and give you some more details about what exactly it is that makes quantum theory weird. This is not only important for understanding the foundations of the subject, but some of the things we identify as being "quantum weirdness" in this lecture are exactly what will be exploited as features in quantum information applications.

I should say that what I'm going to do is a little unfair. I'm going to try to condense a century of some of the deepest thinking in physics and philosophy into one lecture. Unfortunately this means that we'll have to skip over some subtleties, but hopefully this will give you a flavor of the main ideas (and there is further reading material listed at the end).

I. THE DIFFERENCES BETWEEN QUANTUM AND CLASSICAL STATES

You've learned that:

- The "state" of a system is represented by a wave function, or state vector, $|\psi\rangle$, which is a normalized state in a complex vector space. I am specializing here to finite dimensional systems (otherwise the complex vector space would be a normed Hilbert space) and to "pure states" (otherwise the state is more generally represented by a density matrix). The latter specialization is particularly important because it is pure states of systems (*i.e.*, wavefunctions) that we really want to understand – once we understand these, density matrices are easy to grasp.

- If we expand the state vector in a basis of distinguishable (or more mathematically, orthonormal) states, $|x_i\rangle$, as

$$|\psi\rangle = \sum_i \alpha_i |x_i\rangle, \tag{1}$$

  then, if we measure which of these distinguishable states the system is in, the probability of observing $|x_i\rangle$ is $|\alpha_i|^2$.

A lot of the difficulty with quantum mechanics boils down to how we interpret what the "state" of the system, $|\psi\rangle$, means. Much of our intuition for state come from classical mechanics. Consider a system of N classical particles with no internal degrees of freedom. Then the "state" of this system is completely characterized by a vector:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ p_1 \\ x_2 \\ p_2 \\ \vdots \\ x_N \\ p_N \end{pmatrix}, \tag{2}$$

where x_i and p_i are the position and momentum of particle i. This state vector is a *complete* description of the system in the sense that the results of any observation on the system can be calculated from it, and in addition, the dynamics of the system can be written in terms of it (Hamilton's equations). This state also has another property, which is so intuitive and obvious that we often don't even mention it. Namely, each of it's elements (x_i, p_i) corresponds to a physical property in the world that exists independently of whether we write down the state or observe any property of the system. That is, particle 1 has a position and momentum regardless of whatever else there is in the universe, and we simply tabulate these values in the state vector.

Now, contrast this view of the state vector with the quantum mechanical state, $|\psi\rangle$. Recall that a general state of a qubit is:

$$\begin{aligned} |\psi\rangle &= \cos(\theta)\,|0\rangle + e^{i\phi}\sin(\theta)\,|1\rangle \\ &= \begin{pmatrix} \cos(\theta) \\ e^{i\phi}\sin(\theta) \end{pmatrix}, \end{aligned} \tag{3}$$

where in the first line we've used ket notation and in the second line written the equivalent 2×1 vector in the computational basis. Note that this "state vector" differs from the classical state vector in Eq. (2) in at least two ways:

1. It is not completely predictive of the outcomes of observations on the system. For example, if I measure the observable σ_z on this qubit, all this state tells me is that I will get the value $+1$ with probability $|\cos(\theta)|^2$. Similarly, I will get the value -1 with probability $|\sin(\theta)|^2$. Contrast this with the classical state, where if I measure the position of particle 1, I will *deterministically* get x_1. So the quantum mechanical state vector cannot always yield deterministic values for observable outcomes. Let's call this feature of the quantum state, `feature 1`.

2. Imagine that we measure σ_z on this state and get the value $+1$. Then the post measurement state (recall how to determine the post measurement state) is $|\psi_1\rangle = |0\rangle$. Now, what happened to the variable θ and ϕ that determined the state prior to measurement? They have been completely erased by the measurement we performed. Contrast this with the classical state, where if we are clever enough there is nothing preventing us from measuring the position of particle 1 without changing any of the other entries in the state vector. That is the post measurement state can be exactly as the state prior to measurement: $\mathbf{x}$. This is not possible in general for quantum systems, regardless of how clever we are (unless the state is an eigenstate of the observable being measured). So in what sense do the properties θ and ϕ exist? Is there some physical property of the qubit that corresponds to θ? If so, why can't we determine it from any measurement we can do on it (on one copy of it, not many copies) [1]. Let's call this feature of the quantum state, `feature 2`.

`Features 1` and `2` make the quantum state decidedly different from any classical notion of state. And consequently there has been much debate in the history of physics about how to think about the quantum state. An early realization was that that `features 1` and `2` are similar to properties of probability distributions, and the generalized notion of a classical state. That is, if a classical system contains uncertainty then the description of the system is not in terms of a state vector $\mathbf{x}$, but rather as a probability distribution over phase space (in other words, the positions and momenta become random variables).

Let's look at an example. Consider a coin that we toss. In principle, it is possible to know whether it will land heads or tails if we keep track of a lot of information about the force applied during the toss, the air currents, the height of the toss, etc. But when we ignore all of these physical properties, the most we can do is ascribe a probability distribution for the toss outcomes: $\Pr\{H\} = \Pr\{T\} = \frac{1}{2}$, which we could denote as a probability distribution vector:

$$P = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \tag{4}$$

In classical mechanics, this quantity is called a *phase space distribution* and is the generalization of the state of the system when there is uncertainty. The *Liouville equations* specify how the phase space distribution evolves in time, just like Hamilton's equations do for the state vector. Note that this phase space distribution also has `features 1` and `2`; *i.e.*, it only tells us the probability of getting certain outcomes, and it's values (the probabilities) can't be determined by any measurement of a single coin.

Given this resemblance, it was suggested by early quantum physicists that the quantum mechanical state vector is an *incomplete* description of the system, like a phase space distribution. However, a crucial thing to note about phase space distributions is that one can always add more information to convert the description of the system into a state

[1] We have only looked at σ_z measurements, but it is easy to see that no measurement will allow us to determine θ or ϕ with only one copy of the qubit.

vector like Eq. (2). For example, for the coin toss example an alternative description with much more information is

$$\mathbf{x}_{\text{cointoss}} = \begin{pmatrix} \text{face} \\ \text{face}_0 \\ \overrightarrow{F}_0 \\ h \\ x_1 \\ p_1 \\ x_2 \\ p_2 \\ \vdots \end{pmatrix}, \tag{5}$$

where face is the (deterministic) face of the coin after the toss, face_0 is the face of the coin that is up before the toss, $\overrightarrow{F}_0$ is the initial force imparted by the toss, h is the height of the toss, x_i, p_i are the (possibly time-dependent) positions and momemta of all the molecules in the air, etc. With this (hypothetical) *complete* description of the system, the toss outcome can again be prescribed deterministically. But once we average over the values of all the variables except the first one, we end up with a phase space distribution for the first variable, from which only probabilistic predictions for the value of face can be obtained. Similarly, is $|\psi\rangle$ a description that results from averaging over some additional degrees of freedom that we don't have access to?

Given the shared properties of quantum states and phase space distributions, and the above fact that phase space distributions can be augmented to obtain system states, a strategy for understanding what the quantum mechanical state vector represents is to look for how to complete the description that it provides. The hypothetical extra variables one has to include into the quantum state vector description to complete it to get deterministic predictions are called *hidden variables.*

II. EPR AND HIDDEN VARIABLES

The most famous expression of this opinion that hidden variables need to be included to complete the quantum mechanical description of state is a paper by Einstein, Podolsky and Rosen (EPR) from 1935 (see references). In this landmark paper, the authors start with two reasonable definitions:

1. A complete theory is one in which *"every element of the physical reality must have a counterpart in the physical theory"*.
2. For a definition of physical reality, they offer: *"If, without in any way disturbing a system, we can predict with certainty (*i.e.*, with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity"*.

Note that classical mechanics, which is based on the classical notion of a state, *e.g.*, Eq. (2), satisfies the definition of being a complete theory. To show that quantum mechanics does not, EPR consider the scenario with two parties, Alice and Bob, who are given two qubits [2]. Remember that given a single qubit, quantum mechanics says that we cannot determine both it's σ_z value and it's σ_x value; this is discussed above where we defined `feature 2` of quantum mechanics. Then EPR-Bohm specify that Alice and Bob are spatially separated – so that nothing Alice does can effect Bob and vice-versa – and that the state of two qubits is the entangled state

$$\left|\psi^{AB}\right\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \tag{6}$$

Then if Alice measures in the σ_z basis, she will get $+1$ (and post-measurement state $|0\rangle$) with probability $\frac{1}{2}$, and -1 (and post-measurement state $|1\rangle$) with probability $\frac{1}{2}$. But regardless of the result, she will know exactly what the value of Bob's qubit is as well; it will be the same as the post-measurement state that she has. Therefore since she knows with certainty the σ_z value of Bob's qubit, and she acquired this knowledge without in any way disturbing

---

[2] EPR actually analyzed the sharing of a two particles, with position and momentum. This version of the EPR analysis with shared qubits (or spin$-\frac{1}{2}$ systems) is originally due to David Bohm (in *Quantum theory*, Dover Publications, 1951). Hence we will use EPR-Bohm instead of EPR when referring to the authors of the following argument.

Bob's qubit (because she is spatially separated from it), it must be an element of physical reality. Similarly, if Alice chooses to measure σ_x on her qubit, she will get she will get $+1$ (and post-measurement state $|+\rangle$) with probability $\frac{1}{2}$, and -1 (and post-measurement state $|-\rangle$) with probability $\frac{1}{2}$. Again, regardless of the result, she will know exactly what the value of Bob's qubit is as well, it will be the same as her post measurement state. So as before, σ_x of Bob's qubit must also be an element of reality (since Alice can predict it's value with certainty without disturbing Bob's qubit). Now comes the critical step in the EPR-Bohm argument: if both σ_x and σ_z are elements of reality, then a complete theory should specify the value of both. But we know quantum mechanics does not do this, since it says that only one of those values can be precisely determined at any time, and thus quantum mechanics is incomplete.

Exercise: Verify the statements in the last paragraph about measurement results and post-measurement states when the shared state is $|\psi^{AB}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$.

Exercise: Why was it important for EPR-Bohm to use an entangled state shared among spatially separated parties for their argument? Can one make the same argument with a single qubit?

This EPR-Bohm argument is essentially a formalization of the feeling that quantum states are incomplete descriptions of reality. It and the discussion in section I beg the question: how do we complete this description?

III. BELL'S THEOREM AND THE CHSH INEQUALITY

The question of how to complete the description of reality given by quantum mechanics and quantum states is a difficult (and unresolved) one. However, John Bell provided critical direction to this effort with an amazing result that we now refer to as Bell's theorem, which shows that if hidden variables that complete the quantum state do exist, then they have to take a very special form. The statement of Bell's theorem is:

> No theory that respects the locality principle and the reality principle (more concisely, no *local, realistic* theory) can reproduce the results of quantum mechanics.

We will specify what the locality and reality principles are below. The way Bell proved this theorem is to derive an inequality that probabilities arising from any local, realistic theory would have to satisfy and then show that quantum mechanical predictions violate this inequality. Subsequently, experiments that tested this inequality (or more accurately, a related inequality) showed that it is indeed violated. Therefore nature agrees with quantum mechanics, rather than (the more intuitive) local realistic models.

Now, let's see what the above principles are:

1. The *reality principle* states that the values of physical quantities have physical reality independent of whether a measurement of them is made or not. Systems possess intrinsic properties described by states.

2. The *locality principle* states that the results of a measurement are determined by the local state of the system being measured. That is, the results of a measurement should not depend on causally disconnected regions.

Both of these principles seems reasonable and are obeyed by classical mechanics. But Bell's theorem tells us that they can't both be true in a world consistent with quantum mechanics.

Let's see how we derive a result like Bell's theorem. Since Bell's derived his inequality in 1964, there have been several extensions and refinements of it, and one of the simplest versions (that is also more experimentally viable) was derived by Clauser, Horne, Shimony and Holt (CHSH) in 1969. We will reproduce this CHSH inequality in the following.

The setting for deriving the CHSH inequality begins by considering the "game" in Figure 1. In each round of this game Alice and Bob receive an object each from a source in the middle. They both decide to measure one of two properties/observables of the object they receive: P or Q for Alice and R and S for Bob. R and S could be the same properties as P and Q, but we will denote them differently for clarity. For the following it's important that each of these properties/observables has binary outcomes, labeled by -1 and $+1$, and that both parties can only measure one property in each round of the game. Therefore in each round Alice outputs a result ± 1, which we shall call the random variable X, and similarly Bob outputs a result ± 1, which we shall call the random variable Y.

Example 1 A classical example of this setup is where the source is a coin factory and Alice and Bob both receive a coin in each round of the game. The properties they measure could be the color of the coin (blue (-1) or red (+1)) and the shape of the coin (square (-1) or disc (1)). Then $P = R =$ color and $Q = S =$ shape. If Alice and Bob knew the "state" of the factory (given by a huge list of parameters λ)

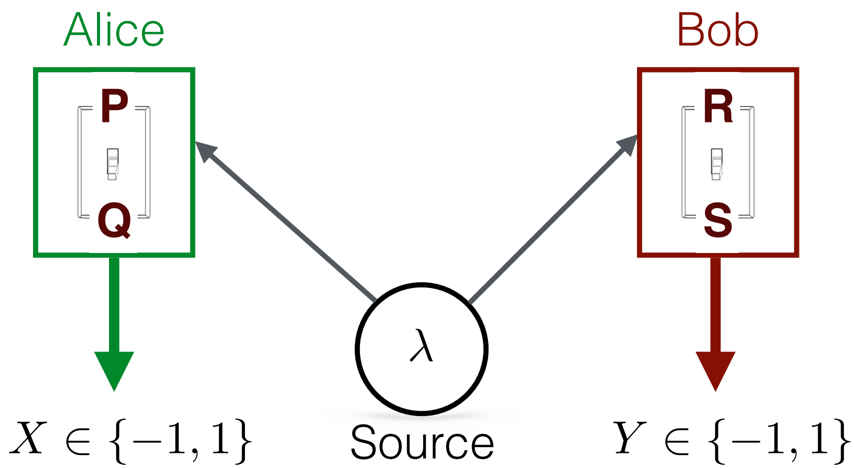

FIG. 1: The CHSH game

then they would know what the properties of the coins that receive in each round is. But they don't and therefore it looks like they are receiving coins with random properties. But for what we will look at below, it's important to consider that there could be correlations between the coins that Alice and Bob receive in each round. For example, if in each round the factory always produces one square blue coin and one disc-shaped red coin, and then send one of these coins at random to Alice and Bob, then the properties that Alice and Bob measure will always be anti-correlated.

Example 2 A quantum example of this setup is where the source produces qubits and send one qubit each to Alice and Bob. Then the properties/observables that they can measure on each qubit they receive is either the σ_z value or the σ_x value. These measurements naturally have ± 1 outcomes. We will say more about this example later.

We want to consider the correlations between what Alice measures and Bob measures. Let $a \in \{P, Q\}$ be the property that Alice chooses to measure in a particular round, and let $b \in \{R, S\}$ be the property that Bob chooses to measure. Then the correlation between the measurement outcomes, given a specific choice by Alice and Bob of properties to measure is:

$$E(a, b) = \int_\Lambda X(a, b, \lambda) Y(a, b, \lambda) \rho(\lambda | a, b) d\lambda, \tag{7}$$

where $\rho(\lambda)$ is a distribution of the unknown hidden variables dictating the behavior of the source, Λ is the domain of these hidden variables. $X(a, b, \lambda) \in \{-1, 1\}$ is a deterministic functions given a, b, λ in this case, and similarly for $Y(a, b, \lambda)$; *i.e.*, once the properties being measured and the hidden variables are specified we can say deterministically what the outcome will be (think about example 1).

As an aside, there is a more general setting we could consider where there is some remaining uncertainty even when the hidden variables are specified. This could be for example, because the channel on which the objects coming from the source is noisy. In a local, realistic model we could incorporate this randomness into a larger set of hidden variables, but if we did not do this, we would get what is called the *stochastic* setting for the CHSH game, in which the correlation between measurement outcomes is given by:

$$E_{\text{stoch}}(a, b) = \int_\Lambda XY P(X, Y | a, b, \lambda) \rho(\lambda | a, b) d\lambda, \tag{8}$$

where now the X, Y are not deterministically determined by specification of a, b, λ, but a probability distribution for them, $P(X, Y | a, b, \lambda)$ is determined. Everything we do below can be repeated for this stochastic setting, the only thing that changes is how we mathematically specify the assumptions of locality and realism changes a little. Therefore, we will restrict ourselves to the deterministic local realistic model here, and take the correlation function to be given by Eq. (10).

Returning to Eq. (10), we will now incorporate the assumptions of locality and realism to simplify the expressions in the integral to:

$$\begin{aligned} X(a,b,\lambda) &= X(a,\lambda) \\ Y(a,b,\lambda) &= Y(b,\lambda) \\ \rho(\lambda|a,b) &= \rho(\lambda) \end{aligned} \tag{9}$$

In the first two simplifications we used locality to say that the outcome on Alice's (Bob's) side only depends on the observable choice made by Alice (Bob). In the last simplification we used realism to say that the distribution over the hidden variables is independent of the choices made by Alice and Bob – these variables (that eventually determine the measurement outcomes) have independent existence, regardless of what Alice and Bob do. Given these simplifications, the correlation function between measurement outcomes becomes:

$$E(a,b) = \int_\Lambda X(a,\lambda)Y(b,\lambda)\rho(\lambda)d\lambda, \tag{10}$$

Now, the quantity CHSH considered is a particular sum of such correlations: $CHSH \equiv E(P,R) + E(Q,R) + E(P,S) - E(Q,S)$. Using the definition of the correlations, this is:

$$\begin{aligned} CHSH &= \int_\Lambda \Big[X(P,\lambda)Y(R,\lambda) + X(Q,\lambda)Y(R,\lambda) + X(P,\lambda)Y(S,\lambda) - X(Q,\lambda)Y(S,\lambda)\Big]\rho(\lambda)d\lambda \\ &= \int_\Lambda \Big[X(P,\lambda)\Big\{Y(R,\lambda) + Y(S,\lambda)\Big\} + X(Q,\lambda)\Big\{Y(R,\lambda) - Y(S,\lambda)\Big\}\Big]\rho(\lambda)d\lambda \end{aligned} \tag{11}$$

Now, since each $X, Y \in \{-1, 1\}$, we know that $-2 \le [\cdot] \le 2$, where $[\cdot]$ is the sum in the square brackets above. Therefore,

$$\begin{aligned} &\int_\Lambda (-2)\rho(\lambda)d\lambda \le CHSH \le \int_\Lambda (2)\rho(\lambda)d\lambda \\ \Rightarrow\quad &|CHSH| \equiv |E(P,R) + E(Q,R) + E(P,S) - E(Q,S)| \le 2 \end{aligned} \tag{12}$$

This is the CHSH inequality. It says that for any model that satisfies the assumptions used in Eq. (9) (locality and realism), this combination of correlations must bounded above by 2.

Exercise: Confirm that if $X, Y, X', Y' \in \{-1, 1\}$, then

$$|XY + X'Y + XY' - X'Y'| \le 2 \tag{13}$$

A. Quantum violation

The CHSH inequality is a general statement about local, realistic theories. Now, let us see if the quantum formalism agrees with this inequality. Consider a variation of example 2 above, where the source produces two qubits, one of which is sent to Alice and Bob each. Let the state of the two qubits that are produced be:

$$|\psi^{AB}\rangle = \frac{1}{\sqrt{2}}(|0^A\rangle|1^B\rangle - |1^A\rangle|0^B\rangle) \tag{14}$$

And let the observables being measured by Alice and Bob be:

$$\begin{aligned} P:&\quad \sigma_z^A \\ Q:&\quad \cos(\frac{\pi}{4})\sigma_z^A + \sin(\frac{\pi}{4})\sigma_x^A \\ R:&\quad \sigma_z^B \\ S:&\quad \cos(\frac{\pi}{4})\sigma_z^B - \sin(\frac{\pi}{4})\sigma_x^B, \end{aligned} \tag{15}$$

where $\sigma_z^A \equiv \sigma_z^A \otimes I^B$ is an observable on Alice's qubit only, and so on. Given these specifications, we can calculate each of the correlations in the CHSH inequality by finding expectation values. Explicitly,

$$\begin{aligned}
E(P,R) &= \langle \psi^{AB} | \, \sigma_z^A \otimes \sigma_z^B \, | \psi^{AB} \rangle = -1 \\
E(Q,R) &= \langle \psi^{AB} | \left[\cos(\frac{\pi}{4})\sigma_z^A + \sin(\frac{\pi}{4})\sigma_x^A \right] \otimes \sigma_z^B \, | \psi^{AB} \rangle = -\frac{1}{\sqrt{2}} \\
E(P,S) &= \langle \psi^{AB} | \, \sigma_z^A \otimes \left[\cos(\frac{\pi}{4})\sigma_z^B - \sin(\frac{\pi}{4})\sigma_x^B \right] | \psi^{AB} \rangle = -\frac{1}{\sqrt{2}} \\
E(Q,S) &= \langle \psi^{AB} | \left[\cos(\frac{\pi}{4})\sigma_z^A + \sin(\frac{\pi}{4})\sigma_x^A \right] \otimes \left[\cos(\frac{\pi}{4})\sigma_z^B - \sin(\frac{\pi}{4})\sigma_x^B \right] | \psi^{AB} \rangle = 0
\end{aligned} \tag{16}$$

Combining these, gives:

$$|CHSH|_{\text{quantum}} = 2.4142 \nleq 2, \tag{17}$$

which is a violation of the CHSH inequality!

What does this mean? Well, if we go back to the derivation of the CHSH inequality, everything following the initial local and realistic assumption was just mathematics. Therefore the only contradiction could be in the assumptions. This is where the conclusion of Bell's theorem come from: one of these assumptions must be false since the predictions of quantum mechanics have been confirmed by experiments. Hence, nature's laws do not obey both the local principle and the realism principle [3].

Exercise: Calculate the CHSH quantity when the state produced by the source is not entangled, *e.g.*, it is $|\psi^{AB}\rangle = |0^A\rangle |0^B\rangle$, and the observables remain as above. Is the CHSH inequality violated in this case?

Exercise: Let the joint state between Alice and Bob as given in Eq. (14), and let the P and R observables be the same as above. But consider the more general Q and S observables:

$$\begin{aligned}
Q: &\quad \cos(\phi)\sigma_z^A + \sin(\phi)\sigma_x^A \\
S: &\quad \cos(\phi)\sigma_z^B - \sin(\phi)\sigma_x^B,
\end{aligned}$$

for $0 \leq \phi \leq 2\pi$. For what values of ϕ is the CHSH inequality violated?

IV. CONCLUSION

In face of Bell's theorem, one has to accept that either the locality principle or the reality principle has to be dispensed with if one wants to complete the quantum state with hidden variables. Each of the many interpretations of quantum mechanics that exist forsake (at least) one of these principles.

I note that Bell's theorem is just one of the results that demonstrates the stark differences between the quantum world and classical intuition. Another example that you might want to read about is the Kochen-Specker theorem, which dispenses with another piece of classical intuition: *non-contextuality*, the notion that if a system has a property, then it does so independently of how that property is measured.

So all of this may seem a little unsatisfactory (or even unsettling) to you. We know certain ways in which quantum theory is different from classical theories, but still cannot frame it in terms of things that are intuitive. One response to this is that there is no reason to expect a law of nature, especially a microscopic law of nature, to be intuitive and that we should just get used to it and use it for calculations. Another view, is that there must be a more intuitive theory lurking behind quantum mechanics, and we just haven't found it yet, although if you are forced to throw away things like locality, realism or non-contextuality, how intuitive could it really be? So more than one hundred years on from the inception of the theory, we must be satisfied with it as a calculation tool.

I note that there is at least one other example where full understanding of a theory came only 200 years after its inception, and this is Newton's universal theory of gravity. Although it was formulated around 1700, and has

[3] Another assumption that went into this derivation was that Alice and Bob have freedom to randomly choose which measurement they make in each run of the CHSH game. This is called the *free will* or *free choice* principle. But giving up this principle is not usually considered an option.

spectacular predictive power, no one understood why the law of gravitation has the form it does. There were many debates about the unnatural qualities of "action at a distance" between two massive bodies. It wasn't until Einstein's general theory of relativity in the early 1900s that we finally understood *why* the law of gravitation looks the way it does. So maybe we have another 100 years of work to do with quantum theory.

V. REFERENCES AND FURTHER READING

1. Nielsen & Chuang. Section 2.6.

2. The original Einstein, Podolsky & Rosen paper, as well as John Bell's papers are fairly readable (although the first contains continuum quantum mechanics, which we haven't covered in this course). These are:

   Einstein, A., Podolsky, B., Rosen, N., & Rosen, N. *Can quantum-mechanical description of physical reality be considered complete.* Phys. Rev., 47, 777 (1935).

   Bell, J. S. *On the Einstein Podolsky Rosen Paradox.* Physics, 1, 195 (1964).

3. Hideo Mabuchi's notes on Bell (CHSH) inequalities, which are the second half of this lecture from a previous version of the class:

   `http://www-inst.eecs.berkeley.edu/~cs191/fa05/lectures/lecture4plusMabuchi_fa05.pdf`.

   These notes are a nice complement to the above discussion because they have a practical take on how you actually go about measuring the correlations functions needed for the CHSH inequality.

4. Chapter 6 of "Quantum theory: concepts and methods" by Asher Peres (Kluwer Academic Publishers, 1995). This is in general a great book to refer to for the physical meaning behind the quantum formalism.

5. A very readable summary of the EPR paradox is *From Einstein's theorem to Bell's theorem: a history of quantum non locality* by H. Wiseman, `arXiv:quant-ph/0509061`.

6. Prof. Rob Spekkens' excellent lectures on the foundations of quantum theory: `http://www.perimeterscholars.org/332.html`