Probabilities in Quantum Mechanics
APMA1930W, Fall, 2023

Lecture notes prepared by S. Geman & Chen Li

Contents

1 Introduction

We will focus on the discrete spin states of elementary particles. Only rarely will we talk about the continuous aspect of a particle's state, such as its position. But we will start off with an exception: the modern-day version of Young's 1801 two-slit experiment. This will give context to the remarkable transition from Newtonian to quantum mechanics that came about during the first half of the twentieth century. At the same time, we will make a case for the utility of complex variables in the state description, and introduce the three most consequential, though apparently unavoidable, experimental conclusions: at the quantum level, particles behave much like waves; the state of a system is best described as a probability distribution rather than an accounting of positions and velocities; observations irreversibly alter the evolution of the state of a system.

1.1 Some examples of waves

Consider this simple example of a one-dimensional wave:

$$\psi(x,t) = A\cos(\omega t - kx) \tag{1}$$

In this representation, A denotes the amplitude of the wave, ω denotes its angular frequency (in radians per second), and $k = 1/\lambda$ denotes its wavenumber (in radians per meter). Recall that there are 2π radians in a cycle, so the period of the wave is $2\pi/\omega$. (If you are a stationary observer, at some location x_o, then the period refers to the time between peaks: $2\pi/\omega$ seconds.)

If we take phase into account, the representation becomes

$$\psi(x,t) = A\cos(\omega t - kx + \varphi) \tag{2}$$

where φ is the phase shift.

Imagine that at some time t_o you hop on the wave at location x_o. You will then be at height $h = A\cos(\alpha)$, where $\alpha = \omega t_o - kx_o + \varphi$. Riding the wave means that α doesn't change. To keep things simple, let's assume that ω, k, and φ are positive, in which case

(i) You will be moving to the right, since $\omega t - kx + \varphi = \alpha$ implies that

$$\frac{dx}{dt} = \frac{\omega}{k} > 0$$

for any fixed value of φ.

(ii) Similarly, in going from equation (1) to (2), the phase shift by φ will move you backwards, since

$$\frac{dx}{d\varphi} = -\frac{1}{k} < 0$$

for any fixed value of t.

With some trigonometric manipulations we can rewrite (2) as a sum of two waves, one of the type in equation (1) but with the amplitude changed to $A\cos(\varphi)$, and the other with the amplitude $-A\sin(\varphi)$ and $\cos(\omega t - kx)$ changed to $\sin(\omega t - kx)$:

$$\psi(x,t) = A\cos(\omega t - kx + \varphi) = A\cos(\varphi)\cos(\omega t - kx) - A\sin(\varphi)\sin(\omega t - kx)$$

There is a much more convenient way of representing this wave, and waves in general, using complex variables and Euler's equation[1]:

$$A\cos(\omega t - kx + \varphi) = \mathrm{Re}(Ae^{i\varphi}e^{i(\omega t - kx)}) \tag{3}$$

where $Re(a + bi) = a$ for any complex number $a + bi$.

A key physical property of waves, as they are used in most applications, is that the intensity (think of the brightness of a beam of light or the impact of a water wave) is proportional to the square of the amplitude. So in both of the two cases, equations (1) and (2), the intensity is proportional to A^2. Another key property is that the sum (or "superposition") of two waves with common angular frequency ω is just another wave with angular frequency ω.

[1] $$e^{i\theta} = \cos(\theta) + i\sin(\theta)$$

1.2 The two-slit experiment

The classical two-slit experiment (see Figure 1) involves letting a plane wave of (classical) light pass through two slits and hit a screen (or camera) on the other side. On the screen, we observe an interference pattern due to the superposition of the two light waves (ψ_1 and ψ_2) emerging from the two slits, with varying intensity (brightness) at different locations on the screen. As will be shown in §1.3, this phenomenon can be explained using the appropriate wave representations. The resulting intensity, at a position on the screen will be proportional to the square of the amplitude of $\psi_1 + \psi_2$ at that position.

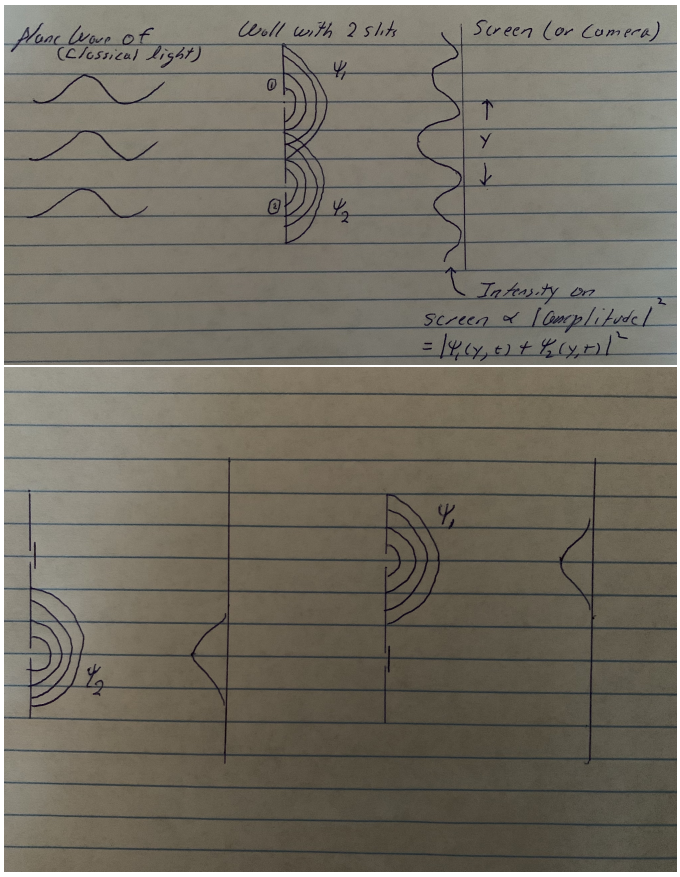


Figure 1: **Thomas Young's classical two-slit experiment, 1801. Top:** At every location and every time, the two waves add up linearly. The result is a degree of cancelation (opposite signs) or reinforcement (same sign). The "interference pattern" observed on the screen (or by the camera) is the recorded intensity (or brightness) at each location on the screen. **Bottom:** Recorded intensities when top (left-hand side) or bottom (right-hand side) slit is blocked. Because of interference, the intensity recorded in the top figure is not the sum of the intensities recorded in the two bottom figures.

As the relevant technologies have improved, the experiment has since been extended in a number of directions. In one direction, detectors are sufficiently sensitive that the strength of the entering plane wave can be reduced by many orders of magnitude. It turns out that for low enough intensity of the entering light wave, the pattern can be shown to emerge from single (temporally separated) discrete recordings, as

would be expected from the photon (or particle) theory of light. In this case, it becomes evident that the intensity of the emergent pattern is proportional to the *probability* that a given photon will land at a given position on the screen (indicated by the value of y). This is verified by suitably discretizing, or "binning," the continuous space of locations, and then computing the relative frequencies of photons arriving in each bin.

The inescapable conclusion is that photons, although particle-like when they strike the screen, are governed by classical wave equations—as in §1.3—during their transit through the slits and then onward to the screen. But what about other particles, like electrons or protons or even whole atoms? Beginning in the 1960s, ever-bigger particles have been shown to obey the same rules: wave-like behavior in transit, resulting in an interference pattern, followed by particle-like detection, and with the probability of detection at any particular location on the screen determined, with precision, by the normalized squared amplitude of the re-combined (post-slit) waves at that location. And if one or the other slit is blocked (as in the bottom frame of Figure 1), then the interference pattern disappears, and hence the result can not be predicted by simply summing the single-slit results. *In fact, any measurement that determines which of the two slits the particle passes through will destroy the interference pattern—the measurement itself forces the particle to follow one path or the other, just as though one were were closed and the other open.* In 2019 the experiment was performed with molecules consisting of more than 2,000 atoms.

1.3 Calculation of the interference pattern

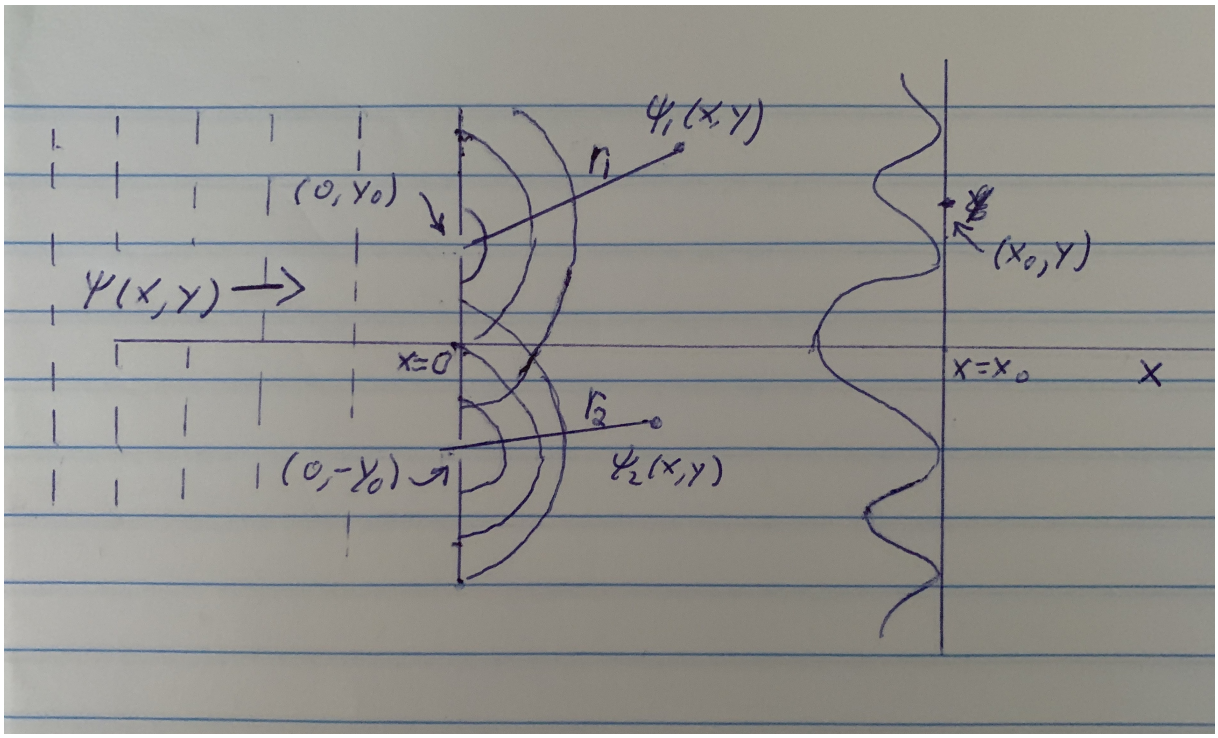


Figure 2: **A simplified "two-slit experiment".** These simplified waves depend only on x, y, and t, (as opposed to x, y, z, and t). The waves enter from the left and travel to the right. On the left side of the barrier (located at $x = 0$), they are "plane waves," meaning that their heights are independent of y. Those that pass through the two slits in the barrier expand circularly and add linearly until they reach the "screen," located at $x = x_o$, along which their intensities (squared amplitudes) exhibit an interference pattern.

With reference to Figure 2, a two-dimensional plane wave enters from the left and travels in the positive x direction. To the left of the slits, the wave can be modeled as

$$\psi(x,t) = A\cos(\omega t - kx)$$

(You can think of ψ, at location (x, y) and time t, as the strength of an electric field, in the case of light, or the density of air, in the case of sound.) We have chosen coordinates such that the vertical wall is situated at $x = 0$, and its two slits are at positions $(0, +y_o)$ and $(0, -y_o)$. If we treat the slits as infinitely narrow, then after the wave passes through one its energy (which is proportional to $|A|^2$) will be evenly spread out in a half circle centered at the slit. Therefore, the amplitude of the wave decreases with r like $1/\sqrt{\pi r}$, where r is the distance from the slit. (If the slits were, instead, small holes, and if the waves were in three dimensions, then the amplitude would decrease like $1/\sqrt{4\pi r^2}$.) Since we will only be concerned with *relative* amplitudes (intensities will be normalized to get probabilities), the two waves coming out of the slits can be modeled as

$$\psi_1(r_1,t) = \frac{A}{\sqrt{r_1}}\cos(\omega t - kr_1) \quad \text{and} \quad \psi_2(r_2,t) = \frac{A}{\sqrt{r_2}}\cos(\omega t - kr_2),$$

where r_1 is the distance from $(0, y_o)$ and r_2 is the distance from $(0, -y_o)$:

$$r_1 = \sqrt{x_o^2 + (y - y_o)^2} \quad \text{and} \quad r_2 = \sqrt{x_o^2 + (y + y_o)^2}. \tag{4}$$

And to further simplify notation we let

$$A_1 = \frac{A}{\sqrt{r_1}} \quad \text{and} \quad A_2 = \frac{A}{\sqrt{r_2}} \tag{5}$$

What we are after is the wave intensity, $I(y)$, for every location y on the screen. And since waves add linearly, to compute I we will need to compute the squared amplitude of $\psi_1(r_1, t) + \psi_2(r_2, t)$. To calculate this, we can take advantage of the complex representation to write the sum as

$$\begin{aligned} W(y,t) &\doteq \psi_1(r_1,t) + \psi_2(r_2,t) \\ &= A_1 \cos(\omega t - kr_1) + A_2 \cos(\omega t - kr_2) \\ &= \text{Re}\left(A_1 e^{i(\omega t - kr_1)} + A_2 e^{i(\omega t - kr_2)}\right) \\ &= \text{Re}\left(e^{i\omega t}\left(A_1 e^{-ikr_1} + A_2 e^{-ikr_2}\right)\right) \end{aligned}$$

For fixed y, $A_1 e^{-ikr_1} + A_2 e^{-ikr_2}$ is a complex variable. Every complex variable $z = a + ib$ can be put into its polar form $z = Re^{i\phi}$, where $R = \sqrt{a^2 + b^2}$ is the modulus (aka magnitude) and $\phi = \arctan b/a$ is the angle (aka phase). Hence we can write

$$A_1 e^{-ikr_1} + A_2 e^{-ikr_2} = R(y) e^{i\phi(y)} \tag{6}$$

in which case

$$W(y,t) = \text{Re}\left(e^{i\omega t} R(y) e^{i\phi(y)}\right) = R(y)\cos(\omega t + \phi(y))$$

where $R(y)$ is the wave amplitude and $I(y) = R(y)^2$ is its intensity.

It remains to compute R^2, which can be done by applying the law of cosines to the expression in (6)–see Figure3:

$$R^2 = A_1^2 + A_2^2 - 2A_1A_2\cos(\pi - k(r_1 - r_2)) = A_1^2 + A_2^2 + 2A_1A_2\cos(k(r_1 - r_2))$$

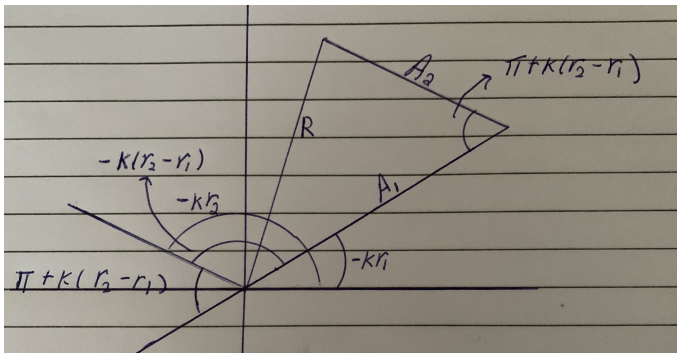


Figure 3: **Law of Cosines** If $\theta = \pi + k(r_2 - r_1)$, then $R^2 = A_1^2 + A_2^2 - 2A_1A_2\cos(\theta)$.

Finally, use in the expressions for r_1, r_2, A_1, A_2 given in equations (4) and (5) to write $I(y)$ $(= R(y)^2)$ as

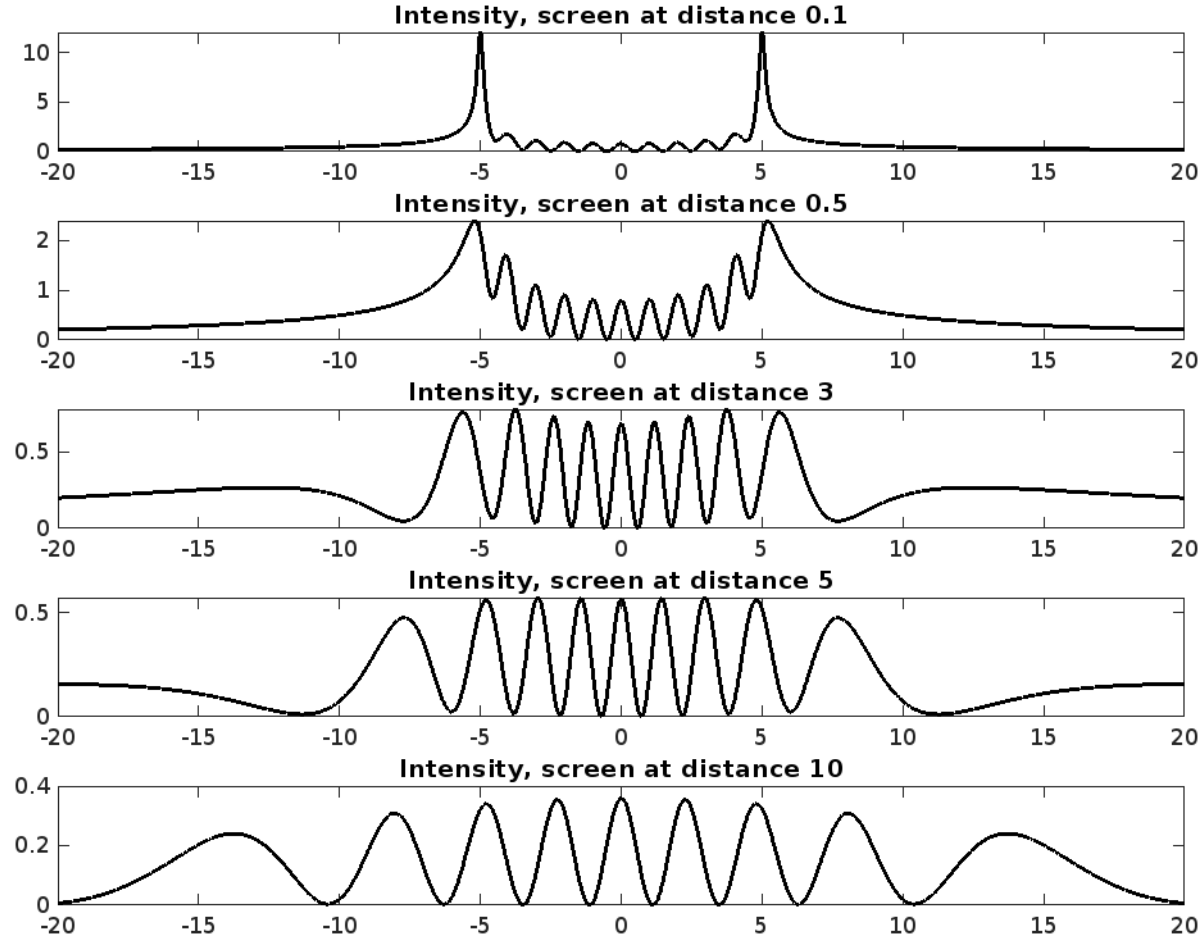


Figure 4: Squared amplitude of the waves as a function of location on screen in the two-slit experiment.

an explicit function of y:

$$I(y) = \frac{A^2}{\sqrt{x_o^2 + (y - y_o)^2}} + \frac{A^2}{\sqrt{x_o^2 + (y + y_o)^2}} + 2\frac{A^2}{\sqrt[4]{(x_o^2 + (y - y_o)^2)(x_o^2 + (y + y_o)^2)}} \cos\left(k\left(\sqrt{x_o^2 + (y - y_o)^2} - \sqrt{x_o^2 + (y + y_o)^2}\right)\right)$$

The plots in Figure 4 show the squared amplitude $I(y) = R(y)^2$ of the superposition of the waves emenating from the two slits, as a function of y, when the slits are positioned at $\pm y_o$, where $y_o = 5$, and the original amplitude is $A = 1$. Each plot corresponds to a different location, x_o, of the screen.

2 Single Particle Spins and their State Space

Elementary particles possess a mysterious property known as spin, or more descriptively, as intrinsic angular momentum. The spin of charged elementary particles (such as electrons, positrons, and muons) will cause them to be deflected when passing through a magnetic field. We can think of this, very loosely, as though the charge were spinning in a loop and creating a magnetic dipole. The Stern-Gerlach experiments were designed to demonstrate and measure the spin of electrons, and, as we shall see shortly, these measurements reinforce the principal lessons of the two-slit experiments: wave-particle duality and the essential roles of probability and the observer in the quantum mechanical world. In §2.1, we will describe the results of these experiments and learn how to predict (probabilistically!) the outcomes of various manipulations of the spin state. In §2.2, we will develop a mathematical description of the spin state that will capture the discrete analog of the linear interference that is at the heart of the two-slit experiment. We will then be in a position to meaningfully formulate the role of an observation, which will be captured, conveniently, using simple (though complex-valued) matrices (§2.3). In §2.4, we will use this formalism to state and prove a general version of the Heisenberg Uncertainty Principle. Finally, in §2.5, we will give a brief derivation of the Schrödinger equation for non-relativistic systems.

2.1 The Stern-Gerlach experiments and their empirical probability rules

The Stern-Gerlach experiment was formulated by Otto Stern in 1921 and subsequently performed by Walther Gerlach in 1922. It tested a prediction of the Bohr-Sommerfeld Theory, which asserted the quantization of an atom's angular momentum. The first tests were performed with silver atoms. The results (see Figure 5) were consistent with the Bohr-Sommerfeld Theory, though it was several more years before the connection was made to the intrinsic angular momentum of electrons. For our purposes, we can pretend like individual electrons, rather than atoms, are used, though this turns out to be a more difficult experiment requiring a more elaborate machine.[2] What is important is that the main conclusions are compelling and well supported by many experimental approaches.

In our version of the experiment, electrons are shot through an inhomogeneous magnetic field resulting in a deflection of their paths that is proportional to the strength of their spins. The actual direction of travel of the deflected particle (e.g. up and down in Figure 5) will depend on the orientation of the machine. Classically, we would expect a continuous range of deflections resulting from a continuous range of strengths of the intrinsic angular momentum. However, carrying out the experiment reveals that there are only two possible outcomes. For a machine oriented in an arbitrary direction, say $\vec{\mathbf{n}}$, the spin is deflected in the $\vec{\mathbf{n}}$ direction by an amount consistent with a magnetic spin strength that is either $+\hbar/2$ relative to $\vec{\mathbf{n}}$ (resulting in an upward deflection at a certain angle) or $-\hbar/2$ (resulting in a deflection at the the same angle but in the opposite direction), where $\hbar$ denotes the reduced Planck constant ($\sim 1.055 \times 10^{-34}\,\mathrm{J}\cdot\mathrm{s}$).

[2] Part of the challenge stems from the charge on the electron, which would cause a sizable deflection with or without spin. This can be compensated, but the device is more complicated.

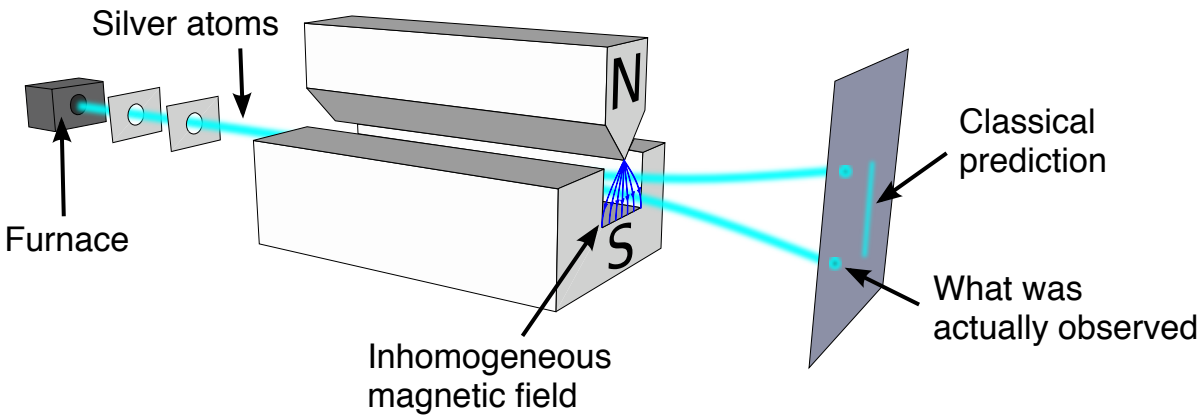


Figure 5: **The setup of the Stern-Gerlach experiment.** Silver atoms were used instead of electrons, since they have the same magnetic moment but are much more massive, making them easier to work with. The surprising result of the experiment is that the beam is bent, up or down, by a fixed amount, rather than the continuum of deflections that might arise from a continuum of spin strengths, as in a more classical picture. We will simplify our discussion by pretending that it is a beam of electrons rather than atoms being directed through the device.

The situation is summarized in Figure 6, where $\vec{\mathbf{n}} \in \mathbb{R}^3$ is a unit vector pointing in an arbitrary direction, and the output is indicated by simply writing $\lambda_{\vec{\mathbf{n}}} = \pm 1$, depending on whether the deflection is in the positive or negative $\vec{\mathbf{n}}$ direction ("spin up" or "spin down") respectively. Going forward, we will usually drop the vector indication and simply write $\mathbf{n}$ instead of $\vec{\mathbf{n}}$, and we will indicate that a vector is a unit vector (length 1) by writing $\mathbf{n} \in \mathbb{R}^3_1$, where $\mathbb{R}^3_1 \doteq \{\vec{\mathbf{n}} \in \mathbb{R}^3 : |\vec{\mathbf{n}}| = 1\}$ (more commonly referred to as S^2).

What would happen if we were to pass the electron through multiple Stern-Gerlach devices? It turns out that if we collect the electrons that were "prepared" in the $\mathbf{n}$ direction at some time t (i.e. $\lambda_{\mathbf{n}}(t) = 1$), and measure them again in the $\mathbf{n}$ direction at some time later, $t+1$, then (assuming that they remained undisturbed) all of them will again deflect in the positive $\mathbf{n}$ direction. Similarly, if the electrons were "prepared" in the $-\mathbf{n}$ direction, then they will remain in that direction (see Figure 7). In summary:

$$\mathbb{P}(\lambda_{\mathbf{n}}(t+1) = 1 \mid \lambda_{\mathbf{n}}(t) = 1) = \mathbb{P}(\lambda_{\mathbf{n}}(t+1) = -1 \mid \lambda_{\mathbf{n}}(t) = -1) = 1 \tag{7}$$

$$\text{and} \quad \mathbb{P}(\lambda_{\mathbf{n}}(t+1) = -1 \mid \lambda_{\mathbf{n}}(t) = 1) = \mathbb{P}(\lambda_{\mathbf{n}}(t+1) = 1 \mid \lambda_{\mathbf{n}}(t) = -1) = 0 \tag{8}$$

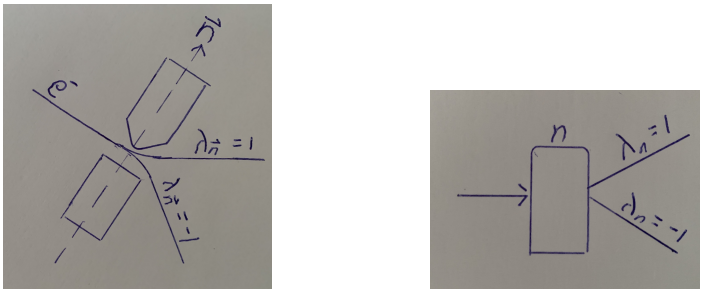


Figure 6: **Representation of a Stern-Gerlach machine. Left:** The machine is oriented at an arbitrary direction in three dimensional space, indicated by the unit-vector $\vec{\mathbf{n}}$. The deflection, which is either in the direction $\vec{\mathbf{n}}$ or $-\vec{\mathbf{n}}$, is indicated by writing $\lambda_{\vec{\mathbf{n}}} = 1$ or $\lambda_{\vec{\mathbf{n}}} = -1$, respectively. **Right:** Simplified representation of the figure on the left.

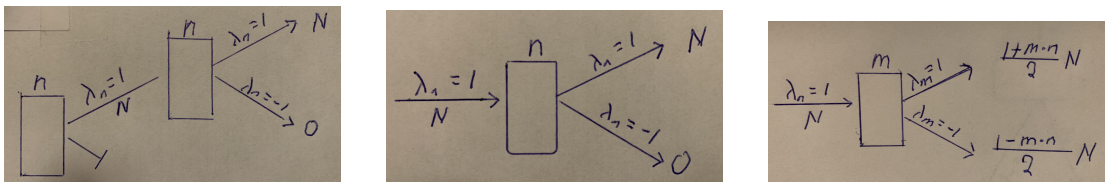


Figure 7: **Two measurements. Left:** N electrons are prepared in the positive $\mathbf{n}$ direction, and then passed through a second Stern-Gerlach machine positioned in the same direction. **Middle:** Same as left, but, more simply, using $\lambda_{\mathbf{n}} = 1$ to represent the N selected ("prepared") electrons. **Right:** N electrons are prepared in the positive $\mathbf{n}$ direction, and then passed through a second Stern-Gerlach machine positioned in the $\mathbf{m}$ direction. The two possible outputs are annotated with the expected number of electrons traveling in each direction.

This result is not particularly surprising, especially given the word "intrinsic" used to describe the magnetic spin of an electron. If the strength of this magnetic spin is a constant of nature, and if the effect of the Stern-Gerlach machine is to re-orient the electron so that it's spin aligns perfectly with $\pm\mathbf{n}$, then it makes sense that a repeated measurement would recover the same alignment. But what happens if we re-orient the second machine to a different direction, say $\mathbf{m} \in \mathbb{R}^3_1$? The result is again quantized, and the direction is again deflected by an amount consistent with a magnetic moment of magnitude $\hbar/2$, but the probabilities of the two outcomes depend on the orientation of the second machine *relative to* the first (see Figure 7, right-hand panel). In particular, the probability of each outcome depends only on $\mathbf{n} \cdot \mathbf{m}$:

$$\mathbb{P}(\lambda_{\mathbf{m}}(t+1) = 1 \mid \lambda_{\mathbf{n}}(t) = 1) = \frac{1 + \mathbf{m} \cdot \mathbf{n}}{2} \quad \text{and} \quad \mathbb{P}(\lambda_{\mathbf{m}}(t+1) = -1 \mid \lambda_{\mathbf{n}}(t) = 1) = \frac{1 - \mathbf{m} \cdot \mathbf{n}}{2} \tag{9}$$

Notice that (7) and (8) are special cases, with $\mathbf{m} = \mathbf{n}$ and $\mathbf{m} = -\mathbf{n}$ respectively. Since we will always be conditioning on the earlier event, from now on we well drop the time notation and, e.g., simply write $\mathbb{P}(\lambda_{\mathbf{m}} = 1 \mid \lambda_{\mathbf{n}} = 1)$ instead of $\mathbb{P}(\lambda_{\mathbf{m}}(t+1) = 1 \mid \lambda_{\mathbf{n}}(t) = 1)$.

Of course we do not actually observe the probabilities themselves, but we can gather the number of times that each path is taken. When N is large, the empirical probabilities match those specified in equation (9) with high accuracy. This is very much analogous to verifying the intensities calculated in §1.3 for the two-slit experiment by tallying bin counts.

Here is a handy observation that might make it easier to think about the results in (9):

$$\mathbb{P}(\lambda_{\mathbf{m}} = 1 \mid \lambda_{\mathbf{n}} = 1) = \cos^2 \frac{\theta}{2},$$

where θ is the angle between $\mathbf{m}$ and $\mathbf{n}$ (see homework). It follows, for example, that $\mathbb{P}(\lambda_{\mathbf{m}} = -1 \mid \lambda_{\mathbf{n}} = 1) = 1 - \cos^2 \frac{\theta}{2}$. Or, consider preparing an electron along the z-axis ($\lambda_{\mathbf{z}} = 1$) and then measuring along the y-axis: Since $\theta = \frac{\pi}{2}$, $\cos\left(\frac{\theta}{2}\right) = \cos\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}}$ and therefore

$$\mathbb{P}(\lambda_{\mathbf{y}} = 1 \mid \lambda_{\mathbf{z}} = 1) = \frac{1}{2}.$$

Notice that these probabilities are not in any way dependent on the particular choice of orthonormal vectors for the coordinate system, but rather they depend only on the angle between $\mathbf{m}$ and $\mathbf{n}$.

If we extend our experiments a bit more, something strange turns up.

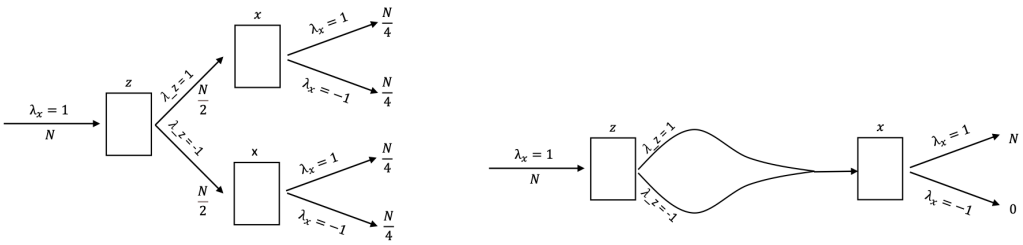


Figure 8: **Strange result from a Stern-Gerlach experiment.** We prepare N electrons along the x-axis, measure them along the z-axis, and then again along the x-axis. All together, about $\frac{N}{2}$ electrons end up with $\lambda_{\mathbf{x}} = 1$ (left-hand figure). But if we *superimpose* the two paths, then *all* of the electrons end up with $\lambda_{\mathbf{x}} = 1$ (right-hand figure).

Suppose we prepare a large number, N, of electrons along the positive x-axis, and then measure them along the z-axis. Then about $N/2$ will exit with $\lambda_{\mathbf{z}} = 1$, and the rest with $\lambda_{\mathbf{z}} = -1$. If we then take those with $\lambda_{\mathbf{z}} = 1$ and measure them along the x-axis, then about $N/4$ will exit with $\lambda_{\mathbf{x}} = 1$ and the other half with $\lambda_{\mathbf{x}} = -1$. Similarly, for the approximately $N/2$ electrons with $\lambda_{\mathbf{z}} = -1$, if we measure them along the x-axis, about $N/4$ of them will come out with $\lambda_{\mathbf{x}} = 1$ and the rest with $\lambda_{\mathbf{x}} = -1$. At the end of the experiment, in total about half of the N electrons will have $\lambda_{\mathbf{x}} = 1$—see left-hand panel in Figure 8. These numbers are all consistent with our earlier observations, as summarized by the formula in equation (9).

However, if by some manipulations (e.g. using an electric field) we were to bend the trajectories of the $\lambda_{\mathbf{z}} = 1$ electrons downward and those of the $\lambda_{\mathbf{z}} = -1$ electrons upward so as to merge the two trajectories into one (as illustrated in the right-hand panel of Figure 8), and then feed the common path into a single $x-$type Stern-Gerlach machine, then we would discover that they *all* deflect in the positive, $\lambda_{\mathbf{x}} = 1$, direction! Why are these two setups so different? Are the electrons going through **both** paths out of the z-axis measurement and then *interfering* with themselves? Evidently, they are.

Feynman's "improved" Stern-Gerlach device. To help us think clearly about these results, Feynman suggested imagining a different kind of Stern-Gerlach device, one that has a built-in option of recombining the two paths that represent the two outcomes of a traditional Stern-Gerlach machine. The machine, and its different configurations, as well as a handy short-hand notation, are depicted in Figure 9.

By orienting Feynman's device in an arbitrary direction $\mathbf{n}$ and blocking bottom or top path we can prepare electrons in either $\lambda_{\mathbf{n}} = 1$ or $\lambda_{\mathbf{n}} = -1$ spin state. We can thus reproduce different aspects of the results reported in Figure 8. Consider for example the two set-ups in Figure 10. It is perhaps a little startling to realize that *more* electrons get through the configuration in the top panel than in the bottom panel. Sometimes, more is less!

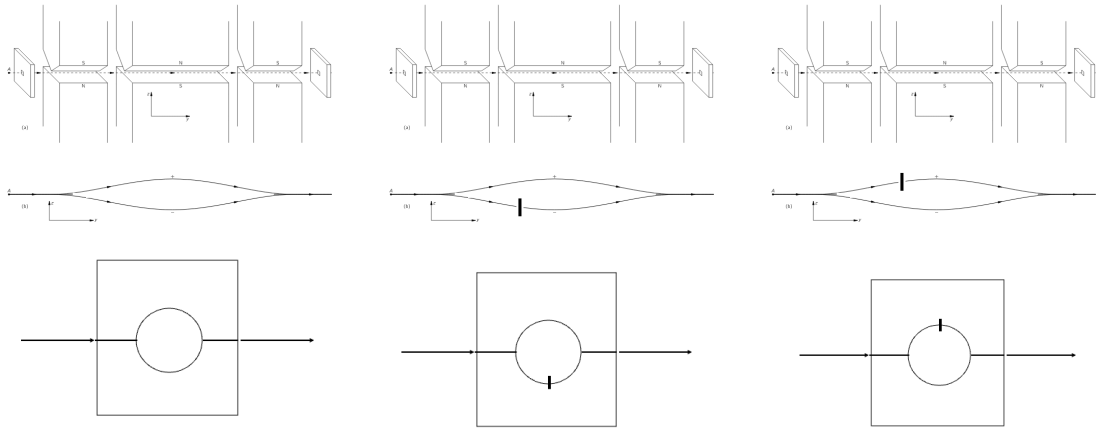


Figure 9: **Feynman's Stern-Gerlach device. Top row:** Each device consists of three pairs of magnets, with alternating orientations, and for which the middle magnet is twice as wide as the outer two magnets. **Middle row:** Blow-ups of the two possible trajectories for electrons. If the particle has a magnetic spin that is oriented parallel to the machine, then the first magnet will bend the trajectory upward, the second will bend it downward, and the third will again bend it upward. The net result is that the particle leaves the machine along the same trajectory at which it entered the machine. By the same reasoning, if the particle has a magnetic spin oriented anti-parallel to the machine, then it will take the lower of the two paths depicted in the drawing. Feynman's (imaginary) machine permits blocking either path: from left to right, both paths open, only top path open, only bottom path open. **Bottom row:** Corresponding symbolic representations of the three configurations.

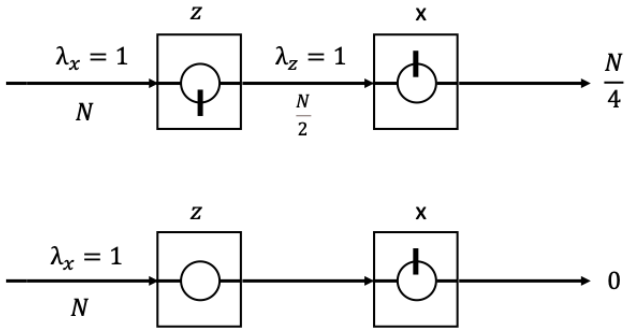


Figure 10: **More is less.** Blocking the lower path of the **z** device results in about $N/4$ electrons passing through the second device, but if we don't block either path of the **z** device then there are no electrons left!

2.2 Formulation of the spin state

The state of a classical system is defined by the positions and velocities of all of its particles. This, plus a complete accounting of charges, fields, and masses, is sufficient to determine the state at any time in the future. In principle, this is easy: just integrate Newton's laws of motion. (In practice, this is hard: consider the infamous "three-body problem.") We can think of the classical state as a minimal sufficient description, in the sense that it is enough to know the future *deterministically*. But in quantum mechanics it appears that the best that we can do is to know the future probabilistically. If we are forced to accept this, and it seems that we are,[3] then what minimal description will serve the purpose? For our simple one-particle spin systems, we already have a solution: Knowing the future probabilistically (i.e. predicting the outcome of any measurement experiment) corresponds to knowing that a particle has been prepared in a particular state, say $\lambda_{\mathbf{n}} \in \mathbb{R}^3_1$. In this case, the probabilities that a future measurement in any direction, say $m \in \mathbb{R}^3_1$, results in $\lambda_{\mathbf{m}} = 1$ or $\lambda_{\mathbf{m}} = -1$, are given by the formulas in equation (9). The set of all possible states (i.e. the state space) is simply $\mathbb{R}^3_1$.

We could declare victory and move on, but that would be a mistake. Among the tasks ahead, virtually all include the interactions ("entanglements") of multi-particle spin systems, and these require elaborate manipulations of the collective state space. For these purposes $\mathbb{R}^3_1$ is much too cumbersome. Instead, we will follow the example set earlier in the treatment of waves (§1) and look for a representation that exploits the conveniences of complex-valued variables.

Consider again the wave intensities displayed in Figure 4. Recall that these intensity functions turned out to be unnormalized probability densities, with one value for every possible location on the screen, i.e. for every $y \in \mathbb{R}$. Of course we can't actually measure all of these values, but instead we use a large number of photons or other particles and accumulate statistics within small, disjoint, bins—each defining a small interval of possible locations. The number of particles falling into a particular bin, divided by the total number of particles received, approaches the probability of a particle landing in the chosen bin. If we bin the vertical axis into, say, twenty-five bins (the lowest and highest would have to extend to $-\infty$ and $+\infty$, respectively), then we can think of the (discretized) intensity function as a vector in $\mathbb{R}^{25}$. What is more, each element of the vector comes from squaring the modulus of an associated complex variable, i.e. squaring the amplitude of the complex-valued representation of the (discretized) wave. Hence, up to discretization, the probability of the observable (i.e. the bin in which the particle lands) is completely described by twenty-five complex-valued numbers. (Of course we would need to have more and more bins, and make them smaller and smaller, in order to approach a continuous complex-valued wave and a real-valued intensity function.)

The spin state is much simpler. For any given orientation of the apparatus, there are only two bins, one for each of the two possible outcomes of a Stern-Gerlach experiment. In other words, the natural vector space for our purpose is simply $\mathbb{C}^2$, i.e. the set of ordered pairs of complex numbers. And this spotlights a challenge: find a correspondence between $\mathbb{R}^3$ and $\mathbb{C}^2$ such that we can reproduce the probabilistic predictions embodied in equation (9), but using only the $\mathbb{C}^2$ representations corresponding to each m and

[3] We will discuss this some more when we talk about the objections of Einstein, Rosen, and Podolsky, and the connections to Bell's inequality (§4.2).

n.

For this purpose, and for the rest of the course, we will make constant use of complex variables, so before going on let's pause to enumerate some of the basic definitions, facts, and notation:

Complex variables, inner products, and Dirac's bra-ket notation.

(1) $\mathbb{C}$**, complex variables**

i. The *complex plane*: $\mathbb{C} \doteq \{\alpha_1 + i\alpha_2 : \alpha_1, \alpha_2 \in \mathbb{R}\}$, where the *imaginary number* i has the property $i^2 = -1$.

ii. The horizontal axis in the complex plane is called the *real* axis and the vertical axis is called the *imaginary* axis

iii. The *real* part of $\alpha_1 + i\alpha_2$ is $Re(\alpha) = \alpha_1$ and the *imaginary* part is $Im(\alpha) = \alpha_2$

iv. The *modulus* or *absolute value* of a complex number: if $\alpha = \alpha_1 + i\alpha_2 \in \mathbb{C}$ then $|\alpha| = \sqrt{\alpha_1^2 + \alpha_2^2}$

v. The *polar representation*: if $\alpha = \alpha_1 + i\alpha_2 \in \mathbb{C}$, then $\alpha = Re^{i\phi}$, where $R \in [0, \infty)$ is the *modulus*, and $\phi \in [0, 2\pi)$ is the *phase*, which is the angle, in the complex plane, between α and the real axis

vi. Euler's formula: $e^{i\phi} = \cos(\phi) + i\sin(\phi)$

vii. The *complex conjugate* of $\alpha = \alpha_1 + i\alpha_2$, denoted α^*, is the complex number $\alpha_1 - i\alpha_2$

viii. (basic operations) If $a \in \mathbb{R}$ and $\alpha, \beta \in \mathbb{C}$, where $\alpha = \alpha_1 + i\alpha_2$ and $\beta = \beta_1 + i\beta_2$, then

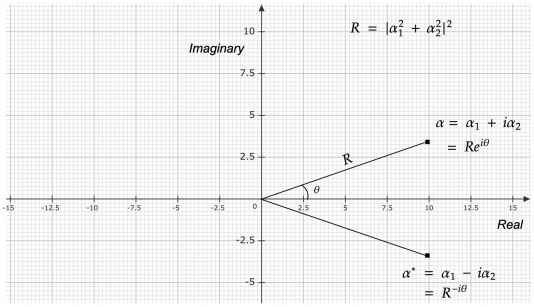
(1) $a\alpha = a\alpha_1 + ia\alpha_2$

(2) $\alpha + \beta = (\alpha_1 + \beta_1) + i(\alpha_2 + \beta_2)$

(3) $\alpha\beta = (\alpha_1 + i\alpha_2)(\beta_1 + i\beta_2) = (\alpha_1\beta_1 - \alpha_2\beta_2) + i(\alpha_1\beta_2 + \beta_1\alpha_2)$

(4) If $\alpha = R_\alpha e^{i\theta_\alpha}$ and $\beta = R_\beta e^{i\theta_\beta}$ then $\alpha\beta = R_\alpha R_\beta e^{i(\theta_\alpha + \theta_\beta)}$

(5) $\alpha^*\alpha = \alpha\alpha^* = |\alpha|^2$



Complex Plane

(2) **$\mathbb{C}^2$, the two-dimensional complex vector space**

As we mentioned earlier, the vector space $\mathbb{C}^2$ will play the role for spin states that was played earlier by wave functions in the two-slit experiment. In fact, we will often use the same terminology: the elements of $\mathbb{C}^2$ will variously be called states, spin states or wave functions.[4]

i. $\mathbb{C}^2 \doteq \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} : \alpha, \beta \in \mathbb{C} \right\}$

ii. If $a \in \mathbb{C}^2$, $a = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, then its *conjugate* is $a^* \doteq \begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix}$ and its *adjoint* (also called its *transpose*) is $a' \doteq a^T \doteq \begin{pmatrix} \alpha & \beta \end{pmatrix}$

iii. If $a \in \mathbb{C}^2$, $a = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, then its *conjugate transpose* (also called its Hermitian adjoint) is $a^\dagger \doteq \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix}$, and $(a^\dagger)^\dagger = a$

iv. *Inner product*: if $a, b \in \mathbb{C}^2$, with $a = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ and $b = \begin{pmatrix} \delta \\ \gamma \end{pmatrix}$, then

$$a \cdot b \doteq a^\dagger b = \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix} \begin{pmatrix} \delta \\ \gamma \end{pmatrix} = \alpha^* \delta + \beta^* \gamma$$

In physics, the inner product is almost never represented using the dot-product notation, $a \cdot b$. The tradition, instead, is to use the highly versatile "bra-ket" notation (pronounced "bra" as in brassiere, and "ket" as in kettle).

(3) **Dirac's bra-ket notation.**

i. We will use $|\psi\rangle$ to denote a generic state in $\mathbb{C}^2$,

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad \text{for some } \alpha, \beta \in \mathbb{C}$$

and $|n^+\rangle$ to denote the specific state that results from observing $\lambda_{\mathbf{n}} = 1$, for a particular $\mathbf{n} \in \mathbb{R}^3_1$. Similarly, $|n^-\rangle$ will correspond to the observation $\lambda_{-\mathbf{n}} = 1$ (which is the same as $\lambda_{\mathbf{n}} = -1$). Generically, $|\psi\rangle$ is called a "ket vector," and the conjugate transpose of $|\psi\rangle$ is the "bra vector" $\langle\psi|$

$$\langle\psi| = \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix}$$

ii. Given $|a\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, $|b\rangle = \begin{pmatrix} \delta \\ \gamma \end{pmatrix}$, with both in $\mathbb{C}^2$, define $\langle a|b\rangle$ ("bracket a,b"):

$$\langle a|b\rangle = \begin{pmatrix} \alpha^* & \beta^* \end{pmatrix} \begin{pmatrix} \delta \\ \gamma \end{pmatrix} = \alpha^* \delta + \beta^* \gamma.$$

[4] Perhaps "function" seems an odd designation, but it is worth getting used to. Of course there's nothing odd when the domain of the function is, say, the real line or plane or some other continuum space, as in the wave and intensity functions discussed for the two-slit experiment. But it is really no different when the real line gets replaced by, say, 25 bins: the function becomes a vector, but the entries of the vector are just assignments of numbers to each of the components. In other words, a vector *is a function*, one which assigns a number to each element in a chosen basis.

In other words, $\langle a|b\rangle$ replaces the dot product notation for the inner product. When $\langle a|b\rangle = 0$ we say that $|a\rangle$ and $|b\rangle$ are orthogonal. (In general, inner products bring the basic elements of geometry to a vector space, including orthogonality, projections, and the generalized parallelogram laws.)

iii. If $|b\rangle = |a\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, then

$$\big||a\rangle\big| \doteq \sqrt{\langle a|a\rangle} = \sqrt{|\alpha|^2 + |\beta^2|}$$

which is called the **norm** of $|a\rangle$. This is the analog of length for a complex vector space.

iv. Verify for yourself that for all $a, b, c \in \mathbb{C}^2$ and all $\alpha, \beta \in \mathbb{C}$,

$$\begin{aligned}
|\langle a|b\rangle| &= |\langle b|a\rangle| \\
\langle a|b\rangle^* &= \langle b|a\rangle \\
\langle c|\alpha a + \beta b\rangle &= \alpha\langle c|a\rangle + \beta\langle c|b\rangle && \text{bracket is } \textit{linear} \text{ in its second argument} \\
\langle \alpha a + \beta b|c\rangle &= \alpha^*\langle a|c\rangle + \beta^*\langle b|c\rangle && \text{bracket is } \textit{conjugate linear} \text{ in its first argument}
\end{aligned}$$

Return now to the problem at hand. We have already established that the spin state is adequately captured by a unit vector in $\mathbb{R}^3$, which is to say an element of $\mathbb{R}^3_1$. Our goal is to translate this representation into a complex-valued wave representation, in which the probabilistic interpretations embodied in (9) are more transparent and convenient.

How will we interpret $|\psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$? In the two-slit experiment, *intensities (squared amplitudes of waves) are proportional to probabilities*. Similarly, in the Stern-Gerlach experiments the square of the moduli of the state vector components, $|\alpha|^2$ and $|\beta|^2$, will be interpreted as "*probability amplitudes*," proportional to the probabilities of an up or down deflection in some direction, depending on the choice of basis in $\mathbb{C}^2_1$.[5] We can always assume that these state vectors are normalized, in which case $|\alpha|^2 + |\beta|^2 = 1$ and we might as well simplify our task by considering only those elements of $\mathbb{C}^2$ that have unit norms. Using notation analogous to $\mathbb{R}^3_1$, it will be enough to find, for every $\mathbf{n} \in \mathbb{R}^3_1$, a corresponding element $|n^+\rangle$ in

$$\mathbb{C}^2_1 \doteq \left\{ \begin{pmatrix}\alpha\\ \beta\end{pmatrix} : \alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1 \right\} \subset \mathbb{C}^2$$

If we know that $|\psi\rangle = |n^+\rangle$ then we know the future, probabilistically.

An important feature of (9), already highlighted, is that the probability of a positive or negative deflection in the direction $\mathbf{m} \in \mathbb{R}^3_1$, given a state prepared in the direction $\mathbf{n} \in \mathbb{R}^3_1$, depends only the inner product $\mathbf{n} \cdot \mathbf{m}$, and hence just the angle between the two unit vectors. This intuitive property has the convenient consequence that the coordinate system is arbitrary (in the sense that we can label any three orthogonal unit vectors as $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$).

[5] As a more general principle, this use of the squared absolute value of wave amplitudes to represent probabilities is known as Born's rule, embodied in the *Copenhagen interpretation* of quantum mechanics.

We will now show that we can preserve this property by suitably constructing the map $\mathbf{n} \to |n^+\rangle$ in such a way that

$$\frac{1+\mathbf{m}\cdot\mathbf{n}}{2} = \left|\langle m^+|n^+\rangle\right|^2 \quad \forall \mathbf{m},\mathbf{n} \in \mathbb{R}^3_1 \tag{10}$$

and hence, in $\mathbb{C}^2_1$,

$$\mathbb{P}(\lambda_m = 1 \mid |\psi\rangle = |n^+\rangle) = \left|\langle m^+|n^+\rangle\right|^2 \tag{11}$$

where we have conditioned on being in the state $|n^+\rangle$, which is the same as saying that we have observed $\lambda_n = 1$. One implication of (11) is that for any two states $|m^+\rangle$ and $|n^+\rangle$, the bigger the value of their inner product the less distinguishable they are: preparation in one of the associated directions predicts the result of an experiment in the other direction with high probability. For this reason, the inner product $\langle\psi_1|\psi_2\rangle$ between any two states $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^2_1$ is sometimes referred to as their "overlap.")

We will go about this construction in two steps. First we will find six vectors $|z^+\rangle$, $|z^-\rangle$, $|x^+\rangle$, $|x^-\rangle$, $|y^+\rangle$, and $|y^-\rangle$ in $\mathbb{C}^2_1$ that correspond to the six cardinal directions $\mathbf{z}$, $-\mathbf{z}$, $\mathbf{x}$, $-\mathbf{x}$, $\mathbf{y}$, and $-\mathbf{y}$ in $\mathbb{R}^3_1$, respectively. In the second step we will extend the correspondence by finding a vector $|n^+\rangle \in \mathbb{C}^2_1$ for every vector $\mathbf{n} \in \mathbb{R}^3_1$ in such a way that the relationship in (10) holds for every pair $\mathbf{m},\mathbf{n} \in \mathbb{R}^3_1$. (Some of the work in the first step, and all of the work in the second step, will be done as part of the assignment.)

We will make a special case of the six cardinal directions by typically writing $|u\rangle$, $|d\rangle$, $|r\rangle$, $|l\rangle$, $|i\rangle$ and $|o\rangle$ ("up", "down", "right", "left", "in", and "out") instead of $|z^+\rangle$, $|z^-\rangle$, $|x^+\rangle$, $|x^-\rangle$, $|y^+\rangle$, and $|y^-\rangle$. This will be convenient, especially since it lines up with the notation used by Susskind and Friedman.

Let us start with the up vector: $|u\rangle$. Since the coordinate system that we adopt in $\mathbb{C}^2$ is arbitrary, we are free to declare

$$|u\rangle = \begin{pmatrix}1\\0\end{pmatrix}$$

As for the down vector, write $|d\rangle = \begin{pmatrix}\alpha\\\beta\end{pmatrix} \in \mathbb{C}^2_1$. Since $z\cdot(-z) = -1$, the constraint in (10) implies

$$|\langle u|d\rangle|^2 = \frac{1+z\cdot(-z)}{2} = 0 \Rightarrow \langle u|d\rangle = 0 \Rightarrow \alpha = \begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix} = \langle u|d\rangle = 0$$

But since $1 = |\alpha|^2 + |\beta|^2 = |\beta|^2 = 1$, β must be of the form $\beta = e^{i\phi}$, for any "phase factor" $\phi \in \mathbb{R}$. Since ϕ is arbitrary, we are free to set it equal to zero in which case

$$|d\rangle = \begin{pmatrix}\alpha\\\beta\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}$$

Next, let us consider the representation of $|r\rangle$, starting with the generic representation $|r\rangle = \begin{pmatrix}\alpha\\\beta\end{pmatrix}$ where $|\alpha|^2 + |\beta|^2 = 1$. There are two constraints:

$$|\alpha|^2 = \left|\begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}\right|^2 = |\langle u|r\rangle|^2 = \frac{1+\mathbf{z}\cdot\mathbf{x}}{2} = \frac{1}{2} \Rightarrow \alpha = \frac{e^{i\phi}}{\sqrt{2}}$$

for some ϕ and, similarly,

$$|\beta|^2 = \left|\begin{pmatrix}0 & 1\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right|^2 = |\langle d|r\rangle|^2 = \frac{1-\mathbf{z}\cdot\mathbf{x}}{2} = \frac{1}{2} \Rightarrow \beta = \frac{e^{i\phi'}}{\sqrt{2}}$$

for some ϕ'. Both phases are free, and we are free to set both to zero:

$$|r\rangle = \begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}$$

Let's do one more. Start with $|l\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ and apply the constraints. The first two constraints on $|l\rangle$ are the same as those on $|r\rangle$, and therefore lead to the same conclusion: $|l\rangle = \begin{pmatrix}e^{i\phi}/\sqrt{2}\\ e^{i\phi'}/\sqrt{2}\end{pmatrix}$. If we were to set both phases to zero we would end up with $|r\rangle$. Evidently, these phases are no longer unconstrained, and that is because equation (10) imposes a third constraint:

$$|\langle r|l\rangle|^2 = \frac{1+\mathbf{x}\cdot-\mathbf{x}}{2} = 0$$

and hence

$$0 = \langle r|l\rangle = \begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}e^{i\phi}/\sqrt{2}\\ e^{i\phi'}/\sqrt{2}\end{pmatrix} = \frac{1}{2}e^{i\phi} + \frac{1}{2}e^{i\phi'} \Rightarrow e^{i\phi'} = -e^{i\phi}$$

Finally then,

$$|l\rangle = e^{i\phi}\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}}\end{pmatrix}$$

where the overall phase ϕ is free and we take it to be zero: $|l\rangle = \begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{2}}\end{pmatrix}$, which is the standard choice.

What is the meaning of these arbitrary phases? It will turn out that every spin representation in $\mathbb{C}_1^2$ has a built in phase ambiguity, which we might have anticipated right off the bat by simply counting parameters. Since every element in $\mathbb{R}_1^3$ determines a distinct state, the number of parameters needed to specify a state is just two (say, the two angles of its polar coordinates). But in $\mathbb{C}_1^2$, we start with four parameters—two for each of the complex numbers α and β in the generic representation $|\psi\rangle = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$—and then we lose one to the normalization $|\alpha|^2 + |\beta|^2 = 1$. This still leaves one free parameter for every direction in R_1^3. Therefore, we expect every state $\mathbf{n} \in \mathbb{R}_1^3$ to correspond to any member of a 1-parameter family of states in $\mathbb{C}_1^2$. That parameter is the overall (multiplicative) phase.

The homework includes computing the remaining two ($|i\rangle$ and $|o\rangle$) representations of the cardinal directions, and the development of an expression for the entire mapping from $\mathbb{R}_1^3$ to $\mathbb{C}_1^2$ that is consistent with the mappings of the cardinal directions and with the inner-product constraints embodied in equation (10). The phase ambiguities are explicitly represented in the final mapping.

The homework also includes two interesting facts about the *average* result of a Stern-Gerlach experiment:

(i) **Expectations.** Expectations in quantum mechanics are just like expectations in probability and statistics: Define $\mathbb{E}\big[\lambda_n \big| |\Psi\rangle = |m^+\rangle\big]$ ("the expected value of λ_n given that $|\Psi\rangle$ is in the state $|m^+\rangle$") to be the weighted average of the two possible values of λ_n:

$$\mathbb{E}\big[\lambda_n \big| |\Psi\rangle = |m^+\rangle\big] = (+1)\mathbb{P}(\lambda_n = 1 \big| |\Psi\rangle = |m^+\rangle) + (-1)\mathbb{P}(\lambda_n = -1 \big| |\Psi\rangle = |m^+\rangle)$$

Then $\mathbb{E}\big[\lambda_n \big| |\Psi\rangle = |m^+\rangle\big] = m \cdot n$

(ii) **Pythagorean Principle for Spins.**

$$E\big[\lambda_{\mathbf{x}}|\,|\psi\rangle\big]^2 + E\big[\lambda_{\mathbf{y}}|\,|\psi\rangle\big]^2 + E\big[\lambda_{\mathbf{z}}|\,|\psi\rangle\big]^2 = 1$$

for every $\psi \in \mathbb{C}_1^2$. (Caution: this can fail, dramatically, when $|\psi\rangle$ is *entangled* with the state of another particle. As we shall see a bit later.)

We conclude this section with some remarks and observations about the $\mathbb{C}_1^2$ wave representation:

1. **More general coordinate systems.** Another way to think about these relationships is to view them without reference to the arbitrary choice of the particular basis $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$:

$$|r\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle$$
$$|l\rangle = \frac{1}{\sqrt{2}}|u\rangle - \frac{1}{\sqrt{2}}|d\rangle$$

And these expressions generalizes immediately to bases consisting of other orthogonal pairs in $\mathbb{C}_1^2$, not just $|u\rangle$ and $|d\rangle$. Had we chosen, say, $\pm\mathbf{m}$ instead of $\pm\mathbf{z}$, then the elements of $\mathbb{C}_1^2$ would be expressed in terms of $|m^+\rangle$ and $|m^-\rangle$ instead of $|u\rangle$ and $|d\rangle$. Any other vector, for example $|n^+\rangle \in \mathbb{C}_1^2$, would have a representation of the form

$$\big|n^+\big\rangle = \alpha\,\big|m^+\big\rangle + \beta\,\big|m^-\big\rangle \tag{12}$$

And you can verify for yourself that in this case $\alpha = \langle m^+|n^+\rangle$ and $\beta = \langle m^-|n^+\rangle$:

$$\big|n^+\big\rangle = \big\langle m^+\big|n^+\big\rangle\,\big|m^+\big\rangle + \big\langle m^-\big|n^+\big\rangle\,\big|m^-\big\rangle \tag{13}$$

(simply multiply both sides of the equation, from the left, by $\langle m^+|$ and $\langle m^-|$ respectively).

2. **Figures 8 and 10, from a different point of view.** Our new formalism provides another way to look at the troubling results depicted in Figures 8 and 10, and perhaps add some new insight. Both figures show pretty much the same thing: if we block the path to the up deflection of a spin-right electron, then 50% of the time the electron will be absorbed and the rest of the time it will emerge as a spin-down electron. And if we block, instead, its downward deflection, then the we get either

absorbed or spin-down, also with equal likelihoods. We might guess, then, that if we do not block either pass, then we will get a spin-up electron or a spin-down electron with equal likelihoods. But instead we get back a spin-right electron. Evidently, when both paths are available it is *not* the case that the electron follows one path or the other, but instead it is taking *both* paths.

This agrees with our new formalism, as follows. The spin-right particle is in the state $|r\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle$, a superposition of the two waves, $\frac{1}{\sqrt{2}}|u\rangle$ which travels through the top path, and $\frac{1}{\sqrt{2}}|d\rangle$, which travel through the bottom path. When nothing is blocked, *both paths are taken* (as opposed to one or the other, with even odds) and then recombined by superposition (i.e. summed linearly), thereby reconstituting $|r\rangle$.

In fact these same considerations apply to any pair $|n^+\rangle$ and $|m^-\rangle$, and whether or not they are orthogonal. From (13), a particle prepared as $|n^+\rangle$ and directed into a Stern-Gerlach machine with $\mathbf{m}$ orientation will be deflected towards $\mathbf{m}$ or $-\mathbf{m}$ with probabilities $|\langle m^+|n^+\rangle|^2$ and $|\langle m^-|n^+\rangle|^2$, respectively. The positive deflection is represented by the wave $\langle m^+|n^+\rangle|m^+\rangle$ and the negative by $\langle m^-|n^+\rangle|m^-\rangle$, and if they are reunited then the result will be the wave $|n^+\rangle = \langle m^+|n^+\rangle|m^+\rangle + \langle m^-|n^+\rangle|m^-\rangle$.

3. **Mixture or superposition?** Here is yet another way to look at the distinction between one or another path being "chosen" as opposed to passing through both and then recombining. Consider again the representation of $|r\rangle$ in terms of the $|u\rangle$, $|d\rangle$ basis: $|\psi\rangle = |r\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle$. When an electron is in state $|\psi\rangle$, if we measure the z direction, we get $\lambda_{\mathbf{z}} = 1$ with probability $1/2$ and $\lambda_{\mathbf{z}} = -1$ with probability $1/2$. A common mistake is to interpret this to mean that $|\psi\rangle$ is a *mixture*[6] of the two states $|u\rangle$ and $|d\rangle$, each of which occurs with probability of $1/2$. But if this were true, then

$$\begin{aligned}
&\mathbb{P}\left(\lambda_{\mathbf{x}} = 1 \middle| |\psi\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle\right) \\
=&\mathbb{P}\left(\lambda_{\mathbf{x}} = 1 \middle| |\psi\rangle = |u\rangle\right) \cdot \frac{1}{2} + \mathbb{P}\left(\lambda_{\mathbf{x}} = 1 \middle| |\psi\rangle = |d\rangle\right) \cdot \frac{1}{2} \\
=&\frac{1}{4} + \frac{1}{4} \\
=&\frac{1}{2}
\end{aligned}$$

when in fact

$$\mathbb{P}\left(\lambda_{\mathbf{x}} = 1 \middle| |\psi\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle\right) = \mathbb{P}\left(\lambda_{\mathbf{x}} = 1 \middle| |\psi\rangle = |r\rangle\right) = 1$$

This is what we mean when we say that $|\psi\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle$ is a *superposition* (as in the superposition of two waves) and *not a mixture* (as in arising from one of two possible states).

[6] Suppose you roll a die, and the outcome, $X \in \{1, 2, \ldots, 6\}$, is governed by the probability vector $p = (1/12, 2/12, 3/12, 4/12, 1/12, 1/12)$, so that $\mathbb{P}(X = x) = p_x$. Then, depending on the outcome (i.e. depending on the value, x) you flip a biased coin with probability of heads equal to $x/6$. If all you see is the result of the coin flip (call it $Y = 1$ for heads and $Y = 0$ for tails), then the observation Y is said to have a mixture distribution. In this case, the distribution is mixed over the possible outcomes of the roll of the die:

$$\mathbb{P}(Y = 1) = \sum_{x=1}^{6} \mathbb{P}(X = x)\,\mathbb{P}(Y = 1 | X = x) = \sum_{x=1}^{6} p_x \frac{x}{6} = \frac{41}{72}$$

and $\mathbb{P}(Y = 0) = \frac{31}{72}$.

2.3 Observables and observations

Of course there are many other types of measurements that we might choose to make, not just a deflection in a Stern-Gerlach experiment or a screen location after passing through a barrier with two slits. If the state "determines the future, probabilistically," then presumably the state determines the probability of any particular result from any particular measurement. But what constitutes a measurement? Let's start by formulating what we know about measuring spins, and then generalize later.

Hermitian matrices. We will associate each orientation of a Stern-Gerlach machine with a "Hermitian matrix." These are a generalization of real-valued symmetric matrices, but allowing for complex-valued components. Here are the essential definitions:

Definition. *(i) Given a matrix $A \in \mathbb{C}^{n_1 \times n_2}$, define $A^\dagger \in \mathbb{C}^{n_2 \times n_1}$ by*

$$(A^\dagger)_{kl} = (A^*)_{lk} \quad \forall k \in 1:n_1,\ l \in 1:n_2$$

where $A^* \in \mathbb{C}^{n_1 \times n_2}$ is the matrix A with every element replaced by its complex conjugate. $A^\dagger$ is called the adjoint *or the* Hermitian adjoint *of A.*

(ii) If $A \in \mathbb{C}^{n \times n}$ and $A^\dagger = A$, then A is called Hermitian.

Although the Hermitian adjoint is defined more generally, a Hermitian matrix is necessarily a square matrix ($n_2 = n_1$) since $A^\dagger = A$ implies that $A^\dagger$ and A are the same size.

For a general two-by-two matrix,

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \implies A^\dagger = \begin{pmatrix} \alpha^* & \gamma^* \\ \beta^* & \delta^* \end{pmatrix}$$

For instance,

$$A = \begin{pmatrix} 2+i & 1+2i \\ -1 & 1+i \end{pmatrix} \implies A^\dagger = \begin{pmatrix} 2-i & -1 \\ 1-2i & 1-i \end{pmatrix}$$

and in this case, since $A \neq A^\dagger$, A is not Hermitian. On the other hand, consider

$$A = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \text{for which} \quad A^\dagger = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = A$$

which is Hermitian.

Hermitian matrices are often called *self-adjoint*, and sometimes simply *symmetric* (though this risks being confused with A simply equalling its transpose, as opposed to the conjugate of its transpose). Self-adjoint matrices and their generalizations in infinite dimensions (called self-adjoint operators) pop up all the time in applied mathematics, statistics, and learning theory. Their utility can usually be traced to one or another version of the *spectral theorem*:

Theorem (Spectral Theorem)**.** *A matrix $H \in \mathbb{C}^{n\times n}$ is Hermitian if and only if it can be written in the form*

$$H = \sum_{k=1}^{n} \lambda_k e_k e_k^\dagger$$

where $e_1, \ldots, e_n \in \mathbb{C}^{n\times 1}$, $|e_k| = 1$, $e_k^\dagger e_l = \delta_{kl}$, and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.

It is easy to check that the vectors $e_1, \ldots, e_n$ are eigenvectors of H with the corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, and the fact that the eigenvalues must be real comes from comparing the diagonal entries of H with those of $H^\dagger$.

As an example, consider the following three ortho-normal vectors in $\mathbb{R}^3$:

$$\mathbf{e_1} = \begin{pmatrix} i/\sqrt{3} \\ 1/\sqrt{3} \\ -i/\sqrt{3} \end{pmatrix} \qquad \mathbf{e_2} = \begin{pmatrix} i/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} \qquad \mathbf{e_3} = \begin{pmatrix} i/\sqrt{6} \\ 1/\sqrt{6} \\ 2i/\sqrt{6} \end{pmatrix}$$

(Check for yourself that $\langle e_i | e_j \rangle = \delta_{i,j}$.[7]) Given any three real numbers, say $\lambda_1 = 2$, $\lambda_2 = -1$, and $\lambda_3 = -1.1$, consider the matrix

$$H = 2e_1 e_1^\dagger - e_2 e_2^\dagger - 1.1 e_3 e_3^\dagger \tag{14}$$

where, for example,

$$e_1 e_1^\dagger = \begin{pmatrix} i/\sqrt{3} \\ 1/\sqrt{3} \\ -i/\sqrt{3} \end{pmatrix} \begin{pmatrix} -i/\sqrt{3} & 1/\sqrt{3} & i/\sqrt{3} \end{pmatrix} = \begin{pmatrix} 1/3 & i/3 & -1/3 \\ -i/3 & 1/3 & i/3 \\ -1/3 & -i/3 & 1/3 \end{pmatrix}$$

Here is another representation of (14):

$$H = \begin{pmatrix} | & | & | \\ e_1 & e_2 & e_3 \\ | & | & | \end{pmatrix} \begin{pmatrix} 2 & & \\ & -1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} — & e_1^\dagger & — \\ — & e_2^\dagger & — \\ — & e_3^\dagger & — \end{pmatrix} \tag{15}$$

where the k'th columns of the left-most matrix is e_k, and the kth row of the right-most matrix is $e_k^\dagger$, and the middle matrix has all zeros off of the diagonal.

To convince yourself that (14) and (15) are the same, you could rework the right hand side of each expression into a single three-by-three matrix and then check that the results are identical. That gets messy. Alternatively, imagine right-multiplying each of the two representations of H by one of the vectors $e_1, \ldots, e_n$, say e_k. Argue that the result is in both cases the same, namely $\lambda_k e_k$. Since the two matrices have the same action on this particular basis, they must have the same action on all linear combinations of these basis elements, which is to say that they have the same action on the entire vector space. Hence they are the same. (And as a byproduct you end up showing that e_k is an eigenvector with eigenvalue λ_k.)

[7] $\delta_{i,j}$ is the "Kronecker delta function," $\delta_{i,j} = 1$ if $i = j$ and zero otherwise.

We will almost always use Dirac's bracket notation, so that the generic Hermitian matrix, H, could be written

$$H = \sum_{k=1}^{n} \lambda_k \, |e_k\rangle\langle e_k|$$

Back to measurements. Let us now illustrate the connection between Hermitian matrices and measurements by considering the generic Stern-Gerlach machine, oriented in the direction $\mathbf{n} \in \mathbb{R}^3_1$. We will associate each $\mathbf{n}$ with the Hermitian matrix (known as the "spin operator")

$$\sigma_n \doteq |n^+\rangle\langle n^+| - |n^-\rangle\langle n^-|$$

which is a 2×2 complex-valued matrix, with eigenvectors $|n^+\rangle$ and $|n^-\rangle$, and corresponding eigenvalues $+1$ and -1:

$$\sigma_n |n^+\rangle = |n^+\rangle\langle n^+||n^+\rangle - |n^-\rangle\langle n^-||n^+\rangle = |n^+\rangle$$
$$\sigma_n |n^-\rangle = |n^+\rangle\langle n^+||n^-\rangle - |n^-\rangle\langle n^-||n^-\rangle = |n^-\rangle$$

The matrix σ_n represents this particular measurement in the sense that its eigenvectors correspond to the definite states and its eigenvalues to the observed deflections that result from sending an electron (or other half-spin charged particle) through the device. Here are the analogous representations (aka spin operators) for the three canonical directions up/down, in/out, and right/left:

$$\sigma_z = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\sigma_y = \begin{pmatrix} 1/\sqrt{2} \\ i/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -i/\sqrt{2} \end{pmatrix} - \begin{pmatrix} 1/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & i/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

$$\sigma_x = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} - \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Each matrix has eigenvalues ± 1, corresponding to the two possible deflections. We will be seeing more of these particular matrices, which are collectively called the *Pauli matrices*.

Then why bother with Hermitian matrices, if all they do is summarize what we already know about a measurement? The answer, for now, is convenience—the same argument we made for making the transition from $\mathbb{R}^3$ to $\mathbb{C}^2$. But pretty soon the formalism, though faithful to the experiments, will lead us to anticipate all manner of bizarre consequences; it has a life of its own. Eugene Wigner, a Nobel Prize in Physics recipient, called this the unreasonable effectiveness of mathematics. [8]

[8] Wigner, E. P. (1960). "The unreasonable effectiveness of mathematics in the natural sciences. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959". Communications on Pure and Applied Mathematics. 13: 1–14.

Many applications revolve around computing expected values of the (real-valued) observations that can result from a measurement, or what we could now call the expected value of the eigenvalues of the associated Hermitian matrix. We will conclude this section, and set up some notation for stating and proving the uncertainty principle, by deriving a formula for these expectations in terms of the associated matrices. We will do this in some generality, and then look at some examples involving spin states.

Given a state $|\psi\rangle$ and a measuring device represented by a Hermitian matrix A, we will think of the eigenvalues of A as random variables. This makes sense since once we are given $|\psi\rangle$ we have at hand the probability of any particular outcome of the measurement, as follows: Suppose that a particular measurement is represented by the Hermitian matrix

$$A = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha \, |e_\alpha\rangle\langle e_\alpha|$$

where $\mathcal{A}$ is an arbitrary finite index set, which is simply playing the role usually played by $\{1, 2, \ldots, n\}$. As usual, we denote the (random) observable as λ_A, so that

$$\mathbb{P}\left(\lambda_A = \lambda_\alpha \mid |\psi\rangle\right) = |\langle e_\alpha|\psi\rangle|^2$$

(the squared overlap) and

$$\mathbb{E}\left[\lambda_A \mid |\psi\rangle\right] = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha |\langle e_\alpha|\psi\rangle|^2$$

(the weighted average, aka the expectation).

Here's another way to calculate this:

$$\begin{aligned}
\mathbb{E}[\lambda_A \mid |\psi\rangle] &= \sum_{\alpha \in \mathcal{A}} \lambda_\alpha |\langle e_\alpha|\psi\rangle|^2 \\
&= \langle\psi| \left(\sum_{\alpha \in \mathcal{A}} \lambda_\alpha \, |e_\alpha\rangle\langle e_\alpha| \right) |\psi\rangle \\
&= \langle\psi|A|\psi\rangle
\end{aligned}$$

which is the general formula for computing the expected outcome of a measurement. As you might have hoped, the measured quantity is real valued: $\langle\psi|A|\psi\rangle \in \mathbb{R}$.

Here are some examples:

$$\begin{aligned}\mathbb{E}[\lambda_z \mid |\psi\rangle = |r\rangle] &= \langle r|\sigma_z|r\rangle = \begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix}\\ &= \begin{pmatrix}1/\sqrt{2} & 1/\sqrt{2}\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{pmatrix} = 0\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\lambda_x \mid |\psi\rangle = |u\rangle] &= \langle u|\sigma_x|u\rangle = \begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1\\ 0\end{pmatrix}\\ &= \begin{pmatrix}1 & 0\end{pmatrix}\begin{pmatrix}0\\ 1\end{pmatrix} = 0\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\lambda_x \mid |\psi\rangle = |l\rangle] &= \langle l|\sigma_x|l\rangle = \begin{pmatrix}1/\sqrt{2} & -1/\sqrt{2}\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1/\sqrt{2}\\ -1/\sqrt{2}\end{pmatrix}\\ &= \begin{pmatrix}1/\sqrt{2} & -1/\sqrt{2}\end{pmatrix}\begin{pmatrix}-1/\sqrt{2}\\ 1/\sqrt{2}\end{pmatrix} = -1.\end{aligned}$$

Finally, you might be tempted to interpret $\sigma_x |\psi\rangle$ as the state *resulting* from the measurement, but in fact it is nothing of the sort. The result of a measurement is a particular *state*, and one that is (definitely) one of the eigenvectors, $|x^+\rangle$ or $|x^-\rangle$, of σ_x. But $\sigma_x |\psi\rangle$, on the other hand, is a *vector*, in this case a linear combination (aka superposition) of $|x^+\rangle$ and $|x^-\rangle$.

2.4 The uncertainty principle

The uncertainty principle is about the uncertainty in the outcomes of two future measurements, given the current state $|\psi\rangle$. A good way to lay the conceptual groundwork is to think about what happens when we make two measurements, in succession, having started in state $|\psi\rangle$.

Two measurements instead of just one. For example, suppose we measure σ_x and then σ_y, having started in $|\psi\rangle \in \mathbb{C}^2_1$. Then we will observe a sequence $(\lambda^{\sigma_x}, \lambda^{\sigma_y})$, and there are four possible outcomes, $(\lambda^{\sigma_x}, \lambda^{\sigma_y}) \in \{-1, 1\}^2$, with respective probabilities given by the overlap rule. If, say, $\lambda^{\sigma_x} = 1$ and $\lambda^{\sigma_y} = -1$, then

$$\mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{\sigma_x}=1]{\sigma_x} |r\rangle \xrightarrow[\lambda^{\sigma_y}=-1]{\sigma_y} |o\rangle\right) = \mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{\sigma_x}=1]{\sigma_x} |r\rangle\right) \mathbb{P}\left(|r\rangle \xrightarrow[\lambda^{\sigma_y}=-1]{\sigma_y} |o\rangle\right)$$

$$= |\langle\psi|r\rangle|^2 |\langle r|o\rangle|^2 = \frac{1}{2}|\langle\psi|r\rangle|^2$$

Or, in simple short hand,

$$\mathbb{P}\left(|\psi\rangle \to |r\rangle \to |o\rangle\right) = \mathbb{P}\left(|\psi\rangle \to |r\rangle\right) \mathbb{P}\left(|r\rangle \to |o\rangle\right)$$

$$= |\langle\psi|r\rangle|^2 |\langle r|o\rangle|^2 = \frac{1}{2}|\langle\psi|r\rangle|^2$$

But if we first observed $|\psi\rangle \to |o\rangle$ and then $|o\rangle \to |r\rangle$, then the probability of the sequence will usually be different:

$$\mathbb{P}\left(|\psi\rangle \to |o\rangle \to |r\rangle\right) = \mathbb{P}\left(|\psi\rangle \to |o\rangle\right) \mathbb{P}\left(|o\rangle \to |r\rangle\right)$$

$$= |\langle\psi|o\rangle|^2 |\langle o|r\rangle|^2 = \frac{1}{2}|\langle\psi|o\rangle|^2$$

More generally, if

$$A = \sum_{k=1:n} \lambda^A_k \, |e_k\rangle\langle e_k| \quad \text{and} \quad B = \sum_{l=1:n} \lambda^B_l \, |f_l\rangle\langle f_l|$$

are the spectral representations of two operators (matrices) acting in $\mathbb{C}^n$, then

$$\mathbb{P}\left(|\psi\rangle \xrightarrow{A} |e_k\rangle \xrightarrow{B} |f_l\rangle\right) = \mathbb{P}\left(|\psi\rangle \xrightarrow{A} |e_k\rangle\right) \mathbb{P}\left(|e_k\rangle \xrightarrow{B} |f_l\rangle\right) = |\langle\psi|e_k\rangle|^2 |\langle e_k|f_l\rangle|^2$$

which is the probability that we observe $\lambda^A = \lambda^A_k$ followed by $\lambda^B = \lambda^B_l$.[9] As for the reverse order, $\lambda^B = \lambda^B_l$ followed by $\lambda^A = \lambda^A_k$, the probability is $|\langle\psi|f_l\rangle|^2 |\langle e_k|f_l\rangle|^2$ instead of $|\langle\psi|e_k\rangle|^2 |\langle e_k|f_l\rangle|^2$, which will typically be different for some starting state $|\psi\rangle$.

[9] Here we are using λ^A to represent the random variable which is the outcome of the measurement associated with A, and λ^A_k to represent a particular one of the n possible outcomes.

This raises an important question: when is the probability of an outcome of a sequence of measurements independent of the order? If it is not, then the two measurements can not be thought of as a single measurement, since it would be order-dependent and therefore ill-defined. Typically, we can not expect to find a single Hermitian matrix that represents the combination of the two measurements, but by comparing $|\langle\psi|f_l\rangle|^2|\langle e_k|f_l\rangle|^2$ to $|\langle\psi|e_k\rangle|^2|\langle e_k|f_l\rangle|^2$ we can glean an exception: If A and B have a common set of eigenvectors, in which case we can arrange the labels so that $|f_k\rangle = |e_k\rangle\,,\ k = 1, 2, \ldots, n$, then

$$|\langle\psi|f_l\rangle|^2|\langle e_k|f_l\rangle|^2 = |\langle\psi|f_l\rangle|^2\delta_{k,l} = |\langle\psi|f_k\rangle|^2 = |\langle\psi|e_k\rangle|^2 = |\langle\psi|e_k\rangle|^2|\langle e_k|f_l\rangle|^2$$

It turns out that the exception is also the rule: measurement order is interchangeable only if A and B share their eigenvalues.

The characterization of commuting measurements. When does a sequence of measurements amount to a single measurement? In yet more evidence for the unreasonable effectiveness of mathematics, it turns out that the answer is already contained in some well-known properties of Hermitian matrices:

Theorem. (Commutation of Finite-dimensional Hermitian Matrices) *Let A and B be two Hermitian matrices, $A, B \in \mathbb{C}^{n\times n}$, $A^\dagger = A$, and $B^\dagger = B$. The following three statements are equivalent:*

(i) AB *Hermitian*

(ii) $AB = BA$

(iii) A *and* B *share a common set of* n *ortho-normal eigenvectors. In other words, there exists* $|g_k\rangle\,,\ k = 1, 2, \ldots, n$ *such that* $\langle g_k|g_l\rangle = \delta_{k,l}$ *and*

$$A = \sum_{k=1}^{n} \lambda_k\, |g_k\rangle\langle g_k| \quad \lambda_k \in \mathbb{R},\ \forall\ k$$
$$B = \sum_{l=1}^{n} \gamma_l\, |g_l\rangle\langle g_l| \quad \gamma_l \in \mathbb{R},\ \forall\ l \tag{16}$$

The relationship defined in (ii) is special and has its own name: Any two square matrices A and B are said to commute if $AB = BA$.

The proof is not particularly hard. For those interested, here is one way to go about it.

Proof

(i) $\Leftrightarrow$ (ii) This one is pretty easy:

$$(AB)^\dagger = AB \Leftrightarrow B^\dagger A^\dagger = AB \Leftrightarrow BA = AB$$

(iii) $\Rightarrow$ (ii) This one is also straightforward. Use the orthogonality condition, $\langle g_k|g_l\rangle = \delta_{k,l}$, to multiply out the representations of A and B in (16):

$$AB = \sum_{k=1}^{n} \lambda_k \gamma_k \, |e_k\rangle\langle e_k| = \sum_{k=1}^{n} \gamma_k \lambda_k \, |e_k\rangle\langle e_k| = BA$$

(ii) $\Rightarrow$ (iii) For each unique eigenvalue λ of A, let S_λ be the subspace spanned by the corresponding eigenvectors:

$$S_\lambda = \{|v\rangle : A\,|v\rangle = \lambda\,|v\rangle\}$$

By assumption, $AB = BA$, and hence for any $|v\rangle \in S_\lambda$

$$A\,|v\rangle = \lambda\,|v\rangle \Rightarrow BA\,|v\rangle = \lambda B\,|v\rangle \Rightarrow A(B\,|v\rangle) = \lambda B\,|v\rangle$$

In other words, $B\,|v\rangle$ is also an eigenvector of A and hence $B\,|v\rangle \in S_\lambda$ for every $v \in S_\lambda$.

Now consider the transformation represented by B but restricted to the subspace S_λ. In general, any linear transformation T on a vector space V is completely determined by the set of scalars $\{\langle u|T|w\rangle\}_{u,w\in V}$, and since

$$\langle u|B|w\rangle = \langle u|B^\dagger|w\rangle$$

for all $u, w \in \mathbb{C}^n$, it is also true for all $u, w \in S_\lambda$. Hence the restriction of B to S_λ is also Hermitian. By the spectral representation theorem

$$B = \sum_{k=1}^{m} \gamma_k \, |g_k\rangle\langle g_k|$$

for some complete orthonormal set $|g_k\rangle,\ k = 1, 2, \ldots, m$ of eigenvectors of B contained in S_λ (m being the dimension of the subspace). Since every element of S_λ is also an eigenvector of A, the actions of A and B in S_λ are represented by the common set of eigenvectors $|g_k\rangle,\ k = 1, 2, \ldots, m$.

All that is left is to follow this same procedure for every other eigenvalue of A.

□

Quantum computing is mostly about not measuring anything, but we do need to observe the result! As we will see later on, these commutation relationships are the key to the final "readout" step.

The uncertainty principle. Given two measurements (Hermitian operators) A and B, define the *commutator* $[A, B] = AB - BA$. When the commutator is zero we say that "A and B commute." As we have already seen, $[A, B] = 0$ signals the existence of a single Hermitian operator (namely AB) representing the *pair* of measurements, taken in either order. On the other hand, if $[A, B] \neq 0$, then no such operator exists.

Nevertheless, given any starting state $|\psi\rangle$ we can still talk about probabilities of outcomes of future measurements of either A *or* B. The uncertainty principle states that if A and B do not commute, then there

is a bound on how accurately we can predict *both* outcomes; knowing either outcome with high certainty implies that the other outcome can not be known with high certainty. This tradeoff is decidedly *not* in force if the measurements commute, since then if we were to start in one of the common eigenstates there would be no uncertainty about the outcome of either measurement.

Before we try to make this precise, it might be a good idea to first review the basic definitions of the mean, variance, and standard deviation of a random variable, since the same definitions are used in quantum mechanics. Recall that if $p_1, p_2, \ldots, p_n$ is a probability distribution on a set of n elements, say $\{x_1, x_2, \ldots, x_n\}$, and if we denote by X a random variable with $\mathbb{P}(X = x_k) = p_k$, then the mean μ and variance γ^2 are defined by

$$\mu = \sum_{k=1}^{n} p_k x_k \quad \text{(the probability-weighted average of the possible observations, } \{x_k\}_{k=1:n})$$

$$\delta^2 = \sum_{k=1}^{n} p_k (x_k - \mu)^2 \quad \text{(the probability-weighted average of the "squared deviations," } \{(x_k - \mu)^2\}_{k=1:n})$$

(In probability and statistics the variance is almost always designated using σ^2 rather than γ^2, but in quantum mechanics σ^2 usually refers to the spin operators, i.e. the Hermitian matrices that represent measurements of spin. Hence δ^2 instead of σ^2.) The standard deviation, which we will denote δ, is the square-root of the variance.

We will apply these same statistical notions to the random outcomes of future observations. Given a state $|\psi\rangle$ and a S-G device oriented at direction $\mathbf{n} \in \mathbb{R}^3_1$, how much "uncertainty" is there in the outcome of a measurement of $\lambda_{\mathbf{n}}$? More generally, given a (Hermitian) matrix A and its associated observable λ_A, write

$$\begin{aligned} & \mu_A \quad \text{for} \quad \mathbb{E}[\lambda_A \mid |\psi\rangle] \ \text{(the mean of } \lambda_A \text{ given the current state } |\psi\rangle) \\ \text{and} \quad & \delta_A^2 \quad \text{for} \quad \mathbb{E}[(\lambda_A - \mu_A)^2 \mid |\psi\rangle]] \ \text{(the variance of } \lambda_A \text{ given the current state } |\psi\rangle) \end{aligned}$$

We can expand δ_A^2 using linearity:

$$\begin{aligned} \delta_A^2 &= \mathbb{E}[(\lambda_A - \mu_A)^2 \mid |\psi\rangle]] \\ &= \mathbb{E}\left[\left(\lambda_A^2 - 2\lambda_A \mu_A + \mu_A^2\right) \mid |\psi\rangle\right] \\ &= \mathbb{E}\left[\lambda_A^2 \mid |\psi\rangle\right] - 2\mu_A \, \mathbb{E}\left[\lambda_A \mid |\psi\rangle\right] + \mu_A^2 \\ &= \mathbb{E}\left[\lambda_A^2 \mid |\psi\rangle\right] - \mu_A^2 \end{aligned} \tag{17}$$

But how do we compute $\mathbb{E}\left[\lambda_A^2 \mid |\psi\rangle\right]$? All we need is to generalize, slightly, the formula that we already worked out for computing $\mathbb{E}\left[\lambda_A \mid |\psi\rangle\right]$ in terms fo A:

$$\begin{aligned}\mathbb{E}[\lambda_A^2 \mid |\psi\rangle] &= \sum_{\alpha\in\mathcal{A}} \lambda_\alpha^2 |\langle e_\alpha|\psi\rangle|^2 \\ &= \langle\psi| \left(\sum_{\alpha\in\mathcal{A}} \lambda_\alpha^2 \, |e_\alpha\rangle\langle e_\alpha|\right) |\psi\rangle \\ &= \langle\psi| \left(\sum_{\alpha\in\mathcal{A}} \lambda_\alpha \, |e_\alpha\rangle\langle e_\alpha|\right) \left(\sum_{\beta\in\mathcal{A}} \lambda_\beta \, |e_\beta\rangle\langle e_\beta|\right) |\psi\rangle \\ &= \langle\psi|A^2|\psi\rangle\end{aligned} \tag{18}$$

This works because A^2 has the same eigenvectors as A, but with the eigenvalues squared. (In fact, for any "smooth" function $f(\lambda)$, $\mathbb{E}[f(\lambda_A) \mid |\psi\rangle] = \langle\psi|f(A)|\psi\rangle$, which is cool and, actually, not that hard to prove—first extend the above calculation to polynomials, and then write the smooth function as a limit of a sequence of polynomials.)

Now put the two expressions, (17) and (18), together:

$$\delta_A^2 = \mathbb{E}[\lambda_A^2 \mid |\psi\rangle] - \mu_A^2 = \langle\psi|A^2|\psi\rangle - \mu_A^2 = \langle\psi|(A^2 - \mu_A^2 I)|\psi\rangle$$

Theorem. (The Uncertainty Principle) *Consider a system prepared in the definite state $|\psi\rangle$ and let A and B be the Hermitian matrices corresponding to arbitrary measurements λ_A and λ_B. Define δ_A and δ_B to be the conditional standard deviations (given $|\psi\rangle$) of λ_A and λ_B (i.e. the positive square roots of the variances, δ_A^2 and δ_B^2). Then*

$$\delta_A \delta_B \geqslant \frac{1}{2} |\langle\psi|[A, B]|\psi\rangle| \tag{19}$$

Given the definite state $|\psi\rangle$, equation (19) defines a subtle relationship between any two measurements whereby a small uncertainty in the outcome of one typically means a high uncertainty in the outcome of the other. The disclaimer, "typically", is necessary, since the amount of uncertainty depend on the extent to which the two measurements fail to commute. Indeed, $\frac{1}{2}|\langle\psi|[A, B]|\psi\rangle|$ can be arbitrarily small, and in fact zero if the matrices commute. Some examples will be worked out after the proof, and some more will be in the HW.

Proof. (May be of interest, but it is not "required reading.") In three parts:

(i) Define vectors $|a\rangle$ and $|b\rangle$:

$$\begin{aligned}|a\rangle &\doteq (A - \mu_A I)\,|\psi\rangle \\ |b\rangle &\doteq i(B - \mu_B I)\,|\psi\rangle\end{aligned}$$

where I is the identity matrix. Notice that the lengths $|a\rangle$ and $|b\rangle$ are equal to the respective standard deviations, δ_A and δ_B:

$$||a\rangle|^2 = \langle a|a\rangle = \langle\psi|A^2|\psi\rangle - \mu_A^2 = \delta_A^2$$
$$||b\rangle|^2 = \langle b|b\rangle = \langle\psi|B^2|\psi\rangle - \mu_B^2 = \delta_B^2$$

The use of i in the definition of $|b\rangle$ amounts to a pretty nifty trick, as we will see shortly.

(ii) Bound $\big||a\rangle\big|\,\big||b\rangle\big|$:

The norm of any vector space with an inner product automatically satisfies the *triangle inequality*. Hence, if $|a\rangle, |b\rangle \in \mathbb{C}^{n\times n}$:

$$||a\rangle| + ||b\rangle| \geqslant ||a\rangle + |b\rangle|.$$

Square both sides to get:

$$\begin{aligned} ||a\rangle|^2 + ||b\rangle|^2 + 2\big||a\rangle\big|\,\big||b\rangle\big| &\geqslant (\langle a| + \langle b|)(|a\rangle + |b\rangle) \\ &= ||a\rangle|^2 + ||b\rangle|^2 + \langle a|b\rangle + \langle b|a\rangle \\ \Rightarrow 2\big||a\rangle\big|\,\big||b\rangle\big| &\geqslant \langle a|b\rangle + \langle b|a\rangle \end{aligned}$$

Now do the same thing, but with $-|a\rangle$ in place of $|a\rangle$:

$$2\big||a\rangle\big|\,\big||b\rangle\big| \geqslant -\langle a|b\rangle - \langle b|a\rangle$$

And if we put these two inequalities together we get

$$2\big||a\rangle\big|\,\big||b\rangle\big| \geqslant |\langle a|b\rangle + \langle b|a\rangle|\,.$$

(iii) Plug in the expressions for $|a\rangle$ and $|b\rangle$:

Finally, use the definitions of $|a\rangle$ and $|b\rangle$. The i in the expression for $|b\rangle$ does its magic by reversing the sign of $\langle b|a\rangle$ relative to $\langle a|b\rangle$:

$$\begin{aligned} \delta_A\delta_B &\geqslant \frac{1}{2}|i\langle\psi|(A-\mu_A I)(B-\mu_B I)|\psi\rangle - i\langle\psi|(B-\mu_B I)(A-\mu_A I)|\psi\rangle| \\ &= \frac{1}{2}|\langle\psi|AB - BA|\psi\rangle + \mu_A\mu_B - \mu_A\mu_B - \mu_A\mu_B + \mu_A\mu_B + \mu_A\mu_B - \mu_A\mu_B| \\ &= \frac{1}{2}|\langle\psi|[A,B]|\psi\rangle| \end{aligned}$$

Which completes the proof. □

Let's work through two examples, using spin states.

(i) Let $|\psi\rangle = |r\rangle = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$. What is the lower bound on the product of the uncertainties in measuring $\lambda_{\mathbf{x}}$ and $\lambda_{\mathbf{y}}$?

$$
\begin{aligned}
A = \sigma_x &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
B = \sigma_y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \\
\Rightarrow AB - BA &= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} - \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = \begin{pmatrix} 2i & 0 \\ 0 & -2i \end{pmatrix} \\
\Rightarrow \delta_A \delta_B &\geqslant \left| \langle r | \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} |r\rangle \right| \\
&= \left| \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} i/\sqrt{2} \\ -i/\sqrt{2} \end{pmatrix} \right| \\
&= |i/2 - i/2| = 0
\end{aligned}
$$

As we should have anticipated, since the prepared state $|\psi\rangle = |r\rangle$ is guaranteed to give the result $\lambda_x = 1$, and hence the variance in outcome, δ_A^2 is zero.

(ii) However, for the same two measurements, but this time from the prepared state $|\psi\rangle = |u\rangle$, tell an entirely different story: Since the measurements are the same, $AB - BA$ is unchanged. But

$$
\delta_A \delta_B \geqslant \left| \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right| = \left| \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} i \\ 0 \end{pmatrix} \right| = 1
$$

Since λ_A and λ_B are both bounded by one, their variances can not be smaller than one. Therefore their variances, and hence also their standard deviations, must be exactly one. Of course this too was to be expected, since, given $|\psi\rangle = |u\rangle$, both λ_x and λ_y are 50/50, ± 1.

In the homework you will look at the general case: For any three orientations $\mathbf{m}, \mathbf{n}, \mathbf{o} \in \mathbb{R}_1^3$, assume that the particle is prepared in the $|\psi\rangle = |o^+\rangle$ state, and consider the standard deviations

$$
\begin{aligned}
\delta_A &= \text{SD}(\lambda_{\mathbf{m}} \mid |o^+\rangle) \\
\delta_B &= \text{SD}(\lambda_{\mathbf{n}} \mid |o^+\rangle)
\end{aligned}
$$

In this case, the uncertainty principle comes down to a surprisingly elegant formula: $\delta_A \delta_B \geqslant \mathbb{V}$ where $\mathbb{V}$ is volume of the parallelepiped defined by the three unit vectors $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{o}$. See Figure 11.

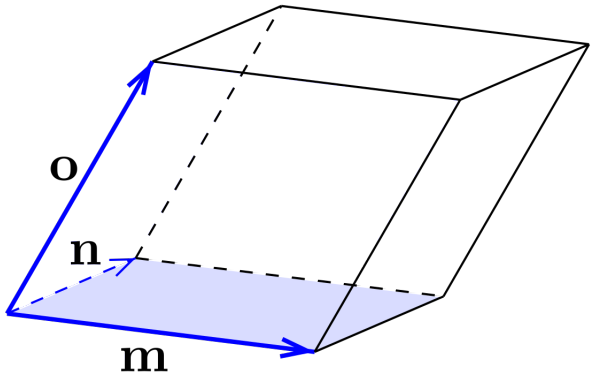


Figure 11: A parallelepiped formed by $\mathbf{m}$, $\mathbf{n}$, and $\mathbf{o}$

2.5 The Schrödinger equation

So far, observations are the only tools that we have for manipulating a spin state. This will always result in a definite state, but unless it is a repeated measurement the outcome will be probabilistic. Is it possible to change a particle's state in such a way as to *deterministically* place it into a new state? For example, if an electron is in the up spin state, $|u\rangle$, can we deterministically manipulate it into the $|r\rangle$ spin state? This is different from *measuring* its spin along the x direction, since the measurement is equally likely to result in the $|l\rangle$ spin state as it is to result in the $|r\rangle$ spin state.

Deterministic manipulations are an essential part of most applications and many of the experiments in quantum mechanics. A simple example is the so-called "not gate" used in quantum computing. Given a two-spin particle, we devise a machine that deterministically changes $|u\rangle$ to $|d\rangle$ and $|d\rangle$ to $|u\rangle$. We can describe the *desired* action precisely with the help of the matrix $U = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which happens to be σ_x, but here we are *not* using it to describe a measurement. Instead we are using it to describe a deterministic alteration of the wave function, $|\psi\rangle \to U|\psi\rangle$. In particular, if $|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ then

$$U|\psi\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$$

This is a "not gate" in the sense that it reverses the up or down "qubit": $\begin{pmatrix} 1 \\ 0 \end{pmatrix} \to \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \to \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

But how is this deterministic transformation of the wave function implemented? What we need is the Schrödinger equation, which gives us the deterministic time evolution of a state when subjected to various influences, such as an electromagnetic field.

Elements of a derivation of the Schrödinger equation. We will assume that we are working in a finite-dimensional state space (such as our spin wave function), and that the particle's environment (as opposed to the particle itself) is not changing with time. Neither assumption is necessary, and the story is largely the same if we are working in the continuum (e.g. position rather than spin) and the particle is subjected to time-varying influences (e.g. an oscillating electromagnetic field).[10]

We will use U_t to represent the transformation that takes a state at time $t = 0$ (say, $|\psi(0)\rangle$) into a state at time t (say $|\psi(t)\rangle$): $|\psi(t)\rangle = U_t|\psi(0)\rangle$.

Considerations.

(1) **Linearity.** As with the propagation of an idealized wave, we will assume that the transformation

[10] Another assumption, or more accurately an approximation, is that relativistic effects are negligible. This too is by no means necessary, cf the *relativistic* Schrödinger equation.

represented by U_t is *linear*:

$$U_t\left(|\psi_1(0)\rangle + |\psi_2(0)\rangle\right) = U_t\,|\psi_1(0)\rangle + U_t\,|\psi_2(0)\rangle$$

Since every linear transformation on a finite-dimensional vector space can be represented by a matrix (given a particular basis), we can proceed under the assumption that U_t is a matrix.

(2) **Conservation of Probability.** This is little more than our definition of the wave function. At each time t we assume that $\psi(t)$ is normalized: $\langle\psi(t)|\psi(t)\rangle = 1$. And since $|\psi(t)\rangle = U_t\,|\psi(0)\rangle$, we conclude that

$$\langle\psi(0)|U_t^\dagger U_t|\psi(0)\rangle = 1 \tag{20}$$

for every $|\psi(0)\rangle$ and $t \geqslant 0$. Of course the relationship in (20) holds for every initial state, $|\psi(0)\rangle$. It turns out that this is enough to conclude that in fact $U_t^\dagger U_t = I$ (the $n \times n$ identity matrix), as we will now show.

For any pair of square matrices A and B, $(AB)^\dagger = B^\dagger A^\dagger$ (as you will show in the HW), and hence $(U_t^\dagger U_t)^\dagger = U_t^\dagger U_t$ for every $t \geqslant 0$. In other words, $U_t^\dagger U_t$ is Hermitian for every t.

For now, fix t, and to ease the notation simply write U to mean U_t. Since $U^\dagger U$ is Hermitian, it has a spectral representation of the form

$$U^\dagger U = \sum_{k=1}^{n} \lambda_k\,|e_k\rangle\langle e_k|$$

where n is the dimension of the wave function, $|e_k\rangle,\ k = 1:n$ is an orthonormal set of eigenvectors, and for each k, λ_k is the (real-valued) eigenvalue associated with the eigenvector $|e_k\rangle$. Going back to equation (20), for any $l = 1:n$ choose the initial condition $|\psi(0)\rangle = |e_l\rangle$:

$$\begin{aligned}
1 &= \langle e_l|U_t^\dagger U_t|e_l\rangle \\
&= \langle e_l|\sum_{k=1}^{n} \lambda_k\,|e_k\rangle\langle e_k|e_l\rangle \\
&= \sum_{k=1}^{n} \lambda_k\,\langle e_l|e_k\rangle\langle e_k|e_l\rangle \\
&= \sum_{k=1}^{n} \lambda_k\delta_{kl} = \lambda_l
\end{aligned}$$

where δ_{kl} is the Kronecker delta function, i.e. one if $k = l$ and zero otherwise. Thus every eigenvalue of $U^\dagger U$ is one, and $U^\dagger U = \sum_{k=1}^{n} |e_k\rangle\langle e_k|$. But this is just another way to write the identity[11]: $U^\dagger U\,|\phi\rangle = \sum_{k=1}^{n} |e_k\rangle\langle e_k|\phi\rangle = |\phi\rangle$, since $\sum_{k=1}^{n} |e_k\rangle\langle e_k|\phi\rangle$ is just $|\phi\rangle$ expanded in the basis $|e_k\rangle,\ k = 1:n$.

Hence $U^\dagger U = I$, and in finite dimensions ($n < \infty$) this also means that $UU^\dagger = I$.[12]

Definition. A matrix $U \in \mathbb{C}^{n\times n}$ for which $U^\dagger U = UU^\dagger = I$ is called *unitary*. These properties, collectively, are called *unitarity*.

[11] This representation is sometimes called "a resolution of the identity."

[12] This is not too hard to show. You might want to try it.

A matrix is the representation of a linear mapping in a particular basis, and the unitary property is basis independent. So we can equally well speak of unitary transformations. Or more commonly "unitary operators" when working in infinite dimensions.

The upshot of all of this, as shown in the homework, is that U itself has a spectral representation of the form

$$U = \sum_{k=1}^{n} e^{i\phi_k} \, |e_k\rangle\langle e_k|$$

where $\phi_k \in [0, 2\pi), \ k = 1 : n$. Notice that U, $UU^\dagger$, and $U^\dagger U$ all have the same eigenvectors.

In addition to preserving probability mass, U_t also preserves "overlap":

$$\langle\phi(t)|\psi(t)\rangle = \langle\phi(0)| \, U_t^\dagger U_t \, |\psi(0)\rangle = \langle\phi(0)|\psi(0)\rangle$$

Interpretation: The probability that a measurement produces the state $|\phi(t)\rangle$, given the prepared state $|\psi(t)\rangle$, is the same as the probability that a measurement produces the state $|\phi(0)\rangle$, given the prepared state $|\psi(0)\rangle$. In fact, this observation is consistent with the best way for you to think about unitary operators in general—think of them as rigid rotations and reflections.

Before returning to the development of the Schrödinger equation, let's look at a few examples of Hermitian and/or unitary transformations.

(a) $\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \Rightarrow \sigma_y^\dagger = \sigma_y$, so σ_y is Hermitian. What's more, $\sigma_y^\dagger \sigma_y = \sigma_y \sigma_y^\dagger = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ $= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, so σ_y is also unitary.

(b) More generally, $\sigma_n = \begin{pmatrix} n_z & n_x - in_y \\ n_x + in_y & -n_z \end{pmatrix}$ is both Hermitian and unitary for all $n \in \mathbb{R}^3$.

(c) $A = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$ is Hermitian but not unitary.

(d) The counter-clockwise rotation (by θ) operator $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is unitary but not Hermitian.

It would be a good idea to check (b), (c), and (d) for yourself.

(3) **State.** This part embodies little more than what we mean when we speak of the **state** of a system. Since there is no explicit dependence on time, the state at time s serves as the initial condition when evolving the state for an additional amount of time t:

$$U_{t+s} \, |\psi(0)\rangle = U_t \, \underbrace{(U_s \, |\psi(0)\rangle)}_{|\psi(s)\rangle}, \ \forall \ |\psi(0)\rangle \, .$$

which is to say,

$$U_{t+s} = U_t U_s.$$

(4) **Continuity.** Assume that

$$\lim_{s\to 0} U_{t+s}\,|\psi(0)\rangle = U_t\,|\psi(0)\rangle\,,\ \forall\ |\psi(0)\rangle\,.$$

(Anything else would be a surprise!)

(5) **Stone's Representation Theorem.** The key ingredient is a nifty piece of mathematics that relates unitary to Hermitian operators:

Theorem. If $U_t \in \mathbb{C}^{n\times n}$ satisfies the properties that we have called unitarity, continuity, and state, then there exists an Hermitian matrix $A \in \mathbb{C}^{n\times n}$ such that

$$U_t = e^{itA} \doteq \sum_{k=0}^{\infty} \frac{(it)^k}{k!} A^k \quad \text{for some Hermitian } A \text{ and for all } t \tag{21}$$

Equation (21) can be written, formally, as

$$\frac{\mathrm{d}}{\mathrm{d}t} U_t = iAU_t$$

And if we multiply both sides on the right by the ket vector $|\psi(0)\rangle$ and then substitute $|\psi(t)\rangle$ for $U_t\,|\psi(0)\rangle$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\,|\psi(t)\rangle = iA\,|\psi(t)\rangle \tag{22}$$

with solution $|\psi(t)\rangle = e^{itA}\,|\psi(0)\rangle$.

(6) **Hamiltonians and the Schrödinger equation.** It turns out that, in the proper units, A represents the "Hamiltonian" of the system described by Schrödinger's equation. The Hamiltonian, being an operator, operates on a state, e.g. $|\psi\rangle$, to extract the total (kinetic plus potential) energy of the particle described by the state. We can make the transformation to proper units by defining $A = -H/\hbar$.

Typically, the state vector (aka wave function) will include both position and time, and hence $|\psi\rangle = |\psi(t,\vec{r})\rangle$, where $\vec{r}$ represents position. Equation (22) becomes

$$i\hbar \frac{\partial \psi(t,\vec{r})}{\partial t} = H\psi(t,\vec{r})$$

which is Schrödinger's equation in its most common form.

H is still Hermitian, but in the continuum it becomes a *differential* operator, usually in the form

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r})$$

where $V(\vec{r})$ is potential energy, m is mass, ∇^2 is the Laplacian ($\nabla^2 = \frac{\partial^2}{x^2} + \frac{\partial^2}{y^2} + \frac{\partial^2}{z^2}$ in rectangular coordinates), and $-\frac{\hbar^2}{2m}\nabla^2$ extracts the contribution of what would be, in the classical setting, the kinetic energy.

When we need to teleport Charlie, or build a not gate for a quantum computer, we start by identifying the unitary operator that would achieve the desired change in the wave function. We then work backwards to find a Hamiltonian such that the solution to the Schrodinger equation, evaluated at some future time, generates the action that is called for. This process of going from a unitary operator to a realizable Hamiltonian is explored in the homework.

Of course, to make all of this precise some conditions would be needed, but what we already have will be good enough for our purposes.

3 Entanglement

The state of multiple-particle systems is formulated through the tensor-product representation. There will be many consequences, most notably the strange business of entanglement. The task at hand is to develop the tensor-product representation. I am hoping that you will find it compelling, perhaps even inevitable given the representation already in place for single particles and given some intuitive notions about independence.

We want to be able to talk about a collection of particles (just two for the time being) in the same way that we have talked about a single particle. In particular, we want the wave function of a collection to be an element of a vector space equipped with an inner product, and we want to represent measurements with Hermitian operators. It is not just a matter of aesthetics. Indeed, the motivation is the same: (i) linearity for superposition (hence a vector space), (ii) overlap for for (pre-) probabilities (hence the inner product), and (iii) expected outcomes by matrix computations (hence Hermitian operators).

3.1 Tensor-product spaces

Consider two particles, one called "particle 1" and the other called "particle 2". Each has a wave-function representation in its own state space. Perhaps both spaces are $\mathbb{C}^2$, but for now we will assume, more generally, that they are some complex inner-product spaces $\mathbb{C}^n$ and $\mathbb{C}^m$. We will refer to these spaces as V, the state space of particle 1, and W, the state space of particle 2. We will use $V \otimes W$ to represent the state space of all such pairs of particles. The goal is to mimic the construction of the state space for spin particles: Formulate $V \otimes W$ as a linear inner product space in such a way that the squared amplitude of an inner product (or "overlap") represents the transition probabilities that predict the outcomes of a measurement.

Construction of $\mathbf{V \otimes W}$. We will proceed incrementally, while attempting to support each step with a logical argument.

1. **(minimal elements)** Imagine two particles, 1 and 2, the first in state $|v\rangle \in V$ and the other in state $|w\rangle \in W$. At the very least, we want to be able to represent an independent pair of such particles, whereby knowledge of the state of one tells us nothing about the state of the other. Maybe they are separated by a large distance, each well out of the radius of influence of the other, and we are proceeding with separate experiments in each location.

   Although the interesting case is when the particles are not independent, we will nevertheless require that the state space $V \otimes W$ accommodates independent pairs. Independence is a kind of default case. We will use the special notation $|vw\rangle \in V \otimes W$ to indicate that the pair of particles with states $|v\rangle$ and $|w\rangle$ are independent.

2. **(overlap)** As in the construction of $\mathbb{C}_1^2$, we will interpret the inner product (aka overlap) as an amplitude and the amplitude squared as a probability: If $|\phi\rangle \in V \otimes W$, then

$$\left|\langle\psi|\phi\rangle_{V\otimes W}\right|^2 = \mathbb{P}\left(|\psi\rangle \to |\phi\rangle\right)$$

the probability of a transition from a starting state to a definite state, as would result from a measurement. Here the subscript $V \otimes W$ is to indicate which inner product is at play. After all, we now have three of them: one for V, one for W, and a new one for $V \otimes W$.

3. **(independence)** Suppose that the pair of particles, 1 and 2, are in state $|vw\rangle$, where $|v\rangle \in V$ and $|w\rangle \in W$, and we perform a pair of independent measurements, say A on v and B on w. What is the probability that these measurements will result in the pair of definite states $\tilde{v}$ and $\tilde{w}$, respectively? If we adopt the formal meaning of independence in probability theory, then by the independence of the initial states and their respective measurements

$$\begin{aligned}\left|\langle vw|\tilde{v}\tilde{w}\rangle_{V\otimes W}\right|^2 &= \mathbb{P}\left(|vw\rangle \to |\tilde{v}\tilde{w}\rangle\right)\\ &= \mathbb{P}\left(|v\rangle \xrightarrow{A} |\tilde{v}\rangle, |w\rangle \xrightarrow{B} |\tilde{w}\rangle\right)\\ &= \mathbb{P}\left(|v\rangle \xrightarrow{A} |\tilde{v}\rangle\right)\mathbb{P}\left(|w\rangle \xrightarrow{B} |\tilde{w}\rangle\right)\\ &= \left|\langle v|\tilde{v}\rangle_V\right|^2\left|\langle w|\tilde{w}\rangle_W\right|^2\end{aligned} \tag{23}$$

4. **(basis)** Let $|e_1\rangle, \ldots, |e_n\rangle$ be an orthonormal set in V and $|f_1\rangle, \ldots, |f_m\rangle$ be an orthonormal set in W. Since V and W have dimension n and m respectively, the sets $\{|e_i\rangle\}_{i=1:n}$ and $\{|f_j\rangle\}_{j=1:m}$ constitute bases for V and W respectively. Now consider the set $\{|e_i f_j\rangle\}_{\substack{i=1:n\\ j=1:m}}$, which is the set of all pairs of independent states where each pair is made up of a basis element in V and one in W. By the "minimal elements" assumption, $|e_i f_j\rangle \in V \otimes W$ for all $i = 1:n,\ j = 1:m$, and by the "independence" assumption

$$\left|\langle e_i f_j|e_k f_l\rangle_{V\otimes W}\right| = \left|\langle e_i|e_k\rangle_V\right|\left|\langle f_j|f_l\rangle_W\right| \Rightarrow \langle e_i f_j|e_k f_l\rangle_{V\otimes W} = \delta_{i,k}\delta_{j,l} \tag{24}$$

for all $1 \leqslant i, k \leqslant n$ and $1 \leqslant j, l \leqslant m$. In other words, $\{|e_i f_j\rangle\}_{\substack{i=1:n\\ j=1:m}}$ is an orthonormal set in $V \otimes W$.

5. **(Occam's razor)** At this point, the smallest possible vector space is the $n \times m$ dimensional space defined by the span of $|e_i f_j\rangle$, $i = 1:n$, $j = 1:m$:

$$\{\sum_{\substack{i=1:n\\ j=1:m}} \gamma_{i,j}\,|e_i f_j\rangle\,|\gamma_{i,j} \in \mathbb{C},\ \forall\ i, j\}$$

Are we done? Is this space big enough? Not if you insist on more dimensions, but they turn out to be unnecessary. What we have is enough to fit the data, i.e. it is consistent with the existing observations. This is no different from stopping at $n = 2$ in our construction of a state space for two-spin particles, and just another instance of a time-honored approach to modeling known as "Occam's razor" or the "principle of parsimony."

A more rigorous and straightforward definition would be to specify, *a priori*, the set of pair states $V \otimes W$, and an associated inner product:

- Let $|e_1\rangle, \ldots, |e_n\rangle$ be an orthonormal basis for V and $|f_1\rangle, \ldots, |f_m\rangle$ an orthonormal basis for W. For every ordered pair $|e_i\rangle \in V$ and $|f_j\rangle \in W$, associate a state in $V \otimes W$ denoted by $|e_i, f_j\rangle$, and then define
$$V \otimes W \doteq \{ \sum_{\substack{i=1:n \\ j=1:m}} \gamma_{i,j} |e_i f_j\rangle \big| \gamma_{i,j} \in \mathbb{C}, \ \forall\, i, j\}$$

- Define
$$\langle e_i f_j | e_k f_l \rangle_{V\otimes W} = \delta_{i,k}\delta_{j,l} \ \ \forall i, k \in \{1, \ldots, n\}, \ j, l \in \{1, \ldots, m\}$$
and then extend by linearity to $\langle \psi | \phi \rangle_{V\otimes W}$ for every $|\psi\rangle, |\phi\rangle \in V \otimes W$: If
$$|\psi\rangle = \sum_{\substack{i=1:n \\ j=1:m}} \alpha_{i,j} |e_i f_j\rangle \quad \text{and} \quad |\phi\rangle = \sum_{\substack{k=1:n \\ l=1:m}} \beta_{k,l} |e_k f_l\rangle$$
then
$$\begin{aligned} \langle \psi | \phi \rangle_{V\otimes W} &= \sum_{\substack{i=1:n \\ j=1:m}} \sum_{\substack{k=1:n \\ l=1:m}} \alpha^*_{i,j}\beta_{k,l} \langle e_i f_j | e_k f_l \rangle \\ &= \sum_{\substack{i=1:n \\ j=1:m}} \sum_{\substack{k=1:n \\ l=1:m}} \alpha^*_{i,j}\beta_{k,l}\delta_{i,k}\delta_{j,l} \\ &= \sum_{\substack{i=1:n \\ j=1:m}} \alpha^*_{i,j}\beta_{i,j} \end{aligned}$$
(As you might have expected, given that $\{|e_i f_j\rangle\}_{\substack{i=1:n \\ j=1:m}}$ defines a basis for $V \otimes W$.)

But what happened to independence, our starting point?

Definition. *(i) For any $\psi \in V \otimes W$, with*
$$\psi = \sum_{\substack{i=1:n \\ j=1:m}} \gamma_{i,j} |e_i f_j\rangle$$
the pair of particles represented by $|\psi\rangle$ will be called independent *if there exists $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_m \in \mathbb{C}$ such that $\gamma_{i,j} = \alpha_i\beta_j$ for all $i = 1:n$, $j = 1:m$.*

(ii) In this case, we say that particle 1 is in the state $\vec{v} \doteq \sum_{i=1:n} \alpha_i |e_i\rangle \in V$ and particle 2 is in the state $\vec{w} \doteq \sum_{j=1:n} \beta_j |f_j\rangle \in W$ and write $v\vec{w}$ (or $|v\rangle \otimes |w\rangle$) for $|\psi\rangle$.

Remarks (some of these will be explored in the HW):

1. If $|\psi\rangle = \sum_{i=1}^n \sum_{j=1}^m \gamma_{ij} |e_i f_j\rangle$, then $\langle \psi | \psi \rangle = \sum_{i=1}^n \sum_{j=1}^m |\gamma_{ij}|^2$.

2. Notice that every element of $V \otimes W$ is characterized by an $n \times m$ complex-valued matrix, $\{\gamma_{ij}\}_{i=1:n,j=1:m}$. Therefore, we will sometimes refer to $V \otimes W$ as $\mathbb{C}^{n\times m}$, or better yet $\mathbb{C}^n \otimes \mathbb{C}^m$.

3. If $|v\rangle, |\tilde{v}\rangle \in V$ and $|w\rangle, |\tilde{w}\rangle \in W$, then

$$\langle vw|\tilde{v}\tilde{w}\rangle_{V\otimes W} = \langle v|\tilde{v}\rangle_V \langle w|\tilde{w}\rangle_W$$

   In other words

$$\mathbb{P}\left(|vw\rangle \to |\tilde{v}\tilde{w}\rangle\right) = \mathbb{P}\left(|v\rangle \to |\tilde{v}\rangle\right) \mathbb{P}\left(|w\rangle \to |\tilde{w}\rangle\right)$$

4. If $|v\rangle = \sum_{i=1}^n \alpha_i |e_i\rangle \in V$ and $|w\rangle = \sum_{j=1}^m \beta_j |f_j\rangle \in W$ are normalized ($\sum |\alpha_i|^2 = \sum |\beta_j|^2 = 1$), then

$$|vw\rangle = \sum_{\substack{i=1:n \\ j=1:m}} \gamma_{i,j} |e_i f_j\rangle$$

   where $\gamma_{i,j} = \alpha_i \beta_j$ for all i, j is also normalized.

5. If $|v\rangle = \sum_{i=1}^n \alpha_i |e_i\rangle \in V$ and $|w\rangle = \sum_{j=1}^m \beta_j |f_j\rangle \in W$, then $|vw\rangle = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \beta_j |e_i f_j\rangle$. (See HW.)

6. If $\mathbb{C}^n \otimes \mathbb{C}^m$ is just a vector space equipped with an inner product, why not write it that way? You can, if you want. (In fact you will, at least if you finish the homework.) But first we need to agree on a linear ordering of the components in the basis, so that $|\psi\rangle = \sum_{i=1}^n \sum_{j=1}^m \gamma_{ij} |e_i f_j\rangle$ can be written as a single vector. A good choice turns out to be the column vector

$$|\psi\rangle = \left(\gamma_{11}, \gamma_{12}, \ldots, \gamma_{1m}, \gamma_{21}, \gamma_{22}, \ldots, \gamma_{2m}, \ldots, \gamma_{n1}, \gamma_{n2}, \ldots, \gamma_{nm}\right)'$$

   This would then uniquely dictate the form of the measurement operators, which become $nm \times nm$ complex valued matrices. This alternative representation can sometimes add perspective to the nature of operators acting on elements of the tensor product space, but at the same time this can be at the expense of complicating the interpretations and formal manipulations. See the HW for some simple examples, and see the next section, §3.2, for the formulation of operators on tensor-product spaces.

So after all of this, what does it mean to say that two particles are entangled?

Definition. *A pair of particles are* entangled *if they are not independent.*

Soon we will have much more to say about entangled particles, but for now let's look at an example:

$$|\psi\rangle = \frac{1}{\sqrt{2}} |ud\rangle - \frac{1}{\sqrt{2}} |du\rangle$$

This particular element of $V \otimes W$ is known as the "singlet" state, often written as $|\text{sing}\rangle$, for reasons that will become clear in the §3.2.

Here are some of its properties, each of which can be derived directly from the connection between overlap and probabilities:

- $\mathbb{P}(\lambda_z^{(1)} = 1 \big| \, |\text{sing}\rangle) = \frac{1}{2}$
- $\mathbb{P}(\lambda_z^{(1)} = \lambda_z^{(2)} \big| \, |\text{sing}\rangle) = 0$
- $\mathbb{P}(\lambda_z^{(1)} = -\lambda_z^{(2)} \big| \, |\text{sing}\rangle) = 1$
- $\mathbb{P}(\lambda_x^{(1)} = 1 \big| \, |\text{sing}\rangle) = \frac{1}{2}$ (and *not* 1 or -1)

Among other things that it is *not*, particle 1 is not like $|l\rangle = \frac{1}{\sqrt{2}}|u\rangle - \frac{1}{\sqrt{2}}|d\rangle$. In fact, it would be a mistake to think of particle 1 as being in *any* state; it is inextricably part of a whole.

To make sense of all of this, the best place to start is with a formulation for what we can actually measure. This takes us back to Hermitian operators, but this time formulated as linear transformations on tensor-product spaces.

3.2 Observations and operators in tensor-product spaces

Given a state $\psi \in V \otimes W$, representing two particles ('1' and '2'), suppose that we want to measure a property of particle 1 corresponding to some Hermitian operator A that acts on V. Let's start by considering the simplest case, $|\psi\rangle = |e_i f_j\rangle$, a basis state in $V \otimes W$ consisting of two independent basis elements from V and W. What is the expected value of the measurement outcome, call it $\lambda_A^{(1)}$? For example, perhaps $V = W = \mathbb{C}^2$, the state space of a two-spin particle, and $A = \sigma_x$, in which case we could write A as $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. How would we compute $\mathbb{E}\left[\lambda_A^{(1)} \mid |\psi\rangle = |e_i f_j\rangle\right]$? Reason this way: Independence implies that

$$\mathbb{E}\left[\lambda_A^{(1)} \mid |\psi\rangle = |e_i f_j\rangle\right] = \mathbb{E}\left[\lambda_A^{(1)} \mid |v\rangle = |e_i\rangle\right] = \langle e_i|A|e_i\rangle = \langle e_i|Ae_i\rangle$$

where $\langle e_i|Ae_i\rangle$ indicates that we can first operate on $|e_i\rangle$ and then take the inner product of the resulting vector $|Ae_i\rangle$ with $|e_i\rangle$.

Alternatively, if we define $A \otimes I\,|e_i f_j\rangle$ to mean $|Ae_i\ f_j\rangle$, then we can express the same computation by

$$\langle e_i f_j|\, A \otimes I\,|e_i f_j\rangle = \langle e_i f_j|Ae_i\ f_j\rangle = \langle e_i|Ae_i\rangle\langle f_j|f_j\rangle = \langle e_i|A|e_i\rangle$$

More generally, consider measuring the product of two properties, λ_A of particle 1 and λ_B of particle 2. In other words, suppose we wish to compute $\mathbb{E}[\lambda_A\lambda_B\big|\,|\psi\rangle]$ for some $|\psi\rangle \in V \otimes V$. (In statistics, the expected value of the product of two random variables X and Y, i.e. $\mathbb{E}[XY]$, is sometimes called the correlation, though the terminology is not entirely consistent and that term could equally well apply to the correlation coefficient, which is different. To make matters worse, in physics the term correlation is more commonly used for $\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$, which is what the statisticians call the covariance! But all of these quantities measure, one way or another, the tendency of the two random variables to move, together, in the same direction.)

Following the example that we have already worked out , let us *define* $A \otimes B\,|e_i f_j\rangle = |Ae_i, Bf_j\rangle$ (which we might also write as $A \otimes B\,|e_i f_j\rangle = A\,|e_i\rangle \otimes B\,|f_j\rangle$), and then work up to the computation of $\mathbb{E}[\lambda_A\lambda_B\big|\,|\psi\rangle]$ for arbitrary $|\psi\rangle \in V \otimes V$, one step at a time.

(i) Extend the domain of $A \otimes B$ from $\{|e_i f_j\rangle\}_{i=1:n, j=1:m}$ to arbitrary $|\psi\rangle = \sum_{i,j} \gamma_{ij}\,|e_i f_j\rangle \in V \otimes W$: Since measurements are described by *linear* operators,

$$A \otimes B\,|\psi\rangle = A \otimes B \sum_{i,j} \gamma_{ij}\,|e_i f_j\rangle = \sum_{i,j} \gamma_{ij} A \otimes B\,|e_i f_j\rangle = \sum_{i,j} \gamma_{ij}\,|Ae_i, Bf_j\rangle$$

(ii) Expectation of $\lambda_A^{(1)}\lambda_B^{(2)}$ for $|\psi\rangle = |e_i f_j\rangle$:

$$\begin{aligned}
\mathbb{E}\left[\lambda_A^{(1)}\lambda_B^{(2)} \mid |e_i f_j\rangle\right] &= \langle e_i f_j|A \otimes B|e_i f_j\rangle = \langle e_i f_j|Ae_i, Bf_j\rangle \\
&= \langle e_i|Ae_i\rangle\langle f_j|Bf_j\rangle \\
&= \langle e_i|A|e_i\rangle\,\langle f_j|B|f_j\rangle = \mathbb{E}\left[\lambda_A \mid |e_i\rangle\right] \cdot \mathbb{E}\left[\lambda_B \mid |f_j\rangle\right]
\end{aligned}$$

(iii) Expectation of $\lambda_A^{(1)}\lambda_B^{(2)}$ for arbitrary $|\psi\rangle = \sum_{i,j}\gamma_{ij}\,|e_i f_j\rangle \in V \otimes W$: By linearity, again.

$$\mathbb{E}\left[\lambda_A^{(1)}\lambda_B^{(2)} \mid |\psi\rangle = \sum_{i,j}\gamma_{ij}\,|e_i f_j\rangle\right] = \left\langle \sum_{i,j}\gamma_{ij}\,|e_i f_j\rangle \middle| \left(\sum_{k.l}\gamma_{kl}\,|Ae_k, Bf_l\rangle\right)\right\rangle$$
$$= \sum_{i,j,k,l}\gamma_{ij}^*\gamma_{kl}\,\langle e_i|A|e_k\rangle\,\langle f_j|B|f_l\rangle$$

The result in (iii) is a little hard to interpret, but the following special case is instructive. Assume that the eigenvectors of A and B coincide, respectively, with the basis vectors of V and W (After all, by the spectral theorem, the eigenvectors of A do provide a possible basis for V, as do those of B for W.) Specifically, assume that

$$A = \sum_{i=1}^{n}\lambda_i\,|e_i\rangle\langle e_i| \qquad \text{and} \qquad B = \sum_{j=1}^{m}\tilde{\lambda}_j\,|f_j\rangle\langle f_j|$$

Then the expression in (iii) simplifies:

$$\mathbb{E}\left[\lambda_A^{(1)}\lambda_B^{(2)} \mid |\psi\rangle = \sum_{i,j}\gamma_{ij}\,|e_i f_j\rangle\right] = \sum_{i,j,k,l}\gamma_{ij}^*\gamma_{kl}\,\langle e_i|A|e_k\rangle\,\langle f_j|B|f_l\rangle$$
$$= \sum_{i,j,k,l}\gamma_{ij}^*\gamma_{kl}\lambda_k\tilde{\lambda}_l\delta_{ik}\delta_{jl}$$
$$= \sum_{i,j}\gamma_{ij}^*\gamma_{ij}\lambda_i\tilde{\lambda}_j$$
$$= \sum_{i,j}|\gamma_{ij}|^2\lambda_i\tilde{\lambda}_j$$

Which we can make sense of: if $|\psi\rangle$ is normalized, then $|\gamma_{ij}|^2$ is the probability (overlap squared) that we find $|\psi\rangle$ in the definite state $|e_i f_j\rangle$ following the measurement. But in this case, the observed quantity is $\lambda_i\tilde{\lambda}_j$. In other words, the expected value is just the sum over all outcomes of the probability-weighted observations.

3.3 Marginals, mixtures, and density operators

Why can't we make sense of the state of a particle that is entangles with another particle? Because if we could than the particle would not be entangled! Reason like this: suppose that particle (1), of a pair of particles (1) and (2), is in the state $|v\rangle \in V \doteq \mathbb{C}_1^n$. Without loss of generality we can then assume that $|v\rangle = |v_1\rangle$, where $|v_1\rangle, \ldots, |v\rangle_n$ is an orthonormal basis for V. If $|f\rangle_1, \ldots, |f\rangle_m$ is an orthonormal basis for $W \doteq \mathbb{C}_1^m$ then we can use $\{|v_i f_j\rangle\}$ as a basis for $V \otimes W$. Let $|\psi\rangle \in V \otimes W$ be the wave function of the pair, particles (1) and (2), where

$$|\psi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} |v_i f_j\rangle$$

Then by assumption, $\mathbb{P}((1) \to |v_1\rangle) = 1$, in which case

$$\sum_{j=1}^{m} |\gamma_{1,j}|^2 = \mathbb{P}((1) \to |v_1\rangle) = 1$$

and this gives the factoring $\gamma_{i,j} = \delta_{i,1}\gamma_{1,j}$, which means that (1) and (2) are independent rather than entangled.

But this is an odd situation. We can certainly measure the first particle (with σ_n, for example), and we even have a formula for the expected outcome

$$\mathbb{E}\left[\lambda_{\sigma_n} \middle| |\psi\rangle\right] = \langle\psi|\sigma_n \otimes I|\psi\rangle$$

Yet we can't talk about its state.

Nevertheless we *can* talk about the probability of observing any particular value from a measurement on (1), and we can use this to develop a very handy representation of our uncertainty about either member of an entangled pair. But before we get started, we should review some common notions in probability and statistics that turn out to be very relevant to the task.

Marginal and mixture distributions in probability. Consider two discrete random variables, X and Y, with joint probability distribution $p(x, y) = p(X = x, Y = y)$ defined by the following table

	$y = 1$	$y = 2$
$x = 1$	$\frac{2}{12}$	$\frac{2}{12}$
$x = 1$	$\frac{5}{12}$	$\frac{1}{12}$
$x = 1$	$\frac{1}{12}$	$\frac{1}{12}$

Definition. *Given a joint distribution $p(x, y)$ on two discrete random variables, X and Y, the distributions*

$$p_X(x) \doteq \sum_y p(x, y) \quad \textit{and} \quad p_Y(y) \doteq \sum_x p(x, y)$$

are known as the marginal distribution on X *and the* marginal distribution on Y*, respectively. The process of deriving the* $p_X(x)$ *is sometimes referred to as* "marginalizing out Y" *or* "marginalizing down to X." *Similar terms are used for the derivation of* $p_Y(y)$.

Applied to the example,

$$p_X(1) = \frac{4}{12} \qquad p_X(2) = \frac{6}{12} \qquad p_X(3) = \frac{2}{12}$$

for p_X, and

$$p_Y(1) = \frac{8}{12} \qquad p_Y(2) = \frac{4}{12}$$

for p_Y.

Definition. *A distribution* $p_X(x)$ *written in the form*

$$p_X(x) = \sum_{k=1}^{n} p_k q_k(x) \tag{25}$$

is called a mixture distribution*, where* $p_1, \ldots, p_n$ *is a probability distribution on* $1, 2, \ldots, n$*, and* $q_k(x)$ *is a distribution on the range of* X *or each* $k = 1 : n$.

Notice that marginal distributions *are always mixture distributions*:

$$p_X(x) = \sum_y p_Y(y) \frac{p(x,y)}{p_Y(y)} = \sum_y p_Y(y) p(x \mid y) \tag{26}$$

$$p_Y(y) = \sum_x p_X(x) \frac{p(x,y)}{p_X(x)} = \sum_x p_X(x) p(y \mid x) \tag{27}$$

In particular, $p_X(x)$ is a mixture of the conditional probabilities given each possible value of Y and $p_Y(y)$ is a mixture of the conditional probabilities given the values of X.

This can be handy, as a kind of divide-and-conquer way of approaching calculations. For instance, concerning the expected value of X, we could use the mixture representation to write it as a sum of expected values, one for each mixing component:

$$\mathbb{E}[X] = \sum_x x \left(\sum_y p_Y(y) p(x \mid y) \right) = \sum_y p_Y(y) \, \mathbb{E}_{q_y}[X] \tag{28}$$

where $q_y(x) = p(x \mid y)$ is the mixture component for the particular value $Y = y$.

Returning to the example, define $q_y(x)$ to be the conditional probability that $X = x$ given $Y = y$:

$$q_1 = \frac{1}{p_Y(1)} \begin{pmatrix} 1/12 \\ 5/12 \\ 2/12 \end{pmatrix} = \begin{pmatrix} 2/8 \\ 5/8 \\ 1/8 \end{pmatrix}$$

$$q_2 = \frac{1}{p_Y(2)} \begin{pmatrix} 2/12 \\ 1/12 \\ 1/12 \end{pmatrix} = \begin{pmatrix} 2/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

where we have represented the two conditional distributions as 3×1 vectors of probabilities (soon to be state vectors in V). Now we can rewrite p_X as a mixture of these two probability vectors:

$$p_X = \frac{2}{3} \begin{pmatrix} 2/8 \\ 5/8 \\ 1/8 \end{pmatrix} + \frac{1}{3} \begin{pmatrix} 2/4 \\ 1/4 \\ 1/4 \end{pmatrix}$$

Mixed states. Return now to the discussion of a pair of particles, (1) and (2), which we will represent generically by $|\psi\rangle \in V \otimes W$, where

$$|\psi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} \langle e_i f_j | e_i f_j \rangle$$

We want to think of the first particle as a mixture of states, one for each state of the second particle. To arrive a formula that is analogous to (26), we will try to rewrite the expected value of an observation A made on particle (1) as a mixture of expected values, as in equation (28):

$$\begin{aligned}
\mathbb{E}\left[\lambda^{(A)} \middle| |\psi\rangle\right] &= \langle \psi | A \otimes I | \psi \rangle = \left\langle \psi \middle| \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} \, |Ae_i, f_j\rangle \right\rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{i,j} \left\langle \sum_{k=1}^{n} \sum_{l=1}^{m} |v_k f_l\rangle \middle| Ae_i, f_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{m} \gamma^*_{k,l} \gamma_{i,j} \langle e_k, f_l | Ae_i, f_j \rangle \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{m} \gamma^*_{k,l} \gamma_{i,j} \langle e_k | Ae_i \rangle \langle f_l | f_j \rangle = \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{n} \gamma^*_{k,j} \gamma_{i,j} \langle e_k | Ae_i \rangle \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n} \sum_{k=1}^{n} \gamma^*_{k,j} \gamma_{i,j} \langle e_k | \, A \, | e_i \rangle = \sum_{j=1}^{m} \left\langle \sum_{k=1}^{n} \gamma_{k,j} \, |e_k\rangle \right| A \left| \sum_{i=1}^{n} \gamma_{i,j} \, |e_i\rangle \right\rangle \\
&= \sum_{j=1}^{m} p_j \left\langle \sum_{k=1}^{n} \frac{\gamma_{k,j}}{\sqrt{p_j}} \, |e_k\rangle \right| A \left| \sum_{i=1}^{n} \frac{\gamma_{i,j}}{\sqrt{p_j}} \, |e_i\rangle \right\rangle \quad \text{(where } p_j = \sum_{i=1}^{n} |\gamma_{i,j}|^2\text{)} \\
&= \sum_{j=1}^{m} p_j \langle v_j | A | v_j \rangle \quad \text{(where } v_j = \sum_{i=1}^{n} \frac{\gamma_{i,j}}{\sqrt{p_j}} \, |e_i\rangle\text{)}
\end{aligned}$$

In summary,

$$\mathbb{E}\left[\lambda^{(A)}\middle|\,|\psi\rangle\right] = \sum_{j=1}^{m} p_j \langle v_j|A|v_j\rangle \quad \text{where} \quad p_j = \sum_{i=1}^{n} |\gamma_{ij}|^2 \quad \text{and} \quad |v_j\rangle = \sum_{i=1}^{n} \frac{\gamma_{i,j}}{\sqrt{p_j}} |e_i\rangle$$

Finally, in analogy to mixtures of distributions, we define any such set of vectors $|v_1\rangle, |v_2\rangle, \ldots, |v_j\rangle \in V$ together with the probabilities $p_1, p_2, \ldots, p_j$ to be a "mixture" or "mixed state," and in the case of the first of a pair of particles in $V \otimes W$, we write

$$M^{(1)} = \sum_{j=1}^{m} p_j \delta_{|v_j\rangle} \tag{29}$$

where the superscript (1) refers to the first of the two particles, and δ_{v_k} is a way of indicating that *if* the k'th component of the mixture were selected (which happens with probability p_k), then particle (1) will act in every regard as though it were actually in the state $|v_k\rangle$. Indeed, once we have identified the mixed state (29) for a particular $|\psi\rangle \in V \otimes W$, we can compute any expectation or specific outcome probability associated with a measurement operator, say O, on V:

$$\mathbb{E}\left[\lambda^{O}\middle|\,|\psi\rangle\right] = \sum_{j=1}^{m} p_j \langle v_j|O|v_j\rangle, \quad \text{and for any eigenvector } \tilde{v} \in V \text{ of } O$$

$$\mathbb{P}\left((1) \xrightarrow{O} \tilde{v}\middle|\,|\psi\rangle\right) = \sum_{j=1}^{m} p_j \,\mathbb{P}\left((1) \xrightarrow{O} \middle|\,|\tilde{v}\rangle\right) = \sum_{j=1}^{m} p_j |\langle \tilde{v}|v_j\rangle|^2$$

Of course we could just as well have made the same computation, but for an observation B made on particle (2), which would lead us to the mixed state $M^{(2)}$ for any $|\psi\rangle \in V \otimes W$:

$$M^{(2)} = \sum_{i=1}^{n} p_i \delta_{|w_i\rangle} \tag{30}$$

where for every $i = 1, \ldots, n$

$$p_i = \sum_{j=1}^{m} |\gamma_{i,j}|^2 \;\text{ and }\; v_i = \sum_{j=1}^{m} \frac{\gamma_{i,j}}{\sqrt{p_i}} |f_j\rangle$$

Expectations and probabilities for particle (2) now follow from the mixed state defined by (30), just as they did for particle (1) using (29).

Remarks.

1. Both $M^{(1)}$ and $M^{(2)}$ can be written in different ways. For example,
$$\frac{1}{2}\delta_{|u\rangle} + \frac{1}{2}\delta_{|d\rangle} = \frac{1}{2}\delta_{|u\rangle} + \frac{1}{2}\delta_{-|d\rangle} \quad \text{and} \quad \frac{1}{3}\delta_{|u\rangle} + \frac{2}{3}\delta_{|u\rangle} = \delta_u$$
2. (see HW) What becomes of the mixed states when particles (1) and (2) are independent, say $|\psi\rangle = |v\rangle \otimes |w\rangle = |vw\rangle$. In this case $M^{(1)} = \delta_{|v\rangle}$ and $M^{(2)} = \delta_{|w\rangle}$, and we say that (1) and (2) are "pure states," as opposed to mixed states. What is more, if either $M^{(1)}$ or $M^{(2)}$ is pure, then both are pure and (1) and (2) are independent.

4 Consequences and Applications

We have now developed enough of the theory to work through a variety of consequences and some of the applications of quantum mechanics. We have chosen five. Each is thought provoking, and each is consequential in its own way.

4.1 Essential ingredients

Before we start, it would be a good idea to revisit the essential ingredients of the theory. In the first section, §4.1, we have organized the important concepts, definitions, and notational conventions, into a list for quick review and access, as needed.

1. **(two-spin particle)** For every measurement direction $n \in \mathbb{R}^3_1$), $\lambda_n \in \{-1, 1\}$, where λ_n is the spin deflection (down or up, respectively) when measured in the n direction.

2. **(state space)** To every direction $n \in \mathbb{R}^3_1$ there is an associated state $|n^+\rangle \in \mathbb{C}^2_1$:

$$\left|n^+\right\rangle \doteq \begin{pmatrix} \cos\left(\frac{\theta}{2}\right) \\ e^{i\phi}\sin\left(\frac{\theta}{2}\right) \end{pmatrix} \in \mathbb{C}^2_1$$

   where ϕ and θ are the polar and azimuth angles of the polar coordinates of n. Write $|n^-\rangle$ for the state associated with the direction $-n$. Since $-n$ has polar angle $\pi - \phi$ and azimuth angle $\theta + \pi$, $|n^-\rangle$ and $|n^+\rangle$ are orthogonal in $\mathbb{C}^2_1$

3. **(spare degree of freedom)** For any η, $e^{i\eta}\,|n^+\rangle$ and $|n^+\rangle$ refer to the same physical state.

4. **(cardinal directions)**

$$|u\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |d\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad |r\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad |l\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{pmatrix} \quad |i\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} \end{pmatrix} \quad |o\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{-i}{\sqrt{2}} \end{pmatrix}$$

5. **(n−spin particles & $\mathbf{C}^n_1$)** In general, discrete wave functions of single particles are formulated as elements of the vector space $\mathbb{C}^n$ with inner product $\langle\psi|\phi\rangle \in \mathbb{C}$. Complex-valued inner products are assumed to be conjugate linear in the first argument and linear in the second: if $|\Psi\rangle = \sum_{i=1}^{a} \alpha_i\,|\psi_i\rangle$ and $|\Phi\rangle = \sum_{j=1}^{b} \beta_j\,|\phi_j\rangle$ then

$$\langle\Psi|\Phi\rangle = \sum_{i=1}^{a}\sum_{j=1}^{b} \alpha_i^*\beta_j$$

   The dimension n is the number of possible outcomes of any given measurement of the state.

6. **(measurements)** A measurement A returns one of n real numbers, $\lambda_k^{(A)}$, $k = 1 : n$, and leaves the measured particle in one of n corresponding states $\left|e_k^{(A)}\right\rangle$, $k = 1 : n$. The set of possible states

that result from a measurement are orthonormal: $\left\langle e_k^{(A)} \middle| e_l^{(A)} \right\rangle = \delta_{k,l}$. (The superscript (A) is rarely needed.)

7. **(definite state)** The state resulting from a measurement is called a *definite state*. But it is just another state, and only "definite" in the sense that repeating the same measurement will give the same answer and will not alter the state.

8. **(Hermitian operators)** Given a measurement A, the set of measured values λ_k and their corresponding definite states $|e_k\rangle$, $k = 1 : n$, can be combined to define a Hermitian matrix, also called A:

$$A \doteq \sum_{k=1}^{n} \lambda_k |e_k\rangle\langle e_k|$$

What's more, every $n \times n$ Hermitian matrix corresponds to a measurement (in principle—some might be difficult to implement) through it's *spectral representation*: If A is an $n \times n$ Hermitian matrix, then there exists scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and orthonormal vectors $|e_1\rangle, \ldots, |e_n\rangle \in \mathbb{C}^n$ such that

$$A = \sum_{k=1}^{n} \lambda_k |e_k\rangle\langle e_k|$$

In both cases, $\lambda_1, \ldots, \lambda_n$ are eigenvalues of A with corresponding eigenvectors $|e_1\rangle, \ldots, |e_n\rangle$.

9. **(spin operators in $\mathbb{C}_1^2$)** For any $n = (n_x, n_y, n_z) \in \mathbb{R}_1^3$, the operator that measures spin in the $\pm n$ direction (which is called σ_n) can be written as

$$\sigma_n = n_x\sigma_x + n_y\sigma_y + n_z\sigma_z = \begin{pmatrix} \cos\theta & e^{-i\phi}\sin\theta \\ e^{i\phi}\sin\theta & -\cos\theta \end{pmatrix}$$

where ϕ and θ are the polar and azimuth angles of n, and σ_x, σ_y, and σ_z are the operators measuring spin in the x, y, and z directions respectively:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

10. **(probabilities & squared overlap)** For any initial state $|\psi\rangle \in \mathbb{C}_1^n$ and measurement A, the probability that the observation results in the definite state $|e_k\rangle$ (and hence yields the value λ_k) is given by the "squared overlap": $|\langle\psi|e_k\rangle|^2$. The inner product $\langle\psi|e_k\rangle$ is sometimes called the probability amplitude.

$$\mathbb{P}\left(|\psi\rangle \xrightarrow{A} |e_k\rangle\right) = |\langle\psi|e_k\rangle|^2$$

In fact, the expression is unambiguous without indicating A on the left-hand side. All that matters is that a measurement was made for which $|e_k\rangle$ is a definite state. Hence we could write

$$\mathbb{P}\left(|\psi\rangle \to |\phi\rangle\right) = |\langle\psi|\phi\rangle|^2$$

for any two states $|\psi\rangle$ and $|\phi\rangle$, and there would be no need to specify a particular measurement.

11. **(probabilities & overlap in $\mathbb{C}_1^2$)** For any $n, m \in \mathbb{R}_1^3$

$$\mathbb{P}(\lambda_m = 1 \Big| \, |\psi\rangle = \big|n^+\big\rangle) = \frac{1 + m \cdot n}{2} = \mathbb{P}(\big|n^+\big\rangle \to \big|m^+\big\rangle) = \big|\big\langle m^+ \big| n^+ \big\rangle\big|^2$$
$$\mathbb{P}(\lambda_m = -1 \Big| \, |\psi\rangle = \big|n^+\big\rangle) = \frac{1 - m \cdot n}{2} = \mathbb{P}(\big|n^+\big\rangle \to \big|m^-\big\rangle) = \big|\big\langle m^- \big| n^+ \big\rangle\big|^2$$

12. **(expectations)** In general, if the measurement A is represented by measurement values λ_k, $k = 1 : n$, and corresponding definite states $|e_k\rangle$, $k = 1 : n$, then given any starting state $|\psi\rangle$ the observable value of the measurement, call it λ^A or just λ, is a random variable with probabilities given by the squared overlap:

$$\mathbb{P}\left(\lambda = \lambda_k \big| \, |\psi\rangle\right) = |\langle\psi|e_k\rangle|^2 \tag{31}$$

Its *expected value* is the probability-weighted average measurement:

$$\mathbb{E}\left[\lambda \big| \, |\psi\rangle\right] = \sum_{k=1}^{n} \mathbb{P}\left(\lambda = \lambda_k \big| \, |\psi\rangle\right) \lambda_k = \sum_{k=1}^{n} \lambda_k |\langle\psi|e_k\rangle|^2 \tag{32}$$

This can be rewritten, conveniently, as a simple matrix-vector operation:

$$\mathbb{E}\left[\lambda \big| \, |\psi\rangle\right] = \langle\psi|A|\psi\rangle \tag{33}$$

13. **(probabilities as expectation)** For any $|\psi\rangle$ and $|\phi\rangle$

$$\mathbb{P}\left(|\psi\rangle \to |\phi\rangle\right) = |\langle\psi|\phi\rangle|^2 = \langle\psi|\phi\rangle\langle\psi|\phi\rangle^* = \langle\psi|\phi\rangle\langle\phi|\psi\rangle = \langle\psi|A|\psi\rangle$$

where A is the particular (rank-one) Hermitian operator $A = |\phi\rangle\langle\phi|$, which has the single non-zero eigenvalue $\lambda = 1$. All the other eigenvalues are zero.

14. **(more expectations)** If H, representing a measurement, is an $n \times n$ Hermitian matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $|e_1\rangle, \ldots, |e_n\rangle$, then we can define a new measurement by changing the observables, i.e. the eigenvalues, without changing the eigenvectors: for any function $f : \mathbb{R} \to \mathbb{R}$

$$\tilde{H} \doteq \sum_{k=1}^{n} f(\lambda_k) \, |e_k\rangle\langle e_k|$$

The relationship between H and $\tilde{H}$ is conveniently summarized by writing $\tilde{H} = f(H)$. This extends the reach of expectations to expectations of functions of the observables:

$$\mathbb{E}\left[\lambda \big| \, |\psi\rangle\right] = \langle\psi|f(A)|\psi\rangle = \sum_{k=1}^{n} f(\lambda_k)|\langle\psi|e_k\rangle|^2$$

(In fact, for well-behaved function f the relationship $\tilde{H} = f(H)$ can be taken literally, e.g. you can check that the matrix HH has the same eigenvectors as H, but the corresponding eigenvalues are squared.)

15. **(Pythagorean principle for spins states in $\mathbb{C}_1^2$)**

$$E\big[\lambda_{\mathbf{x}} \big| \, |\psi\rangle\big]^2 + E\big[\lambda_{\mathbf{y}} \big| \, |\psi\rangle\big]^2 + E\big[\lambda_{\mathbf{z}} \big| \, |\psi\rangle\big]^2 = 1$$

16. **(commutators and the ordering of measurements)** Given any two Hermitian matrices A and B, the measurement A and followed by the measurement B will not typically give the same results and the measurement of B followed by A. In fact, they give the same results for every starting state $|\psi\rangle$ if and only if $[A, B] = 0$, where $[A, B] \doteq AB - BA$ is called the commutator. When $[A, B] = 0$, A and B are said to commute.

17. **(uncertainty principle)** Given a starting state $|\psi\rangle$ and a measurement A, $\lambda^{(A)}$ is a random variable with some mean μ_A and variance δ_A^2, both of which depend on $|\psi\rangle$:

$$\mu_A = \mathbb{E}\left[\lambda^{(A)} \middle| |\psi\rangle\right] \quad \text{and} \quad \delta_A^2 = \mathbb{E}\left[(\lambda^{(A)})^2 - \mu_A^2 \middle| |\psi\rangle\right]$$

The *uncertainty principle* asserts that for any two measurements A and B and any starting state $|\psi\rangle$, we can not expect to simultaneously accurately predict the outcomes of measuring A and B unless A and B happen to commute:

$$\delta_A \delta_B \geqslant \frac{1}{2} |\langle\psi|[A, B]|\psi\rangle|$$

18. **(Hamiltonians and unitary operators)** The *Hamiltonian*, H, is the Hermitian operator corresponding to the measurement of the total energy. This will usually involve continuous variables like time, position and velocity. Unitary operators (matrices in finite dimensions) are defined by the property $U^\dagger U = UU^\dagger = I$. Unitary operators obey a spectral decomposition that is similar to the decomposition obeyed by Hermitian operators: If U is an $n \times n$ unitary matrix then there exists $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ and $|e_1\rangle, \ldots, |e_n\rangle \in \mathbb{C}^n$ such that

$$U = \sum_{k=1}^{n} \lambda_k |e_k\rangle\langle e_k|$$

where $|\lambda_k|^2 = 1$ (i.e. $\lambda_k = e^{i\theta_k}$ for some $\theta_k \in \mathbb{R}$) and $\langle e_k|e_l\rangle = \delta_{k,l}$, for all k and l. The importance of unitary operators stems from the fact that they preserve normalization and inner products: If U is unitary, then

$$\langle (U|\phi\rangle)|(U|\psi\rangle)\rangle = \langle\phi| U^\dagger U |\psi\rangle = \langle\phi|\psi\rangle$$

(Think of a unitary operator as a generalization of a rigid rotation.)

19. **(the Schrödinger equation)** The Hamiltonian defines the continuous evolution of a state $|\psi\rangle$, which will typically depend on both time t and position r, $|\psi(t, r)\rangle$. The Hamiltonian determines the evolution via the continuous evolution of a unitary operator:

$$|\psi(t, r)\rangle = U_t |\psi(0, r)\rangle = e^{-\frac{i}{\hbar} t H(r)} |\psi(0, r)\rangle$$

where $H(r)$ is the Hamiltonian, assumed here to be time-independent, and the exponential of an operator (or matrix) is defined via the spectral representation used earlier to define $f(A)$ for any Hermitian operator A. Or, written as a differential equation

$$\frac{\partial}{\partial t} |\psi(t, r\rangle = \frac{-i}{\hbar} H(r) |\psi(t, r)\rangle$$

The Schrödinger equation can be used to solve for the time evolution of a state, given the Hamiltonian, or in reverse: given a desired unitary transformation and time t, find the Hamiltonian that will generate the desired transformation in the desired time.

20. **(tensor-product space, $\mathbf{V \otimes W}$)** For any orthonormal basis $|e_1\rangle, \ldots, |e_n\rangle$ of V and orthonormal basis $|f_1\rangle, \ldots, |f_m\rangle$ of W, the tensor-product space of $V \otimes W$ can be written as

$$V \otimes W = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \, |e_i f_j\rangle \,\middle|\, \gamma_{ij} \in \mathbb{C}, \ \forall \, 1 \leqslant i \leqslant n, \ 1 \leqslant j \leqslant m \right\} \tag{34}$$

21. **(inner product in $\mathbf{V \otimes W}$)** For any $|\psi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \, |e_i f_j\rangle$ and $|\phi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} \, |e_i f_j\rangle$

$$\langle \phi | \psi \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij}^{*} \gamma_{ij}$$

and hence

$$\langle \psi | \psi \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} |\gamma_{ij}|^2$$

(i) It follows that $\langle e_i j_j | e_k f_l \rangle = \delta_{i,k} \delta_{j,l}$ (take $|\phi\rangle = |e_i f_j\rangle$ and $|\psi\rangle = |e_k, f_l\rangle$)

(ii) The inner product $\langle \phi | \psi \rangle$ is conjugate linear in the first argument and linear in the second.

22. **(independence)** The vector

$$|\psi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \, |e_i f_j\rangle \in V \otimes W$$

represents two *independent particles*, particle 1 in V and particle 2 in W, if and only if there exists $|v\rangle = \sum_{i=1}^{n} \alpha_i \, |e_i\rangle \in V$ and $|w\rangle = \sum_{j=1}^{m} \beta_j \, |f_j\rangle \in W$ such that $\gamma_{ij} = \alpha_i \beta_j$ for all $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant m$. If particles 1 and 2 are not independent, then we say that they are *entangled*.

(i) If follows that $|e_i f_j\rangle$ represents two independent particles (just take $|v\rangle = |e_i\rangle$ and $|w\rangle = |f_j\rangle$).

(ii) We often indicate that the pair of $|v\rangle$ and $|w\rangle$ are independent in $V \otimes W$ by writing their joint state as $|vw\rangle$.

(iii) Another way to write $|vw\rangle$ is to write the pair as a *tensor product*, $|vw\rangle = |v\rangle \otimes |w\rangle$, where the *tensor-product operator*, mapping $V \times W \to V \otimes W$, is bi-linear, and therefore defined by its actions on the basis elements of V and W. The tensor-product notation makes it easy to find the vector in $V \otimes W$ associated with any pair from V and W. For an example, consider $\mathbb{C}^2 \otimes \mathbb{C}^2$ using the up-down basis for both particles:

$$\begin{aligned} |ll\rangle &= \left(\frac{1}{\sqrt{2}} |u\rangle - \frac{1}{\sqrt{2}} |d\rangle \right) \otimes \left(\frac{1}{\sqrt{2}} |u\rangle - \frac{1}{\sqrt{2}} |d\rangle \right) \\ &= \frac{1}{2} |u\rangle \otimes |u\rangle - \frac{1}{2} |u\rangle \otimes |d\rangle - \frac{1}{2} |d\rangle \otimes |u\rangle + \frac{1}{2} |d\rangle \otimes |d\rangle \\ &= \frac{1}{2} |uu\rangle - \frac{1}{2} |ud\rangle - \frac{1}{2} |du\rangle + \frac{1}{2} |dd\rangle \end{aligned}$$

(iv) If $|a\rangle, |c\rangle \in V$ and $|b\rangle, |d\rangle \in W$, then

$$\langle ab | cd \rangle_{V \otimes W} = \langle a | c \rangle_V \langle b | d \rangle_W$$

Remark: To see this, remember that $|ab\rangle$ is short for $|a\rangle \otimes |b\rangle$, meaning that "a" and "b" are independent particles. Similarly, "c" and "d" are independent. Consequently

$$|ab\rangle = \sum_{i=1}^{n} \alpha_i |e_i\rangle \otimes \sum_{j=1}^{m} \beta_j |f_j\rangle$$
$$|cd\rangle = \sum_{k=1}^{n} \tilde{\alpha}_k |e_k\rangle \otimes \sum_{l=1}^{m} \tilde{\beta}_l |f_j\rangle$$

Now compute $\langle ab|cd\rangle$. You get four summations, two of which disappear because of the orthogonality of the two bases, in V and W respectively. One of the remaining sums gives $\langle a|b\rangle_V$ and the other gives $\langle c|d\rangle_W$.

23. **(measurements in $\mathbf{V} \otimes \mathbf{W}$)** If A and B are Hermitian operators (measurements, Hermitian matrices) on V and W then $A \otimes B : V \otimes W \to V \otimes W$ is defined by linear extension from $A \otimes B\, |e_i f_j\rangle = |Ae_i Bf_j\rangle \ \forall i, j$:
for any $|\phi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} |e_i f_j\rangle$

$$A \otimes B = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} |Ae_i Bf_j\rangle$$

Which then also applies to any independent pair, $v \in V$ and $w \in W$: $A \otimes B\, |vw\rangle = |AvBw\rangle$.

24. **(characterization of Hermitian operators on $\mathbf{V} \otimes \mathbf{W}$)** From (23.), it follows that

$$(A \otimes B)^\dagger = A^\dagger \otimes B^\dagger \tag{35}$$

which implies that $A \otimes B$ is Hermitian if and only if A and B are both Hermitian. And from (35) if follows that every Hermitian operator H on $V \otimes W$ can be written as

$$H = \sum_{k=1}^{N} A_k \otimes B_k$$

for some N, N Hermitian operators $A_1, \ldots A_N$ on V, and N more Hermitian operators $B_1, \ldots B_N$ on W.

25. **(three or more particles)** These definitions and their consequences are recursive, e.g. we can use the basis $\{|e_i f_j\rangle\}_{i=1:n, j=1:m}$ for $V \otimes W$ and $|u_1\rangle, \ldots, |u_l\rangle$ for U to define the tensor-product space

$$(V \otimes W) \otimes U = V \otimes W \otimes U = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} \sum_{k=1}^{l} \gamma_{ijk} |e_i f_j u_k\rangle \,\middle|\, \gamma_{ijk} \in \mathbb{C},\ \forall\ 1 \leqslant i \leqslant n\ \ 1 \leqslant j \leqslant m,\ 1 \leqslant k \leqslant l \right\}$$

26. **(probabilities and expectations)** In general, the rules for computing probabilities of outcomes and expectations of observables are the same, which should not be surprising since the tensor-product space is also an inner-product space, but with dimension nm (which id the dimension of $V \otimes W$, and the product of the dimensions of V and W). Thus if H is a Hermitian operator on $V \otimes W$, then H is defined by nm orthonormal eigenvectors, $|\phi_1\rangle, \ldots, |\phi_{nm}\rangle \in V \times W$, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_{nm}$.

(i) The probability, given the current state $|\psi\rangle \in V \otimes W$, that the measurement results in the definite state $|\phi_k\rangle$, for some $k = 1, 2, \ldots nm$:

$$\mathbb{P}\left(|\psi\rangle \xrightarrow{H} |\phi_k\rangle\right) = |\langle\psi|\phi_k\rangle|^2$$

Or, as noted in item (10.), just drop the H and write $\mathbb{P}(|\psi\rangle \to |\phi\rangle) = |\langle\psi|\phi\rangle|^2$ for any pair $|\psi\rangle, |\phi\rangle \in V \otimes W$.

(ii) Consider $\mathbb{C}_1^2 \otimes \mathbb{C}_1^2$ and the singlet state $|sing\rangle = \frac{1}{\sqrt{2}}|ud\rangle - \frac{1}{\sqrt{2}}|du\rangle$, for example. Conclude that

$$\mathbb{P}\left(|sing\rangle \to |ud\rangle\right) = |\langle sing|ud\rangle|^2 = \frac{1}{2} \quad \text{and} \quad \mathbb{P}\left(|sing\rangle \to |du\rangle\right) = |\langle sing|du\rangle|^2 = \frac{1}{2}$$

(iii) Item (i) leads to the expected value of $\lambda^{(H)}$ given $|\psi\rangle$:

$$\mathbb{E}\left[\lambda^{(H)}\middle||\psi\rangle\right] = \sum_{k=1}^{nm} \mathbb{P}\left(|\psi\rangle \xrightarrow{H} |\phi_k\rangle\right)\lambda_k = \sum_{k=1}^{nm} \lambda_k |\langle\psi|\phi_k\rangle|^2 = \langle\psi|H|\psi\rangle$$

which is identical to the expression derived for single spin-state particles.

(iv) Following the same reasoning used in item (13.), for any $|\psi\rangle$ and $|\phi\rangle$ we can write

$$\mathbb{P}\left(|\psi\rangle \to |\phi\rangle\right) = \langle\psi|H|\psi\rangle$$

where $H = |\phi\rangle\langle\phi|$.

27. **(special case: expectations when $\mathbf{H} = A \otimes B$)** If A and B are the Hermitian operators that correspond, respectively, to the measurements of $\lambda^{(A)}$ and $\lambda^{(B)}$, then, for example:

(a)
$$\mathbb{E}[\lambda_A \lambda_B \big| |\psi\rangle] = \langle\psi| A \otimes B |\psi\rangle$$

(b)
$$\mathbb{E}[\lambda_A \big| |\psi\rangle] = \langle\psi| A \otimes I |\psi\rangle$$

(c)
$$\mathbb{E}[\lambda_A + \lambda_B \big| |\psi\rangle] = \langle\psi| A \otimes I |\psi\rangle + \langle\psi| I \otimes B |\psi\rangle$$

28. **(flattening)** Tensor-product spaces are inner-product spaces, i.e. vector spaces equipped with an inner product. As such, it is sometimes convenient to translate ("flatten") them into the familiar vector/matrix representation. This can be done by the following rules: Let

$$V \otimes W = \left\{ \sum_{i=1}^{n}\sum_{j=1}^{m} \gamma_{ij} |e_i f_j\rangle \middle| \gamma_{ij} \in \mathbb{C},\ \forall\ 1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant m \right\}$$

and assume that A is represented in the basis $|e_1\rangle, \ldots, |e_n\rangle$ as an $n \times n$ matrix with components a_{ij}, and B is represented in the basis $|f_1\rangle, \ldots, |f_m\rangle$ as an $m \times m$ matrix with components b_{ij}. Then $A \otimes B$ can be represented by an $nm \times nm$ matrix

$$A \otimes B \to \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ & & \vdots & \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{pmatrix}$$

where $a_{ij}B$ is the $m \times m$ matrix whose k, l component is $a_{ij}b_{kl}$. The corresponding "flattening" of $|\psi\rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} |e_i f_j\rangle$ is then

$$|\psi\rangle \to \begin{pmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1m} \\ \vdots \\ \vdots \\ \gamma_{n1} \\ \gamma_{n2} \\ \vdots \\ \gamma_{nm} \end{pmatrix}$$

29. **(a mixture of states)** We can represent a collection of K states, $|v_1\rangle, \ldots, |v_K\rangle$, chosen randomly from a distribution $p_1, \ldots, p_K$, where $\sum_{k=1}^{K} p_k = 1$, as

$$M = \sum_{k=1}^{K} p_k \delta_{v_k}$$

This is called a mixed state, and it corresponds to the notion of a mixture distribution in probability. A mixed state is more general than a state, in the sense that it can represent a single state, $|v\rangle$, by taking $K = 1$, $|v_1\rangle = |v\rangle$, and $p_1 = 1$. In this case, M is said to be a pure state.

30. **(probabilities and expectations for mixtures)** Given that a mixed state corresponds to randomly choosing a state $|v\rangle$ from the ensemble of states $|v_1\rangle, \ldots, |v_K\rangle$, according to the probabilities $p_1, \ldots, p_K$, equations (31) and (33) have a natural generalization using M in place of $|\psi\rangle$:

$$\mathbb{P}(\lambda_A = \lambda_k \big| M) = \sum_{k=1}^{K} p_k |\langle v_k | e_k \rangle|^2$$

$$\mathbb{E}[\lambda_A \big| M] = \sum_{k=1}^{K} p_k \langle v_k | A | v_k \rangle$$

31. **(mixture vs superposition)** Consider the mixed state $M = \frac{1}{2}|u\rangle + \frac{1}{2}|d\rangle$ and the pure state $|r\rangle = \frac{1}{\sqrt{2}}|u\rangle + \frac{1}{\sqrt{2}}|d\rangle$. Consider one particle that has been prepared in the M state (usually, by virtue of being one of a pair of two entangled particles) and another in the $|r\rangle$ state. How are they different? If $A = \sigma_z$ (measuring up vs down) then both $|r\rangle$ and M yield the same conditional probability distribution on the outcome: $\lambda_{\sigma_z} = \pm 1$ with equal probabilities. But the situation is entirely different for the measurement of right vs left, for which $A = \sigma_x$: The distribution on the outcome starting with the mixture M is again 50/50, ± 1. But if we start in the prepared state $|r\rangle$ then we will always observe $\lambda_{\sigma_x} = 1$. Hence there is no sense in which $|r\rangle$ is a 50/50 mixture of $|u\rangle$ and $|d\rangle$. Instead, it is a superposition of the two wave functions $|u\rangle$ and $|d\rangle$.

32. **(marginals and mixtures)** If particles 1 and 2 are entangled, then it makes no sense to talk about their respective individual states. But you can still calculate the probabilities of outcomes of any

measurement made on one or the other member of the pair. These are best described using mixed states, already introduced in item 29..

Let $\{|e_i\rangle\}_{i=1:n}$ be an orthonormal basis for V, let $\{|f_j\rangle\}_{j=1:m}$ be an orthonormal basis for W, and let $|\psi\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m}\gamma_{ij}\,|e_i f_j\rangle$. Then particle 1 is in the mixed state

$$M^{(1)} = \sum_{j=1:m} p_j \delta_{|v_j\rangle}$$

where, $p_j = \sum_{i=1:n}|\gamma_{ij}|^2$ and $|v_j\rangle = \sum_{i=1:n}\frac{\gamma_{ij}}{\sqrt{p_j}}\,|e_i\rangle$ for each $j = 1,\ldots,m$, and particle 2 is in the mixed state

$$M^{(2)} = \sum_{i=1:n} q_i \delta_{|w_i\rangle}$$

where, $q_i = \sum_{j=1:m}|\gamma_{ij}|^2$ and $|w_i\rangle = \sum_{j=1:m}\frac{\gamma_{ij}}{\sqrt{q_i}}\,|f_j\rangle$ for each $i = 1,\ldots,n$.

In essence, each particle's mixed state is a mixture over each possible state of the other particle.

i. The representation of a mixed state is not unique.

ii. These for expressions for $M^{(1)}$ and $M^{(2)}$ might have fewer than m and n terms, respectively. In fact, the maximum number of terms needed for either particles is $\min(n,m)$

iii. These expressions might contain repeated vectors, e.g. $|v_i\rangle = |v_{i'}\rangle$, $i \neq i'$ or $|w_j\rangle = |w_{j'}\rangle$, $j \neq j'$. These terms can be combined—simply add the respective probabilities.

33. **(probabilities and expectations for marginal distributions)** Continuing with item 32. and following the prescriptions in item 30., but applied to the state $|\psi\rangle$: for any Hermitian operator A acting on V, with spectral representation $A = \sum_{i=1}^{n}\lambda_i\,|a_i\rangle\langle a_i|$,

$$\mathbb{P}(\lambda_A = \lambda_i | M^{(1)}) = \sum_{j=1}^{m} p_j |\langle v_j | a_i\rangle|^2$$

$$\mathbb{E}[\lambda_A | M^{(1)}] = \sum_{j=1}^{m} p_j \,\langle v_j|\, A\,|v_j\rangle$$

4.2 EPR paradox and Bell's Inequality

In 1935, Albert Einstein, Boris Podolsky, and Nathan Rosen proposed the following thought experiment, attempting to argue that the description of physical reality provided by quantum mechanics was incomplete: Suppose that two particles, 1 and 2, are prepared in the maximally entangled single state $|\psi\rangle = \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle)$[13]. Now, even if particles 1 and 2 are separated by 20 light years, measuring the spin of particle 1 in *any* direction $n \in \mathbb{R}^3_1$ would enable us to completely determine the spin of particle 2 in that same direction. After all, two particles in the singlet state have opposite spins in every direction. It would appear that information could be transmitted instantaneously, violating relativity.

To account for this paradox, Einstein, Podolsky, and Rosen argued that the theory was incomplete in the sense that the state of the pair had not been fully specified. A way out of the dilemma would be to assume a *hidden variable*, say h, which represented a property of the particles, outside of the existing theory, sufficient to predetermine the outcomes of these measurements, thereby preserving *locality*. John Stewart Bell set out to demonstrate that a hidden variable of this type could be formulated in such a way as to be consistent with the known properties of quantum measurements. Instead, In 1964 he wound up proving that no such variable could exist. We will present his argument in the section, but let us start by asking what this hidden variable h would look like if it actually existed.

Suppose that h belongs to $\mathcal{H}$, some space of variables, and let $P_{\mathcal{H}}(h)$ be a probability distribution on $\mathcal{H}$. By assumption, h would determine the outcomes of any pair of measurements ($\lambda_r^{(1)}$ and $\lambda_q^{(2)}$), $r, q \in \mathbb{R}^3_1$. Formally,

$$(\lambda_r^{(1)}, \lambda_q^{(2)}) = (\lambda_r^{(1)}(h), \lambda_q^{(2)}(h)) \in \{-1, 1\}^2,$$

where $\lambda_r^{(1)}(h), \lambda_q^{(2)}(h)$ are deterministic functions of h.

As a concrete example, let $|\psi\rangle = \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle)$ (the "singlet state"). For any $r \in \mathbb{R}^3_1$ we could choose $\mathcal{H} = \{1, 2\}$, and $P_{\mathcal{H}}(1) = P_{\mathcal{H}}(2) = \frac{1}{2}$, and then define $(\lambda_r^{(1)}(1), \lambda_r^{(2)}(1)) = (1, -1)$ and $(\lambda_r^{(1)}(2), \lambda_r^{(2)}(2)) = (-1, 1)$. Which would give the correct probabilities for the particular case of measuring both particles in the same direction.

As another example, starting with the same singlet state $|\psi\rangle = \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle)$, consider $r, q \in \mathbb{R}^3_1$, where r is perpendicular to q, and pick $\mathcal{H}, P_{\mathcal{H}}, (\lambda_r^{(1)}(h)$ and $\lambda_q^{(2)}(h))$ such that $P_{\mathcal{H}}(h : \lambda_r^{(1)}(h) = i, \lambda_q^{(2)}(h) = j) = \frac{1}{4}$ for every $i, j \in \{-1, 1\}$. Again, this would be consistent with the known outcomes of these particular Stern-Gerlach experiments.

So what, if anything, goes wrong for a more general set of Stern-Gerlach experiments?

Theorem. *(Bell's Theorem) Let $P_{\mathcal{H}}(h)$ be a probability on $\mathcal{H}$, and for every $r, q \in \mathbb{R}^3_1$ let $\lambda_r^{(1)}(h), \lambda_q^{(2)}(h) \in \{-1, 1\}$ be two functions on $\mathcal{H}$. On the other hand, let $\mathbb{P}$ be the corresponding distribution on $\lambda_r^{(1)}$ and $\lambda_q^{(2)}$ given the singlet state $|\psi\rangle = |S\rangle$:*

$$\mathbb{P}\left(\lambda_r^{(1)} = a, \lambda_q^{(2)} = b \mid |\psi\rangle = |S\rangle\right) \quad \forall\, a, b \in \{-1, 1\}.$$

[13] They actually discussed position and momentum, rather than spin states.

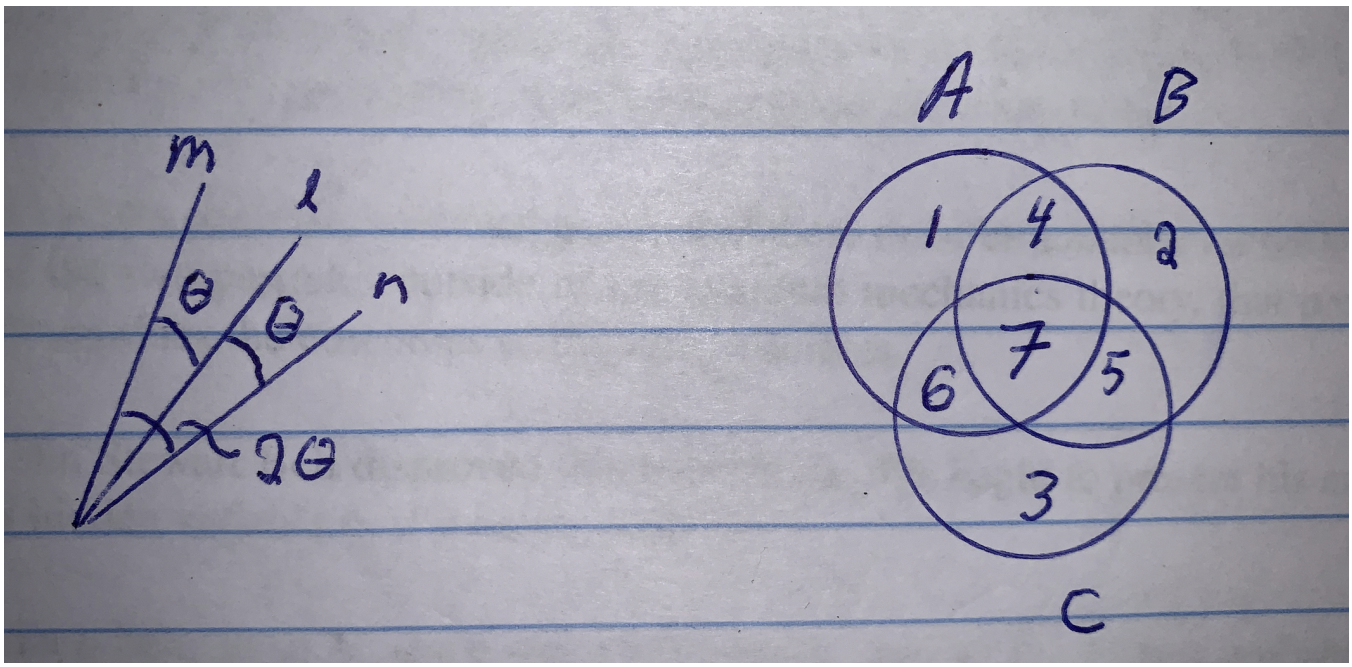


Figure 12: **Left figure.** Three unit vectors ($n, m, l \in \mathbb{R}^3_1$) all in the same plane, and separated by a small angle θ. **Right figure.** Venn diagram with three regions, A, B, and C, in the hidden-variable space $\mathcal{H}$. The union of the three regions define seven disjoint sub-regions, labelled $1, 2, \ldots, 7$.

Then $\mathbb{P}\left(\lambda_r^{(1)} = a, \lambda_q^{(2)} = b \mid |\psi\rangle = |S\rangle\right)$ *cannot be equal to* $P_{\mathcal{H}}\left(\lambda_r^{(1)}(h) = a, \lambda_q^{(2)}(h) = b\right)$ *for all* $q, r \in \mathbb{R}^3_1$ *and* $a, b \in \{-1, 1\}$.

The proof proceeds is by contradiction. Suppose that there is such a $P_{\mathcal{H}}$. With $|\psi\rangle = \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle)$, choose three directions, $n, m, l \in \mathbb{R}^3_1$, all in a plane, and all pointing in nearly the same direction. The direction l is between the directions of m and n, with identical angles between m and l as between n and l. Call the angle θ. See figure 12, left-hand side.

Note that under $|\psi\rangle$, $\lambda_r^{(1)}(h) = 1$ is equivalent to $\lambda_r^{(2)}(h) = -1$ for any direction r. Therefore, if θ is small enough, all three of the probabilities

$$\mathbb{P}(\lambda_n^{(1)}(h) = 1, \lambda_m^{(2)}(h) = 1)$$
$$\mathbb{P}(\lambda_n^{(1)}(h) = 1, \lambda_l^{(2)}(h) = 1)$$
$$\mathbb{P}(\lambda_l^{(1)}(h) = 1, \lambda_m^{(2)}(h) = 1)$$

should be small. Bell's theorem turns out to be a statement about the delicate relationship among these probabilities.

As you will show in the homework, for any pair of directions, $q, r \in \mathbb{R}^3_1$,

$$\mathbb{P}(\lambda_q^{(1)} = 1, \lambda_r^{(2)} = 1) = \frac{1}{2}\sin^2(\varphi/2) \tag{36}$$

where φ is the angle between q and r.

With this critical formula in hand, all we need is this general (though somewhat subtle) fact about probabilities: For **any** probability $P_{\mathcal{H}}$ on $\mathcal{H}$, if $A, B, C \subseteq \mathcal{H}$, then

$$P_{\mathcal{H}}(A \cap \bar{B}) \leqslant P_{\mathcal{H}}(A \cap \bar{C}) + P_{\mathcal{H}}(C \cap \bar{B}) \tag{37}$$

where $\bar{E}$ refers to the complement of the set $E \in \mathcal{H}$. To see why this is true, we will use the Venn diagram depicted in the right-hand side of figure 12. Let us write $P_{\mathcal{H}}(k)$ to represent the probability of the region labeled k, for each $k = 1, \ldots, 7$. Since the seven regions are disjoint,

$$\begin{aligned} P_{\mathcal{H}}(A \cap \bar{B}) &= P_{\mathcal{H}}(1) + P_{\mathcal{H}}(6) \\ P_{\mathcal{H}}(A \cap \bar{C}) &= P_{\mathcal{H}}(1) + P_{\mathcal{H}}(4) \\ P_{\mathcal{H}}(C \cap \bar{B}) &= P_{\mathcal{H}}(3) + P_{\mathcal{H}}(6) \end{aligned}$$

and hence

$$P_{\mathcal{H}}(A \cap \bar{B}) = P_{\mathcal{H}}(1) + P_{\mathcal{H}}(6) \leqslant P_{\mathcal{H}}(1) + P_{\mathcal{H}}(4) + P_{\mathcal{H}}(3) + P_{\mathcal{H}}(6) = P_{\mathcal{H}}(A \cap \bar{C}) + P_{\mathcal{H}}(C \cap \bar{B})$$

Finally, we are ready to derive a contradiction. Let

$$\begin{aligned} A &= \{h : \lambda_n^{(1)}(h) = 1\} \\ B &= \{h : \lambda_m^{(1)}(h) = 1\} \quad (\text{so } \bar{B} = \{h : \lambda_m^{(2)}(h) = 1\}) \\ C &= \{h : \lambda_l^{(1)}(h) = 1\} \quad (\text{so } \bar{C} = \{h : \lambda_l^{(2)}(h) = 1\}) \end{aligned}$$

Plugging into equation (37), we get

$$P_{\mathcal{H}}(\lambda_n^{(1)} = 1, \lambda_m^{(2)} = 1) \leqslant P_{\mathcal{H}}(\lambda_n^{(1)} = 1, \lambda_l^{(2)} = 1) + P_{\mathcal{H}}(\lambda_l^{(1)} = 1, \lambda_m^{(2)} = 1)$$

Now, using equation (36), we have

$$\sin^2(\theta) \leqslant \sin^2(\frac{\theta}{2}) + \sin^2(\frac{\theta}{2}) = 2\sin^2(\frac{\theta}{2})$$

for all θ. But in fact

$$\sin^2(\theta) > 2\sin^2(\frac{\theta}{2}) \quad \forall \theta \neq 0, |\theta| < \frac{\pi}{2}$$

which is the contradiction that proves Bell's theorem.

Interpretation. Any hidden-variable theory would have to be *contextual*—the hidden variable, h, would have to include information about the circumstances (i.e. the context) of the measurement. No *a priori* fixed state, as in a fixed $h \in \mathcal{H}$, could determine the output of an arbitrary measurement.

In contrast, a contextual variable could include, for example, a specification of the measurement that's being made. In this case, the probability rules would be subject to the particular measurement (e.g. including the angle θ) and there would no longer be a contradiction.

For this reason, Bell's argument is called a *local* no-go theorem. It rules out a hidden variable that is intrinsic to the particles, but it does not rule out a context-dependent variable.

4.3 Teleportation and the no-cloning theorem

Sheldon: Here's the problem with teleportation.
Leonard: Lay it on me.
Sheldon: Assuming a device could be invented which would identify the quantum state of matter of an individual in one location, and transmit that pattern to a distant location for reassembly, you would not have actually transported the individual. You would have destroyed him in one location, and recreated him in another.
Leonard: How about that.
Sheldon: Personally, I would never use a transporter, because the original Sheldon would have to be disintegrated in order to create a new Sheldon.
Leonard: Would the new Sheldon be in any way an improvement on the old Sheldon?
Sheldon: No, he would be exactly the same.
Leonard: That is a problem.
Sheldon: So you see it too.

The Big Bang Theory
Season 1 Episode 12

Alice and Bob are in the teleportation business. The typical customer, Charlie, is a single, spin-1/2 particle $|C\rangle = \alpha\,|u\rangle + \beta\,|d\rangle \in \mathbb{C}^2_1$. Charlie and Alice are on Earth, and Bob is on Pluto, which happens to be a popular destination. We will work through the necessary preparations (which in the case of Pluto needs to be done by Alice and Bob well before they have any customers) and then we will explain the mechanics of transporting Charlie. Remarkably, the transportation takes no longer than the time it takes for Alice to radio Bob with a simple two-bit instruction set. (Atoms, molecules, and people require larger instruction sets.)

Preparation. For the purpose of teleportation, a particularly convenient basis for $\mathbb{C}^2 \otimes \mathbb{C}^2$ is the Bell basis:

$$
\begin{aligned}
|\Phi_0\rangle &= \frac{1}{\sqrt{2}}(|uu\rangle + |dd\rangle) \\
|\Phi_1\rangle &= \frac{1}{\sqrt{2}}(|ud\rangle + |du\rangle) \\
|\Phi_2\rangle &= \frac{i}{\sqrt{2}}(|ud\rangle - |du\rangle) \\
|\Phi_3\rangle &= \frac{1}{\sqrt{2}}(|uu\rangle - |dd\rangle)
\end{aligned}
$$

You can check that $\langle \Phi_i | \Phi_j \rangle = \delta_{ij}$, $\forall i, j \in \{0, 1, 2, 3\}$, and since these four vectors define an orthonormal set in the four-dimensional state space $\mathbb{C}^2 \otimes \mathbb{C}^2$, they constitute a basis for $\mathbb{C}^2 \otimes \mathbb{C}^2$.

The "preparation" consists of Alice and Bob creating an entangled pair of particles (call them A and B) in the state $|\Phi_0\rangle$. (Actually, Alice and Bob will likely prepare many copies of particles in the state $|\Phi_0\rangle$, since one pair will be needed for every customer.) Alice keeps A and Bob takes B to Pluto. Now they're ready for their first customer, Charlie. Charlie, being a particle, will be referred to, variously, as Charlie or simply C. Since A and B were prepared independently of Charlie, the state of the *three* particles, A, B, and C, is the tensor product, call it $|\Psi\rangle$, of particles A and B (in $\mathbb{C}^2 \otimes \mathbb{C}^2$) with $|C\rangle \in \mathbb{C}^2$:

$$
|\Psi\rangle_{ABC} = |\Phi_0\rangle_{AB} \otimes |C\rangle_C = \frac{1}{\sqrt{2}} \left(|uu\rangle_{AB} + |dd\rangle_{AB}\right) \otimes \left(\alpha |u\rangle_C + \beta |d\rangle_C\right) \in \mathbb{C} \otimes \mathbb{C} \otimes \mathbb{C}
$$

As we shall see, the subscripts (A, B, and C) are a good way to keep track of the different particles. Keep in mind that A and C are on Earth accompanied by Alice, and B is on Pluto accompanied by Bob.

Now let's do some algebra, with the goal of re-factoring $|\Psi\rangle_{ABC}$ into terms that depend only on A and C, and those that depend only on B:

$$
\begin{aligned}
|\Psi\rangle_{ABC} &= \frac{1}{\sqrt{2}} \left(|uu\rangle_{AB} + |dd\rangle_{AB}\right) \otimes \left(\alpha |u\rangle_C + \beta |d\rangle_C\right) \\
&= \frac{\alpha}{\sqrt{2}} |uuu\rangle_{ABC} + \frac{\beta}{\sqrt{2}} |uud\rangle_{ABC} + \frac{\alpha}{\sqrt{2}} |ddu\rangle_{ABC} + \frac{\beta}{\sqrt{2}} |ddd\rangle_{ABC} \\
&= \frac{\alpha}{\sqrt{2}} |uuu\rangle_{ACB} + \frac{\beta}{\sqrt{2}} |udu\rangle_{ACB} + \frac{\alpha}{\sqrt{2}} |dud\rangle_{ACB} + \frac{\beta}{\sqrt{2}} |ddd\rangle_{ACB} \\
&= \frac{\alpha}{\sqrt{2}} |uu\rangle_{AC} \otimes |u\rangle_B + \frac{\beta}{\sqrt{2}} |ud\rangle_{AC} \otimes |u\rangle_B + \frac{\alpha}{\sqrt{2}} |du\rangle_{AC} \otimes |d\rangle_B + \frac{\beta}{\sqrt{2}} |dd\rangle_{AC} \otimes |d\rangle_B \\
&= \frac{\alpha}{2} \left(|\Phi_0\rangle_{AC} + |\Phi_3\rangle_{AC}\right) \otimes |u\rangle_B + \frac{\beta}{2} \left(|\Phi_1\rangle_{AC} - i |\Phi_2\rangle_{AC}\right) \otimes |u\rangle_B \qquad (38) \\
&+ \frac{\alpha}{2} \left(|\Phi_1\rangle_{AC} + i |\Phi_2\rangle_{AC}\right) \otimes |d\rangle_B + \frac{\beta}{2} \left(|\Phi_0\rangle_{AC} - |\Phi_3\rangle_{AC}\right) \otimes |d\rangle_B \qquad (39)
\end{aligned}
$$

So far, so good. Nothing difficult and nothing surprising.

Finally, we do one more re-factoring, so that each of the four Bell bases from lines (38) and (39) appear only once:

$$\begin{aligned}|\psi\rangle_{ACB} = &\frac{1}{2}|\Phi_0\rangle_{AC} \otimes (\alpha|u\rangle_B + \beta|d\rangle_B) \\ &+ \frac{1}{2}|\Phi_1\rangle_{AC} \otimes (\alpha|d\rangle_B + \beta|u\rangle_B) \\ &+ \frac{i}{2}|\Phi_2\rangle_{AC} \otimes (\alpha|d\rangle_B - \beta|u\rangle_B) \\ &+ \frac{1}{2}|\Phi_3\rangle_{AC} \otimes (\alpha|u\rangle_B - \beta|d\rangle_B)\end{aligned} \tag{40}$$

Bob's particle looks radically different. In fact, you might surmise from (40) that the marginal (mixed-state) for particle 2 is exactly Charlie ($\alpha|u\rangle + \beta|d\rangle$) 25% of time! And you would be right—as you will establish in the homework. But how can this be? Charlie walked in the door, and since then nobody has done anything. And in any case Bob took his particle to Pluto, a long time ago.

Charlie's demise. Evidently (and despite appearances) Charlie is still independent of Alice's and Bob's particles, and Alice's and Bob's are still in the entangled state $|\Phi_0\rangle$. But the system is prepared, and we are ready to teleport Charlie. Naturally, we will first need to collect some information about Charlie. But keep in mind the adage of Susskind and Friedman that in quantum mechanics, experiments are never gentle. In particular, any measurement that involves Charlie will almost certainly change Charlie. In this case, the measurement will involve both Charlie and Alice's particle, resulting in a definite and entangled state for the pair. Since there is no meaningful way to talk about the wave state of a member of an entangled pair (other than as a mixed state), Charlie will cease to exist.

The purpose of the measurement can be gleaned from (40): force A and C into one of the four orthogonal states in the Bell basis, which we can then use to describe the Earth-bound pair and thereby narrow the state of B down to one of four, each of which is already looking suspiciously like Charlie. So we are looking for a measurement that will leave the pair A and C in one of the four states defined by the Bell basis. But *any* measurement described by a Hermitian operator that has those four states as eigenvectors will suffice:

$$\mathcal{O} = \sum_{k=0}^{3} \lambda_k |\Phi_k\rangle\langle\Phi_k|$$

where for each k, $\lambda_k \geqslant 0$. There are plenty of ways to do this, each of which exploits the spin properties of the four Bell states—think of a Stern-Gerlach type machine that re-directs the pair into one of four directions.

Of course Bob and his particle are not available, so the measurement will have to be of the form $\mathcal{O} \otimes I$, involving no measurement of B. There are the four possible outcomes:

1. $|\Psi\rangle \to |\Phi_0\rangle_{AC} \otimes (\alpha|u\rangle_B + \beta|d\rangle_B)$ with probability equal to the squared overlap:

$$\left|\langle\Psi|_{ACB} \left|\left(|\Phi_0\rangle_{AC} \otimes (\alpha|u\rangle_B + \beta|d\rangle_B)\right)\right\rangle\right|^2 = \frac{1}{4}$$

which comes from replacing $|\Psi\rangle_{ACB}$ by the expression on the right-hand side of equation (40), which is a sum of four terms, three of which start with states for the A/C pair that are orthogonal to $|\Phi_0\rangle_{AC}$.

And the identical analysis applies to the remaining three of the possible outcomes:

2. $|\Psi\rangle \rightarrow |\Phi_1\rangle_{AC} \otimes (\alpha\,|d\rangle_B + \beta\,|u\rangle_B)$ with probability

$$\left|\langle\Psi|_{ACB}\,\Big|\big(\,|\Phi_1\rangle_{AC} \otimes (\alpha\,|d\rangle_B + \beta\,|u\rangle_B)\,\big)\Big\rangle\right|^2 = \frac{1}{4}$$

3. $|\Psi\rangle \rightarrow i\,|\Phi_2\rangle_{AC} \otimes (\alpha\,|d\rangle_B - \beta\,|u\rangle_B)$ with probability

$$\left|\langle\Psi|_{ACB}\,\Big|\big(i\,|\Phi_2\rangle_{AC} \otimes (\alpha\,|d\rangle_B - \beta\,|u\rangle_B)\,\big)\Big\rangle\right|^2 = \frac{1}{4}$$

4. $|\Psi\rangle \rightarrow |\Phi_3\rangle_{AC} \otimes (\alpha\,|u\rangle_B - \beta\,|d\rangle_B)$ with probability

$$\left|\langle\Psi|_{ACB}\,\Big|\big(\,|\Phi_3\rangle_{AC} \otimes (\alpha\,|u\rangle_B - \beta\,|d\rangle_B)\,\big)\Big\rangle\right|^2 = \frac{1}{4}$$

Alice has only to record the outcome, which can be summarized with two bits of information, 00, 01, 10, or 11, respectively, and then radio the result to Bob.

Charlie's resurrection. Alice's measurement leaves the three-particle system in one of four definite states:

$$\begin{aligned}
|\Psi\rangle_{ACB} &= |\Phi_0\rangle_{AC} \otimes (\alpha\,|u\rangle_B + \beta\,|d\rangle_B) && \leftrightarrow \text{Alice sends 00}\\
|\Psi\rangle_{ACB} &= |\Phi_1\rangle_{AC} \otimes (\alpha\,|d\rangle_B + \beta\,|u\rangle_B) && \leftrightarrow \text{Alice sends 01}\\
|\Psi\rangle_{ACB} &= |\Phi_2\rangle_{AC} \otimes (\alpha i\,|d\rangle_B - \beta i\,|u\rangle_B) && \leftrightarrow \text{Alice sends 10}\\
|\Psi\rangle_{ACB} &= |\Phi_3\rangle_{AC} \otimes (\alpha\,|u\rangle_B - \beta\,|d\rangle_B) && \leftrightarrow \text{Alice sends 11}
\end{aligned}$$

each of which is of the form $|\Theta\rangle_{AC} \otimes |\Lambda\rangle_B$ where $|\Theta\rangle_{AC}$ is an entangled state of A and C, and $|\Lambda\rangle$ is a state of B. What's more, by the rules for independence, B has been rendered *independent* of A and C.

Four-and-a-half hours later, Bob receives the two-bit message. If the message is 00, there there is nothing to be done; Bob's particle B is already Charlie! Otherwise, Bob needs to subject his particle to an appropriate "rotation," i.e. act on the particle with the right unitary operator. Recall that the Pauli matrices, σ_x, σ_y, and σ_z are unitary, with the properties:

$$\begin{aligned}
\sigma_x\,|u\rangle &= |d\rangle \quad \sigma_x\,|d\rangle = |u\rangle\\
\sigma_y\,|u\rangle &= i\,|d\rangle \quad \sigma_y\,|d\rangle = -i\,|u\rangle\\
\sigma_z\,|u\rangle &= |u\rangle \quad \sigma_z\,|d\rangle = -\,|d\rangle
\end{aligned}$$

It is now an easy matter to check that the action of σ_n, where $n = x$ if Bob received 01, $n = y$ if Bob received 10, and $n = z$ if Bob received 11, will in all cases bring B to the state $\alpha |u\rangle + \beta |d\rangle$. And just like that, Charlie is resurrected on Pluto.[14]

Why not resurrect Charlie on Mars *and* Pluto? It is tempting to try to scale up the operation by bringing on Dave. The idea would be to prepare an entangled triplet, A, B, and D, and send Dave with D to Mars, in addition to sending Bob with B to Pluto. Alice would make a measurement of A and C on earth, and broadcast a suitable code to both Mars and Pluto. Alice's code would instruct Dave to apply a particular unitary operation to D and Bob to apply a potentially different unitary operation on B, both resulting in a resurrected Charlie.

This is called "broadcasting" and it doesn't work (the "no-broadcasting" theorem). The no-broadcasting theorem is one of several results with similar conclusions: you can't make a copy without destroying the original (the "no-cloning" theorem), and in any case you can never make more than one copy (no-broadcasting). As for the proofs, the basic idea is already revealed in the no-cloning theorem, which we will present here and use shortly to assure that quantum encryption is indeed secure.

Imagine that Alice opens a new shop, this time for customers looking to be cloned. She has discovered a very special unitary operator U (and accompanying Hamiltonian) that can transform a particular state $|R\rangle$ (the raw material) into a clone of an arbitrary customer, say Charlie with state $|C\rangle$, without changing the original. Alice's patent demonstrates that for an arbitrary $|C\rangle$ that is independent of the raw material $|R\rangle$, $U |CR\rangle = |CC\rangle$. In words, the operator turns an independent pair of states $|C\rangle$ and $|R\rangle$ into an independent pair of $|C\rangle$ states. One Charlie in, two Charlies out. After a few months, with no customer having actually met their clone, some are becoming a little suspicious. Something is amiss.

Reason, by contradiction, as follows: Suppose that two customers, $|C\rangle$ and $|C'\rangle$, are independent, in which case their collective state is $|CC'\rangle$, and let $|R\rangle$ and $|R'\rangle$ be two independent (but identical) copies of the raw material. If the cloning machine really worked (i.e. if $U |CR\rangle = |CC\rangle$ for arbitrary $|C\rangle$), then

$$\begin{aligned}
\langle C|C'\rangle &= \langle C|C'\rangle\langle R|R'\rangle && \text{(the states are normalized and } R \text{ and } R' \text{ have identical states)} \\
&= \langle CR|C'R'\rangle && \text{(one of the independence rules)} \\
&= \langle CR|\, U^\dagger U\, |C'R'\rangle && (U \text{ unitary} \Rightarrow U^\dagger U = I) \\
&= \langle CC|C'C'\rangle && \text{(since } U|CR\rangle = |CC\rangle \text{ and } U|C'R'\rangle = |C'C'\rangle) \\
&= \langle C|C'\rangle\langle C|C'\rangle && \text{(independence rule, again)} \\
&= \langle C|C'\rangle^2
\end{aligned}$$

from which we conclude that either $|\langle C|C'\rangle| = 0$ or $|\langle C|C'\rangle| = 1$, both of which contradict independence.

[14] Each of the operations can be executed by exposing the particle to an appropriate Hamiltonian for an appropriate amount of time—as you might recall from a previous exercise about connecting a unitary operation to an instance of the Schrodinger equation.

4.4 Quantum encryption

Alice wants to send a message to Bob, securely. The message is binary, made up of M bits, $(m_1, m_1, \ldots, m_M) \in \{0,1\}^M$. Any ordinary encoding has its vulnerabilities, usually revolving around whether or not the encryption or decryption formula can be inferred in a computationally feasible manner from examples. Quantum encryption is more elaborate and difficult, but the issue of decrypting is already settled; it's impossible. At best an adversary (henceforth "Charlie") could make a guess, but the probability of success can be made arbitrarily low.

The algorithm that we will present is one of many variations on the same theme: if Alice creates a large set of maximally entangled particles (which is to say that they are guaranteed to have opposite spins), and can manage to deliver the second member of each pair to Bob, then the spins of Bob's particles contain a uniquely private decryption key for whatever message Alice chooses to send.

The procedure, from preparation to decryption, entails a choreographed sequence of actions and transmissions by Alice and Bob:

(i) **(preparation)** Generically, we will designate an entangled pair of particles A and B,[15] and their corresponding states using $|A\rangle$ and $|B\rangle$. Alice begins by preparing a large number of pairs $(A_1, B_1), \ldots, (A_n, B_n)$ independently from the singlet state, $|S\rangle \doteq \frac{1}{\sqrt{2}}(|ud\rangle - |du\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2$. Depending on the number of pairs, only a small fraction might need to be used for a given transmission. (In principle, the rest could be put aside for ensuing messages.) Instead of using $\lambda^{(A)} \in \{-1, 1\}$ and $\lambda^{(B)} \in \{-1, 1\}$ for the respective outcomes of a pair of measurements, we will use the range $\{0, 1\}$, which is generally more convenient for quantum encryption and quantum computing. Specifically, given a spin-1/2 particle and a direction of measurement $d \in \mathbb{R}^3$, define

$$\gamma = \begin{cases} 0 & \text{if the particle is observed in the positive } d \text{ direction} \\ 1 & \text{if it is observed in the negative } d \text{ direction} \end{cases}$$

Recall that if the pair A and B are in the singlet state, then for *any* d,

$$\mathbb{P}\left(\lambda^{(A)}_{\sigma_d} = \lambda^{(B)}_{\sigma_d}\right) = 1$$

(The singlet state is spherically symmetric.) It follows that

$$\mathbb{P}\left(\gamma^{(A)}_{\sigma_d} + \gamma^{(B)}_{\sigma_d} = 1\right) = 1$$

The preparation is completed once Alice has shipped $B_1, \ldots B_n$ to Bob. Of course the shipment is not tamper proof, and there can be no guarantee that Bob actually receives what Alice sent. But there is the no-cloning theorem, which guarantees that for each k neither Charlie nor anyone else can intercept B_k, make a measurement or make a copy, and then pass an unaltered version to Bob.

[15] The typical choice uses entangled pairs of photons.

(ii) **(choose and broadcast measurement directions)** Alice and Bob make no secret of their intention to use the entangled particles to transmit a message. In the end, Alice and Bob will make measurement of each pair of particles, in some specified direction. Alice will use her observation for encoding and Bob will use his observation for decoding. The guarantees for their approach are based on the assumption that

$$\mathbb{P}\left(\gamma_k^{(A)} + \gamma_k^{(B)} = 1\right) = 1 \qquad \text{for all } k = 1, \ldots, n$$

This means that for each pair, (A_k, B_k), they will have to use the same direction, say $d_k \in \mathbb{R}_1^3$.

But this is a potential vulnerability. If Charlie happens to know the direction, then he can intercept the shipment, measure B_k, rendering it in the definite state $|d^+\rangle$ or $|d^-\rangle$, and then pass on a new particle prepared in that state. Neither Alice nor Bob would be any the wiser.

Oddly, an excellent way for Alice and Bob to protect against this vulnerability is to publicly *broadcast* their intentions! Alice and Bob agree on two orthogonal directions in $\mathbb{R}_1^3$, say p and q with $p \cdot q = 0$, from which each will choose, independently, a measurement direction d_k for each $k = 1, 2, \ldots, n$: let

$$d_1^{A)}, \ldots, d_n^{(A)} \sim \; iid \; \{p, q\} \; 50/50 \qquad \text{and} \qquad d_1^{(B)}, \ldots, d_n^{(B)} \sim \; iid \; \{p, q\} \; 50/50$$

At this point Alice and Bob make public ("broadcast") their respective directions. Keep in mind that Bob has already received his shipment of particles. The selections and broadcasting of directions are done later, when it's too late for Charlie or anyone else to make use of the information.

The remaining steps ("security check," encoding, decoding) will make use of the following subsets of the set of n integers, $\{1, 2, \ldots, n\}$:

$\mathcal{I} = \left\{k \middle| d_k^{(A)} = d_k^{(B)}\right\}$ (the useful set, on which Alice and Bob's measurement directions agree)

$\mathcal{S} \subseteq \mathcal{I}$ (an arbitrary subset of $\mathcal{I}$ such that $|\mathcal{S}| \approx \dfrac{|\mathcal{I}|}{2}$)

$\mathcal{M} = \mathcal{I} \backslash \mathcal{S}$ (i.e. $\mathcal{M} = \mathcal{I} \bigcap \bar{\mathcal{S}}$)

(iii) **(security check)** So far, Alice and Bob have no idea about whether or not Charlie fiddled with the shipment. In fact, he might have simply guessed the directions, measured the B particles accordingly, and then substituted particles in the resulting definite states for the B particles. For this reason, before the encoding and transmission of the message Alice and Bob perform a security check by making use of the particular particles indexed by $\mathcal{S}$:

(a) For each $k \in \mathcal{S}$, Alice measures A_k in the direction $d_k^{(A)}$, and broadcasts the result $\gamma_k^{(A)}$.

(b) Bob makes the same measurements, but on the particles B_k, $k \in \mathcal{S}$, and broadcasts $\gamma_k^{(B)}$.

At this point, Alice and Bob (and everyone else) know whether or not every pair in $\mathcal{S}$ passed the security check. If they didn't, then Alice and Bob declare a breach and start over, presumably taking better care of the new shipment.

But maybe Charlie got lucky. Let's say that he intercepted particles, guessed the directions, and substituted definite states. What is the probability that the particle he sent to B nevertheless passes

the security check? For each $k \in \mathcal{S}$, let D_k be the direction in which Charlie measured B_k, and let $\Gamma_k^{(B)} \in \{0,1\}$ be the result of the measurement. Then for each $k \in \mathcal{S}$

$$\begin{aligned}\mathbb{P}\left(\text{particle } k \text{ passes security check}\right) &= \mathbb{P}\left(\gamma_k^{(A)} + \Gamma_k^{(B)} = 1\right) \\ &= \mathbb{P}\left(\gamma_k^{(A)} + \Gamma_k^{(B)} = 1 \middle| D_k^{(B)} = d_k\right) \mathbb{P}\left(D_k^{(B)} = d_k\right) \\ &+ \mathbb{P}\left(\gamma_k^{(A)} + \Gamma_k^{(B)} = 1 \middle| D_k^{(B)} \neq d_k\right) \mathbb{P}\left(D_k^{(B)} \neq d_k\right) \\ &= 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}\end{aligned}$$

But Charlie needs to be lucky for every one of the $|\mathcal{S}|$ pairs. The probability of that happening is $\left(\frac{3}{4}\right)^{|\mathcal{S}|}$, which can be made negligible with, say, 500 pairs set aside for a security check.

(iv) **(encryption)** We require that $|\mathcal{M}| \geqslant M$, since there needs to be enough pairs left over for the message. For convenience, let us assume that in fact $\mathcal{M} = M$.

If she hasn't already done so, Alice measures the spin $\gamma_k^{(A)}$ of each message particle A_k, $k \in \mathcal{M}$. The intended message, $(m_1, m_1, \ldots, m_M) \in \{0,1\}^M$, is then encrypted using the "exclusive or" function: for every $k \in \mathcal{M}$

$$e_k = m_k \oplus \gamma_k^{(A)}$$

where $\oplus$ represents addition, modulo 2. In other words, e_k is one if exactly one of m_k or $\gamma_k^{(A)}$ is one, and zero otherwise. Although Bob is the only intended reciepient of the message, there is no point in bothering with a secure channel. Alice simply broadcasts the encrypted bits, $e_1, e_2, \ldots, e_M$.

(v) **(decryption)**

By now Bob is supposed to have measured each of the particles $B_1, \ldots, B_M$ in the appointed directions $d_1^{(B)}, \ldots, d_M^{(B)}$. Entanglement ensures that the outcomes, $\gamma_1^{(B)}, \ldots, \gamma(B)_M$ are the opposite of the outcomes observed by Alice: $\gamma_k^{(A)} = 1 - \gamma_k^{(B)}$, for every $k = 1, \ldots, M$. Decryption is easy: for every k,

$$\gamma_k^{(A)} \oplus \gamma_k^{(A)} = 0 \Rightarrow m_k = m_k \oplus \gamma_k^{(A)} \oplus \gamma_k^{(A)} = e_k \oplus \gamma_k^{(A)} = e_k \oplus (1 - \gamma_k^{(B)})$$

Remarks.

(1) After the delivery of $B_1, \ldots, B_n$ to Bob, every action is taken locally and every communication is broadcast. Bob's particles amount to a secret key with which Bob unlocks the encrypted message. The delivery of the key goes by the acronym QKD—Quantum Key Distribution. By far the easiest way to conceive the delivery system is to use photons, passed through air or fiber optic cables. But alternatives are also being explored.

(2) Feasibility has been demonstrated using photons and the Micius satellite, launched by China in 2016, to relay messages between ground stations separated by up to 2,500 kilometers.

(3) There are many variations, involving for example more directions, more or fewer security checks, and other types of particles.

4.5 Free will and the Conway-Kochen Theorem

Why are you reading this sentence right now? Philosophers present two theories for this matter: some say that it was predetermined, while others think that you chose to do so.

John Conway and Simon Kochen believe in free will. The Conway-Kochen Free Will Theorem follows a thought experiment with two entangled particles. Each of these particles is a spin-one particle with three possible outcomes (which we will label $+1$, 0 and -1) for a spin measurement in any direction. (In contrast, we have mostly been discussing spin-1/2 particles, though we changed units so that we could identify their two possible outcomes of a measurement with ± 1.)

The Free Will Theorem relies on three axioms: SPIN, TWIN, and MIN, all of which come from the physical theory of quantum mechanics and relativity. We begin by examining these axioms. We will first outline the representation of the wave functions (states) of the kind of spin-one particles that we will be working with (concrete examples include para and ortho helium). We will then discover that these particles offer a second, and in some ways stronger, version of the no-hidden variable theorem by Bell. After that, we will consider a pair of such particles, using the tensor-product formulation, and we will define a singlet state that is analogous to the one we encountered earlier in $\mathbb{C}^2 \otimes \mathbb{C}^2$, but this time in $\mathbb{C}^3 \otimes \mathbb{C}^3$. Finally, we will use the singlet state to build an even stronger version of the no-hidden variable theorem and, at the same time, prove the Free Will Theorem.

Many of the details will be addressed in the homework.

Before we begin, we would like to present an analogy between experimenting on particles and playing the game *20 questions* with your friends. When the experimenters are performing measurements on particles, it can be viewed as your friends asking you questions pertaining to the object of choice in mind. However, in the game, you can also choose to cheat by not having a predetermined object in mind. You might first answer some questions, and then come up with an object that satisfies your answers. In some sense, the results we will present in this section, namely the Kochen-Specker Paradox and the Free Will Theorem, says that a spin-one particle behaves exactly like a cheating player since there is no "object" (spins) that satisfies "all questions" (measurement of squared spins in a well-chosen set of 33 directions). In other words, the particle does not have a pre-determined spin in every direction.

A single spin-one particle with three definite states. At the quantum level angular momentum only exists in discrete values, much like the magnetic spin of electrons and other particles. Photons, for example, have an angular momentum that can be either $\hbar$ or $-\hbar$, which we usually write as ± 1. Still other, more elaborate, particles have any of three levels of angular momentum: $\hbar$, 0, or $-\hbar$. The Conway-Kochen "Free Will" theorem is about entangled pairs of these kinds of particles, for which we will again make a convenient change of scale and write ± 1 and 0 in place of $\pm\hbar$ and 0, and for which we will continue to use the term "spin state."

The formulation of these and other quantum-level discrete-state systems mimics our development of the

state space and measurement operations for spin-1/2 particles, except that we require three complex-valued components instead of just two, corresponding to the three possible outcomes of a measurement of angular momentum in any chosen orientation. Thus the wave functions will lie in $\mathbb{C}^3_1$ for individual particles, and $\mathbb{C}^3_1 \otimes \mathbb{C}^3_1$ for pairs of particles (independent or entangled).

The analog of using the up and down states, $|u\rangle$ and $|d\rangle$, as a basis for $\mathbb{C}^2$, is to use $|u\rangle$, $|n\rangle$, and $|d\rangle$ as a basis for $\mathbb{C}^3$:

$$|u\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad |n\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad |d\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

where $|u\rangle$, $|n\rangle$, and $|d\rangle$ refer, respectively, to the spin states "up", "neutral", and "down" in the $z \in \mathbb{R}^3_1$ direction. In general, the wave function is a three-component (instead of two-component) normalized complex vector, corresponding to a quantized magnetic moment or angular momentum, in suitable units, of $+1$, 0, or -1.

Notice that the ket vectors $|u\rangle$, $|n\rangle$, and $|d\rangle$ define an orthonormal basis for $\mathbb{C}^3$. Following the half-spin example, we can write the operator representing a measurement of spin in the z direction as

$$\sigma_z = \lambda_u |u\rangle\langle u| + \lambda_n |n\rangle\langle n| + \lambda_d |d\rangle\langle d|$$

with $\lambda_u = 1$ corresponding to spin along the z axis and the definite state $|u\rangle$, and, similarly, $\lambda_n = 0$ and $\lambda_d = -1$ for the definite states $|n\rangle$ (zero spin) and $|d\rangle$ (spin in the $-z$ direction).[16] As always, the observation operator is Hermitian, the eigenvalues (λ_u, λ_n, and λ_d) are the (real-valued) possible measurement outcomes, and the corresponding eigenvectors ($|u\rangle$, $|n\rangle$, and $|d\rangle$) are the possible (definite) states that could result from the measurement. If the initial state is $|\psi\rangle \in \mathbb{C}^3_1$, then the squared overlaps ($|\langle u|\psi\rangle|^2$, $|\langle n|\psi\rangle|^2$, $|\langle d|\psi\rangle|^2$) are the respective probabilities of the different outcomes.

With the appropriate substitutions, we get

$$\begin{aligned} \sigma_z &= (1) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} + (0) \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \end{aligned}$$

If we were to follow the development of representations for half-spin systems, we would find a mapping from $\mathbb{R}^3_1$ into $\mathbb{C}^3_1$ that would again be unique up to an overall phase. We could then use different bases to describe the wave function, and to get operators that measure spin in different directions. Since there is nothing really new that would need to be done, we will forgo most of the details and jump to the various consequences that will be needed for the remainder of the discussion.

[16] The term "definite" is relative. It means definite with respect to a particular measurement, as in $|u\rangle$ is definite for σ_z, since the outcome is deterministic. But $|u\rangle$ is certainly not definite for σ_x. A measurement always leaves a particle in a definite state (an eigenstate, in particular) *with respect to* that measurement.

Were we to find the two vectors in $\mathbb{C}^3_1$ that correspond to the two additional cardinal directions, $x, y \in \mathbb{R}^3_1$, then we could construct the corresponding measurement operators σ_x and σ_y, just as we did, above, for σ_z. These turn out to be

$$\sigma_x = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & \frac{-i}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & 0 & \frac{-i}{\sqrt{2}} \\ 0 & \frac{i}{\sqrt{2}} & 0 \end{pmatrix}$$

From these we can compute σ_m, for arbitrary $m = (m_x, m_y, m_z) \in \mathbb{R}^3$:

$$\sigma_m = m_x\sigma_x + m_y\sigma_y + m_z\sigma_z \tag{41}$$

(Which we already knew for the spin-1/2 operators—HW3, problem 2.)

Our interest here is exclusively in measuring squared spins, rather than the spins themselves, e.g.

$$\sigma_z^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(We don't conjugate anything; we just square the matrix. After all, this is what corresponds to squaring the *real-valued* eigenvalues. We've seen this before, in the set up and proof of the uncertainty principle—§2.4.) Similarly,

$$\sigma_x^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \qquad \sigma_y^2 = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

Now something special happens. In the homework, you will show that the three squared-spin matrices commute:

$$[\sigma_x^2, \sigma_y^2] = [\sigma_x^2, \sigma_z^2] = [\sigma_y^2, \sigma_z^2] = 0$$

In other words, they all commute with each other. Since all three of these operators are Hermitian, we can invoke the theorem, proven earlier (in §2.4), *Commutation of Finite-dimensional Hermitian Matrices*. The consequence consequence is that these three matrices, σ_x^2, σ_y^2 and σ_z^2, can be "simultaneously diagonalized," which is a fancy way of saying that there exists a set of three orthonormal vectors, say $|e_1\rangle$, $|e_2\rangle$, and $|e_3\rangle$, each of which is an eigenvector for each of the three matrices.

Given this information, it is straightforward to find a set that does the trick:

$$|e_1\rangle \doteq \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad |e_2\rangle \doteq \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad |e_3\rangle \doteq \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \tag{42}$$

And then, even easier to show that the associated eigenvalues can be summarized by the following table:

	$\|e_1\rangle$	$\|e_2\rangle$	$\|e_3\rangle$
σ_x^2	1	1	0
σ_y^2	1	0	1
σ_z^2	0	1	1

Notice that each eigenvector has exactly two eigenvalues equal to one and one eigenvalue equal to zero, among the three matrices.

Let us pause to highlight some of the important consequences:

1. We can represent the three measurements using their common set of eigenvectors and the table of eigenvalues:

$$\sigma_x^2 = |e_1\rangle\langle e_1| + |e_2\rangle\langle e_2|$$
$$\sigma_y^2 = |e_1\rangle\langle e_1| + |e_3\rangle\langle e_3|$$
$$\sigma_z^2 = |e_2\rangle\langle e_2| + |e_3\rangle\langle e_3|$$

2. Let λ_x^2, λ_y^2, and λ_z^2 be the observations resulting from the measurements σ_x^2, σ_y^2, and σ_z^2, respectively, and imagine that we have prepared a particle in the state $|\psi\rangle \in \mathbb{C}_1^3$. If we were to make the measurements σ_x^2, σ_y^2 and σ_z^2, sequentially in that order, then with probability $|\langle e_1|\psi\rangle|^2$ the definite state after the first measurement would be $|e_1\rangle$, which is also definite for the remaining two measurements, σ_y^2 and σ_z^2. Since $|e_1\rangle$ is also an eigenvector of σ_y^2 and σ_z^2, the final state will remain $|e_1\rangle$, and the sequence of observations can be read off from the first column in the table: (1,1,0) for (λ_x^2, λ_y^2, λ_z^2). In fact, if any of the three measurements were performed first, then the probability of $|\psi\rangle \to |e_1\rangle$ would still be $|\langle e_1|\psi\rangle|^2$, and the outcomes, $\lambda_x^2 = 1$, $\lambda_y^2 = 1$, and $\lambda_z^2 = 0$ would also be unchanged, regardless of the order.

   The same reasoning applied to the definite states $|e_2\rangle$ and $|e_3\rangle$, leads to the conclusion that the three measurements σ_x^2, σ_y^2, σ_z^2, taken in any order, will result in the definite state $|e_2\rangle$ and the observations $\lambda_x^2 = 1$, $\lambda_y^2 = 0$, and $\lambda_z^2 = 1$ (second column of the table), with probability $|\langle e_2|\psi\rangle|^2$, and the definite state $|e_3\rangle$ and observations $\lambda_x^2 = 0$, $\lambda_y^2 = 1$, and $\lambda_z^2 = 1$ (third column) with probability $|\langle e_3|\psi\rangle|^2$.

3. For comparison, you might want to consider what happens if we measure, instead, λ_x, λ_y and λ_z in one order versus another order. You will find that the result very much depends on the order of the measurements. That is because σ_x, σ_y, and σ_z do not commute.

4. Finally, as already noted during our development of $\mathbb{C}^2$ (§2), since the entire set up traces back to inner products rather than any particular orientation, these same observations would apply had we started with *any* set of three orthogonal vectors in $\mathbb{R}_1^3$, not just the cardinal directions x, y, and z.

We learned in the last paragraphs that particles with spin states -1, 0, and +1 have some special properties. Most useful, for our purposes, is the property that the measurement of the squared spins, in three arbitrary orthogonal directions, can be made, independently of order and will always produce two 1's and a zero. This is what Conway and Kochen call "the SPIN axiom".

> **The SPIN Axiom.** *Measurements of the squared (components of) spin of a spin 1 particle in three orthogonal directions always give the answers* $1, 0, 1$ *in some order.*

A second "no-go" theorem. Given the SPIN axiom, there is already strong evidence suggesting that the spin does not exist before the measurement. For if it does exist beforehand, we would have a function defined on the sphere of all directions that takes each orthogonal triple to $1, 0, 1$ in some order. We call any such function a *101 function*. It is not hard to see that any 101 function has the following properties:

(a) Opposite directions give one and the same answer.

(b) Two perpendicular directions cannot both be 0.

(c) Three perpendicular directions cannot all be 1.

These properties suggested to Kochen and Specker an opportunity for a potentially stronger hidden-variable result than Bell's inequality. They reasoned as follows:

1. Let $\mathcal{L}$ be the set of all lines in $\mathbb{R}^3$ that pass through the origin.
2. Suppose that a hidden variable $h \in \mathcal{H}$ determined the squared spin in every direction: in other words, assume that for every line $l \in \mathcal{L}$ there exists a function $\lambda_l^2(h) \in \{0, 1\}$ such that $\lambda_l^2(h)$ is the squared spin when measured in the orientation l.
3. Then the peculiar property that for every three perpendicular orientations we will always measure two squared spins equal to 1, and one squared spin equal to 0, might be the deterministic consequence of the hidden variable h.
4. To make this precise, let $\mathcal{O}$ be the set of all perpendicular triplets in $\mathcal{L}$:

$$\mathcal{O} = \left\{(l_1, l_2, l_3) : l_i \in \mathcal{L}, i = 1, 2, 3, \& \; i \neq j \Rightarrow l_i \perp l_j\right\}$$

and let $\mathcal{M}$ be a mapping from $\mathcal{O}$ into the set $\mathcal{A} \triangleq \{(1, 1, 0), (1, 0, 1), (0, 1, 1)\}$, so $\mathcal{M} : \mathcal{O} \to \mathcal{A}$. We will say that $\mathcal{M}$ is consistent if

$$\mathcal{M}(l_1, l_2, l_3) = (\lambda_{l_1}^2(h), \lambda_{l_2}^2(h), \lambda_{l_3}^2(h)) \;\; \forall (l_1, l_2, l_3) \in \mathcal{O} \;\text{ and some } h \in \mathcal{H}. \tag{43}$$

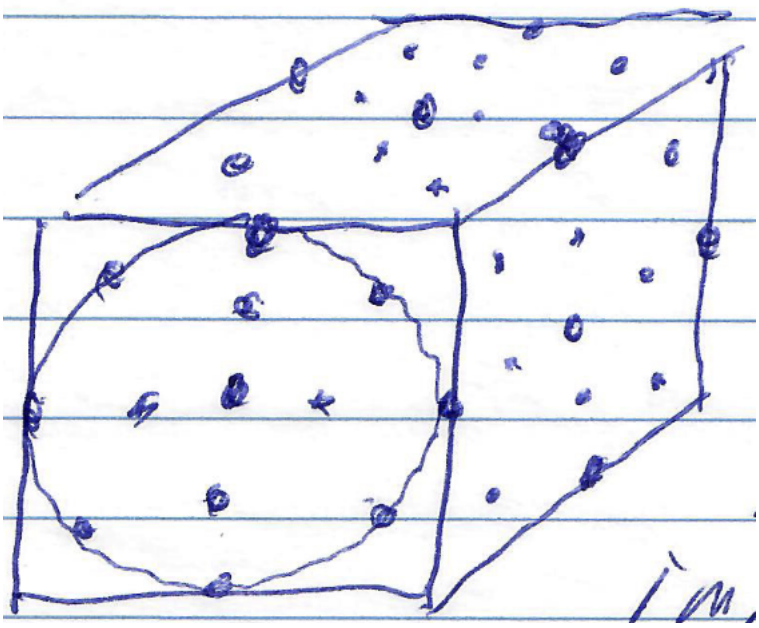


Figure 13: A sphere with 33 lines and 16 orthogonal triplets.

It turns out that there is no consistent $\mathcal{M}$. It will always be possible to find some $l \in \mathcal{L}$ such that l is a member of two different triplets in $\mathcal{O}$ and called upon, by $\mathcal{M}$ to be 0 in one case and 1 in the other.

The proof is by contradiction via a combinatorial argument, using the three properties of 101 functions that we observed above. Given the functions $\lambda_l^2(h)$ and the mapping $\mathcal{M}$ satisfying (43) you construct a finite subset of $\mathcal{O}$ and show that at least one line l is forced to be both 1 and 0. Modern proofs use an a priori subset constructed by Peres (Figure 13), which consists of just 33 (well-chosen) lines (so we can assume that $\mathcal{L} = \{l_1, l_2, \ldots, l_{33}\}$) among which there are only 16 orthogonal triplets (so $\mathcal{O} = \{(l_{o_i^1}, l_{o_i^2}, l_{o_i^3}) : i = 1, \ldots, 16\}$).

This is the Kochen-Specker Theorem, and it is another example of what is called a "no-go" theorem, meaning that it precludes an explanation of a quantum phenomenon that uses hidden variables. It is reasonable to think of it as a stronger result than Bell's Theorem, in the sense that the Kochen-Specker Theorem does not even depend on the probabilities assigned to outcomes.

Theorem. *(The Kochen-Specker Paradox) There does not exist a 101 function for the* 33 *pairs of directions of Figure 13 (the Peres configuration).*

The Kochen-Specker Paradox shows that there does not exist a function defined on the directions that tell us the spins. It is not there unless we measure it. In a sense, the particle tells us the answer "on the fly".

However, this theorem does not rule out the possibility that the particle's answer depends on the day of the week, the conversations between the experimenters, or anything in the history of the universe. The Free Will Theorem strengthens it by asserting that there is no function, not just of a single direction, but also

of everything in the past history of the universe before the measurement, that can possibly determine the spins.

A pair of spin-one three-state particles in the singlet state. Now consider a pair of two particles in the spin-one state. Their wave function is therefore in the tensor-product space $\mathbb{C}^3 \otimes \mathbb{C}^3$. We are interested in a particular maximally-entangled state $|S\rangle \in \mathbb{C}^3 \otimes \mathbb{C}^3$ that is in many ways an analog of the singlet in $\mathbb{C}^2 \otimes \mathbb{C}^2$ that came up earlier. Specifically, we are interested in

$$|S\rangle = \frac{1}{\sqrt{3}} \left(|ud\rangle - |nn\rangle + |du\rangle \right).$$

Recall that $\mathcal{O}$ contains 16 triplets, each made up of three orthogonal lines drawn from $\mathcal{L}$, which is a pool of 33 lines passing through the origin in $\mathbb{R}^3$. Also recall that $\mathcal{A} = \{(1,1,0), (1,0,1), (0,1,1)\}$.

For any $l \in \mathcal{L}$, let σ_l^2 be the single-particle squared-spin operator oriented in the direction of l. To distinguish measurements of particle 1 from those of particle 2, let λ_l^2 be the value measured on particle 1 by σ_l^2, and let γ_l^2 be the value measured on particle 2 by σ_l^2.

The proofs of the following properties are continuations of the calculations we already made on the single spin-one particle, and will be explored in the homework:

1. For any $(l_1, l_2, l_3) \in \mathcal{O}$, the six operators

   $$\sigma_{l_1}^2 \otimes I \quad \sigma_{l_2}^2 \otimes I \quad \sigma_{l_3}^2 \otimes I \quad I \otimes \sigma_{l_1}^2 \quad I \otimes \sigma_{l_2}^2 \quad I \otimes \sigma_{l_3}^2$$

   commute, and therefore can be measured in any order without changing the probability of a given outcome.

2. The results of measuring $\sigma_{l_1}^2 \otimes I$, $\sigma_{l_2}^2 \otimes I$, and $\sigma_{l_3}^2 \otimes I$ always have two 1s and one 0 (i.e. $(\lambda_{l_1}^2, \lambda_{l_2}^2, \lambda_{l_3}^2) \in \mathcal{A}$), and the results of measuring $I \otimes \sigma_{l_1}^2$, $I \otimes \sigma_{l_2}^2$, and $I \otimes \sigma_{l_3}^2$ also always have two 1's and one 0 (i.e. $(\gamma_{l_1}^2, \gamma_{l_2}^2, \gamma_{l_3}^2) \in \mathcal{A}$).

3. $(\lambda_{l_1}^2, \lambda_{l_2}^2, \lambda_{l_3}^2) = (\gamma_{l_1}^2, \gamma_{l_2}^2, \gamma_{l_3}^2)$ ("Twinned," as Conway and Kochen call it.)

This is the TWIN axiom, the other axiom that comes from the quantum theory.

> **The TWIN Axiom.** *For twinned spin 1 particles, suppose Alice performs a triple experiment of measuring the squared spin component of particle 1 in three orthogonal directions (l_1, l_2, l_3), while Bob measures the twinned particle 2 in one direction, l. Then if l happens to be in the same direction as one of l_1, l_2, l_3, Bob's measurement will necessarily yield the same answer as the corresponding measurement by Alice.*

In our proof, we will only consider l to be one of the 33 directions in $\mathcal{L}$, and l_1, l_2, l_3 to be one of 16 particular orthogonal triples in $\mathcal{O}$.

The paradoxes of relativity. Unlike the SPIN and the TWIN axioms, both of which are based on quantum theory, the third axiom, MIN, is based on the theory of relativity.

The tenets of Einstein's special theory of relativity are that the speed of light is a constant independent of the observer and that the laws of physics are the same in all inertial frames of reference. One consequence is that if two events take place far from each other in space, they will happen in one order with respect to some inertial frames, but in the reverse order with respect to other inertial frames. This is what we will use to justify the MIN axiom.

Alice and Bob, both on Earth, prepare particles 1 and 2 in the singlet state. Bob takes particle 2 to Pluto, along with a device that can measure the squared spin of particle 2 in any orientation $l \in \mathcal{L}$.

> **The MIN Axiom.** *Assume that the experiments performed by Alice and Bob are space-like separated. Then Bob can choose any one of the* 33 *particular directions* l*, and particle 1's response is independent of this choice. Similarly and independently, Alice can choose any one of the* 16 *triples* (l_1, l_2, l_3)*, and particle 2's response is independent of that choice.*

According to the theory of relativity, there exists one inertial frame in which Bob's experiment comes only after Alice's, and thus it cannot influence particle 1's response. Similarly, Alice's choice of directions cannot influence particle 2's response because, in some inertial frame, by the time Alice performs her experiment, particle 2's spin has already been observed. Since the future cannot alter the past, it is safe for us to assume the MIN axiom, given that we are willing to accept the theory of special relativity.

The Free Will Theorem. When we say Alice's choice of (l_1, l_2, l_3) is free, we mean that it is not a function of the past history (in any inertial frame).

Theorem. *(The Free Will Theorem) The axioms SPIN, TWIN and MIN imply that if Alice and Bob can freely choose which directions to measure the spin, then the particles' responses are similarly free - meaning that they too are not a function of the past.*

Proof. Just like the proof of the Kochen-Specker paradox, we begin by supposing the contrary that particle 1's response to Alice's measurement $(\lambda^2_{l_1}, \lambda^2_{l_2}, \lambda^2_{l_3})$ is a function $\lambda^2_{(l_1,l_2,l_3)}(h_A) \in \mathcal{A}$, where the triplet (l_1, l_2, l_3) is in $\mathcal{O}$ and h_A (Alice's history) denotes all events prior to the moment that Alice chooses (l_1, l_2, l_3).

Similarly, we suppose that particle 2's response to Bob's measurement can be written as a function $\gamma^2_l(h_B) \in \{0, 1\}$ for $l \in \mathcal{L}$ and h_B is Bob's history (all events prior to the moment that Bob chooses l).

Note that by MIN, the function $\lambda^2_{(l_1,l_2,l_3)}(h_A)$ does not wary with l and likewise $\gamma^2_l(h_B)$ does not vary with (l_1, l_2, l_3).

By the free will assumption, $\gamma_l^2(h_B)$ is defined for all 33 directions in $\mathcal{L}$ since Bob's choice of l does not depend on h_B. Similarly, $\lambda^2_{(l_1,l_2,l_3)}(h_A)$ is defined for all 16 triplets in $\mathcal{O}$.

Now by TWIN we have

$$\lambda^2_{(l_1,l_2,l_3)}(h_A) = (\gamma^2_{l_1}(h_B), \gamma^2_{l_2}(h_B), \gamma^2_{l_3}(h_B)).$$

Then by SPIN we have that $\gamma_l^2(h_B) : \mathcal{L} \to \mathcal{A}$ is a 101 function. But this is impossible, by the Kochen-Specker no-go Theorem. Hence, either Alice and Bob were not free to choose their experiments, or the particles were free to choose their responses.

□

Scientific consequences. Before we proceed to demonstrate some of the implications of the theorem, we contend that determinism cannot be disproved. This is because there is no way to contradict a "second time around" argument: what is to say that we are not performing a second time as the universe is moving? We are being fed exactly the same sense impressions as we were in the original universe so that even if the first time around was free, the second time around is certainly deterministic. Therefore, there is no way to know whether free will definitely exists - it cannot be proved in a deductive manner.

We conclude our presentation with a comment on different opinions of philosophers on free will over the last two millennia.

Nearly 2000 years ago, Roman philosopher Lucretius already suggested that human free will comes from "the slight swerve of the atoms at no determinate time or place". This is perhaps the first account of the notion "free will" in the literature.

However, most scientists, until relatively recently, are determinists. For example, Gottfried Willhelm Leibniz held on to the principle of sufficient reason, which he described as "nothing happens without a reason why it should be so". One consequence of the Free Will Theorem is that if we do have free will, Leibniz's principle of sufficient reason would not be true even in the world of particles.

René Descartes' belief was that of a "disconnected" determinism: a dualism between mind and matter in the sense that everything except for our mind is deterministic, but "the will is by its nature so free that it can never be constrained". Note that the Free Will Theorem directly contradicts Descartes' dualism.

Perhaps this is what the theorem is really about – there is no middle ground between free will and determinism. As the 1999 Nobel Prize in physics winner Gerard 't Hooft puts it: "If you believe in determinism, you have to believe it all the way. No escape possible. Conway and Kochen have shown here in a beautiful way that a half-hearted belief in pseudo-determinism is impossible to sustain."

4.6 Quantum computing

Consider a system of n spin-1/2 particles, represented as an element (or wave function) $|\psi\rangle$ of the n-fold tensor-product space $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$. The typical element is a super-position of 2^n basis states, and the action of a unitary operator will simultaneously alter every one of these states. If we think of the initial state as a collection of 2^n possible initial conditions, and a sequence of unitary operations as a sequence of computations, then there is a sense in which each computation is performing, simultaneously, a given operation on each of 2^n states. If we could organize the unitary operators in such a way as to compute a useful consequence from each initial condition, then we could think of the sequence of operations as representing 2^n computers running in parallel. This is one way to look at the promise of quantum computing.

Feynman viewed it as a possible way to simulate otherwise intractable complex quantum systems; Shore demonstrated the potential to more efficiently factor large numbers made up of two large prime divisors, thereby challenging the security of the commonly used encryption standards; and Grover and others showed that, in principle, many of the most computationally intensive problems (e.g. circuit design for chips, image segmentation, data-base searching, and routing problems) could be solved much faster than what is currently possible using conventional computers. The race is on to build general-purpose quantum computers that can be programmed to find solutions to these and other challenging problems.

A good place to start the study of quantum computing is with Grover's quantum search algorithm, which can be described fairly succinctly using the tools that we have developed for spin systems.

Commuting Hermitian operators. The free-will theorem made use of the fact that commuting Hermitian operators represent measurements that can be performed in an arbitrary sequence without changing the probability of the corresponding outputs. Before developing the Grover search procedure, we will review the connection between commuting operators and sequential observations, and show how to exploit this connection to extract a solution from the final state of a quantum computer, following the execution of an algorithm.

1. **(single measurement and some notation)** Let H be a Hermitian operator with spectral representation
$$H = \sum_{i=1}^{n} \lambda_i^H \, |e_i\rangle\langle e_i|$$
Measurement with H will have result in $|\psi\rangle \to |e_i\rangle$, for some i, and the corresponding measurement value λ_i^H will be read. Let's summarize the result of the measurement using the notation
$$|\psi\rangle \xrightarrow[\lambda^H = \lambda_i^H]{H} |e_i\rangle$$

or just $|\psi\rangle \xrightarrow{H} |e_i\rangle$ if we don't need to emphasize the measurement value. The respective probabilities and expected values are then

$$\mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^H=\lambda_i^H]{H} |e_i\rangle\right) = |\langle e_i|\psi\rangle|^2$$
$$\mathbb{E}\left[\lambda^H \middle| |\psi\rangle\right] = \sum_{i=1}^{n} \lambda_i^H |\langle e_i|\psi\rangle|^2$$

2. **(two measurements)** Suppose now that we have two Hermitian operators, H_1 and H_2:

$$H_1 = \sum_{i=1}^{n} \lambda_i^{H_1} |e_i\rangle\langle e_i|$$
$$H_2 = \sum_{j=1}^{n} \lambda_j^{H_2} |f_j\rangle\langle f_j|$$

What if we try to measure both? There is no meaningful way to talk about measuring both at once, so to make sense of this we will need to imagine measuring them one at a time. Consider measuring H_1 and then H_2, leading to states $|e_i\rangle$ and then $|f_j\rangle$, respectively:

$$\begin{aligned}\mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_1}=\lambda_i^{H_1}]{H_1} |e_i\rangle \xrightarrow[\lambda^{H_2}=\lambda_j^{H_2}]{H_2} |f_j\rangle\right) &= \mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_1}=\lambda_i^{H_1}]{H_1} |e_i\rangle\right) \mathbb{P}\left(|e_i\rangle \xrightarrow[\lambda^{H_2}=\lambda_j^{H_2}]{H_2} |f_j\rangle\right) \\ &= |\langle e_i|\psi\rangle|^2 |\langle f_j|e_i\rangle|^2\end{aligned} \tag{44}$$

But in the other order:

$$\begin{aligned}\mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_2}=\lambda_j^{H_2}]{H_2} |f_j\rangle \xrightarrow[\lambda^{H_1}=\lambda_i^{H_1}]{H_1} |e_i\rangle\right) &= \mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_2}=\lambda_j^{H_2}]{H_2} |f_j\rangle\right) \mathbb{P}\left(|f_j\rangle \xrightarrow[\lambda^{H_1}=\lambda_i^{H_1}]{H_1} |e_i\rangle\right) \\ &= |\langle f_j|\psi\rangle|^2 |\langle e_i|f_j\rangle|^2\end{aligned} \tag{45}$$

We get the same readings, but with a different probability and final state.

3. **(two measurements, commuting operators)** The story is very different if H_1 and H_2 commute, i.e. $[H_1, H_2] = H_1H_2 - H_2H_1 = 0$. The key is the following important theorem:

Theorem. *If H_1 and H_2 are commuting Hermitian operators, then there exists a basis, $|g_1\rangle, |g_2\rangle, \ldots, |g_n\rangle$, of common eigenvectors:*

$$H_1 = \sum_{i=1}^{n} \lambda_i^{H_1} |g_i\rangle\langle g_i|$$
$$H_2 = \sum_{j=1}^{n} \lambda_j^{H_2} |g_j\rangle\langle g_j|$$

Remarks:

(i) The theorem extends immediately to a set of N commuting operators: if $H_1, H_2, \ldots, H_N$ are Hermitian operators such that $[H_i, H_j] = 0$ for all $i, j \in \{1, 2, \ldots, N\}$, then there exists a common set $|g_1\rangle, |g_2\rangle, \ldots, |g_n\rangle$ of eigenvectors.

(ii) The proof is not hard, but a good place to start is with the simplifying assumption that neither H_1 nor H_2 have degenerate eigenvalues. (An eigenvalue is degenerate if it has two or more linearly independent eigenvectors.)

4. **(example from the free-will theorem)**. Recall this example from the free-will theorem:

$$\sigma_z^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

(We don't conjugate anything; we just square the matrix. After all, this is what corresponds to squaring the *real-valued* eigenvalues. We've seen this before, in the set up and proof of the uncertainty principle.) Similarly,

$$\sigma_x^2 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \qquad \sigma_y^2 = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 1 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \qquad \sigma_z^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

For which

$$[\sigma_x^2, \sigma_y^2] = [\sigma_x^2, \sigma_z^2] = [\sigma_y^2, \sigma_z^2] = 0$$

with common eigenvectors

$$|e_1\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \qquad |e_2\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \qquad |e_3\rangle = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

5. **(commuting operators and order-independent measurement)** Let's redo the computations in equations (44) and (45), but this time under the assumption that H_1 and H_2 commute, in which case they have a representation of the form:

$$H_1 = \sum_{i=1}^{n} \lambda_i^{H_1} |g_i\rangle\langle g_i| \qquad H_2 = \sum_{j=1}^{n} \lambda_j^{H_2} |g_j\rangle\langle g_j|$$

If we measure H_1 and then H_2, then for any $|g_i\rangle$ and $|g_j\rangle$

$$\begin{aligned} \mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_1} = \lambda_i^{H_1}]{H_1} |g_i\rangle \xrightarrow[\lambda^{H_2} = \lambda_j^{H_2}]{H_2} |g_j\rangle \right) &= \mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_1} = \lambda_i^{H_1}]{H_1} |g_i\rangle \right) \mathbb{P}\left(|g_i\rangle \xrightarrow[\lambda^{H_2} = \lambda_j^{H_2}]{H_2} |g_j\rangle \right) \\ &= |\langle g_i|\psi\rangle|^2 |\langle g_j|g_i\rangle|^2 \delta_{ij} = \delta_{ij} |\langle g_i|\psi\rangle|^2 \end{aligned}$$

by the orthogonality of the g basis elements. The interpretation is simple: the first measurement left the particle in the definite state $|g_i\rangle$, which is already an eigenstate of H_2, and therefore will not

change. With this in mind, we should not be surprised to find that we get the same probability if we start with H_2:

$$\mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_2}=\lambda_j^{H_2}]{H_2} |g_j\rangle \xrightarrow[\lambda^{H_1}=\lambda_i^{H_1}]{H_1} |g_i\rangle\right) = \mathbb{P}\left(|\psi\rangle \xrightarrow[\lambda^{H_2}=\lambda_j^{H_2}]{H_2} |g_j\rangle\right) \mathbb{P}\left(|g_j\rangle \xrightarrow[\lambda^{H_1}=\lambda_i^{H_1}]{H_1} |g_i\rangle\right)$$
$$= |\langle g_j|\psi\rangle|^2 |\langle e_i|f_j\rangle|^2 = \delta_{ij}|\langle g_j|\psi\rangle|^2$$

Same result in either order. The probability is determined after the first measurement, since the result is already a eigenvector of the remaining measurement.

These results extend, without any change in reasoning, to N commuting operators, $H_1, H_2, \ldots, H_N$: the first measurement determines the definite state, and its probability, for all of the remaining measurements. Since the ensuing state does not change, the ensuing measurement readings are already determined. Nevertheless, they're not actually known until we carry out the remaining measurements.

Grover's quantum search algorithm. A Boolean function is a two-valued (true or false, one or zero) function of several two-valued variables, $f : \{0,1\}^n \to \{0,1\}$. Suppose that we are given the function f, in the sense that we can compute its value, $f(x)$ equal to zero or one, for any given $x \in \{0,1\}^n$. The "satisfiability" problem is the problem of finding a particular value $g \in \{0,1\}^n$ such that $f(g) = 1$, or determining that no such g exists. The best known (conventional) algorithm will, in general, require $O(2^n)$ operations, which amounts to brute-force search. Since the problem belongs to the NP-complete class of problems, any faster solution to satisfiability would in principle translate into equally efficient speedups for the entire class of problems, which includes a very long list of practical problems for which the current best solutions amount to a patchwork of situation-dependent *ad hoc* approaches. Grover's quantum search algorithm requires $O(2^{\frac{n}{2}})$ operations, i.e. the square-root of the number of operations needed when operating with a conventional computer. (Actually building the quantum computer is another matter entirely, although, as you might imagine, plenty of people are trying. And there has been good progress to show for it.)

Here, we will make the simplifying assumption that our job is to find g such that $f(g) = 1$ under the working hypothesis that there is one and only one solution. All of the main ideas for solving more general versions are there, as you will see in the homework.

The description of the algorithm uses notation that is not standard in discussions of spin states, but is equivalent and far more convenient for describing quantum computations. Specifically, $|0\rangle$ is used instead of $|u\rangle$ to describe the spin-up state, and $|1\rangle$ is used instead of $|d\rangle$ to describe the spin-down state, and in this case the particle is called a qubit. In $\mathbb{C}^2$, only the labels are changed: $|0\rangle$ is represented by $\begin{pmatrix}1\\0\end{pmatrix}$ and $|1\rangle$ is represented by $\begin{pmatrix}0\\1\end{pmatrix}$. Measurement in the z direction has the same eigenvectors, $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$, but it is convenient to relabel the observables (i.e. the eigenvalues) as zero and one, respectively. To avoid

confusion, we will call the new operator ω_z rather than σ_z:

$$\omega_z = 0 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \tag{46}$$

The key data structure is a set of n spin-1/2 particles, maintained in a (quantum mechanical) "register" with initial state

$$|\psi\rangle \in \mathbb{C}^{2\otimes n} \doteq \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2 \qquad \text{(the n-fold tensor-product of } \mathbb{C}^2\text{)}$$

The state of the register (meaning the collective state of its n particles) evolves during the computation into a final state, $|\Psi_\tau\rangle \in \mathbb{C}^{2\otimes n}$, which is used in a "read-out" procedure to produce the final output of the algorithm.

Using the $\{|0\rangle,|1\rangle\}$ (up/down) basis for $\mathbb{C}^2$, the basis elements of the entire register are the 2^n vectors $|x_1 x_2 \cdots x_n\rangle$, where $(x_1, x_2, \ldots, x_n) \in \{0,1\}^n$ is a sequence of n zeros and ones. We can make convenient use of the "$|0\rangle\,|1\rangle$" notation by indexing basis elements with vectors $x \in \{0,1\}^n$. If $x = (x_1, x_2, \ldots, x_n)$ then we write $|x\rangle$ as short for the basis element $|x_1 x_2 \cdots x_n\rangle \in \mathbb{C}^{2\otimes n}$. We will use $\mathcal{B}$ to designate the entire basis,

$$\mathcal{B} \doteq \left\{ |x\rangle : x \in \{0,1\}^n \right\} \tag{47}$$

Notice that each basis element $|x\rangle \in \mathbb{C}^{2\otimes n}$ represents a set of n independent spin-1/2 particles, $|x\rangle = |x_1 x_2 \cdots x_n\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_n\rangle$.

A generic state $|\phi\rangle$ is a superposition of basis elements with arbitrary amplitudes,

$$|\phi\rangle = \sum_{x\in\{0,1\}^n} \gamma_x |x\rangle$$

where for each $x \in \{0,1\}^n$, $\gamma_x \in \mathbb{C}$ (so $\gamma_x = \gamma_{(x_1,\ldots,x_n)}$), and where $\sum_{x\in\{0,1\}^n} |\gamma_x|^2 = 1$ (so $|\phi\rangle \in \mathbb{C}_1^{2\otimes n}$).

With these conventions in place, we are ready to construct the Grover quantum search algorithm. We will start at the end (the readout), and then work our way from the preparation of the register's initial state, through a sequence of unitary transformations performed on the register, and to the conclusion that a measurement of the final state of the register, which have called $|\psi_\tau\rangle$, is very likely to be the solution to the search problem:

$$\mathbb{P}\left(|\psi_\tau\rangle \xrightarrow{\text{measure register}} |g\rangle \right) > 1 - \epsilon$$

for a small value of ϵ that we will compute explicitly.

1. **(how to measure the spin states of the particles in the register)** . Quantum computers operate on a quantum register, which is a collection of separated two-state particles ("qubits") held in fixed positions and manipulated by applying physical realizations of unitary operators. The algorithm defines the particular sequence of operations. Read-out is accomplished by the physical realization of

a Hermitian measurement operator, often with the expectation that the most likely resulting eigenstate will hold the solution to the problem at hand. But how can we sample from the superposition of 2^n states? It turns out that this part can be done with just n operations, by simply measuring, in sequence, each element of the register.

The coordinate measurement operators apply a single-particle spin measurement to each individual particle, just as though it could be pulled out of the lattice making up the register, measured, and then re-inserted. We have already studied the action of such measurements, noting that, by definition, $A \otimes B \left|e, f\right\rangle = \left|Ae, Bf\right\rangle$ for any pair of independent particles in the states $\left|e\right\rangle$ and $\left|f\right\rangle$, in which case $A \otimes I \left|e, f\right\rangle = \left|Ae, f\right\rangle$ and $I \otimes B \left|e, f\right\rangle = \left|e, Bf\right\rangle$. Importantly, this implies that $A \otimes I$ and $I \otimes B$ commute:

$$(I \otimes B)(A \otimes I) \left|e, f\right\rangle = I \otimes B \left|Ae, f\right\rangle = \left|Ae, Bf\right\rangle = A \otimes I \left|e, Bf\right\rangle = (A \otimes I)(I \otimes B) \left|e, f\right\rangle$$

The construction for higher-order tensor product spaces is not different. So let us define n z-direction coordinate operators[17]

$$H_k = I^{\otimes k-1} \otimes \omega_z \otimes I^{\otimes n-k} \quad \forall k = 1, \ldots, n$$

where $A^{\otimes m}$ is short for the m-fold tensor product of A with itself (which we might also have written as $\otimes_{i=1}^{m} A$), and note, analogously, that every pair of the set $H_1, H_2, \ldots, H_n$ commutes. Therefore, applying our theorem about commuting Hermitian operators, we conclude that there exists a basis for $\mathbb{C}^{2 \otimes n}$ such that every element of the basis is an eigenvector of every H_k, $k = 1, \ldots, n$.

It's not hard to find such a basis. In fact, we've already identified one, which we labelled $\mathcal{B}$ in equation (47). For any k and any $\left|x\right\rangle \in \mathcal{B}$:

$$H_k \left|x\right\rangle = I^{\otimes k-1} \otimes \omega_z \otimes I^{\otimes n-k} \left|x\right\rangle = \lambda_{x_k} \left|x\right\rangle$$

where λ_{x_k} is zero (formally one) if $\left|x_k\right\rangle = \left|0\right\rangle$ and λ_{x_k} is one (formally minus one) if $\left|x_k\right\rangle = \left|1\right\rangle$. The upshot is that we can measure (actually sample) the final state of the register, $\left|\psi_\tau\right\rangle$, by sequential measurement of the z-direction coordinate operators: for any $\left|x\right\rangle \in \mathcal{B}$ (equivalently, any $x \in \{0, 1\}^n$

$$\mathbb{P}\left(\left|\psi_\tau\right\rangle \xrightarrow[\lambda^{H_1} = x_1]{H_1} \left|x\right\rangle \xrightarrow[\lambda^{H_2} = x_2]{H_2} \left|x\right\rangle \cdots \xrightarrow[\lambda^{H_n} = x_n]{H_n} \left|x\right\rangle\right) = \mathbb{P}\left(\left|\psi_\tau\right\rangle \to \left|x\right\rangle\right) = \left|\left\langle x|\psi_\tau\right\rangle\right|^2 = \left|\gamma_x\right|^2 \quad (48)$$

2. **(prepare register)** The register is initialized as the normalized sum over all of the eigenstates:

$$\left|\psi\right\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} \left|x\right\rangle \quad (49)$$

i.e. the uniform superposition of all basis functions. (It will be shown in the homework that $\left|\psi\right\rangle$ is more simply described as n independent particles in what we used to call the $\left|r\right\rangle$ direction, i.e. $\frac{1}{\sqrt{2}} \left|0\right\rangle + \frac{1}{\sqrt{2}} \left|1\right\rangle$.)

3. **(the action subspace)** By defining the right subspace, it turns out that Glover's algorithm can be viewed as executing repeated small rotations in a *real-valued two-dimensional subspace* of $\mathbb{C}^{2 \otimes n}$. We will first define the subspace and then define the rotations.

[17] Recall that we have re-labelled the eigenvalues of σ_z, and hence its matrix representation, which now call ω_z. See (46).

First, notice that since the solution g is in $\{0,1\}^n$, it specifies a particular element of the basis set $\mathcal{B}$, which we will call $|g\rangle$. Consider now another state $|\psi'\rangle$, also in $\mathcal{B}$, which can be thought of as a small rotation of $|\psi\rangle$ in the direction of $|g\rangle$:

$$|\psi'\rangle = \frac{1}{\sqrt{2^n - 1}} \sum_{x \in \{0,1\}^n, x \neq g} |x\rangle$$

Observe that

i. $\langle \psi' | g \rangle = 0$

ii.

$$\begin{aligned} |\psi\rangle &= \frac{1}{\sqrt{2^n}} |g\rangle + \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n, x \neq g} |x\rangle \\ &= \frac{1}{\sqrt{2^n}} |g\rangle + \frac{\sqrt{2^n - 1}}{\sqrt{2^n}} \frac{1}{\sqrt{2^n - 1}} \sum_{x \in \{0,1\}^n, x \neq g} |x\rangle \\ &= \frac{1}{\sqrt{2^n}} |g\rangle + \frac{\sqrt{2^n - 1}}{\sqrt{2^n}} |\psi'\rangle \end{aligned}$$

iii. And, therefore

$$|\psi\rangle \in \mathcal{S} \doteq \left\{ \alpha |g\rangle + \beta |\psi'\rangle \middle| \alpha, \beta \in \mathbb{R} \right\}$$

which is the two-dimensional real space spanned by the orthonormal vectors $|\psi'\rangle$ and $|g\rangle$.

In fact, if we let $\mathcal{S}_1 = \left\{ \alpha |g\rangle + \beta |\psi'\rangle \middle| \alpha^2 + \beta^2 = 1 \right\}$, then $|\psi\rangle$ is also in $\mathcal{S}_1$, and everything that follows can be thought of as rotating $|\psi\rangle$ clockwise until it is very nearly in alignment with $|g\rangle$—see figure(2).

4. **The rotation operator.**

5. **Lower bound on the probability of success.**