Not Bell's Theorem

Brendan Jackson

April 7, 2009

1 Introduction

The purpose of this lecture is to introduce you to an important result about the foundations of quantum mechanics. This result is embodied in Bell's theorem.

But Bell's original analysis is technical, not at all memorable, and unsuitable for pedagogical purposes. This lecture will *not* be about Bell's theorem (hence the title).

Fortunately, a pedagogical version of the theorem is available. This version is by David Mermin, the result of a long chain of references. By way of an incomplete list, this is my version....

- ... of the refinement by David Mermin... [6]
- ...of the thought experiment by Greenberger, Horne and Zeilinger... [4]
- ...based on Bell's original theorem... [2]
- ...inspired by Einstein, Podolsky and Rosen's original paradox [3].

Of course, the majority of relevent references are omitted.

In the first half of this document, the scene will be set. We'll recall the basics of quantum mechanics, and see a formulation that is not often covered: de Broglie-Bohm mechanics. This is a way of rewriting quantum mechanics so that it deals with particles, moving along well defined trajectories, with well defined position and momenta. And it reproduces the predictions of quantum mechanics.

We'll then look at why this interpretation fails to be compatible with the principle of locality. The next logical step would be to ask if it's possible to do better, and come up with an interpretation that ensures objects have well defined properties, but is still a local theory.

The answer is 'no', and the second half of this document will be proving this, using Mermin's thought experiment.

2 Quantum Mechanics

Completeness

The state of a quantum system is completely specified by the state $|\Psi\rangle$.

Normalisation

The state is normalised such that $\langle\Psi|\ \Psi\rangle = 1$.

Evolution States evolve according to causal evolution of the Schrodinger equation,

$$\hat{\mathcal{H}}\,|\Psi\rangle = i\hbar\frac{\partial\,|\Psi\rangle}{\partial t}. \tag{1}$$

Observables and operators

Observables have a corresponding linear, Hermition operator. For instance, the classical observable of momentum in the x-direction p_x has a corresponding operator $-i\hbar\frac{\partial}{\partial_x}$.

Measurements

The classical observables, q_n associated with an operator are the eigenvalues associated with that operator $\hat{Q}$. If the state is a pure eigenstate of that operator, the measured observable is the eigenvalue.

$$\hat{Q}\,|\Psi\rangle = q\,|\Psi\rangle\,. \tag{2}$$

If the wave-function is a linear combination of eigenstates of the operator, e.g.

$$|\Psi\rangle = \sum_n a_n\,|\phi_n\rangle \tag{3}$$

the measured value is *one* of the eigenvalues, with probability $|a_n|^2$.

So much for standard quantum mechanics.

3 Not Quantum Mechanics: de Broglie-Bohm Theory

3.1 Probability

The final postulate in the last section leaves many people troubled. Very few quantum systems are so fortunately arranged as to be pure eigenstates of the observable we were interested in. The outcome of all other sorts of measurements on a quantum system are probabilistic.

In all other areas of our experience, probability only enters due to our own ignorance. The result of the throw of a dice is random because we do not know the initial position and momentum of the dice well enough to predict its course from the laws of Newtonian mechanics. The chaotic behaviour means we treat the outcome as probabilistic. Classical statistical mechanics likewise ascribes events particular probabilities, even though the theory is a result of Newton's Laws, which are completely deterministic.

3.2 Hidden variables from de Broglie

One view that can be taken is that the probabilistic nature of the predictions of quantum mechanics are a clue that we are ignorant of some important variables in quantum systems. If we knew these 'hidden variables', we would be able to precisely determine the outcome of any experiment. But since our measuring devices are large, clumsy things, that disturb the very objects we attempt to measure, we can never be so precise. The outcome therefore appears probabilistic, governed (to good approximation) by the rules of quantum mechanics. In this view quantum mechanics is an approximate theory.

It may be easy to dismiss such a view out of hand, but a very simple example exists that reproduces all the predictions of quantum mechanics. This interpretation was originally worked out by de Broglie, promptly forgotten, but later rediscovered and developed by David Bohm. It works as follows.

First consider the Schrodinger equation:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + V\psi. \tag{4}$$

We can write ψ as a combination of two real functions, $S(\underline{x},t)$ and $R(\underline{x},t)$ by using plane-polar representation of imaginary funtions.

$$\psi = R\exp(iS/\hbar). \tag{5}$$

Inserting this into the Schrodinger equation, cancelling the common factor of $\exp(iS/\hbar)$, and seperating out the real and imaginary components we find:

$$\frac{\partial S}{\partial t} + \frac{1}{2m}(\nabla S)^2 + V - Q = 0 \tag{6}$$

$$\frac{\partial R}{\partial t} + \nabla R \,.\left(\frac{\nabla S}{m}\right) + R\frac{\nabla^2 S}{2m} = 0. \tag{7}$$

Here we've written

$$Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} \tag{8}$$

to keep the maths tidy.

All we have done so far is re-write Schrodingers equation in terms of these new functions. But it paves the way to an entirely different interpretation. First look at equation (7). Multiplying by $2R$ and rearranging yields

$$\frac{\partial (R^2)}{\partial t} + \nabla.\left(R^2\frac{\nabla S}{m}\right) = 0. \tag{9}$$

Compare this to the continuity equation from classical mechanics:

$$\frac{\partial \rho}{\partial t} + \nabla.(\rho \underline{v}) = 0. \tag{10}$$

Since $R^2 = |\Psi|^2$, the probability density, comparison strongly hints that we associate the particle a velocity $\underline{v} = \nabla S/m$, and that the momentum $\underline{p} = \nabla S$.

If we work with this interpretation, we can apply it to (6). If we neglect Q (or alternatively, let $\hbar \to 0$) we are left with something very similar to the Hamilton-Jacobi equation.

3.3 Hamilton-Jacobi equation

If you have not seen the Hamilton-Jacobi equation before, here is a non-rigorous derivation for it. We will work with a conservative potential $U(\underline{x})$.

Newton's law tells us that:

$$\frac{d\underline{p}}{dt} = -\nabla U = \frac{\partial \underline{p}}{\partial t} + \frac{1}{m}\left(\underline{p}.\nabla\right)\underline{p}, \tag{11}$$

from expanding the full derivative into its set of partial ones. We will assume we can write the momentum as the gradient of some scalar S', so that $\underline{p} = \nabla S'$. From this we find:

$$\nabla U + \nabla\frac{\partial S'}{\partial t} + \frac{1}{2m}\nabla\left(\nabla S'\right)^2 = 0. \tag{12}$$

The last term follows from the identity $\frac{1}{2}\nabla\left(\underline{A}.\underline{A}\right) = \left(\underline{A}.\nabla\right)\underline{A}$ when $\underline{A}$ is curl free. Factoring out the gradient operator,

$$\nabla\left(U + \frac{\partial S'}{\partial t} + \frac{1}{2m}\left(\nabla S'\right)^2\right) = 0, \tag{13}$$

which implies:

$$U + \frac{\partial S'}{\partial t} + \frac{1}{2m}\left(\nabla S'\right)^2 = f(t) \tag{14}$$

Writing

$$S = S' - \int f(t)d\,\mathrm{t} \tag{15}$$

we are left with the Hamilton-Jacobi equation for a particle in a potential U,

$$U + \frac{\left(\nabla S\right)^2}{2m} = H = -\frac{\partial S}{\partial t}. \tag{16}$$

This is a reformulation of Hamiltonian mechanics.

Had we started with this end result, we could even show that $\underline{p} = \nabla S$; this gives some justification for making the assumption earlier (see [5] for a rigorous derivation from canonical transformations of Hamilton's equations).

3.4 The Quantum Potential

Comparing the Hamilton-Jacobi equation with the re-arranged, real part of the Schrodinger equation, and using our interpretation that $v = \nabla S/m$ that was motivated from the imaginary part, we see that the Hamilton-Jacobi equation has been recovered, provided we use a modified potential $V = U + Q$. This modification to the 'true' potential is called the quantum potential.

$$Q = -\frac{\hbar^2}{2m}\frac{\nabla^2 R}{R} \tag{17}$$

We now have an entirely different interpretation of quantum mechanics. A particle may have a definite position and momentum. Were we to know this precisely, its motion would be entirely determined by the Hamilton-Jacobi equation with this additional quantum potential. All that remains is to assert we do not know its position and momentum exactly. Our instruments are large, clumsy, imprecise things that disturb the very objects we seek to measure. We can merely solve the possible paths the particle might take, given our uncertainty. As $R^2 = |\Psi|^2$, then the probability distribution of particles will evolve just as predicted by the Schrodinger equation. The predictions are therefore statistically identical to the predictions of quantum mechanics.

3.5 Can we rule out hidden variables?

It is perfectly consistent, so far, for a person to believe that quantum mechanics can be explained by some (so far unknown) underlying classical-like theory. de Broglie-Bohm theory is an example, but we can at least imagine the possibility of others.

Such a person might make the following statements:

- The state of a quantum system is determined by extra, hidden variables not included in our description of the wavefunction.
- These hidden variables determine the outcome of measurements, but our ignorance and inability to measure these variables results in an appearance of randomness.
- We may remain in hope that one day some deeper theory will emerge and provide us with governing equations for these hidden variables.
- Our instruments may even become sensitive enough to measure them.
- The properties of objects exist even if we do not measure them.

Is there any way we can rule such a world-view out?

Well, we can rule out de Broglie-Bohm provided we include one more assumption.

4 An Important Assumption: Locality

Consider a region of spacetime, and three sub-regions (see figure 4). We'll call them A, B and B'. They are arranged so that A is at a later time coordinate than B', and an earlier coordinate than B. A cannot interfere with B' without sending information backwards in time; likewise B cannot interfere with A.

But a Lorentz transformation can leave A at the origin, and map B onto B'. We are perfectly entitled to consider B actually B' according to some other observer. Thus A cannot interfere with events at B or B', and vice-versa.

Thus, spacelike seperated events cannot interfere with each other. This is the principle of locality.

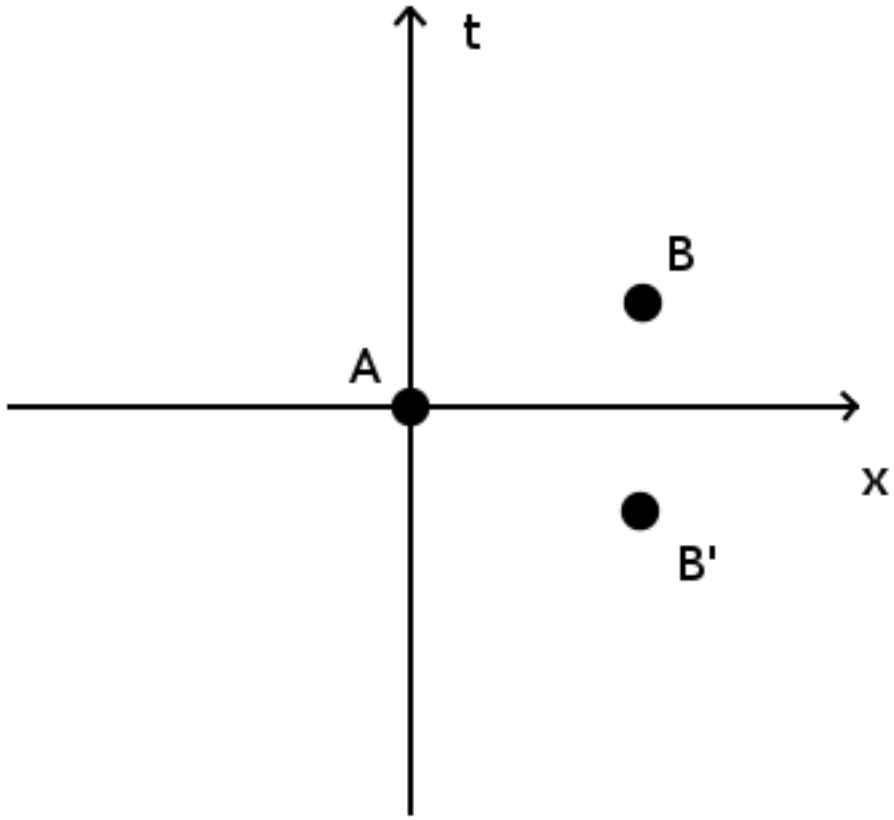

Figure 1: A region of spacetime. Regions A and B, and A and B', are spacelike seperated. A cannot interfere with B' without sending information backwards in time; likewise B cannot interfere with A.

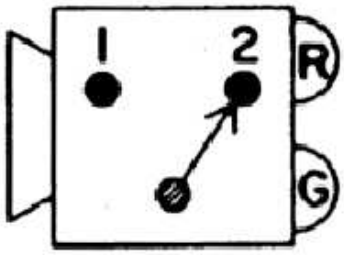

Figure 2: A sophisticated detector, with a setting to measure 1 or 2. On a successful detection, the green light flashes. If no detection is made, the red light flashes. Image from [6].

We can rule out de Broglie-Bohm theory immediately using this argument. The quantum potential is dramatically non-local - a small change anywhere in the universe would influence particles anywhere else.

Can we be cleverer than de Brogle-Bohm theory? Can we find a theory that is both local, and imbues objects with real properties, even if they're not measured?

We cannot. Now we shall see why.

5 Not Bell's Theorem

We are going to proceed as a thought experiment. To keep the element of surprise, exactly what is taking place will be revealed only at the end. We will begin with describing the experiment, the results, and proving no classical theory can explain all aspects of the experimental results.

Three detectors are taken some huge distance apart, equidistant from some origin. They have a person with them (no doubt a poor graduate student). Since their professor has forgotten their names, we will refer to them as A, B and C.

There are two settings the detectors can measure, but the grad students do not know what

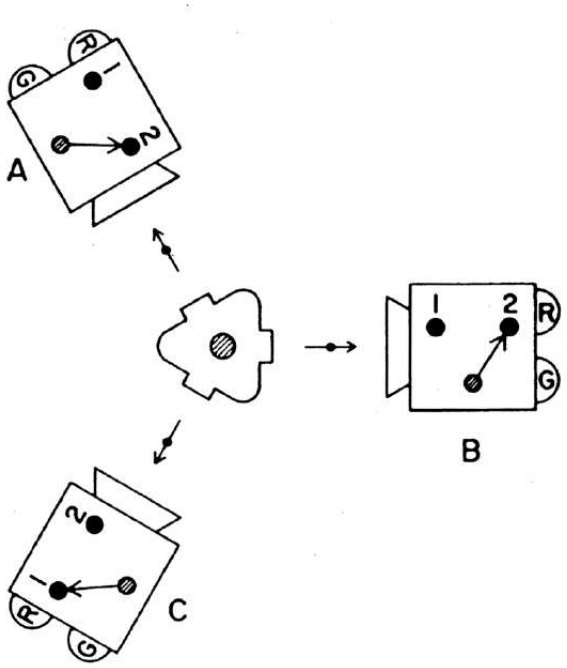

Figure 3: The experimental set-up. The three detectors are sent some large distance away, and are sent samples from the central source. Image from [6].

they are. All they know is they can choose to set the device to measure 1 or measure 2. They don't know what these settings mean. After carrying out a detection, either a green light will flash, or a red light.

The students are sent samples from a central source, setting off from the origin to their respective destinations all at once. These samples are also unknown. The students may be recieving blood samples and are testing either for high blood sugar, or high cholesterol. They may be recieving playing cards and are testing for red or black. The whole thing may be a hoax, and the lights may be flashing randomly or according a pre-programmed computer chip. At this point they (and we) do not know.

The students decide what they will measure (setting 1 or setting 2) randomly. They then write down how their detector flashes in their notebooks. They use -1 to signify a green light flashing, and $+1$ to show a red light flashing. They write it down like this:

$$B_1 = +1 \tag{18}$$

This means *detector B* measured *setting 1* and got a *result* of $+1$ (red light). They write it in this way because they believe they are measuring properties of their samples. Here is an example of some data they might collect.

$$\begin{array}{lll} A_1 = +1 & B_2 = -1 & C_2 = -1 \\ A_2 = +1 & B_2 = -1 & C_1 = -1 \\ A_2 = -1 & B_1 = -1 & C_2 = +1 \end{array}$$

..etc.

After collecting the data and staring at it for a while, two important features of the data are apparent.

- If only one detector is set to 1 (and the others to 2), an even number of green lights always flash. This can be either be all the lights flashing red, or two flashing green. As we've written a green light as -1, multiplying the detector results for this set up will result in a positive number, e.g. $A_1 B_2 C_2 = +1$. You can check this example with the example data I gave above.
- If all three detectors are set to 1, an odd number of red lights (or equally an even number of green lights) is *never* observed to flash.

We are now in a position to prove any classical explanation of this data is simply untenable.

5.1 Short Proof

Let us write down some general results we observed. It is *always* the case that:

$$A_1 B_2 C_2 = +1 \tag{19}$$

$$A_2 B_1 C_2 = +1 \tag{20}$$

$$A_2 B_2 C_1 = +1 \tag{21}$$

Of course, not every 'run' of the experiment included one of the above combinations. But we conclude that *had* this combination occured, the result would hold. Afterall, unless the detectors can communicate with each other faster than the speed of light, any run could potentially end up as one of the above combinations (for what is to stop someone changing the setting at the last second?).

We can multiply these together:

$$(A_1 B_2 C_2)\,(A_2 B_1 C_2)\,(A_2 B_2 C_1) = +1. \tag{22}$$

In a single-run, the value of A_2 is *either* $+1$ *or* -1. And so $A_2A_2 = 1$. The same argument is true for B_2B_2 and C_2C_2 terms, and so we're left with:

$$A_1B_1C_1 = +1. \tag{23}$$

But compare this to the second observation about the experimental results obtained: when all the detectors are set to 1, an even number of green lights is *never* observed to flash. We have concluded the complete opposite.

5.2 Longer Proof

The above argument is extremely quick, and may leave you uncomfortable. Here's a more elaborate way of making the same argument.

The detection events are causally seperated. The switch settings on the detector are chosen randomly, and need not be set in advance - they could be chosen only moments before the detection takes place. Therefore when a sample leaves the detector, it must carry with it all the information for how it will make the lights flash, given a certain switch setting. We'll call this information 'instruction sets', and simply imagine the samples are carrying their instruction sets with them. When they reach the detector, they can look up how they should make the detector behave based on its switch setting.

There is nothing stopping the samples having correlated instruction sets. In fact, since there are correlations between the three detectors, it seems inevitable this is the case. There is nothing mysterious in this.

We'll represent the instructions carried by a single sample as a pair of letters: $\begin{smallmatrix}R\\R\end{smallmatrix}$, $\begin{smallmatrix}R\\G\end{smallmatrix}$, $\begin{smallmatrix}G\\R\end{smallmatrix}$, or $\begin{smallmatrix}G\\G\end{smallmatrix}$. The upper letter specifies the colour that will flash if the switch is set to 1, and the lower letter if the switch is set to 2.

The full instruction set for all three particles in a given run can be written by just listing them in order. For instance, if the instructions carried by the particles were $\begin{smallmatrix}RGG\\GRR\end{smallmatrix}$, this would result in RRR is the switch settings were 122, GGR for 212 and GRG for 221.

Let us once more consider only the first aspect of the data. When only one of the switch settings is set to 1, an odd number of red lights flash. The above instruction set is legal - it produces this result. An example of an illegal instruction set would be $\begin{smallmatrix}RRG\\GRR\end{smallmatrix}$: an even number of red flashes (GRR) would be observed for the setting 212.

What are all the legal instruction sets?

The choices 122,212 or 221 may occur in any run. There must be an odd number of red lights produced in any of these.

First consider 122. This forces:

R - -	R - -	G - -	G - -
- R R	- G G	- R G	- G R

The blank entries are unconstrained by our reasoning so far.

How many ways are there to fill the blank slots? Each of the four sets above already specifies the colour of light flashed at detector B for setting 2. Now consider that the setting 221 must yield an odd-number of flashed lights. If the B_2 setting is R, only if the remaining entries are RR or GG would an odd-number of red lights flash. If the B_2 setting is green, this means the remaining entries must be RG or GR. There are two sets with $B_2 = R$ and two sets with $B_2 = G$. Adding them up gives eight possible settings, and leaves the entry for B_1 still unspecified.

But now we note that 212 must also flash an odd number of red lights. So if A_2 and C_2 are different colours, the colour of B_1 must be G. If they are the same colour, B_1 must be R.

Thus there are a total of eight possible instruction sets, and here they are:

RRR	RGG	GRG	GGR
RRR	RGG	GRG	GGR

RGG	RRR	GGR	GRG
GRR	GGG	RRG	RGR

You can check that every instruction set will always cause an odd number of red lights to flash, when there is a single detector set to 1. As there are only 8 legal sets, and the above 8 are legal, we can rest assured these are the ones we were looking for.

Now consider the last form of run, where all three detectors are set to 1. There is nothing to stop the setting being flipped just before the sample arrives, so these sets must determine the results for these runs also.

But *every* one of the eight allowed instruction sets results in an odd number of red flashes when all three are on setting 1. And the experiment *never* produces an odd number of flashes. Thus the particles *cannot* be carrying instructions for how the detectors flash.

5.3 How is it really done?

Now for what is really taking place. At the source, three spin-1/2 particles are arranged in a curious initial state. Defining the z-axis as the axis along the line of flight for each particle (one to each detector) the state reads:

$$\Psi = \frac{1}{\sqrt{2}} \left(|\uparrow, \uparrow, \uparrow\rangle - |\downarrow, \downarrow, \downarrow\rangle \right). \tag{24}$$

The up-arrow is spin up along the appropriate z-axis for the particle, down-arrow is spin down.

The three detectors actually contain Stern-Gerlach magnetics which measure the vertical (x) component of spin if set to 1, and the horizontal component (y) when set to 2. If spin down is found, the green light flashes. If spin up, the red light flashes.

Let's use units of angular momentum $\hbar/2$, so that the spin operators are just the Pauli spin matrices: $\sigma_x, \sigma_y, \sigma_z$. These are the operators associated with the spin observable. A spin up state in the corresponding axis has eigen-value $+1$, spin down has eigen-value -1.

Consider the measurement 122, which the language of quantum mechanics corresponds to the operator $\sigma_x^a \sigma_y^b \sigma_y^c$. The individual operators commute, so we can consider what happens in any order. First recall,

$$\sigma_x^A \, |\uparrow\rangle = |\downarrow\rangle \, . \tag{25}$$

(If you cannot remember why, see the appendix.) Acting on an eigen-state of spin on the z-axis, with x-axis spin operator, just flips the state. So does acting on the eigen-state with a y-axis spin operator,

$$\sigma_y^A \, |\uparrow\rangle = i \, |\downarrow\rangle \, , \tag{26}$$

but an extra factor of i is picked up. As there are two such factors, we end up with an overall factor of -1. So acting on our initial state with $\sigma_x^A \sigma_y^B \sigma_y^C$ gives back the same state, multiplied by 1. The same is true if A, B and C are swapped about, as the operators commute.

But now consider what happens if the measurement is made with all the detector settings set to 1. This corresponds to an operator $\sigma_x^A \sigma_x^B \sigma_x^C$. Acting with this on the initial state and we see it is eigen-value -1, reproducing the second aspect of the data perfectly.

At this point it should be clear where we made the mistake in the previous sections. We assumed that the properties of the particles existed, even if they hadn't been measured. Standard quantum mechanics does not make this assumption, which is why the conclusion does not hold.

6 Final Comments

In the first part of this, we talked about the possibility that classical ideas about reality could yet return to our explanations. We saw an explicit example of this in de Brogle-Bohm theory.

But now we can conclude that one of our classical beliefs - that objects have definite properties - *must* be wrong if we assume locality, and that the experimental results described above are true. This is essentially Bell's thereom for this experiment.

And then we saw that the experimental results are a prediction of quantum mechanics. If we accept just the *predictions* of quantum mechanics (and not even necessarily the content), we must rule out our classical ideas entirely. There can be no instruction sets. Objects cannot have definite properties. They cannot be spoken of.

Of course, such an experiment has never been carried out. But similar Bell 'inequalities' have been derived for other experiments, more typically with photons instead of electrons, and polarisation instead of spin. These sort of experiments were famously first carried out by [1], and many more have been done since. All so far have agreed with quantum mechanics. The inequality that any hidden-variable theory must satisfy has been violated time and time again, contradicting any possible existence of local, hidden variables.

A Spin eigen-states

The Pauli spin operators have eigen-states. We can represent them in matrix form. We'll use the notation $\uparrow_x$ to indicate a spin-up state along the x-axis, $\downarrow_z$ to indicate spin down along the z-axis, etc.

Here is a representation of the eigen-states:

$$|\uparrow_x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |\downarrow_x\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{27}$$

$$|\uparrow_y\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \quad |\downarrow_y\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix}, \tag{28}$$

$$|\uparrow_z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |\downarrow_z\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{29}$$

Now we can write $|\uparrow_z\rangle$ as a combination of two other spin states,

$$|\uparrow_z\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2}\left[\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} c \\ -1 \end{pmatrix}\right] = \frac{1}{\sqrt{2}}\Big(|\uparrow_x\rangle + |\downarrow_x\rangle\Big). \tag{30}$$

In a similar fashion,

$$|\uparrow_z\rangle = \frac{1}{\sqrt{2}}\Big(|\uparrow_x\rangle - |\downarrow_x\rangle\Big). \tag{31}$$

Now we can evaluate $\sigma_x |\uparrow_z\rangle$.

$$\sigma_x |\uparrow_z\rangle = \frac{\sigma_x}{\sqrt{2}}\Big(|\uparrow_x\rangle + |\downarrow_x\rangle\Big) = \Big(|\uparrow_x\rangle - |\downarrow_x\rangle\Big) = |\downarrow_z\rangle. \tag{32}$$

The first term just gives the eigen-state back, eigen-value $+1$, while the second term gives the eigen-state back multiplied by eigen-value -1.

Similar results can be obtained for acting on the spin-down state along the z-axis, and for σ_y.

References

[1] A. Aspect, P. Grangier, and G. Roger. Experimental realization of einstein-podolsky-rosen-bohm gedankenexperiment: A new violation of bell's inequalities. *Phys. Rev. Lett.*, 49(2):91–94, Jul 1982.

[2] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.

[3] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777–780, May 1935.

[4] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going beyond bell's theorem, 1989.

[5] E. M. Lifshitz and L. D. Landau. *Course of Theoretical Physics : Mechanics (Course of Theoretical Physics)*. Butterworth-Heinemann, January 1982.

[6] N. D. Mermin. Quantum mysteries revisited. *American Journal of Physics*, 58:731–734, Aug. 1990.