ARE THERE HIDDEN VARIABLES?

Wylie Breckenridge

This essay is about the issue of hidden variables in quantum mechanics. First, I will explain what hidden variables are and why they might be thought to exist. Second, I will present one argument for why they do - the EPR argument. Third, I will present a counter-argument against this - Bell's theorem. Finally, I will give careful consideration to what we should conclude from these about the existence of hidden variables.

The Motivation for Hidden Variables

To any physical system qauntum mechanics (QM) assigns a mathematical function called its *state function*. Its physical significance is that for any of the system's measurable quantities, Q, for any possible value of that quantity, q, and for any time, t, the state function gives the probability that a measurement of Q at time t will give the result q. For example, if the system consists of a single electron and Q is the electron's spin in a certain direction, then there are two possible values of Q - 'up' and 'down' - and the state function gives the probability that a measurement of the spin in that direction at a certain time will be up, and the probability that it will be down. In some cases, one of the possible values of Q is assigned probability 1, and all of the others are assigned probability 0. In that case, the state function says that that value of Q *will* be the outcome of the measurement. In other cases, however, the state function assigns positive probabilities to more than one possible value of Q, and so does not say which value of Q *will* be measured, but only which values of Q are *likely* to be measured. It has nothing more to say about the measurement than that.

In cases like the second, in which the state function of a system assigns positive probabilities to more than one outcome of a measurement, nothing that QM tells us about the system is enough to determine the result of the measurement: the QM description of the system is *indeterministic*. If the description is *complete* - if it tells us all that there is to tell about the system - then the system itself must be indeterministic. There cannot be features of the system that determine the result of the measurement, because if there were then the (complete) description would tell us so. In particular, the result of the measurement cannot be a feature of the system before the measurement takes place - it must be brought about by the measurement itself, from a 'fog of possibilities'. Contrapositively, if the system is deterministic then its QM description must be *incomplete* - there must be features of the system (the ones that determine the result of the measurement) that QM fails to describe. There must be, that is, so-called *hidden variables*. If the result of the measurement is a feature of the system before the measurement takes place, then that will do as the hidden variable. If not, then again the result must be brought about by the measurement itself (as it must *always* be if the QM description is complete), but this time in a way that is pre-determined by whatever features they are that are the hidden variables.

So: If QM gives a complete description of every system then some systems are indeterministic. That is, if QM is a complete theory then the world is indeterministic. On the other hand, if every system is deterministic then some systems have hidden variables and their QM description is incomplete. That is, if the world is deterministic then there are some hidden variables and QM is an incomplete theory. For anyone who thinks that the world is deterministic, this is the motivation for thinking that there must be hidden variables.

The EPR Argument for Hidden Variables

Einstein believed that the world is deterministic and that there must therefore be hidden variables. In the 1930s, with B. Podolsky and N. Rosen, he developed the so-called EPR argument, designed to show that there is at least one system whose behaviour can only be explained by positing hidden variables. It goes like this. QM predicts (and experiments have confirmed), that a system consisting of two spin-½ particles (think of them as electrons) can be prepared such that its state function says this: If the spin of particle 2 is measured in a certain direction, without the spin of particle 1 having already been measured in that direction, then the outcome is undetermined - it could be either one of two values. But if the spin of particle 1 has already been measured in that direction, then the outcome *is* determined - it will be the *other* value, the value that was not measured for particle 1. (And similarly with the roles of particles 1 and 2 reversed.) Furthermore, this is true no matter how far apart in space and close together in time the two measurements are made. In particular, it is true even if no causal signal (i.e. one travelling at or below the speed of light) could travel from the first measurement event to the second, so that, according to the locality principle of special relativity, the first event could have no effect on the second.

The EPR argument claims that for this system the outcome of a spin measurement on each particle in any direction must be determined before any measurement (on either particle) is made. Why? Prepare the particles in a lab, then move them sufficiently far apart in space so that no causal signal can travel from the first to the second in less than, say, ten seconds (relative to the lab reference frame). Get person 1 and person 2 together in the lab, synchronise their watches, choose a direction, tell person 1 to measure the spin of particle 1 in that direction at 10:00:05am, tell person 2 to measure the spin of particle 2 in that direction at 10:00:10am (five seconds later), then get them to move to their respective particles (this is done well enough in advance so that they can get to their particles in time to perform their measurements - we can adjust the times if we need to). According to the QM description of the system, as a result of person 1's 10:00:05 measurement, person 2's measurement changes from being undetermined at 10:00:04 to being determined at 10:00:06. If QM is right in claiming that the system itself does not determine the outcome of person 2's measurement at 10:00:04, then the system must physically change, as a result of person 1's measurement, from one that does not determine that outcome to one that does. But that can't be right, because no causal signal can travel between the two particles in that time, so nothing that happens at particle 1 at 10:00:05 can have an effect on what happens at particle 2 at 10:00:06. Thus QM cannot be right, and the outcome of person 2's measurement must have already been determined before person 1's measurement (it was probably determined back in the lab). We can repeat this line of reasoning for any direction in which we choose to measure the spins, and for any particle whose spin we choose to measure first. So the outcome of a spin measurement on each particle in any direction must be determined before any measurement is made on either particle. But the QM description says that it's not. So here is a system that has hidden variables and whose QM description is incomplete. That is the EPR argument.

<u>Bell's Theorem against Hidden Variables</u>

The EPR argument utilises the following property of the two-particle system:

(1) When the particles have their spins measured in the same direction, the results are opposite on every occasion (for one it is up and for the other it is down).

The argument is that this property can only be explained by positing the existence of hidden variables - features of the system that determine, for each of the two particles and for any direction, the result of a spin measurement on that particle in that direction. There is, however, another property that this system has. Instead of considering spin measurements in the same direction, consider spin measurements in any one of three directions equally spaced in the plane perpendicular to the particles's line of flight. QM predicts (and, again, experiments have confirmed) that the system has the following property as well:

(2) When the particles have their spins measured in different directions, the results are opposite on 25 percent of the occasions and the same on 75 percent of the occasions.

In the 1960s, John Bell presented an argument that any hidden variables that might explain property (1) cannot also explain property (2). How does it go? Suppose that the system does have hidden variables that explain property (1). That is, suppose that there are features of the system that determine, for each of the two particles and for each of the three directions, the result of a spin measurement on that particle in that direction, and that explain why the results are always opposite when the particles have their spins measured in the same direction. We don't know what sorts of things the hidden variables might be, nor what sorts of values they might have. But for the purposes of this argument we don't need to - we can assign our own 'values' to them. Let the value of the hidden variables be a pair triples of 'U's and 'D's that represents the spin measurements that they determine. Let the two triples in the pair represent the first and second particles respectively. Let the three positions in each triple represent the first, second and third spin directions respectively. Let the value in the j^{th} position of the k^{th} triple be 'U' if the hidden variables determine that the outcome of a spin measurement in the j^{th} direction on the k^{th} particle will be up, and let it be 'D' if they determine it will be down. Then, for example, ((U,U,D),(U,D,U)) will be the value of any set of hidden variables that determine that the outcome of a spin measurement on particle 1 in direction 1 will be up, that the outcome of a spin measurement on particle 1 in direction 2 will be up, that the outcome of a spin measurement on particle 1 in direction 3 will be down, that the outcome of a spin measurement on particle 2 in direction 1 will be up, and so on.

Property (1) restricts the set of values that the hidden variables can take. The second triple must be the U-D complement of the first (i.e. it must have 'U's exactly where the first has 'D's) - if the hidden variables determine that particle 1 will have spin up in direction 2 then they must determine that particle 2 will have spin down in direction 2, and so on. (As it turns out, then, the example that I gave above is not a possible value for the hidden variables.) There are only eight values that satisfy property (1):

((U,U,U),(D,D,D))
((U,U,D),(D,D,U))
((U,D,U),(D,U,D))
((D,U,U),(U,D,D))
((U,D,D),(D,U,U))
((D,U,D),(U,D,U))
((D,D,U),(U,U,D))
((D,D,D),(U,U,U))

What about satisfying property (2) as well? There are six possible ways of measuring the spins of the two particles in different directions, which we can write as (1,2), (1,3), (2,1), (2,3), (3,1) and (3,2) by taking (j,k) to represent a spin measurement of particle 1 in direction j and of particle 2 in direction k. We can then form the following table:

		Measurements					
		(1,2)	(1,3)	(2,1)	(2,3)	(3,1)	(3,2)
Hidden variables	((U,U,U),(D,D,D))	(U,D)	(U,D)	(U,D)	(U,D)	(U,D)	(U,D)
	((U,U,D),(D,D,U))	(U,D)	(U,U)	(U,D)	(U,U)	(D,D)	(D,D)
	((U,D,U),(D,U,D))	(U,U)	(U,D)	(D,D)	(D,D)	(U,D)	(U,U)
	((D,U,U),(U,D,D))	(D,D)	(D,D)	(U,U)	(U,D)	(U,U)	(U,D)
	((U,D,D),(D,U,U))	(U,U)	(U,U)	(D,D)	(D,U)	(D,D)	(D,U)
	((D,U,D),(U,D,U))	(D,D)	(D,U)	(U,U)	(U,U)	(D,U)	(D,D)
	((D,D,U),(U,U,D))	(D,U)	(D,D)	(D,U)	(D,D)	(U,U)	(U,U)
	((D,D,D),(U,U,U))	(D,U)	(D,U)	(D,U)	(D,U)	(D,U)	(D,U)

Table 1

Table 1 gives us, for each of the eight possible values of the hidden variables and for each of the six possible pairs of spin measurements, the result of those measurements as determined by the hidden variables. For example, the shaded cell tells us that if the hidden variables have value ((U,D,D),(D,U,U)) then if we measure the spin of particle 1 in direction 1 the outcome will be up, and if we measure the spin of particle 2 in direction 3 the outcome will also be up. From Table 1 we can get Table 2 (with punctuation removed to save space):

		Measurements						
		12	13	21	23	31	32	Fraction
Hidden variables	UUUDDD	X	X	X	X	X	X	1
	UUDDDU	X		X				1/3
	UDUDUD		X			X		1/3
	DUUUDD				X		X	1/3
	UDDDUU				X		X	1/3
	DUDUDU		X			X		1/3

	DDUUUD	X		X				1/3
	DDDUUU	X	X	X	X	X	X	1

Table 2

Table 2 tells us, for each of the eight possible values of the hidden variables and for each of the six possible pairs of spin measurements, whether or not the measurement will give *opposite* spins for the two particles (represented by an 'X'). It also tells us (in the last column), for each of the eight possible values of the hidden variables, on what fraction of occasions in which we measure the spins of the particles in different directions in a system whose hidden variables have that value we should expect to obtain opposite spin measurements.

We can read off some important results from Table 2. Suppose that there is only one possible value that the hidden variables can have. If it's the value in the first or last row then we should expect that when the particles have their spins measured in different directions the results will be opposite on every occasion. If it's one of the values in the middle six rows, then we should expect the results to be opposite on 1/3 of the occasions. Both of these values are greater than the 1/4 given by property (2), so there cannot be just one possible value that the hidden variables can have. Suppose that there are two possible values. If they are the first and last, then we should expect the results to be opposite on every occasion. If they are two of the middle six, then we should expect the results to be opposite on 1/3 of the occasions. If they are either the first or the last and one of the middle six, then no matter how often the hidden variables have one value rather than the other, we should expect the results to be opposite on somewhere between 1/3 of the occasions and all of the occasions (inclusive). But, again, these are all greater than the 1/4 given by property (2), so there cannot be just two possible values that the hidden variables can have. We need not go on. Clearly, no matter which of the eight values are possible values for the hidden variables, and no matter how often each possible value is actually taken by them, we should expect the results to be opposite on *at least* 1/3 of the occasions, which is greater than the 1/4 given by property (2). So any hidden variables that can explain property (1) cannot also explain property (2). That is, there cannot be hidden variables that explain both properties (1) and (2) together. That is Bell's theorem.

So are there Hidden Variables?

The EPR argument claims that the two particle system must have hidden variables, because if it didn't it would violate the principle of locality. Bell's theorem claims that it can't have hidden variables, because if it did we wouldn't observe the correlations between spin measurements that we do. So does it have hidden variables or not?

We need to think carefully about this question. Bell's theorem seems to be a valid argument. But a valid argument does nothing more than show that its premises are incompatible with the denial of its conclusion. It might be possible to reject the conclusion by rejecting one of the premises. Can proponents of the EPR argument do this?

It will be easier to think clearly about this question if we switch to a system that is simpler and more familiar (at least in its description, if not its behaviour), but which none the less retains all of its important features:

> <u>Three cups</u>
> There are three cups upside-down on a table. Three-cup theory predicts (and it has been experimentally confirmed) that every time two cups are turned over, a black stone will be found under one and a white stone will be found under the other.

The analogy between this and the two-particle system should be clear: The three cups correspond to the three spin directions; turning two cups over corresponds to measuring the spin of the particles in different directions; three-cup theory corresponds to QM; the prediction of three-cup theory that we always find different coloured stones under the cups corresponds to the prediction of QM that on one in four occasions (in the long run) we will measure opposite spins for the two particles.

Let's see if we can we develop a particular sort of 'hidden variable' theory for the three-cup system - one that explains the prediction of three-cup theory in terms of there being a distribution of stones under the cups prior to two of them being turned over. Any success that we have might help the proponent of hidden variables in the case of the two-particle system.

If three-cup theory had predicted that we will always find a black stone under each of the two cups then this would have been easy - our hidden variable theory could have just said that prior to the cups being turned over there is a black stone under all three of them. But for what three-cup theory *actually* predicts it is much harder. In fact, it seems to be impossible. Why? Let's introduce some notation. Let's represent each possible distribution of stones as a triple (c_1,c_2,c_3) where c_j is "B" if the stone under cup j is black, and "W" if it is white. Thus (B,B,W) represents the distribution of stones in which a black stone is under the first cup, a black stone is under the second cup, and a white stone is under the third cup. There are eight possible distributions:

(B,B,B)
(B,B,W)
(B,W,B)
(W,B,B)
(B,W,W)
(W,B,W)
(W,W,B)
(W,W,W)

If the stones are distributed in the first or last way, then there is a chance that the cups we choose will reveal stones of the same colour (a very good chance, in fact). If the stones are distributed in one of the middle six ways, then there is also a chance that the cups we choose will reveal stones of the same colour (a one-in-three chance). So no matter how the stones are distributed, there is a chance that stones of the same colour will be revealed. But this disagrees with three-cup theory, which predicts that there is *no* chance that stones of the same colour will be revealed. So it is impossible to explain the prediction of three-cup theory with such a hidden variable theory. This is the position that Bell's theorem has left us in for the two-particle system - that it is impossible to explain properties (1) and (2) predicted by QM in terms of features of the system that exist prior to the spin measurements being made. So this approach is no help to proponents of the EPR argument.

But this is not the only type of hidden variable theory that we can try. We can try a *contextual* one, in which we claim that the choice of cups has an effect on how the stones are distributed before the cups are turned over - in which, that is, the values of the hidden variables depend upon the 'context' of the measurement. We can try a theory like this: At some point in time, a decision will be made about which two cups to turn over (it may happen well in advance of them being turned over, or it may happen just moments before, but it will happen). Once made, the decision causes stones to be arranged under the cups such that when the chosen two are turned over there will be a black stone under one and a white stone under the other. Such a theory is easy to find. Suppose cups j and k are chosen, where $j, k \in \{1,2,3\}$ and $j < k$. Put a black stone under cup j, a white stone under cup k, and nothing under the third cup. Then we have a hidden variable theory that agrees with three-cup theory in predicting that we will always find stones of different colours under the cups that we choose. It might be objected that it also predicts that we will always find a black stone under the left of the two cups and a white stone under the right one (assuming that the cups have been numbered from left to right), whereas three-cup theory makes no such prediction. I'm not sure that this is a valid objection, but we can easily respond to it. There are deterministic random number generators that produce rational numbers between 0 and 1 (inclusive). Instead of always putting a black stone under cup j and a white stone under cup k, only do so if the random number generator gives a number less than 1/2; otherwise, put a white stone under cup j and a black stone under cup k. Then there will be no particular pattern to which colour stone we find under which cup, just the fact that they will always be different colours.

We can develop a contextual hidden variable theory even if the context is just which cup is turned over first (that is, even if all that is known before the stones are distributed is which cup will be turned over first, but not which cup will be turned over second). Suppose cup j will be turned over first. Put a black stone under cup j and a white stone under the other two. Then no matter which cup is turned over second, we will always find stones of different colours. Again, we can remove any other patterns by using the random number generator: do it the way described if the number is less than 1/2, otherwise put a white stone under cup j and a black stone under the other two.

Can we mimic this success in the case of the two-particle system? That is, can we develop a contextual hidden variable theory for it as well? Yes, easily. Let's develop one where the context is the direction in which it is chosen to measure the spin on particle 1. For every fourth occasion on which direction 1 is chosen set the value of the variables to ((U,#,#),(D,D,D)), and set it to ((U,#,#),(D,U,U)) on every other occasion. (Here '#' stands for either 'U' or 'D', to be chosen arbitrarily and fixed. I've used # rather than make that choice, to make the first 'U' stand out - it's the important one). Similarly, for every fourth occasion on which direction 2 is chosen set the value to ((#,U,#),(D,D,D)), and set it to ((#,U,#),(U,D,U)) on every other occasion. And for every fourth occasion on which direction 3 is chosen set it to ((#,#,U),(D,D,D)), and to ((#,#,U),(U,U,D)) on every other occasion. This theory will predict both properties (1) and (2): property (1) because no matter which direction is chosen for particle 1, if particle 2 is measured in the same direction then it will always produce the opposite result; property (2) because in the long run exactly one in four values of the hidden variables will produce opposite results when the measurements are made in different directions. Again it may be objected that it also predicts patterns in the measurements that we don't actually observe. For example, it predicts that whenever the directions are the same, particle 1 will be measured as having spin up and particle 2 will be measured as having spin down. This is a more decisive objection than it was for the three-cup system. QM predicts that we don't know the result of the spin measurement on particle 1 (and therefore on particle 2) until the measurement has been performed. But our hidden variable theory predicts that we do - we know that particle 1 will be measured as having spin up on *every* occasion, and if we keep track of how many times each direction has been chosen for particle 1 then we know what the spin measurement on particle 2 will be *every* time. No problem, we can bring in our random number generator. Suppose direction 1 is chosen for particle 1. Get the generator to produce a number. If it is between 0 and 1/8, set the value of the hidden variables to ((U,#,#),(D,D,D)); if it is between 1/8 and 1/4, set it to ((D,#,#),(U,U,U)); if it is between 1/4 and 5/8, set it to ((U,#,#),(D,U,U)); and if it is between 5/8 and 1, set it to ((D,#,#),(U,D,D)) (with appropriate choices about what to do if it is exactly 0, 1/8, 1/4, 5/8 or 1). Then each time we measure the spin of particle 1 in direction 1 we don't know what result we will obtain. But we do know that if we then measure the spin of particle 2 in the same direction it will be opposite. Nor do we know what the result will be if we then measure the spin of particle 2 in a different direction. But we do know that on one in four occasions in the long run we will get the opposite result, and on three in four occasions we will get the same result. That's just what QM predicts. Do the same for the other two directions, and we have a hidden variable theory that works.

Or do we? These contextual hidden variable theories rely on the decision about which measurement to perform being able to have an effect on the hidden variables, which in turn are able to have an effect on the outcome of the measurement. That is, they rely on the choice of measurement being able to have an effect on the outcome of the measurement. This may well be so in the case of the three-cup system, where the cups and the person turning them over are all close enough together in space so that once the person decides which cups to turn over a causal signal can travel to the hidden variables (the stones) and arrive well ahead of the cups actually being turned over. And it may well be so when spin measurements are performed on the two particles in the way that I described earlier. There I imagined that the choice of direction is made in the lab, and carried to the particles in the heads of the two measurers. That allows plenty of time for a causal signal to travel from the decision to hidden variables at each particle, well before any measurements are made. But we can imagine a situation where this is not so. Suppose that person 1 does not choose the direction in which to measure the spin on particle 1 until 10:00:04 (one second before his measurement). Then there is not enough time for a causal signal to travel from person 1 (in the vicinity of particle 1) to the hidden variables that determine the outcome of a spin measurement on particle 2, and then to person 2's measurement (which occurs at 10:00:10 - six seconds after the decision). So our contextual hidden variable theory will not work in this case. Nor will *any* contextual hidden variable theory that relies on person 1's choice of direction as setting the context of the measurements. Nor will any contextual hidden variable theory that relies on *person 2's* choice of direction as setting the context. Locality rules them all out.

The EPR argument relies on the fact that property (1) of the two-particle system holds no matter how the two particles are separated. In contrast, Bell's theorem seems to rely only on a mathematical inconsistency between properties (1) and (2), and not on the fact that they both hold no matter how the particles are separated. So it seems that its conclusion - that the behaviour of the system cannot be explained by hidden variables - follows whether or not locality holds. But our recent discussion shows that this is wrong. Without locality it *is* possible to find a hidden variable theory - a contextual one. So it turns out that Bell's theorem *does* have a premise that proponents of the EPR argument can reject and thereby reject its conclusion - the premise of locality. But locality is a premise of the EPR argument as well. And a crucial premise: without locality, the problem that the EPR argument preys upon disappears. There is no need for hidden variables to explain property (1). It can just be said that the measurement on particle 1 has an instantaneous effect on the outcome of a measurement on particle 2, no matter how far apart they are in space. Without locality, hidden variables are not needed, and the EPR argument breaks down.

So does the two-particle system have hidden variables or not? If we accept locality, then it's not clear what we should say: the EPR argument shows that locality is a problem for anyone who thinks that it *doesn't*; Bell's theorem shows that locality is a problem for anyone who thinks that it *does*. If we reject locality, then it's still not clear what we should say: the EPR argument fails to show that it must; Bell's theorem fails to show that it can't. Either way, neither the EPR argument nor Bell's theorem gives us a decisive answer to the question, "Are there hidden variables"?

Bibliography

I have not referred to them in the essay, but these are the works that helped me to understand the ideas involved:

Lockwood, M.
[1989] *Mind, brain, and the quantum: the compound 'I'*, Blackwell. Chapter 12.

Mermin, N.D.
[1985] “Is the moon there when nobody looks”, *Physics Today*, **38:4**, pp. 38-47; reprinted in Boyd, Gasper and Trout, eds., *The Philosophy of Science* (MIT), Chapter 27.

Price, H.
[1996] *Time's Arrow and Archimedes' Point*, Oxford. Especially Chapter 8.

Sklar, L.
[1992] *The Philosophy of Physics*, Oxford. Chapter 4.