Measurement, EPR, Hidden Variables and Bell's Inequality

Last Update: 16th June 2008

1. The Measurement Problem

The measurement problem is this: the "reduction of the wavepacket", denoted by the symbol $\Re$ in Part1 of these notes (QM1), does not appear to have a physical basis compatible with the manner in which quantum systems evolve at times other than those times denoted "measurements". In QM1 we have seen that quantum systems evolve in time according to,

$$|\psi(t)\rangle = \hat{U}(t)|\psi(0)\rangle \qquad (2.1)$$

where $\hat{U}$ is a Unitary operator determined by the Hamiltonian. For obvious reasons this is called unitary evolution. Unitary evolution is necessary to ensure "conservation of probability", i.e. that the state remains normalised. But the "reduction of the wavepacket", $\Re$, cannot be represented by any operator in Hilbert space – for one thing, its outcome is indeterminate. So $\Re$, and hence "measurement", does not represent unitary evolution.

And yet a measuring device is just part of the physical world. The combined "system + apparatus" is just another physical system, which, if quantum mechanics is correct, must be describable as a quantum system. This means that the "system + apparatus" must evolve unitarily, in accord with Equ.(2.1) where $\hat{U}(t) = \exp\left\{-\frac{i}{\hbar}\hat{H}t\right\}$ and $\hat{H}$ is the total Hamiltonian of the "system + apparatus". So we have a stark contradiction, a paradox. This is the measurement problem.

The Copenhagen Interpretation holds that the existence of devices behaving in a classical, i.e. non-quantum, manner, is essential to measurements being performed. I have never understood how this is supposed to solve the problem, since classical physics is not compatible with $\Re$ either, since classical physics is deterministic. Also, I suspect that Bohr himself did not accept that there was, at any level, a distinction between the quantum and the classical. He would, I think, have accepted that the "system + apparatus" should be treatable, in principle, as a quantum system. As far as I can see, the Copenhagen Interpretation is what is now referred to as the "shut-up-and-calculate" interpretation. In other words, do not enquire further.

The attempt to reconcile the collapse of the wavepacket with the physical nature of the measurement hardware has often lead people to push $\Re$ off to the point at which a measurement outcome is perceived by the human mind. Thus, Wigner (1963) was led to attribute $\Re$ to the operation of human consciousness. Penrose (1989, 1994) also believes that resolving the measurement problem is related to the problem of consciousness, but in a different sense. He believes that both problems have a physical resolution, and that the physics of the two are probably related, but not that they are the same problem.

The Many Worlds interpretation holds that the wavepacket does not actually collapse. There is no $\Re$. Actually all the possible outcomes are realised. The different outcomes are, however, realised in different worlds (or universes). Every measurement causes the world to split into N worlds, where N is the number of non-zero terms in the spectral representation of the observable-operator. Deutsch has claimed to have derived the Born Rule from the Many Worlds perspective (2008, has this been published?), as has Turok (2008). Reactions to the Many Worlds interpretation are polarised. Having been virtually crackpot territory in the 1960's, it has now become one of the mainstream interpretations preferred by many working physicists (especially, it appears, those working in quantum computing). However, others tend to be revolted by the thought of such a dizzyingly vast exuberance of universes popping up all the time. It would seem to be a staggeringly extreme violation of Occam's Razor. Again, others disagree, interpreting Occam's Razor as requiring the minimum number of principles, not the minimum number of physical objects, and hence not being inconsistent with a vast multitude of universes.

One of the difficulties I have with the Many Worlds interpretation is exactly when world-splitting takes place. Since the intention is to avoid the non-unitary $\Re$-process, it presumably happens whenever the $\Re$-process would happen in, say, the Copenhagen interpretation. But that's no good as a definition, is it? In the Many Worlds interpretation there *is* no $\Re$-process. It's no good appealing to some other interpretation, which one is trying to refute, to define the occurrence of the key attribute of the alternative description. Strictly within the Many Worlds interpretation, what is the definition of a World-splitting event? There does not appear to be one. In modern incarnations it seems to be aligned with decoherence. But my understanding is that 'decoherence' is a particular mechanism postulated to explain the $\Re$-process as a real physical process. So once again it seems to me that a competing interpretation is being plundered in order to define the central feature of the Many Worlds interpretation. But my understanding of these things is poor. For a recent review of the status of the measurement problem see Wallace (2007).

2. The EPR Paradox

Einstein famously did not like the indeterminacy of quantum mechanics. He felt the theory was incomplete. The paper by Einstein, Podolski and Rosen (1935) purported to show that either quantum mechanics was incomplete or that relativistic causality was violated. The essence of the argument, though not expressed in this way by EPR, is captured as follows.

Suppose a spinless particle decays into a pair of spin ½ particles, which then move away from each other with great speed (we may suppose their masses add to significantly less than the mass of the original particle, and so have large kinetic energy). Suppose also that due to some selection rule, or by some other means, we know that the spin ½ particles are created in an S-wave state, i.e., without orbital angular momentum. It follows that the particles' spin state must be the singlet state of spin zero, which we can write as $\frac{1}{\sqrt{2}}\left(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2\right)$. This means that, as we would have expected, if we measure particle 1 to be "spin up" then we will certainly measure particle 2 to be "spin down", and vice-versa. But we have no way to tell in advance of an individual measurement which of these two outcomes will be found.

So far there is no problem apparent. The difficulty occurs when we note that the measurements on the two particles can be arranged to be at a spacelike separation, so that no causal connection between them is possible. But we are to believe that before the measurement on particle 1, the spin state of particle 2 could be up or down with equal probability. And yet, after the measurement on particle 1 has been carried out, and this could also result in either an up or down spin, the outcome of a measurement on particle 2 becomes determined despite no causal connection between them being possible. EPR argued that either causality is violated or quantum mechanics must be incomplete. In other words, assuming causality is sound, the state $\frac{1}{\sqrt{2}}\left(|\uparrow\rangle_1|\downarrow\rangle_2 - |\downarrow\rangle_1|\uparrow\rangle_2\right)$ plus the $\Re$-process and the Born Rule interpretation cannot provide a complete description of the system in question. The implication is that there must be some “hidden variable” carried by each particle which determines the spin that will be registered for each particle separately.

Experiments do bear out that the particles behave as anticipated by quantum mechanics. This is sometimes referred to as “spooky action-at-a-distance” (spukhafte Fernwirkungen). But it is rather more subtle than that. It is not in the least spooky that the spins of the two particles are opposite. The same would be found for macroscopic particles that originated from a spinless precursor. That is merely the conservation of angular momentum. The spookiness lies in the fact that, before measuring on particle 2, its spin is supposed to be indeterminate – and yet it nevertheless contrives to always be opposite to that of particle 1. The spookiness is that the two $\Re$-processes, which are supposed to be indeterminate, are nevertheless perfectly correlated without the benefit of any causal connection. On the face of it, one must side with EPR and conclude that these $\Re$-processes are not indeterminate at all, but fully determinate or they could not possibly be so correlated. In other words, there must be some “hidden variables” which would render quantum mechanics complete and the $\Re$-process determinate and physical.

3. Hidden Variables – The Algebraic Years

The counter to EPR from the likes of Bohr and Born was that, in fact, there *is* no violation of causality and the quantum mechanics is unobjectionable. The reason is that the correlation between measurement outcomes at the two spacelike separated events cannot be used to send any information. So nothing acausal happens. If one were able to influence the outcome of the measurement of particle 1, then this *would* constitute an acausal connection with particle 2, since it would provide a means of faster-than-light communication. But the point is that the very indeterminacy of the $\Re$-process at particle 1 prevents any such influence over the outcome of the measurement on particle 1.

If you are not comfortable with this counter-argument, you are in good company. The EPR paradox has been the subject of intense debate for over 70 years now. However, the existence of “hidden variables” is even less credible as the resolution of the paradox now than in 1935.

For several decades after the dawn of quantum theory, the Copenhagen interpretation was dominant. In part this was due to the patriarchal influence of Bohr. But the Copenhagen interpretation was also buttressed by von Neumann’s purported proof that hidden variables could not exist. This proof was published in 1932, before the EPR paper, so either EPR were unimpressed by it, or else they were unaware of it. The latter seems unlikely. The ‘proof’ is fatally flawed, as pointed out by Bell (1966), as well as by other people much earlier. When I was learning quantum mechanics in the 1970s there were still people who used von Neumann’s book as a text, and most people still believed his ‘proof’ to be valid. It is therefore worth examining what went wrong with it. Rather than reproducing von Neumann’s original ‘proof’ we shall use the much simpler argument of Bell (1966) based on the same false premise.

The false premise is that the expectation value of a linear combination of observables equals the same linear combination of their individual expectation values. Expressed algebraically, $\langle aP + bQ \rangle = a\langle P \rangle + b\langle Q \rangle$. In other words, “expectation value” is a homomorphism. Quantum mechanics has this property, of course, because it is simply a re-writing of the linearity of the Hilbert space operators. But von Neumann imposed this condition on the hypothetical hidden variable theories as well. Now it is easy to show via a counter-example that this eliminates deterministic theories. Consider a spin ½ particle and the observable $\hat{Q} = \overline{\beta} \cdot \overline{\sigma}$, where $\overline{\sigma}$ are the Pauli matrices and $\overline{\beta}$ is an arbitrary real 3-vector. Upon measurement, this observable[1] can only take the value $|\overline{\beta}|$ or $-|\overline{\beta}|$, where $|\overline{\beta}| = \sqrt{\beta_x^2 + \beta_y^2 + \beta_z^2}$. For a given fully specified state, if a deterministic theory existed then the expectation value would either be $|\overline{\beta}|$ or $-|\overline{\beta}|$ depending on the state, since one or other of these results would be definite. But this contradicts the homomorphism requirement, since this requires that the expectation value be linear in each component of $\overline{\beta}$, namely $\langle \hat{Q} \rangle = \overline{\beta} \cdot \langle \overline{\sigma} \rangle$. von Neumann concluded that deterministic states, i.e. hidden variables, could not exist. This was an unjustified claim.

Actually what von Neumann demonstrated was that, if deterministic, hidden variable, theories existed, then their expectation values for fully defined states could not respect the homomorphism property. Bell (1966) demonstrated that there was no call to assume that deterministic states with specified hidden variables would obey the homomorphism property. He constructed a simple deterministic model in terms of a hidden variable λ

[1] This observable must be understood to mean, “measure the spin in direction $\hat{\beta}$”. If you insist on expanding $\overline{\beta} \cdot \overline{\sigma}$ in terms of a sum over its x, y and z components then you will have trouble seeing what the possible measurement outcomes might be. This is because the eigenvectors of the sum are not eigenvectors of any of the three terms (in general), because the three terms do not commute. Actually, this is precisely the point being made.

which always produced one of the outcomes $|\overline{\beta}|$ or $-|\overline{\beta}|$ on measurement, and hence clearly violated the homomorphism condition as regards expectation values for fully specified states, i.e. for a given value of λ. But, on averaging over all possible values for λ, results for the expectation values in agreement with quantum mechanics were obtained. Thus, the homomorphism property holds once the hidden variables have been averaged-out.

If we write the quantum state as $|\psi\rangle$, and the deterministic state in the hidden variable theory as $|\psi,\lambda\rangle$, then the matter can be summarised as follows,

Quantum mechanics: $\langle\psi|\overline{\beta}\cdot\overline{\sigma}|\psi\rangle = \overline{\beta}\cdot\overline{s}$, where $\overline{s}$ is the real vector $\overline{s} = \langle\psi|\overline{\sigma}|\psi\rangle$ (2.3.1)

Hidden variables: $\langle\psi,\lambda|\overline{\beta}\cdot\overline{\sigma}|\psi,\lambda\rangle = |\overline{\beta}|$ or $-|\overline{\beta}|$ depending on λ (2.3.2)

but, $\int\langle\psi,\lambda|\overline{\beta}\cdot\overline{\sigma}|\psi,\lambda\rangle d\lambda = \overline{\beta}\cdot\overline{s}$ (2.3.3)

and hence the hidden variable theory agrees with quantum mechanics, but has deterministic outcomes for measurements. [NB: It may look odd that the RHS of (2.3.3) is in terms of the spin direction, $\overline{s}$, because this information seems to have been lost on the RHS of (2.3.2). However, $\overline{s}$ is actually codified in the choice of sign on the RHS of (2.3.2)].

Bell (1966) also discusses alternative algebraic ‘proofs’ that hidden variable theories cannot exist, due to Jauch & Piron (1963) and a corollary of a result by Gleason (1957). He shows that both of these suffer from a similar defect. Innocent looking algebraic conditions are assumed to hold for the candidate, deterministic, hidden variable theories which are unduly restrictive. These conditions hold in quantum theory, and hence are required to hold for averages over the hidden variables, but there is no reason to assume they hold for fully specified deterministic states (i.e. for specified λ). On the contrary, there are physical reasons why this should not be expected, as Bell discusses.

In 1967, Kochen & Specker proved a truly remarkable theorem. Assume that observables can be represented by self-adjoint operators in Hilbert space. Assume that a unique number, the ‘possessed value’, can be assigned to every observable, Q. Call it Value(Q). So this is effectively saying that there is a deterministic underlying theory. And finally assume the homomorphism property holds between possessed values, i.e., obtaining the “Value” commutes with functions, f, i.e., $\mathrm{Value}(f(Q)) = f(\mathrm{Value}(Q))$. Kochen & Specker then showed that, for Hilbert spaces of dimension greater than two, these assumptions would result in a contradiction. In other words, it is not possible for self-adjoint operators on Hilbert space to be assigned unique numerical values which also obey the homomorphism property. I interpret this to be a confirmation of Bell’s position that it is the assumptiom of the homomorphism property which would kill off the possibility of hidden variables. But hidden variable theories need not respect this homomorphism. For the proof of Kochen & Specker see the original paper or Redhead (1987).

4. Hidden Variables – Bell's Inequality

A great deal has been said on the subject of hidden variables since Bell's 1966 paper, and I expect there are still competing points of view. But, for my part, I now regard all purely algebraic proofs of "no hidden variables" as dead. Ironically, it was the man who threw the counter-punch, Bell, who also delivered the killer blow to hidden variables. Unlike previous attempts, Bell left the decision in the hands of the experimentalists. In 1964 Bell derived a remarkable inequality between expectation values which, he claimed, must be obeyed by any local, realistic, hidden variable theory, i.e., loosely speaking, any classical deterministic theory. The crucial point is that quantum mechanical expectation values do *not* obey Bell's inequality. Thus, the issue of hidden variables became a truly scientific one – that is, a question which could be decided by experiment. Would experiments respect Bell's inequality? Many experiments of increasing precision have now been conducted (References to be added). In all cases, to my knowledge, the results are consistent with quantum mechanics and violate Bell's inequality. It is now generally accepted that experiment has ruled in favour of quantum mechanics and against local, realistic, deterministic theories. Thus, algebra and experiment between them have killed hidden variable theories[2].

Since Bell's 1964 paper there have been many stronger versions of such inequalities published. However, the original version suffices for illustrative purposes. Bell considers the EPR situation, re-cast in modern spin terminology, as is now usual. Thus a pair of spin ½ particles emerge from the decay a spinless precursor, and hence in the singlet state. Bell envisages the spin of one particle being measured in a direction given by unit vector $\hat{a}$, and the other spin being measured in direction $\hat{b}$. These vectors can be oriented arbitrarily in 3D space. At issue is the correlation between the two spin measurements, for which it suffices to consider the expectation value of their product. The quantum mechanical expectation value of the product is,

$$\left\langle (\bar{\sigma}_1 \cdot \hat{a})(\bar{\sigma}_2 \cdot \hat{b}) \right\rangle = -\hat{a} \cdot \hat{b} = -\cos\theta_{ab} \qquad (2.4.1)$$

In the particular case that $\hat{a} = \hat{b}$ this produces an expectation value for the product of -1, i.e. the spins are always opposed.

Bell argues that if the outcome of an individual spin measurement is determined by a hidden variable λ, then the first particle could be predict with certainty to have a spin of $A(\hat{a}, \lambda) = \pm 1$ if λ were known. Similarly, the second particle also has determinate spin, $B(\hat{b}, \lambda) = \pm 1$ (NB: we are measuring spins in units of $\hbar / 2$). The expectation value of the

[2] No doubt some people would still disagree. When I last looked at the experimental status, the objectors were claiming that measurements were still being conducted at timelike separations, i.e. that there could in principle be a physical influence between the two EPR particles.

product of spins, averaged over many measurements, is thus, according to the hidden variable theory,

$$E(\hat{a},\hat{b}) = \int A(\hat{a},\lambda)B(\hat{b},\lambda)d\lambda \qquad (2.4.2)$$

Bell shows that (2.4.2) is inconsistent with (2.4.1). More generally, he derives an inequality which must be respected by any such hidden variable theory, i.e.,

$$1 + E(\hat{b},\hat{c}) \geq \left| E(\hat{a},\hat{b}) - E(\hat{a},\hat{c}) \right| \qquad (2.4.3)$$

But the quantum mechanical expectation value, (2.4.1), does not obey (2.4.3). For example, consider $\hat{a}$ and $\hat{c}$ to be perpendicular with $\hat{b}$ at 45^o to both. Then the inequality would require,

$$1 - \frac{1}{\sqrt{2}} \geq \left| -\frac{1}{\sqrt{2}} - 0 \right| = \frac{1}{\sqrt{2}}$$

But the LHS = 0.292 whereas the RHS = 0.707, so the inequality is clearly false. Note that it is not even a close miss. The quantum expectation value disrespects the Bell inequality quite radically. This is important because it means that experiments to discriminate between the two need not necessarily be of very great precision.

As mentioned above, to date all experiments have been consistent with quantum theory and demonstrate violation of the Bell inequality (or some generalisation thereof), see (References to be added).