The Quantum Measurement Paradox: The "Orthodox" Solution

In 1974, the physicist-philosopher Bernard d'Espagnat introduced a discussion of the paradox with the following statement: "The problem of measurement in QM is considered as non-existent or trivial by an impressive body of theoretical physicists and as presenting almost insurmountable difficulties by a somewhat lesser but steadily growing number of their colleagues." If anything, this trend has continued over the last four decades.

The name "quantum measurement problem" ("paradox") is actually something of a misnomer: the problem is basically about how the *realization* (actualization) of a particular state of affairs comes about in QM, and this is much more general than those situations we would naturally describe as "measurements". Nevertheless, the measurement situation is a particularly clear-cut example of the problem, and it has become traditional to discuss it in these terms.

What, exactly, do we mean by "measurement"? In classical physics, the concept can be applied at various different levels. E.g., one talks about measuring the temperature of a bowl of water, the mass of a C atom or the energy of a muon (mu-meson) in a bubble chamber, and a little thought shows that the implications, and perhaps even the meaning of the word, are somewhat different in the three cases (e.g., the mass of a C atom is normally assumed to be an inherent and inalienable property of the atom merely by virtue of its being carbon, while both the temperature of the water and the energy of the muon are characteristics only of that particular water or muon at that particular time). Generally speaking, the sense of "measurement" in which we shall be interested in the present context is closest to that of the muon: we want to ascertain the value of some characteristic (energy, spin orientation, polarization...) of an individual microscopic particle (muon, electron, photon...) that is not inherent to it. If we are going to discuss the problem in terms of QM, then, since the latter is (in most interpretations) a description only of ensembles, the microscopic system in question must have been drawn from some large ensemble whose initial preparation is specified (e.g., the ensemble of photons emitted in the decay of H atoms excited according to a particular prescription). In such a case, we would, of course, normally measure the quantity of interest for a large number of individual microsystems and plot the results in the form of a histogram (distribution), which could then be compared with the statistical predictions of QM. However, it is crucial that while we obtain this information on the ensemble as a whole, we get a great deal more than this: we get a value of the measured quantity for each individual microsystem we measure. (Of course, it is necessary to define exactly what we mean by the statement that a particular microsystem "is found to have" a given value. But this can usually be done operationally: e.g., in the case of polarization, we set up a polarizer set in the x-direction with a detector behind it, and define a given photon to have polarization in the x-direction if it triggers the detector, y-polarization otherwise.)* It is necessary to emphasize this point, because in the recent literature

*Note, however, that this definition assumes that each photon either does or does not trigger the

the word "(weak) measurement" has on occasion been applied to operations that give information only about, e.g., average values for the ensemble, and not precise values for individual members of it. Although this operation is very useful in certain contexts, I believe that the use of the term "measurement" to describe it may be problematic.

The above discussion implicitly assumes that the ensemble we are interested in is composed of *many physically distinct* entities (generally microscopic). This is certainly true of the classic beam experiments: the atoms of an atomic beam issuing from an oven are indeed physically distinct, as are the photons in a typical polarization measurement. However, in recent years there has appeared the possibility of taking a single physical system such as a *single* atom in a "Penning trap" and repeatedly preparing it according to the same prescription. This is sometimes called a "time ensemble". The measurements on genuinely macroscopic systems such as superconducting devices, which we will have to discuss later, are almost invariably of this type. Orthodox quantum mechanics makes no distinctions between "time ensembles" and the more familiar type, and there is no particular need to discuss the former separately below. For simplicity, I will discuss in the following only measurements of quantities that are found in experiment to take only "discrete" values, such as polarization (the photon either triggers the detector or it does not) or the angular momentum of an atom (which is found to take one of the values $n = 0, 1, 2 \ldots$).[†] It is perfectly possible to discuss also the case of a continuous variable (e.g., position on a screen) but this adds nothing in the present context.

If we are going to able to "read off" the value of the measured quantity (call it A) for individual microsystems of the ensemble, this implies that we have to produce an output of the measuring device that is sufficiently "macroscopic" that it can be seen directly with the naked eye, recorded in a computer, etc.; and, moreover, that this macroscopic output is *different* for each of the possible values of A. For example, in the case of polarization we should get an audible click or visible flash from our detector when it detects a photon. In general, this will require some kind of amplification process (cf. below). Note that it is *not* implied that every different value of A induces a macroscopic reaction in the measuring device: in many cases (including the above one), the setup is such that for one particular value of the quantity A, the device is left in its original state – a so-called "ideal-negative-result" experiment. (In the particular case in question, if the photon is rejected by the polarizer [and we make no arrangement to detect rejected photons], we get *no* response from the detector. Of course, for this to constitute a valid "measurement", we would need an independent means of knowing there was a photon there in the first place.)

Two points about the above characterization of "measurement" should be noted. First, the statement "this particular microsystem had the particular value a_i of the quantity A" can always be *operationally* defined (by the clicking or not of a detector, etc.), and at this stage there is no need to go into the question of whether the microsystem in question actually "had" the value a_i *before* the measurement took place. Secondly, it is emphatically not implied that the measurement is "noninvasive", that is, that it

detector – an apparently innocuous assumption that we shall nevertheless have to reconsider later.

[†]We could define this statement operationally if required (by a so-called Stern-Gerlach experiment).

leaves the measured quantity unchanged (i.e., that for the microsystem in question, A has the value a_i, *after* the measurement). In fact, the majority of real-life measurements are strongly invasive, particularly when they involve photons: in a typical measurement of photon polarization, the photon is actually absorbed in the detector, so that it makes no sense to talk about its polarization *after* the measurement![‡] Since in the general case one is not sure that the "measured" value of a_i actually characterizes the microsystem in question either before or after the measurement, it is clear that the classical notion of measurement has been at least somewhat attenuated.

Let's briefly analyze the ingredients of a typical measurement setup.[§] We first need to prepare our ensemble, and if we are going to try to describe it by QM we had better do so in a way that enables us to assign it a QM state description (wave function).[¶] In many textbooks, one gets the impression that the way we do this is by conducting an initial set of measurements on the systems of the ensemble. In real life, this is practically never true: we almost always assign the QM state description either on the basis of thermodynamic considerations, or as a result of some "filtering" process, or both. For example, my ensemble might be a gas of He atoms emerging from an oven at some temperature of the order of a few hundred degrees. Suppose I am interested in the QM state of the *electrons* in the atoms of such a beam. I know that the energy necessary to excite the lowest excited state of He is several tens of eV (or in temperature units, several hundred thousand degrees) – i.e., much greater than the thermal energy of the oven – and on this basis I conclude that the electrons are overwhelmingly likely to be in the atomic groundstate. I therefore assign to them the QM wave function that I (or rather my predecessors!) have *calculated* for this groundstate. No preliminary "measurement" is involved anywhere in this process (unless one counts the measurement of the temperature of the oven, which is essentially a classical operation). Another way of preparing an ensemble whose QM state description is known is by "filtering": for example, I take an initial beam of photons whose polarization is unknown and make it impinge on a polarizer set in the x-direction. Then I know that the ensemble formed by the part of the beam which is transmitted certainly "has" polarization in the x-direction, and similarly the reflected beam is described by an ensemble polarized in the y-direction, and I can then write down the appropriate QM state. Note that the "filtering" process described is *not* equivalent to a "measurement": while I know that any photon that is transmitted must have x-polarization, etc., I do not know at this stage which of the individual photons have been transmitted and which reflected. At this stage, what we

[‡]In the literature, those relatively rare measurements that have the property of "if a value a_i of A is measured, then the microsystem in question has the value a_i after the measurement" are called "ideal" or "of the first kind"; correspondingly, other types are said to be "non-ideal" or "of the second kind".

[§]The use of the word "typical" to some extent begs the question. The setup that I will describe is probably typical of a number of classic experiments on microscopic entities, such as measurement of the polarization of a photon or the intrinsic angular momentum (spin) of an atom, and it is overwhelmingly on the basis of these that the notion of "quantum measurement" has traditionally been analyzed. However, it is not at all clear that it corresponds to some of the most interesting examples in recent physics, and I will comment on this aspect when relevant.

[¶]Actually, in real-life cases, a more complicated QM description is often necessary, but we shall ignore this complication here.

have accomplished is the *preparation* of a well-defined QM state (i.e., of an ensemble of microsystems possessing this QM state description). From now on, the details of how the preparation was achieved can be forgotten.

Suppose now that we wish to measure on the systems of this ensemble the value of some quantity A. In general, the state in question will not correspond to a unique value of A (for example, in the case of an "x-polarized" ensemble of photons we might decide to measure the x' [45°] polarization). How do we go about doing the measurement? In most cases, the first stage is to separate the systems "having" different values of a_i *spatially*. E.g., as we have already seen, in the case of a measurement of x'-polarization, we put the beam of photons through a polarizer set in the x'-direction; then those with x'-polarization are transmitted and those with y'-polarization reflected. Similarly, the classic measurement of the spin component of an atom (Stem-Gerlach experiment) involves a stage in which the atom is deflected by an amount proportional to the spin component in question. Some people (in particular the de Broglie school) hold that *any* realistic measurement must involve a stage at which the systems to be measured are spatially separated according to the value of A, and that therefore the "position" variable is in some sense more fundamental than any other. However, while this is certainly true for most, if not all, measurements on genuinely microscopic systems (including some, such as, e.g., the measurement of the "strangeness" of a K-meson, where the measured quantity has prima facie no geometrical significance at all), it seems artificial to make this claim with respect to experiments, e.g., on some macroscopic electronic systems, where the prima facie measurement may be of a quantity such as voltage, and any *spatial* separation occurs, if at all, only at a very late stage (e.g., in the experimenter's fingertips!).

We still have not made a "measurement"; all we know is that if (e.g.) a photon has an x'-polarization, it will be in the transmitted beam; if a y'-polarization, in the reflected one (cf. the discussion of "filtering" above). We still need to find out, for a particular photon, whether it is indeed in (say) the transmitted beam. We therefore put, behind the polarizer, a detector of some kind. What this detector has to do is to convert a *microscopic event* – the arrival of a single photon – into a *macroscopic* event, such as an audible click, a visible flash or (more realistically) a current pulse that can trigger a computer memory element. Typical examples of such a detector are a Geiger (or proportional) counter, a photomultiplier tube or the retina of the human eye. Almost inevitably, the working of such a detector has to involve a macroscopic amount of irreversibility.[‖]

After these rather lengthy preliminaries we are in a position to formulate the quantum measurement paradox. We will assume, for the moment, that we take the "orthodox" point of view that QM is in principle a complete description of the behavior not just of single electrons, photons or atoms, not just of complicated biomolecules, but of the whole of the physical Universe. If this is so, then while it may not be *necessary* to give an account of the working of the measurement process in QM terms, it should be at

[‖] But note the "measurement" of the passage of a cosmic ray by the interstitials produced. In this case, the irreversibility only takes place at the "readout" stage.

least *in principle possible* to do so. So how do we do this? In the following, I implicitly assume that all relevant QM descriptions are in terms of wave functions ("pure" QM states). This is not actually technically correct, but the generalization to take account of this does not (at least in my opinion!) in any way affect the force of the argument, and merely leads to unwanted notational complications.

Let's describe the initial ("ready") state of the measuring device (we can include in this, if we wish, not only the detector but the polarizer and any ancillary bits of equipment) schematically by some wave function $|\Psi_0\rangle$, and the final state that indicates that the value a_i of A has been observed by $|\Psi_i\rangle$. (In the case of an "ideal-negative-result" experiment, one (but no more than one) of the $|\Psi_i\rangle$ may be equal to $|\Psi_0\rangle$.) In order that we can unambiguously "read off" the result a_i, the states $|\Psi_i\rangle$ must represent *macroscopically distinct* states of the device (e.g., "click-versus-no-click": this is an essential point!). Now, one thing we know for sure is that if the measuring device is to work as advertised, then whenever the microsystem in question is drawn from an ensemble that possesses a *definite* value a_i of A (i.e., is in an "eigenstate" of A corresponding to eigenvalue a_i) then the device must end up in the state $|\Psi_i\rangle$. Let us denote the QM state (wave function) of such an ensemble by $|\psi_i\rangle$ and the initial state of the "universe" (system + device) by $|\psi_i, \Psi_0\rangle$. Moreover, for notational simplicity (only!) let us assume that the microsystem disappears in the process of measurement (as usually happens for the case of photon polarization; cf. above). Then, schematically, we can represent the measurement process for this case by the notation $|\psi_i, \Psi_0\rangle \rightarrow |\Psi_i\rangle$. In words: "The ensemble of 'universes' (coupled system plus measuring device complexes) for which the initial quantum state of the system is $|\psi_i\rangle$ and that of the device is $|$"ready"$\rangle$ ($|\Psi_0\rangle$) invariably ends up after measurement in the quantum state $|\Psi_i\rangle$ (system disappeared, device in macroscopic state i)." This must be true for each possible value of i. It is quite easy to construct "toy" models of measuring devices that have this property, i.e., the application of the standard rules for the evolution of the (coupled systems + device) wave function (which, recall, is purely deterministic!) yields the transition expressed by the above rule. To be sure, real-life measuring devices are complicated enough that the corresponding analysis would be extremely messy, but virtually no one doubts that in principle it could be carried out. This stage of the argument, then, is not controversial.

Now the crunch: What happens if the initial state ψ of the micro system does *not* correspond to a definite value a_i of the to-be-measured quantity A? In this case, according to the standard prescriptions of the formalism, we can always express the quantum state ψ as the *superposition* of the states $|\psi_i\rangle$ introduced above:

$$|\psi\rangle = \sum_i c_i |\psi_i\rangle$$

where the c_i are some "weighting" coefficients (amplitudes).** For example,in the case of photon polarization we have (cf. appendix of lecture 19):

$$|\psi\rangle = \cos\theta |\uparrow\rangle + \sin\theta |\rightarrow\rangle$$

**The numbers c_i are in general complex, but this is a subtlety that need not concern us here.

The initial state of the "universe", namely $|\psi_i, \Psi_0\rangle$, can then be written in the form $\sum_i c_i|\psi_i, \Psi_0\rangle$, which is just a superposition of the states $|\psi_i, \Psi_0\rangle$ discussed above. But we have already stressed (lecture 17) that the principle of superposition, which in effect says that we can calculate the time development of the different "elements" of a superposition separately and add the results, is absolutely fundamental to QM: to question it would be like questioning the postulate of the invariance of the speed of light in special relativity – we would essentially be talking about a different theory. So we apply this principle and obtain the result:

$$|\psi, \Psi_0\rangle \equiv \sum_i c_i|\psi_i, \Psi_0\rangle \rightarrow \sum_i c_i|\Psi_i\rangle.$$

e.g., for the photon polarization case,

$$|\Psi\rangle = \cos\theta|\text{ click}\rangle + \sin\theta|\text{ no click}\rangle$$

But this is a disaster! Recall that for our measuring device to work as advertised, the states $|\Psi_i\rangle$ had to represent *macroscopically distinct* states of the device (e.g., "click versus no click"). What we have found, then, is that the final state of the ensemble of "universes" whose initial state was the (perfectly ordinary-looking) state $|\psi, \Psi_0\rangle$ is a *superposition of macroscopically distinct states*! And if we interpret the superposition as we did at the microscopic level (more on this below), this means that a definite macroscopic outcome has not been selected – it is impossible to say, for each individual member of the ensemble, either that the counter has clicked or that it has not clicked; both possibilities are still in some sense represented. This seems to run clearly counter to common sense and everyday experience. That, in a nutshell, is the quantum measurement paradox.

A particularly vivid instance of the paradox was given by Schrödinger in a famous 1935 paper. In the example he considered, the system was a single radioactive atom, which may be regarded as a member of the ensemble of such atoms which are known to be undecayed at time zero. Then if we consider the quantum state of the atom at a time t_0, equal to the radioactive half-life, it is a linear superposition of the "undecayed" state $|u\rangle$ and the "decayed" state $|d\rangle$ with equal weights:

$$\psi = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle)$$

The experimental setup is such that if the atom decays, the emitted (α-particle, etc.) triggers a "hellish device", which results in the death of a cat in a closed box; whereas if the atom remains undecayed, the cat stays alive and healthy: i.e., schematically:

$$|u\rangle \rightarrow |\text{cat alive}\rangle, \quad |d\rangle \rightarrow |\text{cat dead}\rangle$$

Then, applying the linear laws of QM, we find that at the time in question the appropriate description of the quantum state of the cat is

$$|\text{cat}\rangle = \frac{1}{\sqrt{2}}(|\text{cat alive}\rangle + |\text{cat dead}\rangle)$$

I.e., according to the usual interpretation, the cat is neither definitely alive nor definitely dead – both possibilities are still represented! Yet it seems commonsensical that if at this stage, we were to take the lid off the box, we should find our cat either definitely alive or definitely dead, but not both.

In a slightly more elaborate version ("Wigner's friend") of the above "Schrödinger's cat" paradox, the cat is replaced by a human observer and death is replaced by something less drastic, say a slight tingling sensation. Then an observer external to the "box" will still have to describe the state of the in-box observer as a superposition of the "tingling" and "not-tingling" states. On the other hand, when this latter observer is released, he will presumably either report that at the time in question he felt the tingling or that he did not!

In the remainder of this lecture I will describe and criticize what I call the "orthodox" solution to the quantum measurement paradox. I call it "orthodox" because it is essentially the resolution that most practicing physicists, on being pressed and after thinking about it for a bit, would probably give (in particular, it is the one that Rohrlich seems to embrace – though he is not very explicit about it, and his wording also has traces of the original Copenhagen point of view). It is rediscovered, and republished as new in the literature, with depressing regularity. In its barest essentials, the orthodox solution consists of two steps. The first consists in the demonstration that, in a "Schrödinger's cat" type of situation, it is in practice impossible to see the effects of interference between the two (or more) macroscopically distinct states represented in the wave function of the "universe". The second step is to argue that "therefore", one may conclude that a particular outcome is realized by this stage: Schrödinger's cat *really is* either alive or dead, as common sense would indicate.

Let's try to make step II of the argument a bit more explicit. It is asked: Why do we have to assume that (e.g.) in a Young's slits experiment, it is incorrect to assume that each individual electron went either through slit 1 or through slit 2? Answer: Because if we do so (and do not make supplementary assumptions which most people would regard as so pathological as to be quite implausible), then it appears impossible to explain the phenomenon of interference (in particular, of total destructive interference). But, if it is indeed impossible to see the effects of interference between macroscopically distinct states, then by the time we get to the macroscopic level this argument no longer "bites" and all experimental (statistical) predictions are exactly those which we would get by assuming (e.g.) that each individual Schrödinger's cat of the ensemble is either dead or alive, with probability 50% each. Consequently, it is argued, it is indeed perfectly legitimate to state that in each experiment a single definite macroscopic outcome is realized.

In assessing the orthodox solution, let's first briefly look at step I. Why is it claimed that it is in practice impossible to see quantum interference between states that are macroscopically distinct? The details of the argument are too technical to go into here, but the essential reason usually given is that macroscopic systems are so messy and complicated, and interact so strongly with their environment, that the relative phase

of the various components is ill-defined:[††] crudely speaking, the + between the states $|\text{cat alive}\rangle$ and $|\text{cat dead}\rangle$ is as likely to be replaced by a −. Under these circumstances, it is easy to show that the constructive and destructive interference tends to cancel, and the statistical predictions for all possible experiments are just as if there were no interference at all, which is to say that they are exactly what one would get by assuming that each individual system of the ensemble realizes one macroscopic state or the other. It is not necessary to go into this question further, because it is almost universally agreed (irrespective of belief in the orthodox solution or not) that step I of the argument is indeed technically valid, in the sense that, at least in a typical "measurement" situation, interference between macroscopically distinct states is "for all practical purposes" indeed unobservable. (Whether the statement is more generally true, and in particular necessarily applies in circumstances very *different* from a typical measurement situation, is something we shall have to return to at the end of the next lecture.)

The general phenomenon invoked above – the effect of the environment, and of the "messiness" of the system etc., in destroying the possibility of interference between macroscopically distinct states – has acquired in the recent literature the name of "decoherence", and become a major area of technical research in its own right.

So far, we have considered only the question of the technical success of the "decoherence" approach, i.e., of stage I of the "orthodox" program. However, we still have to face up to stage II. The conceptual problem is: given that we have established to our own satisfaction that terms indicating interference between different "outcomes" are unobservably small and will remain so, how does it come about that a particular outcome is actually realized? Or as Bell put it, how do we convert an "and" into an (exclusive) "or"? At the microscopic level of electrons, neutrons and photons, the "orthodox" interpretation of QM says that the description of a given ensemble by a linear-superposition wave function, $a|\psi_1\rangle + b|\psi_2\rangle$, does *not* imply that each of the systems of the ensemble is definitely either in state ψ_1 or in state ψ_2, with a fraction a^2 and b^2 in $|\psi_1\rangle, |\psi_2\rangle$, etc., respectively. How do we know that this is not a viable interpretation? Because of the phenomenon of interference between the two possibilities (e.g., in a Young's slits experiment, as reviewed in lecture 16). Indeed, the phenomenon of interference makes it extremely tempting to go further: that is, to make not merely the negative statement that a particular system of the ensemble has not definitely "realized" either of the states ψ_1 and ψ_2, but also the positive statement that for each system there is some kind of "element of reality" associated with each of ψ_1 and ψ_2. However, most interpretations of the QM formalism at the microlevel, and in particular the Copenhagen "interpretation", warn us strongly not to impose any such metaphysical superstructure on the formalism at this level. But even the most ardent adherents of the Copenhagen interpretation would presumably agree with the negative statement that in the situation considered, it is incorrect to say that a particular micro system has definitely realized one of the two alternatives.

[††] A technically more correct statement is that the device does not have its own wave function at all – the only correct description is in terms of the coupled system (device + environment). The consequences are the same as indicated.

Now let us consider what happens when the microscopic difference is amplified to the macro-level, as in the Schrödinger's cat thought experiment or in a typical measurement, so that the state of the "universe" (lower case!) is very schematically of the form

$$\psi = a|\Psi_1\rangle + b|\Psi_2\rangle$$

with $|\Psi_1\rangle, |\Psi_2\rangle$ corresponding to macroscopically distinct states. If we concede the arguments of the orthodox solution at stage I, then it is agreed that there is now no possibility in practice of observing the effects of interference between $|\Psi_1\rangle$ and $|\Psi_2\rangle$. Then, for each individual system, according to stage II of the orthodox solution – lo and behold, a particular one of the macro states is realized!

In my opinion, stage II of the argument embodies a major logical non sequitur, in that it confuses the *meaning* of the QM formalism with the *evidence* that that assignation of meaning is correct. At the micro-level, we assign a meaning to the formalism that definitely denies the realization of a particular alternative; the evidence for this is the (relatively easily observable at this level) phenomenon of interference. By the time we get to the macro-level, the evidence has disappeared: yet the formalism of QM is a seamless whole, and there is no point in the transition from micro- to macro-level where it undergoes any change. Does the vanishing of the evidence then warrant a reinterpretation of the meaning? Consider an analogy with a certain murder trial: let's say that during the trial, a vital piece of evidence that would certainly have led the jury to convict suddenly disappears; does that mean that the accused then immediately becomes *innocent*?