Quantum Mechanics and "Reality"

What do we actually know?

Let's start by reviewing what exactly it is that we know more or less directly from experiment, *without* invoking quantum theory itself. Later, we will investigate which further conclusions we can draw if we assume, with the majority of currently practicing physicists, that quantum theory is in principle a complete description of the world and will continue to work in situations very far from those in which it has up to now been tested.

What we know more or less directly from experiment relates to:

(a) the simple Young's slits experiment and more abstract generalizations of it using microscopic objects (plus, of course, a lot of circumstantial evidence that the QM predictions are working to a very high degree of accuracy quite generally at the atomic level);

(b) the EPR-Bell setup; and

(c) some recent experiments on the QM of a macroscopic variable.

As (c) is a recent development that has failed to make it into most of the textbooks, I postpone it to the end of this lecture.

What we verify more or less directly from experiment in the Young's slits and related experiments can be summarized briefly as follows: there exist situations in which the same final outcome can be obtained by following two (or more) different paths (which need not necessarily be in geometrical space), which are such that it seems very difficult to imagine that the opening or blocking of one path should affect passage down the other. Nevertheless, in some such situations, we find that while "observation" of the path followed always yield a definite ("yes/ no") result for each individual member of the ensemble observed, the statistical distribution of final outcomes for the ensemble under conditions when the path is not observed shows the phenomenon of *interference* ($N_{A+B}(x) \neq N_A(x) + N_B(x)$). Particularly interesting in this context are the "delayed-choice" experiments, in which the decision as to whether to measure interference or rather "which-way" information is taken *after* the particle in question (normally a photon) is forced to "choose" (or not!) between paths A and B.[*]

(Note: The above formulation does in a sense beg some questions, since the concept of a "path" might be argued already to smuggle in the implicit assumption that there is "something there" that follows the path – an assumption that the more rigorous adherents of the CI, at least, would presumably not concede. It would certainly be possible to reformulate the "raw" experimental results in more directly operational terms without introducing the concept of a "path", but the description would be extremely clumsy and the exercise hardly seems worthwhile in the present context.)

---

[*] The most nearly "ideal" experiment of this type is probably that of Jacques et al., *Science* **315**, 966 (2007).

In the case of the so-called "EPR-Bell" (photon-polarization correlation) experiments, what we know directly from experiment (without invoking quantum theory as yet) can be summarized as follows: There exist experimental situations in which the statistical distribution of the actually obtained data on the correlation of polarizations is incompatible with the predictions of any theory of the "objective local" type defined by the conjunction of the following three postulates:

1. microscopic objectivity ("realism") or macroscopic counterfactual definiteness
2. local causality
3. induction

One caveat is in order here: there exist a number of so-called "loopholes" in the inference from the raw data of the experiments to the conclusion that they exclude the class of objective local theories. While each of the loopholes (1-3) below had by itself been excluded in at least one set of experiments, until November 2015 there existed, to my knowledge, no experiment that simultaneously excludes all of them (and simultaneously maintains the condition of spacelike separation between the relevant "events"). The principal "loopholes" identified in the existing literature are the following:

1. Because of various technical problems associated with, e.g., the rather poor efficiency of photon counters in the optical regime, existing experiments do not in fact rigorously exclude OLTs ("objective local types", as defined above) unless a supplementary assumption about the working of the counters is made. If violation of this assumption is explicitly assumed, it is possible to construct examples of OLTs that reproduce the experimental data in the experiments conducted to date; however, most physicists regard such theories as at least as artificial as the "Lorentz-Fitzgerald" attempts to explain the Michelson-Morley experiment by length contraction, etc. This loophole by itself has been plugged in experiments using atoms rather than photons[†] (the detection efficiency for atoms can be made close to 100%), but in these experiments, the detection events were not spacelike separated.

2. In applying the notion of local causality to the experimental setups used in real life (in the experiments of Aspect and coworkers), it is implicitly assumed that the "decision" as to which measuring device to switch a given photon 1 into has been taken appropriately late that it cannot be subluminally communicated to the device that will effect measurements on photon 2. In the original experiment, the switching was done by two piezoelectric transducers driven at different frequencies, and a determined advocate of OLTs could argue that the switching of 1 is therefore foreseeable and the information available about it could influence the outcome of the measurement on 2 without violation of the principles of SR. Again, such an

---

[†] E.g., Rowe et al., *Nature* **409**, 791 (2001).

explanation has an air of extreme pathology or "conspiracy" about it; by itself it has been blocked (as far as seems possible) by experiments in which the triggering (or not) of the switches was done by a quantum random number generator.[‡]

3. When one says that (pace the loopholes raised in 1 and 2) the experimental data "contradict" the predictions of any OLT, one is of course using the phrase in the way conventional in statistical experiments: that is, given that an OLT is the correct physical description of the behavior of the ensemble in question, the probability of obtaining the data which were in fact obtained in the experiment is so tiny that it is legitimate to neglect it entirely. Obviously, in principle a truly diehard defender of OLTs could claim that this is not so and that the totality of the data obtained in all the experiments to date constitutes a giant statistical fluke. I don't think anyone has in fact made this argument, and indeed in any other context it would immediately be laughed out of court despite its soundness from a purely logical standpoint. However, it is interesting (and surprising) that even this potential loophole has, by itself, been blocked in experiments of the "GHZ" type, involving the correlations of four photons; in such experiments, for certain combinations of settings, the predictions of any OLT contradict those of QM not only statistically, but for *individual events* – i.e., for the outcomes of certain (combinations of) experiments OLT's predict with certainty "yes", while QM predicts with equal certainty "no". In these experiments,[§] the QM predictions were verified within the error bars and the OLT predictions were excluded.

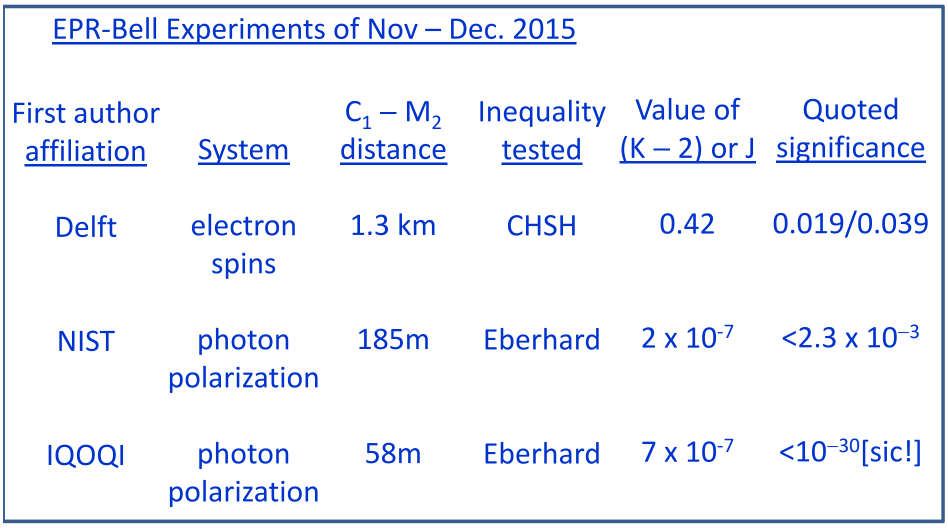
The above summarizes the situation prior to November 2015. In that month, three experiments[¶] were published, each of which is which is claimed by its authors to block all the loopholes simultaneously. The first ("Delft") starts with the standard EPR-Bell correlations for photon papers, but then transfers the entanglement to a pair of atomic systems, thus combining the easy long-distance transport of the former with the high-detection efficiency possible with the latter; it tests the original CHSH inequality for the quantity K defined in lecture 20. The other two experiments use only photons, but exploit an ingenious variant of the CHSH setup due to Eberhard: if we count only the events in which the photon in question is detected in the counter behind the polarizer (call this event "+"), so that the absence of a click (denoted "0") may indicate either the complete absence of the relevant photon or its reflection from/ absorption in the polarizer, and denote by (e.g.) $\mathrm{p}(+0|ab)$ the probability that with polarizers set with transmission axes a, b the (transmission) detector behind a clicks while that behind b does not, then it is possible to prove, using the same premises as for CHSH, the "Eberhard inequality":

$$J = p(++|ab) - p(+0|ab') - p(0+|a'b) - p(++|a'b') \leq 0$$

[‡] The most recent experiment of this type (A. Scheidl et al., *PNAS* **107**, 19708 [2010]) simultaneously blocks the "spacelike separation" loophole.

[§] Z. Zhao et al., *Physical Review Letters* **91**, 180401 (2003).

[¶] B. Hensen et al., *Nature* **526**, 682 (2015); L. K. Shalm et al., *Phys. Rev. Letters* **115**, 250402 (2015); M. Giustina et al., *Phys. Rev. Letters* **115**, 250401 (2015).



EPR-Bell Experiments of Nov – Dec. 2015

First author affiliation	System	$C_1 - M_2$ distance	Inequality tested	Value of (K – 2) or J	Quoted significance
Delft	electron spins	1.3 km	CHSH	0.42	0.019/0.039
NIST	photon polarization	185m	Eberhard	2×10^{-7}	$<2.3 \times 10^{-3}$
IQOQI	photon polarization	58m	Eberhard	7×10^{-7}	$<10^{-30}$[sic!]

Table 1: EPR-Bell experiments, November–December 2015

This inequality holds for arbitrary objective local theories, and is violated by the predictions of QM for detector efficiencies $> 67\%$ (as opposed to 82% for CHSH, hence the experimental advantage vis-à-vis the latter).

The significant features of the data (the discrepancy of $K - 2$ and J from the CHSH and Eberhard limits respectively) are shown in Table 1. Note the minuscule value of the discrepancy for J in the IQOQI experiment, which is however made highly statistically significant by the huge number of trials (several billion).

I have tried to summarize, above, what the raw data (without any QM or other interpretation) tell us in the case of the "EPR-Bell" experiments. It is equally important to be clear what they do *not* tell us. *Contrary* to a statement made repeatedly, and very misleadingly, by Herbert, they do *not* tell us (or at least did not tell us at the time he was writing; cf. below) that the *macroscopic* world must be "nonlocally connected". To discuss this question, we must, of course, first define precisely what we mean by "macroscopic", and this is, of course, a somewhat arbitrary procedure. Herbert characterizes the "macroscopic" world as "the world of cats and bathtubs": we can actually afford a considerably more liberal definition and still make the point. Let's consider what happens where a photon passes a polarizer and is registered in a detector (I have deliberately used "classical" language in the sense of Niels Bohr). We first have the "event" of passage (or non-passage) of the polarizer, and shortly afterwards the "event" (assuming it passes) of absorption in the cathode of the photomultiplier tube that constitutes the detector. The photomultiplier typically has about a dozen stages, and the output of the final stage is the production of a current pulse (typically μA) at the photomultiplier anode. This pulse then travels down some electronics, where it is suitably amplified and perhaps recorded. At this stage, the "events" corresponding to the detection of the two photons are still separated by the length of the laboratory, but eventually, if we are to obtain any information on the all-important correlations, they must be physically brought together in some kind of correlator or coincidence counter. It is the *correlation* data so obtained that are typically stored and later analyzed.

What is the first event in the above sequence that we should call "macroscopic"? Certainly not the passage of the polarizer or the absorption of the photon in the photomultiplier cathode – here only one photon, and in the latter case one electron, is involved. On the other hand, by the time the full amplification of each pulse individually has taken place it seems difficult to deny that we are at the macroscopic level, since already at this stage the individual pulses can be (and sometimes are) recorded in permanent form. Let us then be somewhat generous and define the first "macroscopic" event of the sequence as the production of the current pulse at the anode.

Suppose now that we try to apply the Bell analysis to these "events" rather than the "events" of passage or non-passage of the photon through the polarizer or its absorption in the photomultiplier cathode. Now there enters a crucial consideration that has gone almost totally overlooked in the literature of the subject: the photomultipliers actually take a finite time, and indeed a rather surprisingly long time, to work! In fact, the time for the kind of photomultiplier actually used in the experiments is sufficiently long (~ 30 nsec) that even in the most spectacular cases of the first twenty years of the subject, the current-pulse events (or at least a fair fraction of them) were not spacelike separated! Thus, pace Herbert, he *cannot* use the Bell argument plus the experimental data as they stood in 1985 to exclude an "extended" OLT, which requires realism only at *this* (the "macroscopic") level. Needless to say, this objection can be blocked, at the level of the anode-current pulses, simply by sufficiently extending the length of the laboratory and hence the spatial distance between the events; and, in fact, in recent years (but only in recent years!), experiments have been conducted in which the distance between the detection events, divided by the speed of light (this is the minimum time for a signal to propagate between them) is much longer than the time needed to amplify the detection event to a macroscopic level and even to record it.[‖] However, it is clear that by pushing back the definition of "macroscopic", e.g., to the *amplified* pulses, we can resurrect it. Indeed, as a reductio ad absurdum, we note that in order to obtain any information on the correlations (the fundamental point of the experiment!), we must eventually "reunite" the relevant events in space, at which point invocation of property 2 of an OLT (local causality) completely ceases to "bite".

Many physicists would probably regard the considerations just advanced as on a par with points 1-3 above – that is, as an irritating "loophole" that permits evasion of the conclusions of the Bell argument at the expense of assumptions that to a reasonable person would seem pathological or "conspiratorial". I believe that this point of view is fundamentally mistaken. Objections 1-3 rest on technical limitations (such as the imperfect efficiency of currently available detectors) that are contingent and almost certainly transient. By contrast, the argument just given implicitly relies for its strength on a consideration that is inherent in all current versions of the quantum formalism, namely that *there is no point at which one can definitely state that "actualization" has*

[‖] This experiment blocks loophole 2 (at least to an extent), but leaves 1 open. See in particular G. Weihs et al., *Phys. Rev. Letters* **81**, 5039 (1998).

occurred. In other words, contrary to the impression given in almost all the literature on the EPR paradox and the related Bell's theorem work, one's views on this subject cannot be logically independent of one's solution (or lack of it) to the other great classical paradox, that of Schrödinger's cat. I return to this point below.

What are the options?

The above discussion attempts to summarize the conclusions that we can legitimately draw from the "raw data" *without* necessarily assuming that QM is a valid description even at the microlevel (let alone at the macrolevel!). Given this state of affairs, what are the possible reactions as to the picture of "reality" or lack of it that is thereby implied?

Let's recall once more the assumptions whose conjunction defines the class of OLTs that are refuted (modulo loopholes 1-3) by the experimental data. They are:

1. realism at the *microscopic* level (or "MCFD", cf. below)
2. local causality in the sense of SR
3. induction

Evidently, one of them (at least!) has to go. Let's discuss them in reverse order, which turns out, sociologically speaking, to be the order of increasing "popularity".

Let's first contemplate the possibility of giving up the *principle of induction* – a solution favored by a (very small) minority of physicists. What we would then be saying is that: although the particular set (subensemble) of photon pairs on which (say) the variable A was measured on photon 1 and C on photon 2 was drawn ostensibly "at random" from an ensemble prepared according to a certain uniform specification, the correlations between A and C measured on this subensemble are not characteristic of the ensemble as a whole.

It is necessary to analyze this claim quite carefully. It is not enough to suppose that for each individual photon pair, the value of (say) A measured on photon 1 depends on whether it is C or D that is simultaneously being measured on photon 2 at the other end of the laboratory; such an assumption would violate the postulate of local causality (see below). Rather, one would have to suppose that while the value of A for any one pair is perfectly definite and independent of whether C or D is simultaneously measured, the source as it were "knows in advance" which pairs are going to have A and C measured, which A and D, etc., and emits the relevant subensembles with correlations different from those of the overall ensemble. The most natural context for such an assumption would be a theory that rejects conventional notions about the "arrow of time": in such a theory an event in the "future" (e.g., the switching of a particular photon into counter P_a rather than P_b) could perfectly well influence the statistics of the photons emitted in the "past".

If the rejection of induction inherent in theories of this type were to extend not just to special situations like the EPR-Bell setup but to physics (or science) in general, it is clear that most experimental programs would grind to a halt (as would much of the insurance industry!) Such a prospect is naturally, highly deterrent to most physicists, which is probably the major reason why this solution has not proven to find widespread favor. However, it is less clear that theories in which the normal direction of the "arrow of time" can be broken locally and temporarily are a priori unviable. In such a theory, one might, speaking very crudely and pictorially, view those "local" violations as analogous to the reversed-spin domains that can occur in a magnetic material but can never, as it were, take over much of the sample.

Next let's examine the consequences of relaxing postulate 2, local causality. Specifically, in such an approach, one would allow the possibility that the value of A obtained for the photon 1 of a particular pair might depend on the choice of whether to measure C or D on photon 2 (the switching "event"), despite the fact that the two events in question are spacelike separated. It is a matter of some sociological interest that this situation, while embraced with enthusiasm by many writers of books for the layman (including particularly Herbert!), is firmly rejected by most practising physicists. As with the rejection of induction, this is probably because it would seem prima facie to open the flood gates to the rejection of local causality *in general* – a development which would subvert the foundations of both special and general relativity, a prospect too awful to contemplate for most practising physicists, since special relativity is probably the best-attested "fundamental" theory in the whole of science. To be sure, as in the case of induction, one could raise the question whether a future theory that preserved local causality at the classical level but allowed its violation at that described by QM, would be necessarily ruled out by a priori considerations. While it is impossible to give a definite answer in the absence of a specific instance of such a theory, one thing that should be noted is that the application, in quantum field theory, of the principle of local causality to essentially QM phenomena has had some quite remarkable and unexpected payoffs, such as the (proof of the) "spin-statistics" theorem. For this reason, many would probably hold that the evidence for local causality is even stronger in a quantum picture of the world then it is in a classical one.

If, then, we are unprepared to reject either postulate 2 or postulate 3, the only remaining option open to us is to reject postulate 1, namely the postulate of microscopic realism. Contrary to the impression given by Herbert, this is in fact the solution that is probably embraced by the majority of practising physicists who have thought seriously about the problem. It is, indeed, very much in the spirit of the Copenhagen interpretation: if single microscopic entities such as photons do not possess definite properties in the absence of a specific measuring apparatus – if they are merely, as Bohr claims, links between a macroscopic preparation device and a macroscopic detection device – then why should pairs of such microscopic entities have any different status? The "realization" of a definite outcome takes place only at the level of the macroscopic measuring devices, and there is simply no meaning to the question "what goes on in between?" – all relevant information is given by the (correlated) quantum state of the photon pair and the statistical predictions that can be made from it.

At first sight, then, the advantages and disadvantages of the CI with respect to the EPR-Bell situation are not qualitatively different from those of its applications to the elementary Young's slits problem. And probably if we stick quite rigorously the experiments that had actually been performed up to (say) 1996, this is true. However, suppose that the polarization–correlation measurements are repeated under conditions where the space-time intervals not only between between "microscopic" events at the two ends of the lab (switchings etc.), but also between what are by some reasonable criterion *macroscopic* events (anode-current pulses etc.) are spacelike (as in the experiment of Weihs et al. mentioned above); and, moreover, that we assume that at this level a definite outcome ("click" or "no click") is realized. Following Stapp and others, we can then argue as follows: We previously invoked the idea that each photon "actually possessed" a value of various microscopic quantities, such as the polarization along a, polarization along b, etc., even when these quantities are not measured. But this idea is actually alien to the Copenhagen interpretation, since we should not ascribe "values" in the absence of a measuring device. So let us consider a particular photon that is not in fact switched into the polarizer P_a, and replace the statement (e.g.) "this particular photon 1 has polarization along direction A" which is, at least ostensibly, a statement about the real state of the photon 1, by the statement "had this particular photon been switched into P_a, the (macroscopic) detector placed behind the polarizer would have clicked (produced an anode-current pulse, etc.)". This is a *counterfactual* statement, i.e., a conditional statement whose antecedent is, in fact, false. Now it is clear that we can simply replace, in the definition of what we mean by "$A = +1$", the former statement by the latter, and the argument leading to Bell's inequality goes through as before. Thus, the experimental data, extrapolated in this way, exclude not only OLTs in which postulate 1 refers to microscopic realism, but also versions in which this is replaced by the postulate that counterfactual statements about *macroscopic* events have a definite truth-value. Such a postulate is called *counterfactual definiteness*, and we thus see that if we accept that definite outcomes are realized at the level of clicks, etc., then we can exclude the conjunction of counterfactual definiteness, local causality, and induction.

In other words, if we exclude the possibility of rejecting either local causality or induction, then in the succinct phrasing of Asher Peres, "unperformed experiments have no results" – even when the "results" are defined in *purely macroscopic* terms! Note carefully, however, that the Bell argument *does not* force us to reject the hypothesis that *in the situation that actually obtains*, the macroscopic counters are at all times in definite macroscopic states – indeed, such a hypothesis is an essential ingredient in the argument, since to define A, we have to postulate implicitly that for each photon no. 1 switched into counter P_a, the detector either did ($A = +1$) or did not ($A = -1$) register. To put it briefly, the CFD argument gives us no reason to challenge the ontological status of "does" (click, etc.) – only of "would have"!

Let's sum up our conclusions so far: If we rely only on experimental data actually obtained up to 1996, and ignore the loopholes 1-3, we can conclude that we must reject at least one of the postulates: induction, local causality and microscopic realism. If we go a little further and invoke the results of more recent experiments, then we can also exclude theories in which, in the list of postulates, "microscopic realism" is replaced by "macroscopic counterfactual definiteness". Neither of those statements need particularly worry an advocate of the Copenhagen interpretation, and indeed most them seem remarkably unworried (cf. Rohrlich).

However, there is one major implicit assumption that runs through the argument developed so far, namely that because at the end of the day it is values of quantities like polarization (or the clicks in counters, etc., related to them) that we actually "read off", any picture that is to count as "realistic" must attribute real or hypothetical values *to these quantities.* However, it is not at all clear that such an interpretation exhausts the meaning of "realism", and indeed, in some sense, the formalism of QM itself would seem to indicate rather strongly that it need not. Suppose then that we experiment with the idea that what is "real" in the EPR-Bell situation is not "values" of the quantities that we will eventually read off, but the *probability amplitudes themselves.* And suppose furthermore that, in the spirit of the argument leading to the Schrödinger's cat paradox, we refuse to allow any "events" at all to occur, at least up to a very late stage in the experiment (say after the pulses from the two separate detectors have been combined in the coincidence counter). What then?

It is clear that since there are no "events" until a stage so late as to be irrelevant to arguments about locality, the Bell reasoning has as it were nothing to take hold of. So one can not exactly exclude "realism". However, the question now arises as to the very meaning of concepts like "realism", "locality", etc., under these circumstances. There is indeed a "reality" – the QM wave function (state vector, set of probability amplitudes) – that is formally well-defined and continuously and smoothly evolving. But it has some very strange properties – a "wave amplitude" that is defined as a function of two actually abstract variables corresponding to points distant in ordinary space from one another! It is exceedingly difficult to visualize such a thing in terms of our familiar concepts of waves on water, etc. On the other hand, one might argue that this failure of the ability to visualize is a symptom of our limitations as human beings inhabiting the macroscopic world, rather than a criticism of the model as such.

Before we proceed, a short digression: in this context, there has been a fairly recent development,** which is quite interesting. Pusey et al. consider exactly the above question, "is the QM wave function real?" Of course, to discuss this question meaningfully, one needs to define what it would mean to call it "real", and Pusey et al. (following the work of Hannigan and Spekkens) come up with what they appear to regard as both a necessary and sufficient condition for this, namely that if (as they assume) there exists an underlying level of description characterized by some variable λ, then the value of λ should uniquely fix the wave function ψ; or in other words, that the overlap of the values of λ, which can correspond to two different (not necessarily mutually orthogonal) values of ψ, should not overlap. (They claim that if they do overlap, then the wave function must at best be a statistical description of our ignorance as in the "statistical" interpretation of QM; such an interpretation is nowadays sometimes given the name of "psi-epistemic"). The meat of their paper then consists in showing (under a few very general assumptions about the independence of independently prepared systems, etc.)

Pusey et al., *Nature Physics* **8, 475 (2012).

that any model in which the overlap is nonzero must make experimental predictions different from those of QM (note this is a stronger statement than the CHSH or related theorems). What exactly one concludes from this observation seems rather ambiguous; Pusey et al. rather give the impression that it makes the "psi-epistemic" interpretation of the wave function implausible, but it seems to me that the latter is quite consistent with the lack of overlap between different ψ, and indeed, if one is a believer in either the Copenhagen or the statistical interpretation, one would presumably deny that the λ-level exists at all.

Returning now to our "what then?" question, it is clear that, since there are no "events", that we have avoided – in a sense – the EPR paradox only at the cost of sharpening and reinforcing further the Schrödinger's cat paradox. If, up to a very late stage – one certainly populated by quite macroscopic events – the wave function has not been "collapsed", how eventually does this happen, and what does it mean? Evidently, at this stage, the first choice to make is between theories in which QM is of universal validity and those (such as GRWP) that attempt to modify the formalism so as to produce (eventually) definite macroscopic results. If one chooses the former, one has a further choice between interpretations in which the wave function is nothing more than a calculational convenience, and those in which it represents "something" in the real world. To make the former choice, as is done by the "statistical" interpretation, would seem to imply a very radical abandonment of "realism" at all levels, including the macroscopic: since the only events that are regarded as predicted by the theory are direct observations by human beings, man is indeed "the measure" of all things – yet a human being is presumably physically composed of atoms! If one makes the latter choice, one is thereafter confronted with a choice between various more or less distasteful accounts of how the wave function finally gets reduced (fatalistic, mentalistic, MWI…).

At this point, one may well ask: Can't we actually *tell experimentally* whether QM is indeed applicable at all levels of reality, however macroscopic? For many decades there was almost universal prejudice in the field, based essentially on the phenomenon of "decoherence" (though it had not yet acquired that name!), to the effect that such experiments would be in practice impossible, so that it made no sense even to attempt them. Since about 1980, however, it has become increasingly realized that certain kinds of macroscopic systems – superconducting devices par excellence, but there are other possibilities – may avoid the "decoherence" argument to the extent that meaningful tests using them may be possible, and since 1999 a number of such experiments have actually been done. I now review them.

Other experiments since 1999

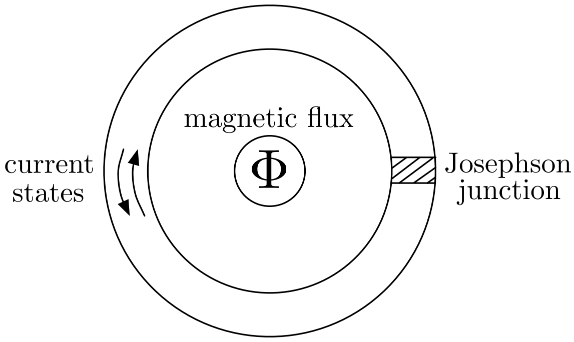
The basic idea underlying recent experiments has been to design a situation where, if we apply the standard formulation of quantum mechanics, a system that is by some reasonable definition "macroscopic" will be described as being at some stage in a superposition of two (or more) states that are, again by some reasonable criterion, "macroscopically distinct", and where furthermore we can *verify*, by their mutual interference, the existence of the superposition (the precise sense of this claim is explained below). Let us call this program the search for "quantum interference of macroscopically distinct states" ("QIMDS" for short). It should be emphasized that, if QM is indeed the whole truth, the creation of a quantum superposition of macroscopically distinct states is a relatively trivial problem. For example, were we willing to actually implement the setup described in Schrödinger's original paper, the resulting final state of the universe is indeed described by a superposition of states of the universe in which the cat is alive and dead respectively; the difficulty is to *verify* that this is what we have, since even if we could realistically do meaningful experiments on the cat by itself, the relative sign[††] of the two states would almost certainly have been randomized by decoherence. In other words, one must try to ensure that, despite the fact that the states of the system are macroscopically distinct, decoherence is not effective. For many years, the prevailing orthodoxy was that this is impossible; it is only in the last ten years or so that this prejudice has been definitively shown to be incorrect.

One question that is frequently and legitimately asked in this context is what exactly we mean when we call two states "macroscopically distinct"? My own view is that most people have a relatively strong intuitive sense of the difference between (say) superpositions of the excited and ground state of an atom on the one hand, and that of the living and dead states of a cat on the other, and that there is limited value in trying to define it too precisely. If one nevertheless demands a quantitative definition, let us consider two numbers, say Λ and D; Λ is the difference, in appropriate atomic units, of the average values of some extensive physical quantity in the two states in question, and D is the number of "elementary particles" (electrons, protons, etc.) that are in some sense "behaving differently" in the two states. Thus, for example, in a standard Young's slit experiment, we would take Λ to be the distance between the slits – or better, that multiplied by the mass of the system involved, divided by (say) a typical atomic size times the mass of a single proton – and D to be the number of electrons + nucleons composing the diffracted particles. We would then describe a given experiment as involving a superposition of macroscopically distinct states if Λ and/ or D exceed some arbitrarily specified threshold value N_c, say 10^6 or 10^{12}. In experiments involving interference in a more abstract space, the concept of "behaving differently" may be a little more ambiguous, as we shall see.

The experiments on QIMDS that are closest in spirit to the standard examples of interference at the atomic level are of the canonical Young's slit type, the difference being that the objects that are being diffracted through the slits are not single electrons or photons, or even H atoms or molecules,[‡‡] but fullerene (C_{60} and C_{70}) molecules. Although the interference pattern seen in the experiment is not as sharply defined as for photons or electrons, it is nevertheless reasonably convincing. In this case, the value of D is simply the total number of electrons and nucleons (neutrons and protons) in a fullerene molecule, i.e., about 1200, while the value of Λ as defined above is something like a million. Thus, the experiment may or may not count as QIMDS, depending on one's definition of the threshold N_c.

[††] Or more accurately, the relative (complex) phase.

[‡‡] A (single-slit) diffraction experiment with H_2 molecules was done as long ago as 1930.



There are a number of other systems in which evidence for QIMDS has been found or claimed, including magnetic biomolecules and ultracold atomic gases. However, the most systematic, and arguably most spectacular, set of experiments has been done on a particular kind of superconducting device known as "rf SQUID". This is a superconducting ring, typically with dimensions ranging from few microns to ~0.15mm, interrupted by a Josephson junction, which is a sort of "gate" for electrons (see figure on next page). The two "macroscopically distinct" states in that case correspond to a circulating current of a few microamps in the clockwise and counterclockwise directions respectively. If for those states one defines Λ in terms of magnetic moment, it is about $10^9 - 10^{10}$. The question of the value of the second number, D, is a bit trickier and depends on one's precise definition of "behaving differently" (as applied to the electrons carrying the circulating current);* depending on this, it can range from 10^4 to 10^{10}. As in the fullerene Young's slit experiment, one detects the existence of a superposition not by a direct measurement at the time it occurs, but by its effect on the later behavior; in fact, if superposition is indeed occurring, one expects periodic oscillations in the probability of an observation, finding the system in the clockwise and counterclockwise circulating-current state, while if decoherence has completely randomized the relative sign, one would expect this probability to be simply a constant. Quite surprisingly, not only are the oscillations seen experimentally, but they last (in the most recent experiments) for over 300 cycles before being damped out. Thus, it is consistent to assume that quantum mechanics is still working at the level of SQUIDS.

However, it is important to be clear exactly what the experiments on QIMDS (whether on fullerenes, on SQUIDS, or whatever) do and do not prove. What they establish is that *if we interpret the raw data in QM terms*, then QIMDS occurs – i.e., the relative sign is not randomized by decoherence. However, this is very far from establishing that QM is the unique explanation of the data. Indeed, it is a purely logical statement that no finite set of experiments could ever establish the truth of QM at this or any other level. What one may be able to do instead is the analog of what has been done in the area of EPR-Bell experiments, that is, to establish that a well-defined class of theories alternative to QM is false. In the present case, the natural class of theories is what

---

*For a comparison with (thought-) experiments that most people would probably assign to the macroscopic side of the divide, see AJL, 'Note on the "size" of Schroedinger cats', arXiv:1603.03992.

may be called "macrorealism"; a possible definition is by the conjunction of three postulates, of which one (induction) is common to the EPR-Bell argument while other two are different:

1. Macrorealism *per se*: a macroscopic system that has available to it one or more macroscopically distinct states must at (almost) all times be in either one or the other of these states.
2. Noninvasive measurability: it is possible in principle to determine whether a system is in one state or the other without affecting the state or the subsequent motion of the system.
3. Induction (ensembles defined by initial conditions).

It can be shown[†] that given these three postulates, one can prove a theorem that is the exact formal analog of (the CHSH version of) Bell's theorem. Taking the SQUID implementation for definiteness, if the current is measured at time t and the value is found to be clockwise (anticlockwise), define the measured value of the variable $Q(t)$ to be $+1$ (-1). Then the theorem states that the experimentally measurable quantity

$$K_{\text{exp}} \equiv \langle Q(t_1)Q(t_2)\rangle_{\text{exp}} + \langle Q(t_2)Q(t_3)\rangle_{\text{exp}} + \langle Q(t_3)Q(t_4)\rangle_{\text{exp}} - \langle Q(t_1)Q(t_4)\rangle_{\text{exp}}$$

has, in any macrorealistic theory, an upper bound of 2. Thus, there seemed to be a reasonable chance of confronting the predictions of QM with those of a macrorealistic theory.

Actually, it turns out that this proposal, as it stands is, as it were, unnecessarily complicated; a simpler protocol will do the job. In fact, rather than measuring 2-time correlations, we can take $Q(t_1)$ to be the starting configuration and simply check directly how far measurement (not necessarily noninvasive) affects $\langle Q(t_3)\rangle = \langle Q_3\rangle$ for (a) the two different macroscopically-distinct states, and (b) for their (putative) quantum superposition. In detail:

Define for *any* state σ at $t = t_2-$:

$$d_\sigma = \langle Q_3\rangle_M - \langle Q_3\rangle_0$$

where the subscript M indicates that at time t_2, $Q(t_2)$ was measured in as noninvasive a way as we can manage, and subscript 0 indicates that no measurement of any kind was made. We first carry out a couple of sets of ancillary measurements in which the state σ is known to be respectively $+1(= |+\rangle)$ and $-1(= |-\rangle-)$; in each case, we take the statistics of Q_3 under conditions M and 0 and thus calculate the quantities d_+ and d_-. (Of course, if the measurement is indeed ideal, then both these quantities are zero.) Then we conduct the main experiment: we now choose σ to be (predicted by QM to be)

[†] AJL and Anupam Garg, *Physical Review Letters* **54**, 857 (1985).

the superposition $2^{-} - \frac{1}{2}(|+\rangle + |-\rangle) = \rho$, again measure Q_3 under conditions M and 0, and calculate the quantity d_ρ. Finally, we define

$$\delta = d_\rho - \min(d_+, d_-)$$

The macrorealistic prediction for δ is $\delta > 0$, whereas QM predicts, under appropriate conditions, a negative number whose precise value depends on the conditions.

This experiment has been done recently,[‡] with the result $\delta = -0.063$, corresponding to a standard deviation of 83. Thus, we can apparently conclude that even at the level of Josephson devices, Nature does not like realism!

[Further discussion in lecture if time allows.]

[‡] Knee et al., ncomms13523 (2016).