QUANTUM MECHANICS AND REAL EVENTS

O.Penrose
Department of Mathematics
Heriot-Watt University
Riccarton, Edinburgh EH14 4AS
Scotland, U.K.

February 19, 2015

To appear in 'Quantum chaos — quantum measurement' ed. Cvitanović et al. (Kluwer, 1992)

Abstract

A way of incorporating real events into the evolution of a quantum-mechanical system, without altering the usual laws of quantum mechanics in any way, is suggested. A simple model is considered, which shows how in an observing apparatus the recorded observations appear as real events occurring with the appropriate probabilities. A characteristic feature of this model is that before the observation the recording apparatus is in a metastable state.

1 Introduction

Although quantum mechanics is wonderfully successful for predicting the results of experiments done in physics laboratories, it has some features that are hard to make sense of if the theory is to be regarded as a theory of how the world as a whole works. The standard interpretation of quantum mechanics takes a dualistic view, dividing the world into two parts, often called the system and the environment, or the quantum part and the classical part. The classical part is described in terms of the familar ideas of everyday life, according to which the world consists of objects which have definite positions at all times (even though we may not know these positions). The quantum part, on the other hand, is described in quite different terms, using Hilbert-space vectors and operators that act on them. Standard quantum mechanics gives no clear guidance about how the line between the two parts of the world is to be drawn.

In addition, a puzzling dualism affects the time evolution of the quantum part of the world. Most of the time the state vector changes according to the unitary evolution rule implied by Schrödinger's equation; this time evolution is deterministic, continuous, and symmetric under time reversal. But when an observation or measurement is made, the state vector changes according to a different rule which is probabilistic, discontinuous, and asymmetric under time reversal. As long as we stay in the physics laboratory, we know what a 'measurement' is and so it is easy to know which is the right rule; but most physical processes take place outside the laboratory and are not observed by anybody. How can the right rules for them be unambiguously formulated?

These questions have been much discussed, for example in J.S.Bell's excellent book[1] and many proposals have been put forward for replacing quantum mechanics by a different theory which avoids the ambiguities. However the alternative theories often suffer from difficulties of their own, such as trouble in formulating them in terms of relativistically invariant concepts.

The object of the present paper is to suggest a possible way of overcoming the ambiguities without making any modification of the usual laws of quantum mechanics, if instead we look for a more unified way of doing quantum mechanics itself. The point of view we shall take is that the reason for the dualistic features of quantum mechanics is an inherent dualism of Nature — in any physical process, two things are going on at the the same time. One of them is described mathematically by the state vector, with the usual deterministic evolution according to Schrödinger's equation. But at the same time there are taking place *real events*, whose occurrence is controlled by a probabilistic law. The two ingredients are linked together, since the probabilities of the real events and the times when they can occur are determined by the state vector, whilst the choice of the state vector and the way it evolves in time are determined by the real events that have already happened. We shall see that in quantum systems of suitable structure both the state vector and the real events can be included in an unambiguous and unified way, without any need either for interactions with an outside 'classical' environment or for modifications of the usual laws of quantum mechanics.

An important component of the theory to be described here is the idea that not every self-adjoint operator is to be regarded as being observable. In this, we differ from Dirac[2] who specifically assumes that every operator whose eigenvalues form a complete set can somehow or other be observed: indeed he uses the word 'observable' as his usual term for such an operator. Our justification for making this distinction beween observable and unobservable operators arises from the fact that an observation implies an interaction between the observed system and some observing apparatus. Since only a limited class of interactions (e.g. local interactions) are available, it seems reasonable to suppose that only a limited class of self-adjoint operators correspond to physical quantities that can be observed.

To have a hope of explaining in detail what is special about this class of 'observable' self-adjoint operators the theory should also provide a model or representation of the act of observation, regarded as a process taking place entirely within the system rather than being imposed on it by some outside agent as in the usual treatments. We shall show that such a representation of the act of observation is possible, at least in a simple model system. In order for an observation to take place, it appears to be necessary for the observing part of the system to start in a metastable state, but there does not appear to be any need for it to be macroscopic (i.e. large) except in so far as a system may have to be large in order to have metastable states.

Only a small part of the programme sketched above has actually been carried out. We confine ourselves here to setting out a postulate giving the conditions under which a real event may occur and showing how these ideas work in the special case of a very simple system consisting of an object system with just two states together with an idealized instrument for detecting which of the two states it is in.

2 Interpretation of the state vector

Although the main part of this theory concerns the conditions under which it makes sense to say that real events occur in a quantum system, we begin by outlining the interpretation that will be put on the state vector so as to show how the occurrence of state vector reduction can be understood without any modification to the unitary evolution implied by Schrödinger's equation.

The state vector will be interpreted as a thing similar to a probability distribution, though more complicated. It is like a probability distribution in the following ways:

(i) it is a statistical quantity, that is to say it can be measured by suitable statistical experiments, involving a large number of identically prepared replicas of the system, but it cannot be measured by a single experiment on a single system. The collection of replicas of the system generated by such a statistical experiment is often called an *ensemble*.

(ii) it depends on how the system was prepared and is therefore conditional on past real events. For this reason, it changes discontinously if new real events are incorporated into the conditions, in just the same way that the probability of a 62-year-old man's living to see his 75th birthday changes discontinuously if it is discovered that he has a bad heart. This discontinuous change, in the case of the state vector, is sometimes called the collapse of the state vector, but this collapse is not a physical event like the collapse of, say, a bridge. All that happens is a switch to a new ensemble for the calculation of probabilities for future events; this switch is convenient, though not logically necessary, because, as far as the future is concerned, all probabilities are conditional on the event that has just happened, so that the probability of this event will always have the value 1.

On the other hand there are important differences between state vectors and probabilities:

(i) the numbers used to represent the state vector are complex whereas probabilities are real (and non-negative.) For this reason, the state vector cannot be measured simply by counting as in the case of probabilities; and moreover interference effects such as those found in the two-slit experiment can occur, in which different possible ways of achieving a given outcome can cancel one another out instead of reinforcing as they always do in the case of probabilities.

(ii) the results of observations are not to be thought of as measurements in the sense of finding out something that is already there[3]; rather, it is as if they come into being as a part of the process of observation itself.

(iii) in classical mechanics it is possible to 'explain' the observed probability distributions by relating them, via the laws of mechanics, to plausible probability distributions at the beginning of time; but in quantum mechanics such a programme, which is tantamount to a search for 'hidden variables' comes up against the difficulty of Bell's inequality.

3 Superselection rules

Now we turn to our main topic, which is to identify the circumstances under which real events can occur in a quantum system, without any interaction with the outside world. In general the events occur at particular times, but to introduce the ideas we consider first a case where there is no time evolution.

Consider a quantum system without time evolution, whose Hilbert space $\mathcal{H}$ of the system is the direct sum of two orthogonal subspaces:

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 \tag{1}$$

such that

$$(\psi_1, A\psi_2) = (\psi_2, A\psi_1) = 0 \tag{2}$$

holds for all ψ_1 in $\mathcal{H}_1$, all ψ_2 in $\mathcal{H}_2$, and all A in the set $\mathcal{A}$ of observable operators.

A condition of the form (2) is called a *superselection rule*. Superselection rules were originally considered in quantum field theory[4] where the two subspaces may correspond, for example, to different values for the total electric charge. Their relevance to the quantum measurement problem has been noted by Wakita[5] and Zurek[6] .

The point of the definition is that if the state ψ of the system is written in the form

$$\psi = c_1\psi_1 + c_2\psi_2 \tag{3}$$

with ψ_1 in $\mathcal{H}_1$ and ψ_2 in $\mathcal{H}_2$, (c_1 and c_2 being complex numbers and all the ψ vectors normalized), then the expectation of a measurement of any observable A made at any later time t is

$$(\psi(t), A\psi(t)) = |c_1|^2(\psi_1(t), A\psi_1(t)) + |c_2|^2(\psi_2(t), A\psi_2(t)) \tag{4}$$

Here $\psi(t)$ denotes a time-dependent state vector, evolving according to the unitary evolution associated with Dirac's[2] 'Schrödinger picture' and equal to the initial state vector ψ at the initial time. The definitions of $\psi_1(t)$ and $\psi_2(t)$ are analogous.

Eqn (4) tells us that the expected result of any measurement is precisely the same as if, even before it had been decided which observable to measure, the system had irrevocably chosen one of the two subspaces, $\mathcal{H}_1$ or $\mathcal{H}_2$, with probabilities $|c_1|^2$ and $|c_2|^2$ respectively. Therefore, when the superselection rule (1,2) holds, it is consistent with the predictions of ordinary quantum mechanics to assume that, as soon as the system is set up in a definite state, a real event occurs, either (with probability $|c_1|^2$) the event $\mathcal{E}_1$ which corresponds to the state vector being in subspace $\mathcal{H}_1$, or (with probablity $|c_2|^2$) the event $\mathcal{E}_2$ corresponding to $\mathcal{H}_2$. The basic postulate of this paper is that *under these conditions a real event, either $\mathcal{E}_1$ or $\mathcal{E}_2$, actually does happen*, and furthermore that it does so at the earliest possible time – i.e. in this case as soon as the system is prepared in state ψ. If the real event $\mathcal{E}_1$ occurs, then the appropriate state vector for future calculations is ψ_1, the projection of the original state vector ψ into $\mathcal{H}_1$; and likewise if $\mathcal{E}_2$ occurs, then the new state vector is ψ_2.

4 Superselective subspaces

So far we have assumed that the superselection rule operates throughout the history of the system, so that the real event may be held to occur as soon as the system comes into existence. We are more concerned, however, with events that happen at later times, such as the times when observations occur. To describe these, we shall use a generalization of the idea of superselection rule.

The generalization has two new features. First, we no longer require that the two orthogonal subspaces have as their direct sum the entire Hilbert space; it will instead be only a susbspace. And secondly, we require the two orthogonal susbspaces to be invariant only under forward time evolution, not under the full two-way time evolution of the system. So, our new definition is this:

Consider a system whose Hilbert space $\mathcal{H}$ has a subspace $\mathcal{F}$ which is the direct sum of two orthogonal subspaces:

$$\mathcal{F} = \mathcal{F}_1 + \mathcal{F}_2 \tag{5}$$

both invariant under forward time evolution, i.e.

$$\psi(t_1) \in \mathcal{F}_i \Rightarrow \psi(t_2) \in \mathcal{F}_i \ (i = 1, 2) \tag{6}$$

whenever $t_1 \leq t_2$. Suppose further that for every observable A we have, as in eqn (2), the superselection rule

$$(\psi_1, A\psi_2) = (\psi_2, A\psi_1) = 0 \tag{7}$$

for all ψ_1 in $\mathcal{F}_1$ and ψ_2 in $\mathcal{F}_2$. When (5, 6, 7), hold we shall say that $\mathcal{F}$ is a *superselective subspace* with components $\mathcal{F}_1$ and $\mathcal{F}_2$. In a system whose time evolution operator is not invariant under time translation, it may happen that eqn (6) holds only for times t_1 after some special fixed time T; in that case we shall say that $\mathcal{F}$ is a superselective subspace after time T.

Theorem Suppose that at some time t_0 the state vector ψ lies in a superselective subspace $\mathcal{F}$ with components $\mathcal{F}_1$ and $\mathcal{F}_2$, so that it can be written

$$\psi = c_1\psi_1 + c_2\psi_2 \tag{8}$$

with ψ_1 in $\mathcal{F}_1$ and ψ_2 in $\mathcal{F}_2$. Then for all $t > t_0$ we have

$$(\psi(t), A\psi(t)) = |c_1|^2(\psi_1(t), A\psi_1(t)) + |c_2|^2(\psi_2(t), A\psi_2(t)) \tag{9}$$

where $\psi(t)$ denotes the time-dependent state vector that is equal to ψ when $t = t_0$, and the definitions of $\psi_i(t)\ \ (i = 1, 2)$ are analogous.

This theorem, whose proof is obvious, has the following consequence, analogous to the one we obtained for the simpler superselection rule in section 2: *If at some time t_0 the state vector of the system is in a superselective subpace $\mathcal{F}$ with components $\mathcal{F}_1$ and $\mathcal{F}_2$, then it is consistent with the predictions of ordinary quantum mechanics to assume that at time t_0 a real event occurs, either (with probability $|c_1|^2$) the event $\mathcal{E}_1$ which corresponds to the state vector's being in subspace $\mathcal{F}_1$, or (with probablity $|c_2|^2$) the event $\mathcal{E}_2$ corresponding to $\mathcal{F}_2$.* It then follows from the postulate enunciated in the preceding section that under these conditions a real event does occur, at the moment when the state vector first enters the subspace $\mathcal{F}$: either the event $\mathcal{E}_1$ with probability $|c_1|^2$ or the event $\mathcal{E}_2$ with probability $|c_2^2|$.

5 A model of a quantum measurement

The following model is intended to illustrate the operation of the above definitions and postulates and to show how a theory of measurement can be constructed within the framework we have described.

Consider a system consisting of two subsystems. One is an 'object system' whose state space is spanned by just two vectors, ϕ_1 and ϕ_2 . These vectors can be thought of as 'spin up' and 'spin down' states of the object system. The other part of the system is a 'detector' whose state space is spanned by an infinite set of basis vectors: a vector μ which is to be thought of as a metastable state, and an infinite set $\{...\lambda_{-2}, \lambda_{-1}, \lambda_0, \lambda_1, \lambda_2, ...\}$ which together make up the stable equilibrium macro-state of the detector. Since this equilibrium macro-state comprises an infinite number of micro-states, its entropy is infinite in this model. The state space of the composite system is the tensor product of the state spaces of the two subsystems.

Time will be assumed to be discrete, taking integer values only, and the normal time evolution rule taking the state vector at time t into the state vector at time $t+1$ is

$$\lambda_n \rightarrow \lambda_{n+1} \ (n = ..., -2, -1, 0, 1, 2, ...) \tag{10}$$

States not mentioned in this rule are left unchanged by the normal time evolution.

However, we postulate a special time T at which a different rule applies: at a time which we shall call $T-0$ immediately before the operation of the normal rule (10) at time T, we apply the following additional transformation, which has the interpretation that if (and only if) the object system is in the state ϕ_1 the states μ and λ_0 of the detector change places. In physical terms, the detector, started in the metastable state μ, is set up so as to be kicked out of the metastable state in the same way that a silver atom in a photographic emulsion might be kicked out of the molecule it was previously in by the arrival of a suitable photon. Mathematically, the rule (to be applied at time $T-0$ only) is

$$\begin{aligned} \phi_1 \times \mu &\rightarrow \phi_1 \times \lambda_0 \\ \phi_1 \times \lambda_0 &\rightarrow \phi_1 \times \mu \end{aligned} \tag{11}$$

The set of observables will be taken to be the set of operators that affect the detector only when it is in or adjacent to its metastable state, i.e. the only basis vectors it affects are $\phi_1, \phi_2, \mu, \lambda_0$. The subspaces $\mathcal{F}$, $\mathcal{F}_1$ and $\mathcal{F}_2$ are taken to be

$$\mathcal{F} \ = \ (\text{ object system space }) \times (\text{ span of } \mu, \lambda_1, \lambda_2, ..)$$

$$\begin{aligned}
\mathcal{F}_1 &= \text{(object system space)} \times \text{(span of } \lambda_1, \lambda_2, ..) \\
\mathcal{F}_2 &= \text{(object system space)} \times \mu \qquad (12)
\end{aligned}$$

Then it can be checked, using the definition in section 4, that the subspace $\mathcal{F}$ is superselective after time T, with components $\mathcal{F}_1$ and $\mathcal{F}_2$.

Initially, we take the system to be in a state where the object system is in a linear combination of its 'up' and 'down' states, while the detector has been prepared in its metastable state ready to detect whether or not the object system is in the 'up' state. Then the time evolution proceeds as follows:

$$\begin{aligned}
\psi(0) &= [c_1\phi_1 + c_2\phi_2] \times \mu \\
\psi(1) &= \text{(same)} && \text{by (10)} \\
&.... \\
\psi(T-1) &= \text{(same)} \\
\psi(T-0) &= c_1\phi_1 \times \lambda_0 + c_2\phi_2 \times \mu && \text{by (11)} \\
\psi(T) &= c_1\phi_1 \times \lambda_1 + c_2\phi_2 \times \mu && \text{by (10)} \\
\psi(T+1) &= c_1\phi_1 \times \lambda_2 + c_2\phi_2 \times \mu \\
&..... && (13)
\end{aligned}$$

In all lines of this array from the $\psi(T)$ line onwards, the first term on the right hand side is in $\mathcal{F}_1$ and the second is in $\mathcal{F}_2$. Hence we can apply the theorem in the preceding section, and the consequences noted there, to conclude that (according to our basic postulate) a real event occurs at time T. With probability $|c_1|^2$ the system chooses subspace $\mathcal{F}_1$ and the event is that the detector has found the object system to be in state ϕ_1 and has itself gone into its true equilibrium state. With probability $|c_2|^2$ the system chooses subspace $\mathcal{F}_2$ and the event is that the detector remains in its metastable state, with the implication that the particle was not in state ϕ_1 and is therefore in state ϕ_2. The 'measurement' is complete, and the occurrence of the real event automatically puts the object system into the new state ϕ_1 or ϕ_2 without any departure from the unitary Schrödinger evolution.

6 Discussion

It may be helpful to compare the theory proposed here with some of the well-known interpretations of quantum mechnics. The present theory is like the standard 'Copenhagen' interpretation in combining a quantum description of the world in terms of state vectors with a 'classical' one in terms of real events; but in the present theory the two descriptions are applied simultaneously to the same system instead of being applied to two different parts of the world. The present theory is like the Bohm-de Broglie[7] and stochastic mechanics[8,9] theories in that they also have real events (the changing configurations of the particles) superimposed on the state vector evolution, and our interpretation of the 'collapse of the state vector' is virtually the same as in those theories; but in that theory real events occur continuously, whereas in ours they only occur when certain conditions are satisfied. For example in the two-slit experiment, the Bohm-de Broglie theory says that the electron really does go through one slit or the other, whereas ours does not say anything about how it got from the emitter to the detector. The present theory is like the 'many worlds' theory[10] in that it envisages, in principle, a state vector for the whole Universe, which splits up according to the various ways that the quantum world may jump ; but it differs from that theory in that it has only one real world rather than many, and in that it seeks to make precise the conditions under which the splitting can occur. The present theory is like the theories of state-vector reduction, such as that of Ghirardi, Rimini and Weber[11], in seeking to elucidate the conditions under which events can occur; but it differs from those theories in that it seeks to do this in a way that is exactly compatible with standard quantum mechanics.

As presented here the theory is obviously incomplete, and of course the possibility exists that it cannot be completed in a fully satisfactory way. For example, the type of superselection rule envisaged in section 4 can be shown to imply that the matrix elements of the Heisenberg operator $A(t)$ corresponding to the observable A in eq. (4) should be exactly zero for all positive values of t; but in a continuous-time theory this is[12] not compatible with the usual type of Hamiltonian with a spectrum which is bounded below. This suggests that a realistic continuous-time version of the present theory can be developed, if at all, only for infinite systems, where the concept of a Hamiltonian does not play such a central role as it does in the theory of finite systems.

7 Acknowledgements

This work was done for the conference on Quantum Measurement Theory held at the Niels Bohr Institute, June 3-5 1991. I am grateful to NORDITA, Copenhagen for their financial support while the work was done, and to the organizers of that conference, particularly I.C.Percival, for making it possible. I am also grateful to many colleagues, including Y.Aharonov, P.Coveney, S.Goldstein, P.Pearle, R.Penrose, I.C Percival, E.Squires, R.Thatcher and B. Tóth for helpful discussions or remarks.

8 References

1. J.S. Bell, Speakable and unspeakable in quantum mechanics (Cambridge 1987)
4. P.A.M. Dirac, The principles of quantum mechanics (Oxford, 1947)
3. J.S. Bell, Against 'measurement', pp 17-31 of 'Sixty-two years of uncertainty', ed. A.I Miller (Plenum, New York, 1990)
4. G.C. Wick, A.S. Wightman and E.P. Wigner, Intrisic parity of elementary particles, Phys.Rev. 88: 101-105 (1952).
5. H. Wakita, Measurement in quantum mechanics, Prog. Theor. Phys. 23: 32-40 (1960).
6. W.H. Zurek, Phys. Rev. D 26, 1862-1880 (1982), Environment-induced superselection rules.
7. D. Bohm, A suggested interpretation of the quantum theory in therms of 'hidden' variables, Phys. Rev. 85: 166-179, 180-193 (l952).
8. I. Fenyes, Eine wahrscheinlichkeitstheoretische Begründung und Interpretation der Quantenmechanik, Zeits. Phys. 132: 81-106 (1952)
9. S.Goldstein, Stochastic mechanics and quantum theory, J. Stat. Phys. 47: 645-667.
10. B.S. de Witt and N.Graham (editors), The many-worlds interpretation of quantum mechanics (Princeton, N.J. 1973)
11. G.Ghirardi, A.Rimini and T.Weber, Unified dynamics for microscopic and macroscopic systems, Phys Rev D34: 470-491 (l986)
12. G.Ghirardi and P.Pearle, private communication