A Conceptual Analysis of Quantum Zeno; Paradox, Measurement, and Experiment

D. Home*

Department of Physics, Bose Institute,
Calcutta 700009, India

and

M. A. B. Whitaker†

Department of Pure and Applied Physics, Queen's University,
Belfast BT7 1NN, Northern Ireland

Received December 27, 1996

Arguments on controversial points concerning quantum measurement theory and the quantum Zeno effect are presented. In particular it is argued that (1) the quantum Zeno effect is a genuine result of quantum theory and current quantum measurement theory, independent of the projection postulate; (2) the effect is of very general nature and rests on analogous arguments to those involved in Bell's theories; (3) the term “quantum Zeno effect” may usefully be restricted to experiments where a measuring device exerts a nonlocal negative-result effect on a microscopic system, mere inhibition of a transition by a directly interacting device not qualifying; (4) since no decay is truly exponential, theoretically all decay phenomena should exhibit the quantum Zeno effect under observation, continuous or discrete. A detailed study is made of the experiments claiming to demonstrate the effect; it is found that they do not meet our criterion above.

1. INTRODUCTION

The “quantum Zeno effect” or “quantum Zeno paradox” has become a topic of great interest in the last few years. Following initial discussion in the field of radioactive decay [1, 2], it has been further discussed in the areas of polarised light [3], the physics of atoms and atomic ions [4–6], neutron physics [7], quantum tunneling [8–10], and quantum optics [11]. Since 1990, around 80 papers have discussed the effect, compared with probably fewer than 15 before that date. As well as theoretical suggestions and analysis, both specific and general, there have been a number of experimental papers [12, 13] claiming to have detected the effect—again in very different areas of physics.

* E-mail: d.hom@boseinst.ernet.in.
† Communicating author. E-mail: a.whitaker@qub.ac.uk.

© 1997 Academic Press

All rights reserved.
Various questions arise—and are still matters of considerable controversy. First, what are the fundamental aspects of quantum theory that give rise to the effect, and its claimed occurrence in so many areas of physics? Indeed, to what extent is it a genuine result of quantum theory? What is the relationship between the effect and the “collapse” of wave-function? To what extent do the various theoretical discussions of the effect actually relate to the more “paradoxical” aspects of the original suggestions? And do the experiments actually demonstrate the fundamental elements of the theoretical proposals?

Questions such as these cannot be addressed without the use, in addition to standard quantum theory, of what may be called “quantum measurement theory.” Some would deny that such a theory exists. Peres, for example, in the preface to his excellent book [14, p. xiii], claims that there can be no quantum measurement theory—only quantum mechanics. Either one uses quantum mechanics, he argues, or some other theory. He points out that a measurement is not a primitive notion; it is a physical process involving ordinary matter and subject to the ordinary physical laws. Yet the hard fact is that a straightforward application of the quantum formalism to measurement situations will not lead, in the general case, to unique measuremental results (as stressed, for example, by Bell in many of his papers [15] on this topic). It seems that some extra features must be introduced.

In fact various strategies—for example, treatment of the apparatus as “classical,” collapse of wave-function—have been used. (Peres [14, p. 373], for example, describes the measuring instrument as being typified by “ambivalence”—it must be treated as a quantum system when it interacts with the measured object, but subsequently as classical.) These strategies have enabled many aspects of measurement to be handled reasonably successfully from a pragmatic point of view, although always standing apart from, in the strictest terms, in contradiction to, standard quantum theory.

In this paper we discuss some of the approaches to quantum measurement theory and their application to the quantum Zeno effect. We consider that, for reasons already implied, all such discussion must be tentative in nature. On the one hand, it is well worth following through the various strategies developed for quantum measurement theory, discovering their effects in novel situations of importance; certainly interesting and genuine phenomena may be and have been predicted in this way. On the other hand, though, a more fundamental aim of such analysis must always be to understand more about the measurement process itself, in particular to aim at eventually elucidating the way in which quantum theory may break down (or be developed) so as to predict the emergence of actual experimental results and, thus, to explain the considerable success of the rather ad hoc strategies.

We take the opportunity at several points to comment on recent papers on the quantum Zeno effect by Fearn and Lamb [8, 16]. These papers may be said to be based on an approach to quantum measurement initiated by Lamb [17] in a famous 1969 paper and developed in many articles cited in Refs. [8, 16].

Indeed we may usefully conclude this section by contrasting our own tentative approach with comments of Fearn and Lamb [16] on the well-known original
Zeno paradox (although we would not wish to suggest that this original Zeno paradox and the modern effect discussed in this paper have much more in common than a superficial broad similarity in results; see Section 8). For Fearn and Lamb, Zeno's original argument was merely “erroneous,” and simple experiments could easily have disposed of the paradox and so ended the discussion. In fact, though, it must have been as obvious to Zeno as to anybody else that he was, in fact, able to complete his afternoon walk! His purpose in devising the argument must surely have been, not to claim a new result, but to question currently held ideas on motion and to attempt to ascertain how they might be improved or replaced in order to avoid such results.

In this paper our approach is analogous. We exploit the puzzling, perhaps contradictory, results of current quantum theory and quantum measurement theory to analyse the quantum Zeno effect and so to attempt to shed light on the many controversial aspects of the field. A more fundamental, although longer-term, aim is to discover where and how fundamental change might be required so as to put the analysis on firmer foundations.

2. THE QUANTUM MEASUREMENT PROBLEM

The great success of quantum theory (exemplified by its approach to two-slit diffraction) was its ability to handle wave-functions or state-vectors which have the form of a sum of individual terms representing quite different macroscopic situations. This success is a direct result of the linearity of the Schrödinger equation. For example in the two-slit diffraction case, one has wave-functions of the form

\[ \phi_{\text{sum}} = (1/\sqrt{2})(\phi_{\text{up}} + \phi_{\text{down}}), \]

(2.1)

where \( \phi_{\text{up}} \) and \( \phi_{\text{down}} \) represent wave-functions which have passed through one or other of the two slits.

There is an associated penalty when one turns to the analysis of measurement. Suppose a microscopic entity with initial wave-function \( \psi_i \) interacts with a macroscopic measuring device with initial wave-function \( \psi_d \). This device performs a good measurement of observable \( O \), represented by Hermitian operator \( \hat{O} \), and we assume that \( \hat{O} \) has a complete set of eigenfunctions \( \xi_n \) with associated eigenvectors \( A_n \), so that

\[ \hat{O}\xi_n = A_n\xi_n, \]

(2.2)

We first consider two distinct cases in which the functions \( \psi_i, \psi_{ii}, \) and \( \psi_{i2} \), are nondegenerate eigenfunctions of \( \hat{O} \), say, \( \xi_1 \) and \( \xi_2 \). Following the interaction between measured system and measuring device, we assume that the final
wave-function of the system is \( \phi_{i1} \) and \( \phi_{i2} \) in the two cases and that of the measuring device is \( \psi_{j1} \) and \( \psi_{j2} \). We may represent the process as

\[
\begin{align*}
\alpha_1: \psi_i &\rightarrow \phi_{i1}\psi_{j1} \quad (2.3a) \\
\alpha_2: \psi_i &\rightarrow \phi_{i2}\psi_{j2}. \quad (2.3b)
\end{align*}
\]

Because the initial wave-functions in the two cases are nondegenerate eigenfunctions of \( \hat{O} \), and since the device performs a good measurement of \( O \), \( \phi_{i1} \) and \( \phi_{i2} \) must represent the measuring device as having registered specific, macroscopically distinct outcomes. A detector has registered a particle or not registered it; a black mark has been produced at point \( P \) on a screen or at point \( Q \) (a macroscopic distance from \( P \)).

Final wave-functions \( \psi_{j1} \) and \( \psi_{j2} \) of the measuring device suggest that the initial values of \( O \) for the microscopic system were \( A_1 \) and \( A_2 \), respectively. Nothing specific has been said about the nature of \( \phi_{i1} \) and \( \phi_{i2} \); in general they need not be related either to \( \phi_{i1} \) and \( \phi_{i2} \), or to \( \psi_{j1} \) and \( \psi_{j2} \). The right-hand sides of Eqs. (2.3a) and (2.3b) have, though, been written as simple products of system and apparatus states, that is to say \emph{not} as entangled sums of such products. This is the case because \( \phi_{i1} \) and \( \phi_{i2} \) are eigenfunctions of \( \hat{O} \).

The processes described in Eqs. (2.3a) and (2.3b) are regarded as following directly from the Schrödinger equation. Thus they may be considered as directly physical, and so as presenting no conceptual problem at all. As Peres quite rightly insists in his remarks already quoted, these processes present measurement as a straightforward interaction between the two systems.

However this conceptual simplicity is immediately lost if \( \phi_i \) is \emph{not} an eigenvalue of \( \hat{O} \). Suppose, in fact, we have

\[
\phi_i = (1/\sqrt{2})(\alpha_1 + \alpha_2). \quad (2.4)
\]

Then the linearity of the Schrödinger equation leads to the following outcome for the measurement process:

\[
(1/\sqrt{2})(\alpha_1 + \alpha_2); \psi_i \rightarrow (1/\sqrt{2})(\phi_{i1}\psi_{j1} + \phi_{i2}\psi_{j2}). \quad (2.5)
\]

For the general case in which

\[
\phi_i = \sum_n c_n \alpha_n \quad (2.6)
\]

the measurement process becomes

\[
\phi_i; \psi_i \rightarrow \sum_n c_n \phi_{i_n} \psi_{j_n}. \quad (2.7)
\]
In complete contrast to Eq. (2.3), we find that Eq. (2.5) and, for the general case, Eq. (2.7) do not predict that the measuring device is left in either of the macroscopically distinct situations represented by a wave-function such as $\psi_{11}$ or $\psi_{22}$, and corresponding to a distinct experimental result such as $A_1$ or $A_2$. Rather it seems to be in a totally unphysical linear combination of several such states. This is in stark contrast with the seemingly obvious fact that, following a good measurement, we observe a unique result; indeed the very purpose of a good measurement would seem to be to obtain such a result.

This conflict between the result of the linearity of the Schrödinger equation and what seems to be the fundamental fact about any measurement process constitutes what may be called the measurement problem of quantum theory.

Any discussion of quantum measurement will have two aspects. The first is the replacement of the schematic nature of Eq. (2.3) by a formal quantum mechanical description of the interaction. The second is to face up to the measurement problem, to attempt to provide some means of escape from it, or conceivably to avoid encountering it in the first place.

The first aspect is certainly of great general interest and importance. It is important to understand the mechanism of the interaction—to appreciate what is special about such an interaction to make it fulfill the role of a “good measurement.” For reasons to be explained, it is particularly interesting to know how $\phi_f$ (about which nothing has been said so far, and indeed about which nothing can be said for the general case) are related to the $\phi_i$, or, what amounts to the same thing, to the $\psi_f$. And, of course, it is interesting to know how we might measure any particular observable.

Interesting though this certainly is in practice, it must be remembered that this first aspect has no conceptual difficulties; in Eq. (2.3) all that is required is to apply the Schrödinger equation. In contrast, the second aspect presents a major conceptual problem—one has to disobey the Schrödinger equation and to provide some coherent justification for one’s procedure. It follows that any account of quantum measurement that concentrates on the first aspect and avoids the second or fails to notice it or treats in a perfunctory way by a conventional strategy has taken a very limited view of the difficulties of the area. We feel that, to a large extent, the approach of Ref. [17] comes into this category.

3. ATTEMPTS TO SOLVE THE MEASUREMENT PROBLEM

There have been many attempts to solve the measurement problem. The first was Bohr’s complementarity [18, 19]. In a sense this denied the truth of Eqs. (2.3), (2.5), and (2.7) by refusing to allocate a wave-function to a measuring device, considering it rather to be classical in nature. Complementarity allows discussion of the value of an observable only in the context of an apparatus being set up to measure that observable, thus avoiding many of the difficulties of quantum measurement,
allowing, in fact, quantum measurement to parallel classical measurement. Complementarity was a successful strategy to avoid the initial difficulties of quantum measurement, but it did not answer the challenge from the Einstein–Podolsky–Rosen (EPR) argument very convincingly, and, since it requires an arbitrarily positioned “cut” between quantum and classical regimes, it cannot be thought of as a rigorous solution to the quantum measurement problem [19].

The so-called “orthodox” solution to the problem is that of von Neumann [20]. It should be pointed out that there are two components of von Neumann’s ideas, corresponding to the two aspects of quantum measurement discussed in the previous section; here we tackle the component directly relevant to the measurement problem, and we discuss the other component in the following section.

Von Neumann’s idea is as follows. Rather than the system being left in the entangled state given by the right-hand side of Eq. (2.5) or (2.7), a second process takes place following the interaction process described in these equations. In this second process, the wave-function of the combined system collapses down to a single product term. For the case of Eq. (2.5) the process

$$\frac{1}{\sqrt{2}}(\phi_f^1 \psi_f^2 + \phi_f^2 \psi_f^2) \rightarrow \phi_{fm}^1 \psi_{fm}$$

occurs with probability $\frac{1}{2}$ for $m = 1$ or 2. In the more general case, the process

$$\sum_{n} a_n \phi_{mn} \psi_{mn} \rightarrow \phi_{fm} \psi_{fm}$$

occurs with probability $|a_m|^2$.

This “collapse postulate” or “projection postulate” or “reduction postulate” should be regarded as an ad hoc strategy rather than a genuine component of quantum theory, because in fact it is directly in conflict with quantum theory itself. As von Neumann himself stressed, if one accepts collapse, one must assume that systems develop in two distinct ways, usually following the Schrödinger equation, but at a measurement following the collapse postulate. One type of process cannot be made equivalent to the other. They differ thermodynamically, the first type of process being reversible, a collapse process irreversible.

They also differ fundamentally from a mathematical point of view. Prior to the collapse process it is always possible to work with a pure state, and the Schrödinger equation maintains the purity of such a state, the corresponding density-matrix being idempotent. At a measurement, though, the pure state becomes mixed. Let us consider the case of Eq. (2.5) with basis-states $|f_1\rangle$ and $|f_2\rangle$ in an obvious notation. In density-matrix notation, Eq. (3.1) becomes

$$\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \rightarrow \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}$$

so an idempotent density-matrix becomes non-idempotent; a pure state becomes mixed.
So clearly the two types of process are totally distinct. And it was Bell [21], in particular, who stressed that the concept of measurement is, in any case, not a fundamental notion capable of precise definition, and so it should not appear in a fundamental theory of physics. The measurement process is presumably built up of straightforward interactions between measuring and measured systems, which should therefore obey the Schödinger equation.

So the collapse postulate should certainly not be regarded as a formally acceptable solution to the measurement problem. Nevertheless it has to be admitted that its application does seem, in practice, to allow the measurement process to be described successfully. This is why it may be entirely permissible for Braginsky et al. [22] to use the postulate systematically in their account of quantum non-demolition measurements and to refer to it as part of “the quantum theory,” which they “presume ... to be correct throughout [their] article.” At the same time and at a deeper level, there should be a desire to understand more about the collapse postulate, why it often works well, its limitations, and how it should eventually be adapted or replaced.

There have been many other attempts to solve the measurement problem which may only be summarised here. There are several interpretations that assert, in total contradiction to what has been taken for granted so far, that the right-hand sides of Eqs. (2.5) and (2.7) can represent the result of a measurement. They thus deny that a collapse process, or any equivalent strategy, is necessary. Among these, the many worlds interpretation [23] says that the two terms of Eq. (2.5), for example, may both be present—but in different worlds. The relative-states interpretation [24] merely retains the total right-hand sides, thus losing one-to-one correspondence between wave-function and state of the system (even at the macroscopic level). Ensemble interpretations [25, 26] retain the full superpositions of Eq. (2.5) and (2.7), but relate them to the state, not of an individual combined system, but to a hypothetical ensemble of such systems.

Decoherence interpretations [27] consider that coupling to the environment produces an analogous effect to that of collapse, while the method of consistent or decoherent histories [28–30] attempts to establish a consistent quantum approach to the development of the Universe and, in particular, to establish its classical nature at the macroscopic level.

Other attempts to solve the measurement problem go beyond the strict Schrödinger equation by adding extra elements. Hidden variable theories (the most important of which is that of Bohm [31]), state that the wave-functions on the right-hand sides of Eqs. (2.5) and (2.7) must be supplemented by extra variables which essentially determine to which branch the system belongs. Spontaneous collapse theories [32] add a term to the Schrödinger equation to produce collapse, not, like von Neumann, at an ill-defined “measurement,” but at all times, and with a probability determined by the size of the system involved.

All these ideas have potential; all of them also have difficulties and drawbacks [19]. In our consideration of the quantum Zeno effect later in this paper, we shall consider both theories that involve collapse and those that do not. In fact we shall
succeed in demonstrating general principles independent of the occurrence of collapse.

4. INTERACTION BETWEEN MEASURING AND MEASURED SYSTEMS

We now turn to what we called the first aspect of the measurement process, where we merely study the interaction of the measuring and measured systems. This process is conceptually straightforward, merely following the Schrödinger equation. We may limit consideration to the special cases of Eq. (2.3), since the more general cases of Eqs. (2.5) and (2.7) follow directly using the fundamental principles of quantum theory.

While the task is straightforward in principle, in practice it is complicated and almost certain to be beyond our powers to carry out rigorously. The measuring system will be macroscopic; even such an apparently simple task as ascertaining the position of a particle must, if we are to trace the physics through rigorously, involve analysing the detailed mechanism of the interaction of a particle with a fluorescent screen, an electronic counter, or a bubble chamber.

And at least in principle we must not shirk such study of the behaviour of the measuring apparatus. Bohr [33] and Bell [34] have both stressed that the results of a quantum measurement must be regarded as the joint product of “system” and “apparatus,” the complete experimental setup, not merely as the registration of some preexisting property of the measured system.

In practice, though, we cannot analyse these processes rigorously and almost certainly will restrict ourselves to schematic coupling of measuring device and measured system. We may, in fact, doubt whether the details of the various processes mentioned above, although obviously important from the point of view of the practical study of measurement, are central to the theory of quantum measurement. It should be noted that any detailed description, following the Schrödinger equation, would inevitably be reversible, while, as a result of the macroscopic nature of the measuring device, the conventional description of a measuring process would be that it is irreversible.

Several of the treatments to be discussed in the following section [20, p. 442; 35; 36] consider the measuring system as a pointer with, on the one hand, either one or two degrees of freedom, and explicitly obeying the Heisenberg principle, yet, on the other hand, macroscopic and classical, and able to be “read” in a straightforward way. The early approach of Lamb [17], in contrast, regards the coupling between measuring and measured systems as noncontroversial. In this scheme the measured system is acted upon by a highly ingenious set of steady fields and transient pulses; the scheme is well-defined mathematically but still somewhat vague from the strictly physical point of view. It is not clear how any permanent registration of measurement result takes place.

A very important question regarding the interaction described by Eq. (2.3) is the relationship between the initial and final wave-functions, $\phi_i$ and $\phi_f$, of the measured
system. (In Eq. (2.3), $\phi_i$ has been put equal to an eigenfunction of $\hat{O}$, $\pi_1$ or $\pi_2$, $\phi_i$ is directly linked to $\psi_f$, so the question may be expressed as—what is the relationship between $\phi_f$ and $\psi_f$, the final wave-functions of the measured and measuring systems?

This question is intimately linked to discussion of repeated measurement. What happens if a second measurement of $O$ is performed immediately after the first one? This question is important from a fundamental point of view and also for the quantum Zeno effect.

It is clear that no general answer can be given to these questions. Different types of measurement will lead to different kinds of relationships between initial and final wave-functions. Indeed, it was perhaps first pointed out by Margenau [37] that for many important measurements, the question of repeated measurement does not arise. If a photon or electron is detected by creating a blackened mark on a suitably sensitive screen, clearly the photon or electron has been assimilated into the measuring device, and is not available for a repeated measurement. Of $\phi_f$, all that can be said is that it also has been assimilated into $\psi_f$, the final wave-function of the measuring device.

Yet there are certainly also cases where an independent $\phi_f$ does exist and measurement may be repeated. One such is discussed by Bell [21]; it just relates to the $\pi$-particle tracks produced in Wilson cloud chambers. At each of the stack of photographic plates, the $\pi$-particle excites certain atoms and molecules and, hence, causes a darkened spot. Each of these events may be termed an approximate measurement of the position of the $\pi$-particle. The fact that the measurement is approximate means that the uncertainty in direction of the momentum may also be fairly low, and the direction of the $\pi$-particle may be practically constant. For this particular type of measurement, the wave-functions of the $\pi$-particle and the measuring system after the measurement are directly correlated; that is to say that $\phi_f$ at that moment is localised in the region of the darkened spot. Together with the approximately constant nature of the momentum, this results in the practically straight tracks well known from experiment.

Bell and Nauenberg [36], though, give another simple example where $\phi_f$ certainly exists after the measurement, but is definitely not directly related to $\psi_f$ (or $\phi_i$, despite the fact that we are still in the case discussed in this section, where $\phi_i$ is an eigenfunction of $O$). Here the momentum of a neutron is measured by observing a recoil proton. The momentum of the neutron must be reduced in the process, perhaps to zero, so $\phi_i$, and $\phi_f$ must be totally distinct.

Measurements where the result gives information about the final, as well as the initial, wave-function are called “measurements of the first kind” [38] or “moral” measurements [36, 39], while the more general type of measurements, where the result gives information only about the initial wave-function, are called “measurements of the second kind” [38] or “immoral” measurements [36, 39]. (While for the case of Eq. (2.3), a moral measurement will, of course, give total information about the final state of the wave-function, at least if degeneracy is ignored, for the general measurement of Eq. (2.7) followed by a collapse, obtaining
the result $A_m$ merely gives the information that the coefficient of $x_m$ in $\psi_i$ is nonzero.

Because measurements of the first kind have interesting properties, they are often paid special attention, as we shall do for most of the rest of this paper. Indeed complaints are sometimes made [8, 16] that the “orthodox” view is that all measurements are of this type. To some extent, unfortunately, this is the case, as will be discussed in the following section.

5. MEASUREMENTS OF THE FIRST KIND

The most famous proponent of the position that all measurements should be of the first kind has been von Neumann [20]. This view is almost invariably taken, together with the collapse idea of the previous section, as constituting a single approach to quantum measurement theory. Here we have been at pains to show that such coupling of ideas need certainly not be automatic. The collapse idea is one way of attempting to solve the basic dilemma of quantum measurement; the view of measurement discussed in this section is of much less significance, and, while it has some nice properties, it is to a considerable extent arbitrary.

If Eq. (2.3), which is for the special case where the initial wave-function of the measured system is an eigenstate of $\hat{O}$, is rewritten as a measurement of the first kind, we obtain

$$x_m; \psi_i \rightarrow x_m \psi_{jm}. \quad (5.1)$$

Clearly an immediately repeated measurement of $A$ will yield the same result as the first, $A_m$.

For the general case of Eqs. (2.6) and (2.7), we obtain

$$\psi_i; \psi_i \rightarrow \sum_n c_n \psi_{jn} \quad (5.2)$$

with each term on the right-hand side representing a correlation between eigenfunctions of the measured system and the measuring device, the total wave-function being, in general, entangled. If one follows this stage of the measurement process with a second collapse stage, one will be left with a single product term, say $x_p \psi_{jp}$.

This implies that the first measurement has yielded the result $A_p$, the presence of the $x_p$ implies that a second and any further immediately repeated measurements will yield the same result. Thus the two parts of von Neumann’s thesis together ensure—as was his motivation in selecting them—that an immediately repeated measurement of the same observable will yield the same result as the first one.

Is there any justification for requiring this idea, or at any rate for feeling comfortable with it? We may start from the case where $O$, the observable measured, is position. In this case it does indeed seem required that the two measurements should give the same result, for otherwise the system would have moved a nonzero
distance in zero time, clearly in conflict with special relativity (at least if the measurement is considered instantaneous—a point we return to briefly below).

Indeed some authors [40] make measurement of position a special case for which alone measurement must be of the first kind. Of course, measurement of position must be special—especially awkward, in fact—in another way, particularly if one thinks of an exact measurement. For an exact position implies a totally indeterminate momentum. So, although an immediately repeated measurement may be said to give the same result, a measurement, even an extremely short time, \( At \) later, could give a result anywhere up to a distance \( cAt \) away from the initial result.

It seems that the concept of an exact measurement of position must inevitably present conceptual problems, and it would seem sensible to restrict consideration of measurements of position to statements such as: the \( x \)-coordinate of a particle has been measured to lie within a range \( Ax \). Of course the Heisenberg principle still implies an uncertainty in momentum following measurement of order \( \hbar/AX \), so the problem previously described becomes large as \( Ax \) becomes small. (Any objection that the argument for repeatability assumed instantaneous measurement, which may be unphysical, is also answered by restricting position measurements to measurement within a range of values.)

For measurement of other observables, the desirability of an immediately repeated measurement yielding the same result seems less obvious. For Dirac [41], though, who regarded it as a basic requirement of measurement theory, it was a matter of “physical continuity.”

6. SCHEMATIC MEASUREMENTS OF THE FIRST KIND

Leaving aside its desirability, can the process represented by Eqs. (5.1) be realised physically? Starting at the most schematic end, von Neumann himself [20] (see also Ref. [36]) constructed a mathematical interaction scheme which would achieve this. Suppose the wave-function of the apparatus, \( \psi \), depends on a single parameter, \( q \), and its initial value is given by

\[
\psi_i = \psi(q_0).
\] (6.1)

Suppose also that the interaction lasts for a time \( t_0 \), and is given by

\[
\hat{H}_{int} = (-1/t_0)(-i\hbar \partial \psi/q) \quad (0 < t < t_0).
\] (6.2)

We assume that the other terms in the Hamiltonian may be neglected during this period, either because of the strength of the interaction, or because of the high mass of system and apparatus (or a combination of both factors).

The time-dependent Schrödinger equation during this period is

\[
i\hbar \frac{\partial \Psi(x, q, t)}{\partial t} = (i\hbar/t_0) \hat{\partial} \frac{\partial \Psi(x, q, t)}{\partial q} \Psi(x, q, t), \quad 0 < t < t_0,
\] (6.3)
and the solution is

$$\Psi(t) = \phi_m(\mathbf{q}_0 + A_m t/\tau_0), \quad 0 < t < \tau_0.$$  (6.4)

Using Eq. (5.3) with \(q_0 = 0\) and putting \(t = \tau_0\), we obtain

$$\Psi(\tau_0) = \phi_m(A_m)$$  (6.5)

which is equivalent to Eq. (5.1). For the more general starting wave-function of Eq. (2.6), the final wave-function of the system is

$$\Psi(\tau_0) = \sum_n c_n \phi(A_n)$$  (6.6)

which is precisely the required result of the interaction.

While this model considers the measurement interaction as occurring over a period of time \(\tau_0\), following Bell and Nauenberg [36], we may consider an interaction with

$$\hat{H}_{\text{int}} = \hat{\delta}(t)(-i\hbar \hat{\mathcal{O}} \hat{\partial}/\partial q),$$  (6.7)

rather than Eq. (6.2). This will give rise to an instantaneous interaction at \(t = 0\), leading to the result of Eq. (6.6) but with \(\Psi(\tau_0)\) replaced by \(\Psi(0+)\). Clearly in this case one does not require assumptions about the masses of the interacting systems.

But it is of course the choice of Eq. (6.2) or (6.7) which has made this possible. Essentially it represents a direct coupling between operators representing the observable being measured and the momentum of the pointer of the measuring device. More realistic forms of interaction would, of course, have to treat the measuring device in a much more complicated fashion; more importantly from our point of view, they may give rise to measurements being of the second, rather than of the first, kind. Thus, these calculations demonstrate that it is certainly in order to consider measurements of the first kind, but not that all measurements must be of this kind.

Arthurs and Kelly [35] developed the von Neumann schematic analysis of interaction to an approximate measurement of the position \(x\) and momentum \(p\) of a system. This was an immensely clever paper, implicitly introducing such topics as squeezed states, fuzzy measurements, and positive operator-valued measures (POVMs). Unfortunately the original paper was also exceptionally terse; several recent publications [3, p. 418; 42; 43] have devoted much attention to explaining and elaborating its methods.

Since \(\hat{x}\) and \(\hat{p}\), the relevant operators for the measured particle, clearly do not commute, Arthurs and Kelly introduce a second system or meter with canonical variables \(X\) and \(P\). Then if

$$\hat{x}_1 = \hat{x} - \hat{X}, \quad \hat{p}_1 = \hat{p} + \hat{P}$$  (6.8)
The corresponding observables can therefore be measured simultaneously, and such measurement may be regarded as a fuzzy or unsharp measurement of $x$ and $p$.

Considerable mathematical intuition or sophistication is required \cite{35, 43} to deduce the initial wave-function of the meter required in order that a successful measurement may be accomplished and, further, that such measurement may be optimal in terms of balance between indeterminacies in $x$ and $p$ after the measurement. In fact, because of the indeterminacies in $X$ and $P$, the minimum product of ranges of $x$ and $p$ as measured is equal to $\hbar$, which is just twice the value given by the Heisenberg principle.

While, of course, one cannot talk of the Arthurs–Kelly scheme as a measurement of the first kind, since there are no common eigenfunctions of $\hat{x}$ and $\hat{p}$ to be the equivalent of the $\mathfrak{a}$ in Eqs. (5.1) and (5.2), it may be said that it is the closest to such a measurement allowed by the Heisenberg principle. If the initial wave-function is $\phi(x)$, the distribution of pairs of meter readings for an ensemble of initial particles is given by $\phi(x)$ broadened by Gaussian windows in $x$ and $p$, so as to provide the increases of distributions of values of $x$ and $p$ mentioned above.

The equivalent of the right-hand side of Eq. (5.2) is an integral of terms, each of which corresponds to a particular set of measurement results, $x_0$ and $p_0$. Each term is not in this case a product of a function of meter state variables and a function of $x$, so the equivalent process to collapse must be to substitute particular values of $x_0$ and $p_0$ into the entire wave-function and then to renormalise \cite{35}. Alternatively Stenholm \cite{43} points out that the probability distribution corresponding to the final wave-function of the combined system factorises as

$$S(x, x_0, p_0) = S(x_0, p_0) \, S(x \mid x_0, p_0)$$

with the first term on the right-hand side corresponding to the probability of the experiment yielding the values $x_0$ and $p_0$, and the second giving the conditional probability distribution of $x$ given that the results are $x_0$ and $p_0$. This is essentially the square of the collapsed wave-function, and this wave-function itself is given by

$$\phi(x \mid x_0, p_0) = (2\pi b^2)^{-1/4} \exp\left[-\frac{(x - x_0)^2}{4b^2}\right] \exp(ip_0 x/\hbar).$$

Here $b$ is a so-called “balance parameter” for the measurement. If $b = 2^{-1/2}$, we have “symmetrical balancing”; the measurement has been optimum in the sense that the product of variances in the measured ranges of $x$ and $p$ is a minimum. If $b > 2^{-1/2}$, the measurement values of $p$ form a narrower distribution and those of $x$ a wider distribution than the optimum case, and if $b < 2^{-1/2}$, the reverse is true.

In all cases, the wave-function of the system after the measurement is the Gaussian packet which is most compatible with the results of the measurement for the particular value of $b$. An immediate repeat of the measurement will come as close to repeating the same results as is possible, given the constraints of the uncertainty principle; in this sense it may be said that the Arthurs–Kelly scheme is as near to a measurement of the first kind as is possible, subject to these constraints.
The very great advance of Arthurs and Kelly over the von Neumann scheme is, of course, just that it does handle a joint measurement of incompatible observables. Other than that, it is as schematic as the earlier approach; there is very little genuine physics, the form of the interaction being specifically chosen to lead to the desired result, rather than to mirror, even simplistically, an actual physical process. And, of course, it handles only the conceptually straightforward aspect of the measurement process, the equivalent of Eq. (2.7). It does not attempt to solve the second aspect which constitutes the actual measurement problem and is forced to resort to the physically unacceptable collapse idea.

Yet Fearn and Lamb [8, 16], although they are discussing a measurement of only one quantity, position (within a given range), apparently consider the Arthurs–Kelly scheme the basis of a model of an actual physical process. It does not seem clear that their own measurement procedure is more physical and less schematic that an approach along the lines of von Neumann, which would essentially equate the probability of a measurement of position being in a particular range, to the integral of probability density throughout that range. And certainly they use the equivalent of the collapse postulate, so in no sense are they relating to the real problem of measurement.

7. MEASUREMENTS OF THE FIRST KIND—MORE DETAILED CONSIDERATIONS

Having considered schematic measurements, let us move to real measurements and see if any should be regarded as being of the first kind. We have already reviewed the $x$-particle track discussion of Bell. Here it is easy to see that each measurement is (approximately) of the first kind because the momentum of the $x$-particle is high enough that, in interacting with the particles of each photographic plate, it is able to disturb them sufficiently to create a permanent blackening, but is itself (approximately) undisturbed.

A different and slightly more complicated example is as follows. A fairly massive spin-$\frac{1}{2}$ particle travelling at rather high speed passes through a series of $N$ Stern–Gerlach apparata, each with magnetic field along the $z$-axis. At each exit from the first apparatus, a photographic plate of the type discussed by Bell allows the particle to be detected but, to a good approximation, for its own path to be undisturbed. Each path passes on to the entrance of a second Stern–Gerlach apparatus. There are thus two apparata at the second stage of the process, and in a similar fashion there is a third stage with $2^2$ apparata, and so on until there are $2^{N-1}$ at the final stage. At both exits of each apparata there are photographic plates to perform what are essentially position measurements.

While it is quite in order to describe the scheme as performing a series of position measurements, it is much more natural to say that it performs successive measurements of $S_z$, and clearly one must expect that each measurement will give
the same result. In this case, though, it would not be true to say that the measurement process does not disturb the system. Clearly the particle is being disturbed by being directed on the path through the Stern–Gerlach apparatus. But what is significant is that the quantity being measured, $S_z$, is not being disturbed.

Another example of a measurement of the first kind is that of survival or decay of a radioactive nucleus. Suppose the nucleus is surrounded by detectors so that, if it decays, the decay particle is certain to be detected. If at a particular time, no detector event has been registered, then an immediate repetition of the examination of the detectors will yield the same result. If, on the other hand, a first examination shows that a decay particle has been registered, so the nucleus has decayed, a second examination will again yield the same result.

As a last example, suppose a trapped ion is known to have two states—either in one particular energy-level or oscillating for a considerable period between two other energy-levels and emitting a stream of photons. If, in a particular time interval, we observe a number of photons, there will be a high probability that, in an immediately subsequent interval, we will also observe photons. The same applies to a lack of photons, so in both cases immediately repeated measurements would yield the same result (although in practice repetition cannot be immediate so there will be a small probability of a change in result).

If in all these cases we assert that the observation does not disturb the quantity being measured, it must be understood that this refers only to the case where the system starts in an eigenstate of the operator for the observable being measured (whether because we imagine a collapse after a previous measurement, or conceivably for some other reason), so we have the case of Eq. (5.1). For other starting wave-functions, even for a measurement of the first kind, we have the equivalent of Eq. (5.2), which leads directly to the measurement problem.

These results are interesting, though, because they establish the physical basis for measurements of the first kind; several of the types of measurement have been made use of in discussions of the quantum Zeno effect for which measurements of the first kind are required.

Last, in this section we discuss an approach due to Daneri et al. [44, 45] which attempts to treat the measurement interaction in a far more detailed and elaborate fashion. In the interaction, the energy of the apparatus is enormously larger than that of the measured system, but one wishes the latter to be undisturbed, to as good an approximation as possible, during the interaction. One achieves this by creating the initial state of the measuring device in a metastable state, as defined in statistical mechanics. The interaction then triggers an irreversible process during which the state of the apparatus evolves towards a stable state, the nature of which depends on the initial state of the observed system.

In some cases, for example cloud and bubble chambers, the metastable nature of the initial state of the measuring apparatus is obvious. In others, such as spark chambers and scintillation counters, the metastability is a result of the system possessing a second macroscopic constant in addition to the energy. The interaction with the measured system induces a transition from an equilibrium state with one
value of this constant, to a nonequilibrium state with a different value, and this gives rise to the irreversible process which is studied in detail in Refs. [44, 45].

The measurement process is of the first kind, since if the initial wave-function of the measured system is an eigenfunction of $\hat{O}$, the interaction follows Eq. (5.1). Interestingly, though, Daneri et al. claimed that their method could achieve more than this. In the case where the initial wave-function of the measured system is a superposition of a number of eigenfunctions of $\hat{O}$, as on the left-hand side of Eq. (5.2), they claimed that the final combined wave-function of system and apparatus consisted of only a single term, the case $\pi_m \psi_m$ being obtained with a probability $|c_m|^2$. Thus they claimed that their process transformed a pure state to a mixture; in effect they claimed that they had solved the measurement problem.

Yet, as Jauch et al. [46] and Bub [47] pointed out, such a suggestion violated the linearity of the Schrödinger equation. (See also the remarks in d’Espagnat’s book [48], and the comment on them at the end of Ref. [45].) Loinger [49] replied, saying that the claim was actually not that a mixture had been obtained, but that the superposition of states that was obtained, equivalent to the right-hand side of Eq. (5.2), was practically or experimentally indistinguishable from a mixture.

It is, hence, reasonable to include this interpretation with the decoherence ideas of Zurek [27] under the title “effective incoherence interpretations” of quantum theory [50], and this broad approach has become very popular [51, 52]. The philosophy behind the approach has also, though, been heavily criticised [26, 53, 54] and it has also recently been argued [50, 55] that, even at the pragmatic level, the approach cannot achieve its goals.

8. QUANTUM ZENO EFFECT—INTRODUCTION

The quantum Zeno effect has been discussed for over 60 years, earlier being called “Turing’s paradox” [56, 57]. For most of this time it has been felt to be difficult, strange, and possibly “paradoxical.” On the other hand, most people have perhaps not worried about it too much, as it has been felt to be (1) probably the result of rather simplistic assumptions, (2) confined to rather special areas of physics, and (3) dependent on a particular interpretation of quantum theory.

In this paper we aim to show that the effect is indeed difficult, strange, and possibly paradoxical. But in addition it is (1) robust as to particular assumptions, (2) likely to occur in many areas of physics, and (3) totally independent of what interpretation of quantum theory is being used, being dependent only on the most standard ideas of quantum measurement theory for measurements of the first kind.

Even compared to the other problems of the interpretation of quantum theory—Einstein–Podolsky–Rosen (EPR), Schrödinger’s cat, and so on [19]—the quantum Zeno effect has been felt to be isolated, and perhaps not giving much significant information about the fundamental nature of quantum theory (as contrasted, say, with the physics surrounding the Bell inequalities [15, 19], which
has been felt to be of great fundamental significance). In contrast we show that what may be called the Zeno and Bell areas of analysis have many analogous features, and we argue that the quantum Zeno effect should be regarded as being very much in the mainstream of discussion of the nature of quantum theory and "quantum measurement."

We briefly mention two verbal matters. Should the effect be called "quantum Zeno"? The original Zeno paradox was expressed in different forms [14, p. 394; 19, p. 308] both of which are based on the difficulty of building up an idea of "motion" from a series of instantaneous snapshots. The quantum Zeno effect is based on the idea of measurement freezing change. Thus there are rather superficial similarities but rather more deep-seated differences between the sets of ideas. More fundamentally, Newton's second law is second order in time, while Schrödinger's equation is first order, so there may certainly be no direct comparison of these processes. Provided this is recognised, the use of similar terminology should do no harm. Some prefer the name "watched-pot effect" or "paradox" [58].

Should the effect be described as a "paradox"? In part this is a purely verbal matter. The word "paradox" may be defined as "contrary to accepted opinion" or "exposing a logical fallacy" [19, p. 225], and this ambiguity in definition makes discussion of whether EPR, Schrödinger's cat, and so on are really "paradoxes" rather devoid of meaning. It should be admitted, though, that the effect predicted by those who initially worked in the area [1, 2, 59, 60] was strange and difficult to believe. What seems important, and this will be emphasised later in this paper, is that the strangeness must not be diluted by extending the name "quantum Zeno effect" to effects which are analogous in some ways, but not particularly startling.

9. INITIAL TIME-DEPENDENCE IN CLASSICAL AND QUANTUM PHYSICS

We first consider classical physics and the evolution of populations in a very general context. One specific case would be chemical reactions [61]. Another would be the evolution of populations in magnetic resonance [62] under the constraints of a steady field along one axis, and an oscillating field along another; it may be natural to discuss explicitly, not populations, but some property of the system, such as magnetism, which gives, in fact, analogous information.

In such cases, leaving out specific details, the behaviour of the system as a function of time is expected to be a sum of exponential terms relating to various rate processes, tending towards a situation of thermodynamic equilibrium, or, in the case of magnetic resonance, a pseudo-equilibrium, or thermodynamic equilibrium in a suitably rotating frame. Initially, then, changes in populations or related parameters will be proportional to \( t \), or, in other words, the initial rate of change will be independent of time.
Along with this exceedingly general remark, we discuss a very specific situation. A population level or chemical state has occupancy \( P_0 \) equal to unity at \( t = 0 \). Subsequently it decays via a simple rate relation

\[
\frac{dP_0}{dt} = -\lambda P_0.
\]  

We assume there is no regeneration; the states to which the system has decayed cannot in turn decay to the initial state. The solution to Eq. (9.1) is, of course trivial:

\[
P_0(t) = P_0(0) \exp(-\lambda t).
\]  

This simple exponential behaviour has the semigroup property

\[
P_0(t_1) P_0(t_2) = P_0(t_1 + t_2),
\]  

so that decay for a time \( t_1 \), followed by an interruption (or measurement), followed by decay for a further period \( t_2 \), is equivalent to uninterrupted decay for a period \( t_1 + t_2 \).

This last point makes this classical behaviour very easy to appreciate and understand. Thus it was particularly pleasing for physicists in the years after 1900 when it seemed that all radioactive processes were precisely exponential and could, thus, be described as classical stochastic processes driven by equations of the form of the form of Eq. (9.1).

Unfortunately it is well known that theory does not totally back up these experimental results. For short times, the decay will usually be a quadratic function of \( t \) [63]; for long times, it will be an inverse power law, possibly with an oscillation [64]. In between there may be a long period for which the exponential law is followed to a good approximation [65]. Peres [66] has provided a detailed survey of the whole problem, while Nakazato et al. [67] have recently given a full discussion.

While radioactive decay is, of course, extremely important, and is the area where the quantum Zeno effect was originally discussed, we stress that it is merely one example of a much more general—practically universal—result. (Indeed Fearn and Lamb [16] suggest that it is not perhaps a very good example, because a rigorous treatment would require quantum field theory.)

The general result concerns the initial time-dependence of a state in circumstances where it may make transitions to, or decay to, a second state; the initial probability of transition or decay is proportional to \( t^2 \), so the relevant rate is itself proportional to \( t \).

A simple proof is as follows [2; 14, p. 416]. If \( \phi_+ \) is the initial state of the system, then after a time \( t \) it will have evolved into \( \exp(-iHt/\hbar)\phi_+ \). The probability of survival is thus

\[
P_+(t) = |\langle \phi_+ | \exp(-iHt/\hbar) | \phi_+ \rangle|^2 = 1 - (\Delta H)^2 t^2/\hbar^2 \ldots
\]
where

\[(\Delta H)^2 = \langle \phi_s | H^2 | \phi_s \rangle - (\langle \phi_s | H | \phi_s \rangle)^2. \tag{9.5}\]

It is assumed that \( H \) does not depend explicitly on time.

A more powerful result is known as Fleming’s rule [68]. It states that

\[P_s(t) \geq \cos^2\left(\frac{(\Delta H) t}{\hbar}\right) \quad (t < \pi\hbar/2\Delta H). \tag{9.6}\]

A convenient proof [69] starts from the generalised uncertainty principle [70, p. 68]

\[\Delta \hat{A} \Delta \hat{B} \geq \frac{i}{\hbar} \langle [\hat{A}, \hat{B}] \rangle, \tag{9.7}\]

where

\[(\Delta \hat{A})^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2 \tag{9.8}\]

and similarly for \( \hat{B} \). With \( \hat{B} \) equal to \( \hat{H} \), the total Hamiltonian of the system, and using

\[i \hbar \frac{d}{dt} \langle \hat{A} \rangle = [\hat{A}, \hat{H}], \tag{9.9}\]

we obtain

\[\Delta \hat{A} \geq [\hbar/2(\Delta H)] |d\langle \hat{A} \rangle/dt|. \tag{9.10}\]

With

\[\hat{A} = |\phi_s\rangle \langle \phi_s| \tag{9.11}\]

we have

\[\langle \hat{A} \rangle = \langle \hat{A}^2 \rangle = P_s(t). \tag{9.12}\]

Then

\[\left[ P_s(1 - P_s) \right]^{1/2} \geq [\hbar/2(\Delta H)] |(dP_s/dt)|. \tag{9.13}\]

We then perform the integral in the inequality

\[\int_0^1 dP_s/\left[ P_s^2(1 - P_s) \right]^{1/2} \leq 2(\Delta H) t/\hbar \tag{9.14}\]

to obtain Eq. (9.6). Fleming’s rule is significant for small values of \( t \), for which the \( \cos^2 \) decay must be slower than any decay proportional to \( t \). Also, from Eq. (9.13), recognising that \( P_s = 1 \) when \( t = 0 \), we obtain \((dP_s/dt) = 0 \) when \( t = 0 \), and hence there can be no linear term in \( P_s(t) \).
While our language of “survival” and “decay” is still reminiscent of radioactive decay, it is stressed that the context may be much wider. Consider, for example [7, 71], a spin-$\frac{1}{2}$ particle such as a neutron, in a magnetic field along the $z$-axis, with two states $\zeta_+$ and $\zeta_-$. If the neutron is in state $\zeta_+$ at $t=0$, we may identify

$$\zeta_+ = \phi_s; \quad \zeta_- = \phi_d,$$

although the $s$ and $d$ subscripts, signifying “surviving” and “decayed” must not be taken as indicating a “one-way” process; at times the coefficient of $\phi_s$ must increase and that of $\phi_d$ decrease. If spin-flips are induced between $\zeta_+$ and $\zeta_-$ via a magnetic field along the $x$-axis, the equation for the dependence of the system may be written as

$$i\hbar \frac{d}{dt} \begin{bmatrix} a_s \\ a_d \end{bmatrix} = \begin{bmatrix} \varepsilon & V \\ -V & -\varepsilon \end{bmatrix} \begin{bmatrix} a_s \\ a_d \end{bmatrix}$$

(9.15)

with

$$\varepsilon = |\mu| B_0/2$$

(9.16)

$$V = |\mu| B_1/2,$$

(9.17)

where $\mu$ is the magnetic moment of the neutron and $B_0$ and $B_1$ are the magnetic fields along $z$- and $x$-axes.

Then, with the initial condition stated, the time-dependence of the system may be written as

$$|\psi(t)\rangle = a_s(t) |\psi_+\rangle + a_d(t) |\psi_d\rangle$$

(9.18)

with

$$a_s(t) = \cos(\omega_0 t) - (i\varepsilon/\hbar \omega_0) \sin(\omega_0 t)$$

(9.19)

$$a_d(t) = - (iV/\hbar \omega_0) \sin(\omega_0 t)$$

(9.20)

where

$$\omega_0 = (1/\hbar)(\alpha^2 + V^2)^{1/2}.$$  

(9.21)

Then

$$P_s(t) = 1 - (V^2/\hbar^2 \omega_0^2) \sin^2(\omega_0 t).$$

(9.22)

For small times,

$$a_s(t) = 1 - i\epsilon t/\hbar - 3\omega_0^2 t^2/2\cdots$$

(9.23)

$$a_d(t) = - iVt/\hbar \cdots$$

(9.24)

$$P_s(t) = 1 - (V^2/\hbar^2) t^2 \cdots.$$  

(9.25)
The behaviour is, of course, coherent, Eq. (9.25) demonstrates clearly the initial $t^2$-dependence, and Eq. (9.22) both agrees with the Fleming inequality and shows that one may move to the limit of that inequality as $\varepsilon \to 0$, since

$$AH = [(\varepsilon^2 + V^2) - \varepsilon^2]^{1/2} = V$$ (9.26)

A related example [4, 12] consists of an ion in a magnetic field, in which it has two energy levels, with eigenfunctions $\phi_1$ and $\phi_2$, and energies $E_1$ and $E_2$. If the wave-function is equal to $\phi_1$ at $t = 0$, and a perturbation is then applied with angular frequency $(E_2 - E_1)/\hbar$, then we may identify

$$\phi_1 = \phi_s; \quad \phi_2 = \phi_d$$ (9.27)

and we may use (9.18) with

$$a_s(t) = \cos(\Omega t/2)$$ (9.28)

$$a_d(t) = \sin(\Omega t/2),$$ (9.29)

where $\Omega$ is the Rabi frequency, the product of $\gamma$, the magnetogyracic ratio of the ion, and $B$. Clearly the $t^2$-decay and the obedience to Fleming’s rule follow as in the previous example.

Having said that the $t^2$-dependence is very general, we should mention that it is not universal. There are assumptions in the proof of Eq. (9.4), especially that $\phi_s$ belongs to the domain of definition of $H$; it is thus assumed that $(\Delta H)$ is finite.

A recent paper [72] studies a solvable dynamic model in the weak-coupling limit and obtains exponential behaviour at all times. But it is the case that an initial decay proportional to $t^2$ should be regarded as the quantum mechanical norm.

There is, thus, a fundamental mismatch between quantum physics and classical physics, both in the very specific point of the $t^2$-dependence versus the $t$-dependence for short time, and in the more general point that genuine decay-like processes are exponential in the classical case, but not in the quantum one. For the latter point, it is important that, even in the intermediate quantum region, which one may speak of as the exponential region, the notion of exponential decay is always an approximate one. This is very important for the theory of the quantum Zeno effect.

Theoretically, Peres [66] is only prepared to say that the argument shows that in this intermediate region, the exponential law may be approximately valid. Experimentally, however, it is difficult to obtain departures from the exponential. The $t^2$-region is too short, in the final region too few species probably survive to allow accurate study, and in the intermediate region it does seem that conditions for near exponential behaviour are obeyed, although Peres points out the exception of neutral kaon decay [73], where there are strong deviations from the exponential law.

In terms of calculations, while existence of the $t^2$-region is easy to demonstrate in principle, in practice it is difficult to estimate its duration. Recently, though, Serot et al. [74] have produced beautiful computations of $\pi$-decay from $^{212}P_{0}$; the
results show a $t^2$-dependence of decay (rate of decay proportional to $t$) for about $10^{-21}$s. Following a very short period of adjustment, one then moves to a region of $t$-dependence of decay (constant rate of decay).

Last, in this section we consider Fermi’s golden rule [69], as it is often felt that this constitutes an exception to the $t^2$ rule. In usual notation, one usually works towards the Fermi rule from the formula for a time-dependent perturbation

$$P(t) = (1/h^2) \left[ \int_{-\infty}^{\infty} |\langle m| V|k\rangle|^2 \rho_k \times \sin^2[\{(E_k - E_m - \hbar\omega)/2\Omega\} t/2\hbar]\right] dE_k$$

(9.30)

which gives the probability of transitions from state $m$ to a group of states $k$ occurring with density $\rho_k$, under a perturbation $V$.

The standard textbook treatment argues that, since $(\sin^2 \theta)/\theta^2$ is small apart from over a very restricted range of $\theta$, it is possible to consider $\rho_k$ and the matrix-element constant over this range. One thus reaches the expression

$$P(t) = 2\pi \rho_k |\langle m| V|k\rangle|^2 t/\hbar$$

(9.31)

with a proportionality to $t$ that satisfies classical preconceptions.

However, this argument is itself an argument for long times, in the sense that the shorter $t$ is, the larger the range of $(E_k - E_m - \hbar\omega)$ for which the integrand in Eq. (9.30) remains large, and the approximation that $\rho_k$ and the matrix-element constant must break down at some point.

A suitable short time approximation is, in fact,

$$P(t) = (1/h^2) \left[ \int_{-\infty}^{\infty} |\langle m| V|k\rangle|^2 \rho_k dE_k \right] t^2,$$

(9.32)

giving, of course, a $t^2$-dependence.

10. ANALOGY BETWEEN $t^2$ AND $\theta^2$ FACTORS

Let us imagine a spin-$\frac{1}{2}$ particle in the state $|sz = \frac{1}{2}\rangle$ or just $|+\rangle$. Clearly a measurement of $sz$ will give the result $\hbar/2$ with probability unity. If, though, we measure $s_{\theta z}$, the spin component along a direction displaced by $\theta$ from the $z$-direction, the probability of obtaining the result $-\hbar/2$ is $\sin^2(\theta/2)$. For small $\theta$ this is proportional to $\theta^2$.

Similarly, we may consider the two-particle singlet state which may be written as

$$\psi = (1/\sqrt{2})(|+\rangle - |\rangle - |+\rangle)$$

(10.1)

in an obvious notation. Measurements of the components of the two spins along the same direction will certainly yield perfect anticorrelation. If, though, one of the
directions is rotated by $\theta$, the possibility of positive correlation becomes again $\sin^2(\theta/2)$, proportional to $\theta^2$ for small $\theta$.

This dependence on $\theta^2$ follows from the identity \[1 - |\langle \psi_1 | \psi_2 \rangle|^2 \leq \| \psi_1 - \psi_2 \|^2. \] (10.2)

Broadly we may contrast the $\theta^2$-dependence with the $\theta$-dependence that may be described as the “normal” behaviour for models which are classical or semiclassical or “realistic.” (The word “normal” is used in a sense which will be strictly qualified for some cases, but which may be made firm in others.)

To introduce the ideas, let us consider the case of a spin-$\frac{1}{2}$ particle in a state with $s_z = \frac{1}{2}$ and a requirement to measure any spin component. A natural (in some sense) hidden variable model of this state might be to represent an ensemble of systems by an array of vectors of constant density throughout the hemisphere with $z$ positive. A measurement of any component of spin would give a result of $+h/2$ for a system represented by a vector with positive component along that direction and $-h/2$ if the component is negative. The model clearly gives the correct quantum expressions for the rather trivial cases of $\theta$, the angle the measurement direction makes with the $z$-axis, equal to $0$, $\pi/2$, or $\pi$, the probabilities of getting a negative result being 0, 1/2, and 1, respectively. However, for general $\theta$ the probability is $\theta/\pi$, compared with the quantum probability of $\sin^2(\theta/2)$. For small $\theta$, the $\theta^2$-dependence required by quantum theory is not matched by the $\theta$-dependence of the hidden variable (semiclassical) approach.

Of course this does not mean that a hidden variable approach is not available for this problem; historically it was Bell's provision of such a model (in 1964, although the paper containing the result was not published until 1966 [75]) that made totally clear the fact that von Neumann's “proof” of 1935 [20] that hidden variable models were impossible was quite wrong.

While Bell's model was initially expressed in algebraical terms, he later expressed it geometrically [76], using the kind of model we have been discussing. Essentially, as compared to our previous model, the vector giving the direction of the measurement is rotated towards the $z$-axis until it makes an angle $\theta'$ with the $z$-axis, where $\theta'/\pi = \sin^2(\theta/2)$. (10.3)

It is $\theta'$ rather than $\theta$ that is used for dividing the vectors into the two categories. One may say that the classical model is “squashed” to make it give the correct quantum results and in that sense is somewhat contrived or artificial.

Let us now move to the two spin-$\frac{1}{2}$ system, and the state of Eq. (10.1). Peres [77; 14, p. 160] has described an analogous classical situation in which a bomb explodes into two asymmetric parts with equal and opposite momenta, $J_1$ and $J_2$. Observers 1 and 2 measure whether $J_1$ and $J_2$ have positive or negative components along arbitrary directions $r_1$ and $r_2$. A positive/negative component is given the value $+1/-1$, and a correlation coefficient may be calculated. Clearly if $r_1$ and $r_2$ are in
the same direction, the (anti-)correlation is perfect, or in other words the correlation coefficient is $-1$. If $\theta$ is the angle between $r_1$ and $r_2$, then, for arbitrary $\theta$, the coefficient is equal to $-1 + 2\sin(\theta/\pi)$, and clearly the dependence on $\theta$ goes as $\theta$ itself, not as $\theta^2$ as quantum theory demands. The quantum theory value for the coefficient is just $-\cos \theta$, varying as $\theta^2$ at the value $\theta = 0$.

Let us construct a hidden variable model for this situation. The natural choice will be a set of pairs of equal and opposite vectors, 1 and 2, all vectors commencing at the origin, with the distribution of vectors uniformly dense over the whole sphere. If measurements are made of the components of spin along directions $r_1$ and $r_2$, and for a particular pair of vectors, the results will be $+\hbar/2$ or $-\hbar/2$ for the first measurement, depending on whether vector 1 has a positive or negative component along $r_1$, and analogously $+\hbar/2$ or $-\hbar/2$ for the second measurement, depending on whether vector 2 has a positive or negative component along $r_2$.

Clearly if $r_1$ and $r_2$ are in opposite directions (the EPR case) we will always get different answers for the two measurements (perfect anticorrelation). But if one of the directions is rotated by an angle $\theta$, the fraction of positive matchings is $\theta/\pi$, compared with the usual quantum $\sin^2(\theta/2)$. As usual the classical dependence is on $\theta$ and the quantum dependence on $\theta^2$.

In this case, unlike that of a single spin, there can be no “squeezing.” Whereas for the single spin case, the $z$-axis was special, because the spin was in an eigenstate of $s_z$, in this two-spin case there is no special axis about which any squeezing might be performed. At least this is the case if one insists on locality; as Bell points out [76], in principle one may use the direction of measurement of the second spin to “adjust” the model to give correct quantum results for the first spin, but, of course, introducing nonlocality in so doing.

If one does retain locality, though, any hidden variable model must be “unsqueezed” or spherically symmetric. Of such models, the classical model of Peres [77] and the “natural” hidden variable model just considered give a $\theta$-dependence.

For the case of the latter, the dependence is a result of the vector nature of the model, the constant density of the vectors reflecting the spherical symmetry of the model.

For the vector model, the following argument is relevant. Let us start from the same measurement direction for both vectors of a pair and, thus, perfect (anti-)correlation of results. Now let the measurement direction of the second vectors be rotated by (small) angle $\alpha$. There will thus be a probability $P(\alpha)$ that a different measurement result will be obtained for the second vector than had $\theta$, the difference between the measurement directions for the two vectors, been zero, and the same probability that there will be a correlation rather than an anticorrelation.

Let us now imagine $\theta$ being increased by a further angle $\alpha$, so that the measurement direction for the second vector makes an angle $2\alpha$ with that for the first set. Because of the constant density of the vectors, the probability that a different measurement result will be obtained for $\theta$ equal to $2\alpha$ than if it were $\alpha$, is again $P(\alpha)$. The specific nature of the model means that changes caused by the first rotation cannot be undone in the second rotation. (The only requirement for this is that
Thus the changes are cumulative, and the probability of a correlation becomes \(2P(\alpha)\). But this should just be the probability of a correlation caused by a single rotation of \(2\alpha\). Thus for this model

\[
2P(\alpha) = P(2\alpha).
\]

If we assume that for small values of \(\theta\),

\[
P(\theta) \sim \theta^n,
\]

then we obtain \(n = 1\), or \(P(\theta) \sim \theta\), as we already know.

That result, however, is specific to the vector model. Let us imagine a more general argument [78, 79] that produces a less specific result, but one still directly in conflict with the quantum mechanical \(\theta^2\)-dependence. Here we use only the spherical symmetry of the hidden variable structure. Starting again from anticorrelation, a rotation of the second measurement direction again introduces a probability \(P(\alpha)\) of correlation. (Herbert [78] relates \(P(\alpha)\) to an “error rate.”)

Spherical symmetry tells us that when \(\theta\) is increased from \(\alpha\) to \(2\alpha\), the probability of a different measurement result for the second set of vectors being caused is again \(P(\alpha)\). (Again we must resist any temptation to question this on the grounds that the first measurement is at \(\theta = 0\) which might seem to make the first change different from the second, because, of course, that is exactly what locality forbids us to take account of.)

However, in this general case, as distinct from the simple vector model, it is possible that some of the changes in measurement result caused by increasing \(\theta\) from \(\alpha\) to \(2\alpha\) undo changes caused by increasing it from \(0\) to \(\alpha\). The cumulative change, then, may be less than \(2P(\alpha)\), and Eq. (10.4) must be generalised to

\[
2P(\alpha) \geq P(2\alpha).
\]

If again we assume Eq. (10.5) for small \(\theta\), we obtain [77, p. 440]

\[
0 \leq n \leq 1
\]

again, of course, in total contrast to quantum theory for which \(n\) must be 2.

Equation (10.6) may be described as a special case of Bell’s inequality, and inequalities of this type were obtained from more orthodox proofs of Bell’s theorem by Wigner [80]. In a paper titled “Non-locality from an Analogue of the Quantum Zeno Effect,” Squires et al. [81] have reviewed the arguments fully and rigorously.

In this section, then, we have developed a general analogy between the two fundamental discords between classical (or semiclassical) and quantum physics. By emphasising the general nature of the \(\theta^2\)-dependence in quantum theory, we have shown that this fundamental aspect of quantum Zeno processes has a broad significance, certainly not being restricted to radioactive decay. We have also shown that the quantum Zeno effect, rather than being considered intrinsically esoteric and isolated from more common problems of quantum measurement—EPR, Bell’s
Theorem and so on, is, in fact, very much connected in origin with the type of mathematics involved in producing the Bell inequalities.

Indeed, just as in our discussion in Section 9, where we described the very specific contrast between the initial $t$-dependence of classical physics and the initial $t^2$-dependence of quantum theory, and also the more general contrast between the usual exponential classical result, and a range of nonexponential quantum forms, so here we may discuss the very specific contrast between $\theta^2$- and $\theta$-dependences, but also general calculations for general values of $\theta$, where Bell’s theorem shows that quantum theory does not respect the results of local hidden variable theory.

While the Bell work commenced in the general region [76], only subsequently discussing the more specific point [78, 80], the quantum Zeno discussions have usually concentrated on the initial dependence of decay. A recent paper, though, attempts to explore theoretically a larger region of time, considering the quantum mechanical two-state oscillator and examining whether the predictions of quantum theory may be mimicked by a realistic model in which an individual system is at any instant definitely in one or the other state. This was found to be possible in the absence of measurement, provided transition rates varied with time, but any measurement was necessarily invasive.

So broadly analogous features may be found in the two areas of study, although it is not, of course, implied that the parallels in concepts or experiment are direct. The fundamental cause of the conflict between classical and quantum is the same in the two areas; contrary to “undounded quantum superstition,” as Peres [14, p. 162] puts it, correlations between adjacent states, whether separated in $t$ or $\theta$, are always greater for the quantum case than for this classical one.

11. THE QUANTUM ZENO EFFECT

The basic structure of the quantum Zeno effect, as described in the original papers [1, 2, 59, 60] is extremely well known. We consider an array of radioactive nuclei and a decay rate proportional to $t^n$ for short times, so that the probability of survival to time $t_0$ is

$$P_s(t_0) = 1 - k t_0^n.$$ (11.1)

We now imagine, though, that at $t = t_0/2$, a measurement of survival or decay is made. (We postpone, for the moment, discussion of what that might mean in practice.) At time $t_0/2$ clearly the probability of survival is given by

$$P(t_0/2) = 1 - k(t_0/2)^n.$$ (11.2)

If (temporarily and for convenience) we use the projection postulate, immediately after the measurement we must restart the decay process for the surviving nuclei. If
we now allow decay for a further period of $t_0/2$, we find the survival probability will be

$$P_s(t_0) = [1 - k(t_0/2)^m]^2$$  \hspace{1cm} (11.3)

$$= 1 - k t_0/m + k^2(t_0/2)^m.$$  \hspace{1cm} (11.4)

Assuming $t_0$ is small, the last term may be ignored, and we deduce that, compared to Eq. (11.2), the probability of decay, given for the two cases by $1 - P_s(t_0)$ and $1 - P'_s(t_0)$, respectively, has been multiplied by $1/2^{m-1}$.

For $m = 1$, the probability of decay is unchanged. This is not surprising, as this case, for which $P_s(t_0)$ is equal to $1 - k t_0$ for small $t_0$, may be said to be the small-$t_0$ form of the exponential, $\exp(-k t_0)$, for which, as has been stressed already, the semigroup law holds, so the intermediate measurement can have no effect.

If $m > 1$, a range of values which includes, of course, the typical quantum case of $m = 2$, the probability of decay at time $t_0$ is reduced by a factor of $2^{m-1}$, or just 2 if $m = 2$, by the intermediate measurement. If the number of (equally spaced) measurements in period $t_0$ is increased to $n$, the probability of decay is reduced by $n^{m-1}$, or just $n$ for the case $m = 2$. As $n$ tends to infinity, the mathematics tells us that the probability of decay is reduced to zero. This is the bare bones of the quantum Zeno effect.

If, though, $m < 1$, clearly the probability of decay is increased by the intermediate measurements; indeed the rate of decay is predicted to tend to infinity as $n$ tends to infinity. This is not, as far as we know, a relevant case physically, but this argument does demonstrate a misleading comment in Ref. [16]. In that paper it is noted that the projection postulate, taken together with the fact that the measurement is of the first kind, means that, if a particle is found not to have decayed, its wave-function will be restored to the undecayed state. The quantum Zeno argument is then said to rely on the fact that this effect could be enhanced by making measurements in rapid succession.

But such a suggestion ignores the balance of probabilities. The projection postulate argument also suggests that the wave-function of a particle that is found to have decayed becomes that of a decayed particle. If $m$ is equal to 1, the classical case, the two effects cancel out. It is the fact that $m$ is 2 for the typical quantum case that leads to the quantum Zeno prediction.

What are the central features of the above argument? We have, of course, used the projection postulate, but later (Section 13), we shall show that such use is convenient but not required to obtain the quantum Zeno result. We have also used the fact that the measurement is of the first kind. Indeed the same mathematics will apply to any system where there is a measurement of the first kind (provided, of course, there is a suitable value of $m$), and we have discussed a range of such systems in Section 7. Equally, of course, the argument will not work for a system where measurements are not of the first kind. But, contrary to the implications of Ref. [16], there are enough of the former type of measurement to make study of the quantum Zeno suggestion interesting.
The fact that many systems may display effects \textit{mathematically} similar or even directly analogous does not mean that all such effects are \textit{conceptually} analogous. In the following section we consider what makes the original quantum Zeno argument rather special and particularly makes the use of the word “paradox” not unreasonable.

12. WHAT IS “PARADOXICAL” ABOUT QUANTUM ZENO?

The quantum Zeno prediction as originally put forward \cite{1, 2, 59, 60} was felt to be a strange prediction, difficult to accept, perhaps “paradoxical.” We now proceed to examine what makes the prediction so puzzling. In other words, the question is—what is the \textit{essence} of the quantum Zeno effect or “paradox”? This is an important question, because, as already noted, many theoretical suggestions and some experimental papers have claimed to be discussing “quantum Zeno.” What particular features must a theory or experiment contain in order to justify such a claim?

Obviously a central part of such discussion is the use of the $t^2$ factor (or, at least in principle, $t^m$ with $m > 1$) to produce a slowing down of decay by measurement. A variant of the above, suggested by Pascazio and Namiki \cite{83} is that when the number of measurements is finite and the rate of decay \textit{is slowed}, we have the quantum Zeno \textit{effect}; it is only when the number becomes infinite and the decay is \textit{totally frozen} that we have the quantum Zeno \textit{paradox}.

This argument may then be tied in with suggestions \cite{83–86} that the frequency of measurements cannot, in fact, be infinite. The suggestions may come from simple physical arguments \cite{83}, use of the Heisenberg principle \cite{84}, the finiteness of the appropriate decoherence time within an environment-induced decoherence theory \cite{85}, or the instability under continuous measurement of the Fubini–Study metric defined on the projective Hilbert space of the quantum theory \cite{86}. These two suggestions may then be put together to suggest, in effect, that the “paradoxical” result is physically impossible.

But these arguments (or, perhaps, more precisely, these uses of words) seem to us to miss the central point about what is difficult to accept about the original quantum Zeno suggestion. The really “paradoxical” point is that a system of decaying atoms is predicted to have its decay affected by the mere presence, a macroscopic distance from the decaying system itself, of a macroscopic measuring device.

There are a number of strange factors involved. First, the predicted effect is non-local. Of course, from the Bohm theory, from the EPR argument, and from Bell’s work, we are well-used to nonlocality in quantum theory, but usually between microscopic systems and only subsequent to an initial interaction. In the quantum Zeno case, one of the systems is macroscopic and so, generally, if loosely, may be felt to be obliged to behave in a classical way and there is in any case no history of interaction to provide a coupled or entangled wave-function.
Furthermore, as has already been suggested, the property of being a measuring device should not be regarded as being fundamental; the behaviour of the device should in principle be explicable in terms of its constituent atoms, and it is difficult to see how the special quantum Zeno effects could emerge. (This argument should perhaps be qualified; a macroscopic device which detects electrons certainly does have a special relationship with an electron, but it is still not clear that something as bizarre as quantum Zeno should emerge from the relationship.)

Last, the prediction is of a negative-result variety [87–89]. In such a prediction, an experimental result is obtained not by the occurrence of a physical event, but by the absence of such an event. The conceptual difficulty of these experiments is that from the quantum mechanical point of view it must be accepted that the wave-function of the system has been changed by the measurement, but with no apparent interaction between measured and measuring systems.

In the quantum Zeno example, this is most obvious in the extreme case where decay is prohibited altogether. Clearly the detector affects the decaying system without itself being affected in any way in its capacity of measuring device. Even in the case where decay is merely reduced rather than eliminated altogether, so some detections are recorded, clearly, in a statistical sense, atoms are affected in the period prior to their decay by the presence of the detector, although there is no direct interaction between atom and measuring device in this period.

It is for this reason that we consider the essence of the quantum Zeno effect is that it is a nonlocal negative-result effect between a microscopic system and a macroscopically separated macroscopic measuring device, and we believe the term should best be restricted to such processes. It may seem to be heavy-handed to attempt to deny the term “quantum Zeno effect” or “paradox” to any experiment not fulfilling this condition. However we would suggest that any experiment or theory claiming to explain or explain away quantum Zeno that does not have this feature, does not in fact address the central dilemma of the founding papers, and so the claim should be considered liable to mislead.

Experiments where there is a direct interaction between measuring and measured systems, so that the presence of a $r^2$-region, together with the performance of repeated measurements, result in a freezing or, in practice, partial freezing of decay, seem conceptually far less challenging. The results are certainly of great interest, both in their own right and as a proof of the existence of the $r^2$-region, but it would scarcely be claimed that anything very strange, anything remotely “paradoxical,” has occurred.

It is of such experiments that Peres [14, p. 394] speaks when he says that the quantum Zeno effect “has nothing paradoxical.” He considers that the quantum system is merely “overwhelmed by the meters which continually interact with it.” But we do not consider such an explanation makes the original quantum Zeno predictions any easier to understand. We return to a more complete analysis of these experiments in Section 15.

Incidentally both Block and Berman [90] and the present authors [91] have previously made a rather different suggestion on terminology. This was that what
we have called here the “real” quantum Zeno effect could be termed the quantum Zeno paradox, while that resulting from direct interaction between measuring and measured systems could be termed the quantum Zeno effect. Whether this clarification could become general remains to be seen.

Let us now take the point of Pascazio and Namiki [83]. We accept, of course, that the total freezing of decay is the most dramatic, in some ways the most surprising, feature of the predictions. Any experimental demonstration of such an effect would deservedly attract considerable attention. Yet if this were done by continuous direct interaction between measured system and measuring device, such an effect would not, in reality, be difficult to understand or appreciate. It is the effect—of any nature—on a system by the mere presence of a macroscopic detecting device that is the possibly “paradoxical” element of quantum Zeno.

We might add that for these latter measurements, it is not nearly so clear as for measurements involving physical interaction that there is an upper limit to their frequency. Despite this, the great majority of experimental and theoretical work uses situations where measurement occurs at discrete intervals, and there may indeed be other reasons why a perfect quantum Zeno effect should not be expected; for example, the measurements may be only approximately of the first kind. There is every reason why what Peres and Ron [92] call a “partial Zeno effect” should be the norm.

A good example is the work of Fearn and Lamb [8]. These authors use measurements of position to study the quantum Zeno effect (or, more accurately, to obtain support for their conviction that such an effect cannot occur [93]). Their use of position measurements was obviously open to the problem described above, that, while in principle measurements of position would be expected to be of the first kind, in practice the Heisenberg principle means that accurate measurements of position lead to large uncertainties in momentum, and hence, the possibility of large deviations in position for a measurement repeated after a short time.

In fact, in Ref. [8], Gaussian meters were used to discover whether a particle lay in a particular range. The authors used a variety of meter widths, $\Delta x$, and times between measurements, $\Delta t$, and claimed to find no evidence for a quantum Zeno effect. Indeed they state that the opposite occurred; the system consisted of a double potential well, and it was found that measurement helped the particle to move across the barrier.

Their conclusions were questioned by Home and Whitaker [94], and by Gagen et al. [9]. The former authors pointed out that the type of measurements considered in Ref. [8] were not ideal for discovering a quantum Zeno effect, so absence of such would not, in fact, imply that the whole quantum Zeno concept was flawed. Nevertheless they also suggested that for some values of $\Delta x$ and $\Delta t$, what they called a “weak quantum Zeno effect” might be predicted, and they demonstrated that Fearn and Lamb’s results did indeed show clear signs of such an effect.

Gagen et al. [9] showed that, for certain parameter regions, measurements similar to those of Fearn and Lamb do produce “strong evidence of the quantum Zeno effects.” Yet Fearn and Lamb dismiss these results on the grounds that the
idea of a “weak quantum Zeno effect” is not mentioned in Ref. 2 [93], and that, even in the results of Gagen et al. [9], the system is not totally frozen. Such arguments seem totally unconvincing. It is not necessary for a decay to be frozen; any influence on it by a macroscopically separated measuring device is strange enough to qualify as a quantum Zeno effect (or “paradox”).

Indeed we may push the point somewhat further. Until now we have assumed we are working in a \( t^2 \)-region, and one is thinking of a considerable slowing, if not actual freezing, of decay. Yet we may recall from Section 9 that it is only for a precise exponential decay that an intermediate measurement would be predicted to have no effect at all on the decaying system, and, of course, even though decays are approximately exponential once one has traversed the initial \( t^2 \)-region, they are not identically so. Thus we may define a generalised quantum Zeno effect [95], referring not just to the rather spectacular decrease in, or total elimination of, decay, but to any effect, however small, of a measuring device on a macroscopically separated quantum system. It is stressed that such an effect is always present.

In a sense this argument shifts the discussion from total concentration on a search for an experimental demonstration of the effect, to conceptual study of the fact that the standard laws of quantum theory, together with the normal accepted rules of quantum measurement theory, tend inevitably to an effect which appears very difficult to accept or understand.

13. RELEVANCE OF THE PROJECTION POSTULATE

As already stated, the original discussions of the quantum Zeno effect used the projection postulate. Of course, from a logical point of view this implied only that the postulate (together with other arguments) was a sufficient condition, not that it was a necessary condition, so occurrence of the effect would certainly be no proof of the truth of the postulate.

Despite this, it would be fair to say that, until comparatively recently, it has been almost universally accepted that the prediction of the quantum Zeno effect relied on the projection postulate. (See, for example, Ref. [16].) To deny the postulate was to deny the effect; to accept the effect was to accept the postulate.

Perhaps the clearest manifestation of this belief is in Ballentine's well-known text [79, p. 237]. Ballentine, a great supporter of an ensemble interpretation and opponent of the projection postulate [25], describes the argument leading to the quantum Zeno effect as “amusing” but “of course, false.” For Ballentine, the prediction depends on the projection postulate; he argues that, because continuous measurement does not, in fact, prevent motion, both the quantum Zeno effect and the projection postulate are disproved. He further suggests that this argument disproves the common suggestion that rival interpretations of quantum theory cannot be experimentally distinguished.

An interesting development was initiated by an extremely important paper by Inagaki et al. [7] from Namiki's group in Waseda University. This paper discussed
how a quantum Zeno effect—a genuine effect, as defined in the previous section—could be obtained using measurements involving neutron spin-flipping. This was a context in which the $t^2$-dependence was clear and obvious and emerged much more easily from the basic physics than for the case of radioactive decay; the paper thus helped to emphasise the general nature of quantum Zeno.

On the question of the projection postulate, this paper argued in favour of the essential part the postulate played in demonstration of the quantum Zeno effect. The paper aimed “to clarify how [the effect] is deeply rooted in the wave function collapse at every step.” Indeed much of the discussion was an attempt to clarify the experiment of Itano et al. [12], which will be described in Section 15. The worry of Ref. [7] was that Itano et al. did not perform a detection at every stage of their experiment. Inagaki et al. [7] analysed two types of experiments with neutron spins, those which did require a collapse at every stage and those that, in their view, did not. They discovered that the results were identical, which they described as a “very peculiar property.”

The present authors had long been convinced that the projection postulate was not required in order to obtain a quantum Zeno effect. There were three strands to this argument. First was detailed analysis of the problem with interpretations which do not require a collapse—the ensemble interpretation [69] and the many worlds interpretation [96]. This suggested that the same effective separation between decayed and undecayed parts of the wave-function that is provided by collapse in the earlier derivations of the effect, is provided by the very act of measurement in the former case, the separation of the various worlds in the latter.

From a more general point of view, the same result is obtained for any interpretation of quantum theory by introducing explicitly the wave-function of the measuring device [71, 91, 97], a point of view which, in the general case, has, from the very first days of quantum theory, been regarded as fundamental.

The analysis uses the simplest type of system, where quantum Zeno results may be expected, “surviving” and “decayed” states being represented by wave-functions $\phi_s$ and $\phi_d$ with corresponding energies $E_s$ and $E_d$. For decaying atoms this will be correct in the area we are concerned with, the $t^2$-region; in the region of exponential decays of course, one would have to use a continuum of states for the decayed system. For experiments using neutron spins [7] or atomic ions [12] which are genuinely two-level, the formalism will be correct, although the terms “surviving” and “decayed” should not be taken too literally. There are merely two states of equal status but differing in energy, and in particular one should not assume that an individual atom may not make a “regeneration” transition from “decayed” to “surviving” even after a measurement.

We take a convenient zero of energy at $(E_d + E_s)/2$, with $E_s = -E_d = \varepsilon$ with respect to the new zero, and write the combined wave-function of observed system and measuring device up to the time of the first measurement as

$$\Psi(t) = \{a_s(t) \phi_s + a_d(t) \phi_d\} \psi,$$  \hspace{1cm} (13.1)
where $\psi$ relates to the measuring device. The time-evolution of the system is then as in Eq. (9.15), where the off-diagonal matrix-element, $V$, may be taken to be real for convenience. With $a_s = 1$ and $a_d = 0$ at $t = 0$, the time-dependence is given by Eqs. (9.19)–(9.25); these equations show clearly that the decay probability for small $t$ is proportional to $t^2$ as required.

If no measurement is performed at $t_0$, and the system evolves for a further period of $t_0$, there will be two contributions to the coefficient of $\phi_s$ in $\psi(2t_0)$. The first comes from further evolution of $a_s(t_0)$, according to Eq. (9.15). This contribution to $a_s(2t_0)$ is $[a_s(t_0)]^2$. There is an additional contribution corresponding to evolution from $a_d(t_0)$ to $a_s(2t_0)$, again according to Eq. (9.15). Solution of this equation with initial conditions $a_s = 0$ and $a_d = 1$ shows that this contribution to $a_s(2t_0)$ is equal to $[a_d(t_0)]^2$. This second contribution may be described as a regeneration term, but since it is negative, as $a_d(t)$ is imaginary in Eq. (9.24), its effect is to reduce, rather than to increase, $P_s(2t_0)$.

It is easy to see that

$$[a_s(t_0)]^2 + [a_d(t_0)]^2 = a_s(2t_0),$$

(13.2)

as given by Eq. (9.23) but with $t_0$ replaced by $2t_0$, and, of course, this is merely required for consistency. In particular,

$$P_d(2t_0) = 1 - (V^2/\hbar^2)(2t_0)^2 \ldots$$

(13.3)

If, though, the regeneration term is removed by any means, then

$$P_d(2t_0) = |a_s(t_0)|^4 = 1 - (1/2)(V^2/\hbar^2)(2t_0)^2$$

(13.4)

so, to order $t_0^2$, the decay probability is reduced by a factor of 2, which is, of course, the mathematics of the quantum Zeno effect.

Suppression of the regeneration term may be achieved by any means of effectively separating $\phi_s$ and $\phi_d$ at time $t_0$. An obvious way is to perform a measurement at $t_0$, and to impose a collapse at that time, as in the original treatments of quantum Zeno.

However, collapse is not the only way of achieving this effect. Restricting ourselves for the moment to the case of radioactive decay, and with a measurement at $t_0$, but now taking explicit account of the measuring device, we have

$$\psi(t_0) = a_d(t_0) \phi_s + a_s(t_0) \phi_d \psi_d,$$

(13.5)

where $\psi_s$ and $\psi_d$ are wave-functions for the measuring device corresponding to the presence or absence of a permanent record of decay having taken place.

Now we let the combined system represented by Eq. (13.5) at time $t_0$ continue to evolve. Since $\psi_d$ represents a permanent record, it is clear that the second term in Eq. (13.5) cannot evolve to produce a term representing survival. Thus the regeneration term is inoperative, Eq. (13.4) is obtained, and a quantum Zeno result is predicted. It is stressed that the result as obtained so far would not necessarily
apply to other systems, where it is not obvious that a measurement of “decay” at
time $t_0$ is incompatible with a measurement of “survival” at time $2t_0$.

Indeed G. Kunstatter and I. D. Laurie (unpublished and private communication)
have discussed the further evolution of the second term in Eq. (13.5). They argue
that, to explain fully the noncontribution of this term to “survival” at time $2t_0$, one
must introduce the radiation field corresponding to the decay particle.

There is, though, a more general argument [71] for the quantum Zeno result
without collapse—the third strand of the argument mentioned earlier in this sec-
tion. Let us start from Eq. (13.5) and allow the system to evolve for a further time
$t_0$. We obtain

$$
\Psi(t_0) = [a_s(t_0) \phi_s + a_d(t_0) \phi_d] \psi_s 
+ a_d(t_0) a^*_s(t_0) \phi_d \psi_s + [a_d(t_0)]^2 \phi_s \psi_d.
$$

(13.6)

The last term is the one liable to cause awkwardness for a further measurement
at time $2t_0$, at least for the case of radioactive decay; it would suggest that a
measurement of decay at $t_0$ is followed by a measurement of survival at $2t_0$. For
other cases, it might seem that the term might demonstrate the nonexistence of the
quantum Zeno effect, but this is not the case.

All the terms in Eq. (13.6) are mutually orthogonal. Thus, in any contribution to
$P_s(2t_0)$ and $P_d(2t_0)$, they act individually; there may be no interference between
the second and third terms, or between the first and last terms, as there would be if the
wave-functions of the measuring device were not included. When one calculates the
contributions to $P_s$, that from the first term is of order unity, those from the second
and third terms are of order $t^2$, while that from the fourth is of order $t^4$. Since we
require working only to order $t^2$, the fourth term may be omitted without further
discussion, leading to a quantum Zeno prediction. The probability of survival is
given by $P_s(2t_0)$ as in Eq. (13.4), rather than by $P_d(2t_0)$ as in Eq. (13.3).

There must, of course, be a compensating effect in the probability of decay; the
fact that the second and third terms of Eq. (13.6) contribute individually to the
probability of decay, rather than jointly, means that the probability of decay at
time $2t_0$ is given by

$$
P_d(2t_0) = |a_s(t_0) a_d(t_0)|^2 + |a_d(t_0) a^*_s(t_0)|^2
$$

(13.7)

and

$$
P_d(2t_0) = 2V^2t^2/\hbar^2,
$$

(13.8)

rather than as $1 - P_s(t)$ from Eq. (9.25), in agreement, of course, with Eq. (13.4),
and this is another statement of the quantum Zeno result.

Over the last few years, several papers have been published supporting the view
that the quantum Zeno effect does not require the projection postulate. Several
papers by Namiki and co-workers (including Pascazio and Rauch) [83, 98–101],
for example, have argued that the quantum Zeno effect is a purely dynamical process, always governed by strictly unitary evolution (which is to say, no collapse).

Various other authors have recently discussed or questioned the relevance of the projection postulate for quantum Zeno. Altenmuller and Schenzle [10] have discussed the effect for position measurements using what they call real measurements rather than the projection postulate. Beige and Hegerfeldt [101] have discussed the applicability of the projection postulate for the results of Itano et al. [12], concluding that the postulate, although not required to obtain the results, acts as an excellent pragmatic tool for a quick and simple, although approximate, understanding of the results. Luis and Perina [102] have also questioned the relevance of the postulate.

Mention of Itano results requires attention to a possible ambiguity. In Section 15, we will discuss these experiments in detail and examine the view that they do not genuinely address the quantum Zeno argument as we have defined it in Section 12. For those who take the view that the collapse is a requirement for a genuine quantum Zeno effect, it is possible to argue that results such as those of Itano cannot demonstrate the effect because they do not require use of the projection postulate, but may be explained by straightforward use of the Schrödinger equation.

Such an argument is not available to us. Recognising that neither “quantum Zeno effect” nor “quantum Zeno paradox” as defined in Section 12 require use of the projection postulate, our distinction between the two must be based on different arguments as explained in that section.

14. THE QUESTION OF CONTINUOUS MEASUREMENTS

In Section 12, we met the various arguments [83–86] that there is a maximum rate for performance of measurements. Fearn and Lamb [8], too, consider that the concept of continuous measurement is not well defined.

This may certainly be the case for interactive measurements, where the measuring device clearly disturbs the system under investigation in a “physical process.” It is not so obvious that it applies for the negative-result measurements at the heart of what we have called the true quantum Zeno effect (Section 12). It may be said that any device which detects particles has a finite response time, so that if it detects a particle at time $t_0$, it may be inoperative for a short time following $t_0$. While this is the case, it is not so clear that for a negative result measurement, if the detector fails to detect a particle at time $t_0$, there is a “dead time” before it can fail to detect a particle at a shortly later time.

Nevertheless there is a general unease concerning the type of measurement involved in the quantum Zeno argument. In a proper measurement, it may be said that the experimenter takes the initiative and investigates the system at a particular time, while in the measurements involved in quantum Zeno, the experimenter (or the measuring device) passively waits for events to occur or not to occur. It may
seem that one is forced into some sort of continuous measurement and yet remain uncertain as to whether the concept has genuine meaning. As Peres [103] asks, does a mere "observation" really have the properties of a quantum mechanical measurement?

A scheme has been suggested [91, 95] which makes quantum Zeno take more the form of conventional quantum measurements, the experimenter being proactive. In this thought-experiment, the decaying system is at the centre of a sphere, the inner surface of which is covered with detecting material. It is arranged that any detection event on the surface gives rise electronically to a macroscopic and permanent black mark on a recording strip. Thus the presence of such a mark at time $T$ indicates that the atom has decayed at time earlier than $T - r/v$, where $r$ is the radius of the sphere, and $v$ is the speed of the decay product, assumed unique for convenience. The absence of such a mark indicates that decay has not taken place by that time.

In order to make the experiment proactive, it is assumed that the surface of the sphere is made of a balloon-like material, the radius of which may be changed extremely rapidly, while maintaining the shape of the sphere and the position of its centre.

If we wish to test whether particles have decayed or survived at times $T/n$, $2T/n$, ..., $T$, the radius of the detector balloon may be held at a value rather greater than $Tv/n$ from $t = 0$ to $t = T/n$, when it is reduced rapidly to a very small value, and then immediately increased to $Tv/n$ again. (It is assumed that the elements of the balloon can move much faster than $v$.) No detection can be made while the radius is $Tv/n$; any detection registered during the sweep-in of the balloon will be interpreted as a decay in the period before $T/n$. Similar contractions and immediate expansions may be performed at $t = Tm/n$, where $m = 2, 3, ..., n$. Thus decay statistics for quantum Zeno may be built up. The whole procedure may be repeated for different values of $n$, and the prediction of the generalised quantum Zeno effect is that the results should depend on the value of $n$.

It will be noted that any suggestion of continuous measurement is absent from the thought-experiment. Indeed the experiment is strongly analogous to a repeated Stern–Gerlach experiment in which a detector is placed at the end of only one beam of the first Stern–Gerlach apparatus. A particle which is not received at that detector (corresponding, say, to $s_z$ being $+\frac{1}{2}$), will be certain to be detected in the $s_z = -\frac{1}{2}$ channel of a second Stern–Gerlach apparatus. It is not clear, though, from a realist point of view, how it obtains the value of $-\frac{1}{2}$ during its passage through the first apparatus.

Similarly a decaying nucleus which is not detected at time $T/n$ must be able to recognise that a measurement has been undertaken at time $T/n$, and its decay profile should recommence with a $r^2$-region, even though it has not interacted with the detector.

Both experiments are of negative result type; both consist of a series of discrete measurements. This is another example of what is one of the main aims of this paper—to establish the quantum Zeno effect very much as part of the mainstream
of discussion of quantum measurement theory, rather than as an isolated and perhaps less important problem.

15. EXPERIMENTAL SUGGESTIONS AND INVESTIGATIONS

We now turn to the various experiments, performed or suggested, which are claimed to relate to the quantum Zeno effect. It should be stated at the outset that all these suggestions are highly ingenious and these experiments actually carried out are invariably both brilliant in execution and important in their implications. In some cases, though, we feel that the original analysis associated with the experiments deserves reexamination.

We commence with the very interesting experiments of Dehmelt [104, 105] on the so-called continuous Stern–Gerlach effect. This effect is contrasted with the conventional or transient Stern–Gerlach effect, where the motion of the centre of mass of an atom is used for a measurement of its spin, by observation of changes in the particle trajectory. In the continuous effect, experiments are performed on a single electron; what is observed is the frequency of oscillation of the electron in a Penning trap, augmented with a weak magnetic bottle. This frequency is very slightly different for the cases where the electron spin is along or opposed to the direction of the magnetic field, and this frequency difference leads to one or the other of two macroscopically distinct readings being obtained on a meter (although this may be obscured by noise, as discussed below). Dehmelt stresses that the continuous effect is nondestructive, whereas all realisations of the transient effect to date have been destructive.

Dehmelt regards the continuous effect as an example par excellence of the textbook description of a quantum measurement. Measurement is carried out on an essentially free individual particle, and, since the measurement is nondestructive, it may be repeated as often as is required, the same result being obtained on each occasion.

Dehmelt interprets the latter result in terms of a state reduction. He comments that, because of intrinsic (zero-point) noise in the detection circuit, the actual reading will be statistically distributed about the expected reading for the particular spin direction. To be (effectively) sure of what the reading should be, the actual reading must be integrated over a period of time $T^*_m$ which Dehmelt terms a minimum measurement time.

Dehmelt identifies $T^*_m$ with the state reduction time $T_r$. We would question this identification. The analysis of the noise is entirely based on standard quantum theory and, indeed, is actually semiclassical in nature. In principle, then, it seems unrelated to any state reduction, which, by definition, stands outside standard quantum theory. We may also note that $T^*_m$ is the same whether the initial direction of spin is along or perpendicular to the magnetic field direction, that is to say, whether or not a state reduction is appropriate. It seems reasonable to suggest that the time $T^*_m$ cannot really have anything to do with state reduction. (Of course,
from a logical point of view, and if one wishes to use the concept of state reduction, since the reduction process must take place within $T_m^*$, it is clear that one has $T_r \leq T_m^*$, but replacing “equals or is less than” by “equals” appears unmotivated.)

Dehmelt [105] describes state reduction as a process taking a time $T_m^*$. We would suggest that there is a process on this time-scale, but, rather than state reduction, it is the process

$$|S_z = \frac{1}{2}; A_i\rangle \equiv (1/\sqrt{2})(|S_z = \frac{1}{2}; A_i\rangle + |S_z = -\frac{1}{2}; A_i\rangle)$$

$$\rightarrow (1/\sqrt{2})(|S_z = \frac{1}{2}; A_{f1}\rangle + |S_z = -\frac{1}{2}; A_{f2}\rangle).$$

(15.1)

Here $|A_{f1}\rangle$ and $|A_{f2}\rangle$ are states of the whole system linking the spin being observed to the meter, thus including in particular the noise-producing load resistance. $|A_{f1}\rangle$ corresponds to $|S_z = \frac{1}{2}\rangle$; $|A_{f2}\rangle$ corresponds to $|S_z = -\frac{1}{2}\rangle$.

The process described by Eq. (15.1) is indeed part of the measurement. It is the process in which apparatus and system states become correlated. As is clear from Dehmelt’s description, it is a process obeying the laws of physics, in particular the Schrödinger equation. And it cannot, therefore, be a state reduction which must lie outside the Schrödinger equation. If one wishes to use the state reduction concept, it must be regarded as a completely separate process. It is, of course, difficult in Dehmelt’s experiment, where the measurement process may be said to be “continuous” but not “instantaneous,” to assert when the state reduction occurs, or how long it might take, but this is totally in accord with our discussion of the projection postulate earlier. The postulate cannot be made rigorous or axiomatic; its one saving grace is that, in a general if crude way, it seems to work tolerably well.

Dehmelt addresses the quantum Zeno effect in two ways. In Ref. [104], he says that the effect relies on the consideration of continuous measurement, while Dehmelt claims that his own work, following Pauli [38], demonstrates the need of a finite time for measurement. In Section 12, in particular, we argued that the conceptually difficult aspects of the quantum Zeno effect in no way rest on the possibility of instantaneous measurement, so we do not believe that this argument of Dehmelt resolves the quantum Zeno problem, as he claims.

In Ref. [105] Dehmelt presents a more specific argument. He considers the electron initially in the $|S_z = -\frac{1}{2}\rangle$ state, and a single $\pi$ pulse of amplitude $H_1$ and duration $t_1$ is applied. The pulse commences at $t = 0$. We must have

$$t_1, H_1 = \text{const.}$$

(15.2)

We may imagine that at any time during this process, at a time $t$, at which the angle may be said to have rotated through angle $\theta$, or in other words,

$$\theta = \pi t / t_1,$$

(15.3)
there is a probability of the spin making a transition,

$$|S_z = \frac{1}{2}; A_i\rangle \rightarrow c_+ |S_z = \frac{1}{2}; A_{f_1}\rangle + c_- |S_z = -\frac{1}{2}; A_{f_2}\rangle,$$

(15.4)

where we have

$$c_+ = \sin^2(\theta/2) \quad c_- = \cos^2(\theta/2).$$

(15.5)

If $t_1$ is very short compared with $T_m^*$, the characteristic time of the process of Eq. (15.4), then this process may be ignored, and, at time $t_1$, the state of the electron will be $|S_z = +\frac{1}{2}\rangle$, as is experimentally confirmed.

For $t_1$ comparable with $T_m^*$, though, the process of Eq. (15.4), acting continuously through the period of time $t_1$, will ensure that there is a nonzero probability of the electron being detected in state $|S_z = -\frac{1}{2}\rangle$ at the conclusion of the pulse.

For $T_m^*$ much shorter than $t_1$, this probability becomes practically unity. A crucial role in the argument is played by the fact that $c_+$ in Eq. (15.5) is proportional to $\theta^2$ for small $\theta$, not to $\theta$ itself.

It will be reasonable to model the behaviour of the system as a series of rotations each for a period $T_m^*$, each being followed by a process of the type of Eq. (15.4) in which there is a probability of $\sin^2(\theta/2)$ of what may be described as the electron making a transition to the $|S_z = +\frac{1}{2}\rangle$ state.

At the end of the first period of rotation, the probability of transition will be given by

$$P_{\text{trans}}(T_m^*) \approx n^2 T_m^* / t_1^2.$$

(15.6)

The squared nature of the short quantity $(T_m^*/t_1)$ implies that $P_{\text{trans}}$ will be very small.

Because this quantity is very small, it will be a good approximation to assume that at the conclusion of each period, there is the same probability of transition. Equally, if the system does make a transition to the $|S_z = +\frac{1}{2}\rangle$ state, it will be a good approximation to assume that at each subsequent stage it returns to that state. (Processes ignored in this approximation are of high order in a small quantity.)

Since there are $(t_1/T_m^*)$ periods in this model treatment, the overall probability of transition will be given by

$$P_{\text{trans}}(t_1) \approx (n^2 T_m^* / t_1^2) \times (t_1/T_m^*),$$

(15.7)

so overall,

$$P_{\text{trans}}(t_1) \approx T_m^*/t_1,$$

(15.8)

in agreement with Dehmelt's result.
Our treatment leads to a number of comments. First, it is pointed out that, were \( c_+ \) in Eq. (15.5) proportional to \( \theta \) for small \( \theta \), our argument would not work, and there would be a substantial probability (roughly 50\%) of a transition being made at some time during the pulse.

This in turn shows that the result is very much connected with the usual mathematics of the quantum Zeno effect. However, from the conceptual point of view, it is clear that the situation is totally distinct from that constituting the “paradoxical” nature of the original suggestions, as discussed in Section 12. There is no action at a distance by a macroscopic measuring device on a microscopic system; the effect is not negative-result in nature. Indeed Dehmelt’s effect, although it is certainly exceptionally ingenious, is not at all difficult to understand and appreciate. Dehmelt claims it to be the “unspectacular resolution” of “Zeno’s paradox,” but we submit that it in no way addresses the genuinely puzzling aspects of the “paradox.”

The next experiment to be discussed is the justly famous one due to Itano et al. [12]. This paper was titled “Quantum Zeno Effect,” and the experiment was brilliantly designed to demonstrate in a clear way the quantum Zeno mathematics. The technique was based on a proposal of Cook [4] and involved a single trapped ion in a three-level atomic configuration. The ground state may be called level 1. Level 2 is an excited metastable state, its spontaneous decay to level 1 being negligible.

During the experiment, the 1–2 transition is driven by a \( \pi \)-pulse at the appropriate frequency and of duration \( T \). During this pulse, though, at times \( T/n, 2T/n, \ldots \) a series of so-called “measurement pulses” are applied. These are optical pulses at the frequency of the 1–3 transition, where level 3 is connected by a strongly allowed transition to level 1, but cannot decay to level 2. Following any “measurement pulse,” there may be a series of photons emitted at the frequency of the 1–3 transition, or there may not. It is at the very least tempting to say that the “measurement pulse” has projected the state-vector of the system into the state corresponding either to level 1, when photons will be emitted, or to level 2, when they will not.

Further, because, as shown in Section 9, and especially Eq. (9.29), a measurement at short time \( t \) after the beginning of the \( \pi \)-pulse would have a probability proportional to \( t^2 \) of detecting the ion in level 2, and, as shown in Section 7, the measurement would be of the first kind, the quantum Zeno mathematics is appropriate, and, as \( n \) is increased, the probability of reaching level 2 by the end of the pulse decreases. Let us restrict ourselves to the large-\( n \) case, so that, at the time of the first “measurement pulse,” the probability that the ion is found to be in state 2 may be written as

\[
P_2(T/n) = \sin^2(\Omega T/2n) = \sin^2(\pi/2n) = \pi^2/4n^2. \tag{15.9}
\]

After all \( n \) pulses, the probability becomes

\[
P_2(T) = \pi^2/4n. \tag{15.10}
\]

The probability that a transition is made during period \( T \) decreases as \( n \) increases, tending to zero as \( n \) tends to infinity.
These predictions were confirmed in the excellent experiments of Itano et al. [12]. In their theory these authors used freely the terminology “measurement pulse,” “collapse of wave-function,” “projection,” and, of course, “quantum Zeno effect,” and our description above shows clearly how attractive these forms of words must be. In particular it has been commonplace in many publications, of which Refs. [106–108] are merely typical examples to assert that Itano et al. have demonstrated the quantum Zeno effect (or paradox).

There has also been, though, a considerable amount of discussion concerning the interpretation given by Itano et al. for their results. Initially most of the criticism concerned the use of the idea of “collapse of wave-function,” and to a lesser extent that of “measurement pulse.” Peres and Ron [92], Ballentine [109, 110], and Frerichs and Schenzle [111] showed that the experimental results of Itano et al. may be explained without the use of the idea of “collapse” or giving a special role to the idea of measurement, but merely by analysing the full behaviour of the three-level system. These authors would appear to accept that the experiments do indeed demonstrate the quantum Zeno effect, which, therefore, they are claiming to show, does not require collapse.

The position of other authors—Petrosky et al. [112, 113], Block and Berman [90], and Fearn and Lamb [8]—would appear to be that because analysis of the experiments does not require the collapse postulate, they are not demonstrating the quantum Zeno effect (although the distinction of Ref. [90], mentioned above, between quantum Zeno “effect” and “paradox,” allows the statement that Ref. [12] demonstrates the former but not the latter).

Itano et al. [114] have replied to Ref. [110], suggesting that the two interpretations—with and without collapse—both explain the experimental data and may be regarded as different, equally valid interpretations. They agree that their results and theory should not be regarded as verifying the notion of collapse, a point several of the authors referred to above had been keen to establish. Several authors [92, 115] have suggested adaptations to the scheme of Ref. [12] whereby the “measurements” would become uncertain, and the “collapse” become partial, thus rendering the Itano terminology less appealing.

Another important offshoot of Ref. [12] was the paper by Inagaki et al. [7]. This paper and further developments by these authors and others [83, 84, 98–100] have been discussed in Section 13.

The position of the present paper [91, 116] is quite different from those mentioned above. Since we do not believe collapse is required for a quantum Zeno effect, we certainly do not regard an explanation of the results of Ref. [12] without use of the postulate as a proof that the effect is not demonstrated.

But clearly the experiment does not meet our criteria given in Section 12 for a genuine quantum Zeno effect or “paradox.” The Itano experiment does not demonstrate a nonlocal negative-result effect between a microscopic system and a macroscopic measuring device. Quite in contrast, the “measurement” is achieved by direct contact between the “measurement pulse” and the ion. The experiment is certainly brilliantly conceived, but it is in no way difficult to appreciate in fairly realistic terms.
Let us again stress that the argument is not a pedantic one, or a mere quibble over words. It may not matter much if the concept of quantum Zeno effect is extended to such experiments, but it matters greatly if the impression is given that the original problems of Chiu and Sudarshan and others have been solved, or, worse still, shown to be little more than misunderstanding; the latter is most definitely not the case!

It should nevertheless be stressed that the Itano experiment and other similar ones, apart from their general great interest, do play one important part in the study of the quantum Zeno effect. They demonstrate clearly that one aspect of the theory is absolutely correct. This is the initial $t^2$-decay and the concept of causing successive breaks in the decay and, thus, successive restarts of the $t^2$-development. (Where, of course, we feel that papers of the Itano type do not match the quantum Zeno idea in the nature of the interaction that causes the breaks in the decay process.)

Another important paper claiming to demonstrate, indeed to make use of, the quantum Zeno effect, is by Kwiat et al. [13]. This paper develops the idea of interaction-free measurement, as suggested by Elitzur and Vaidman [117, 118]. In the latter papers, a Mach–Zehnder interferometer is arranged so that an incident photon is certain to exit from one particular port, say port A, never from port B. However, the insertion of an absorbing object in one arm of the interferometer creates a different situation, and there is now a nonzero possibility that the photon exits from port B. The existence of the object may therefore be ascertained, but the measurement is interaction-free; the photon cannot have interacted with the object, for if it had done so it would have been absorbed. Under this scheme, the maximum probability for any photon to exit from port B is $\frac{1}{2}$.

In a brilliant development of this idea, Kwiat et al. [13] use $N$ such interferometers in series. The reflectivity of each of the beam splitters is selected to be equal to $\cos^2(\pi/2N)$, so that, in the initial experimental arrangement, during the passage of the photon through the interferometer its amplitude in the upper section of each interferometer steadily increases and the photon is certain to exit through the upper port, which we may again call port A. However, if what Kwiat et al. [13] call “detectors” are inserted at each stage, this transfer of amplitude to the upper sections does not occur and there is a nonzero probability of exit through lower port B. This probability may, in fact, be made as large as is desired by increasing $N$; as $N$ tends to infinity the probability tends to 1 (in contrast to its maximum value of $\frac{1}{2}$ for the scheme of Refs. [117, 118]).

The mathematics of Kwiat et al.’s scheme is that of quantum Zeno. Let us consider the amplitude of the photon travelling downwards from the second beam splitter; it has two components. The first, $(C_u)_2$, has been reflected at the first mirror in the lower half of the system; it has been reflected twice at beam splitters, and in all three times, giving a factor of $i^3$, so, with the reflectivity arranged by Kwiat et al., we have

\[
(C_u)_2 = -i \cos^2(\pi/2N) = -i(1 - \pi^2/4N^2 \cdots).
\]
The second, \((C_b)_2\), is for light which has been reflected at the first mirror in the upper half of the system and transmitted twice at beam splitters, giving a factor of \(i\), so that

\[
(C_b)_2 = i \sin^2(\pi/2N) = i(\pi^2/4N^2 \cdots).
\]  

(15.12)

So the total intensity is

\[
I_2 = |(C_a)_2 + (C_b)_2|^2 = \cos^2 \pi/N = 1 - \pi^2/N^2 \cdots
\]  

(15.13)

The “detectors” suppress \((C_b)_2\), so when they are present, the intensity becomes

\[
I_2 = |(C_a)_2|^2 = \cos^4 \pi/2N = 1 - \pi^2/2N^2 \cdots.
\]  

(15.14)

To generalise, the downwards intensity after the \(m\)th beam splitter is

\[
I_m = \cos^2(\pi m/2N)
\]  

(15.15)

without the detectors, and

\[
I'_m = \cos^{2m}(\pi/2N)
\]  

(15.16)

with them.

After the last stage, where \(m = N\), \(I_N\) is zero, while \(I_N\) is unity. This is certainly the mathematics of the quantum Zeno effect. The crucial point is the opposite signs of \((C_a)_2\) and \((C_b)_2\). \((C_b)_2\) is essentially a “regenerating” term, of opposite sign to the “survival” term, \((C_a)_2\). The suppression of \((C_b)_2\), as in the analogous case of working with Eq. (13.4) rather than Eq. (13.3), gives the result of Eq. (15.16) rather than Eq. (15.15).

But while the mathematics of the quantum Zeno effect is obtained, the physics of the effect is not involved. The “detectors” do not play any role in detecting or monitoring the photon. They merely block it and prevent its further progress. The suppression could be achieved in a number of other ways. One could scrap the upper mirrors altogether, so no light could be reflected there, or one could drop strict equality of path lengths in the wings of the interferometer, making it correct to add intensities rather than amplitudes, and thus causing the regenerating term to have negligible effect for large \(N\).

Certainly detection at intermediate stages of the experiment, which would entail formation of a correlation between states of the photon and those of the detector, plays no part in the eventual result. Clearly collapse is not required, as suggested in Ref. [13]. Also the original quantum Zeno argument, as we discussed it in Section 12, depended fundamentally on detection. Each of the coherent superposition of atomic surviving and decayed states becomes correlated at detection with a macroscopic detector state, and the regenerating term is rendered inoperative.
The experiment described in Ref. [13], although highly ingenious and fascinating, is not nonlocal, nor is it a negative-result experiment, both characteristics of the original quantum Zeno effect. It is explained readily from a realistic position. Indeed the experiment itself is totally classical, in the sense that a nineteenth-century physicist would have no difficulty in understanding and explaining the results in terms of nineteenth-century physics. (Of course, the interpretation of the measurement as being interaction-free depends on the idea of photons, unless, at least, one allows $N$ to tend to infinity.)

Just as in the cases of the two previous experiments described and discussed, we feel that the experiments of Kwiat et al. [13] do not meet the criteria of the definition of "quantum Zeno effect."

A recent extremely interesting paper by Plenio et al. [6] attempts to meet the challenge of those (they mention our Refs. [90, 110–113]) who suggest that the Itano experiment [12], in particular, does not verify the quantum Zeno effect as originally defined, because it is a coherent Rabi oscillation that is arrested, rather than a decaying state. These authors discuss in detail various multilevel atomic systems where the inhibition of decay may be investigated.

It may be pointed out that our own criterion for a "true" quantum Zeno effect is not that attributed by Plenio et al. to the listed references. We are keen to apply the term "quantum Zeno effect" to any quantum process with the appropriate initial $t^2$-dependence (as was stressed in Section 9). Our criterion is the nature of the measurement device and measurement process; the measurement should not constitute a direct interaction with the system undergoing the Zeno process.

The schemes of Plenio et al. [6] do indeed centre on an atom making a spontaneous transition between two levels, or, to put things another way, the decay of the population of one level. They discuss ingenious arrangements for increasing the length of the $t^2$-region at the beginning of the decay from around $10^{-16}$s for a normal spontaneous decay by up to eight orders of magnitude.

In this scheme, though, the measurement is undertaken in a way similar to that in the Itano experiment; the decaying level, level 1, is coupled for a short period to an auxiliary level which decays to level 1 exclusively, and the presence or absence of photons corresponding to this transition provides the required information. As with the Itano case, it does not seem that the measurement merely measures passively, but plays an active role in the physics of the experiment. The proposed experiment of Plenio et al. [6] is a very interesting development, and we hope it will be performed, but it does not seem that it meets our precise criteria for a true quantum Zeno effect.

Last, we turn to a recent paper by Panov [119], who takes the area of discussion right back to radioactive decay. The paper may be regarded as having an experimental aspect, inasmuch as it suggests that the suppression of the decay of an excited level of $^{235}$U in a silver matrix is related to the quantum Zeno effect. In this decay, which is an internal conversion process, the 6p-electrons of uranium play the main role. The energies of these electrons are affected by the nature of the chemical bonding from the surrounding atoms, and this process influences the decay rate.
substantially. In conventional experiments, the decaying atoms are collected on
the surface of conducting materials, and the decay rate is found by counting the
electrons produced by the conversion process. Chemical effects of the order of
2-3% of the usual decay rate are reported.

In recent experiments of Kol'tsov and Rimsky-Korsakov [120], the decaying
atoms are embedded in a silver matrix for a given period and then extracted. The
proportion of atoms still in the excited state is then determined by counting the
conversion electrons subsequently produced. It is found that there is a suppression
of the decay while the atoms are in the matrix, which is an order of magnitude
greater than the chemical effects while they are on a surface.

Panov suggests that, when the atoms are in the matrix, the strong interaction
between the decay particles and the atoms of the matrix must be taken into account
explicitly. He attempts a first principles treatment, an exact calculation of the
system of nucleus, decay particle, and lattice, and his results suggest the possibility
of a strong influence of the presence of the lattice on the decay rate; he describes
this as a new class of quantum Zeno effect.

Panov's results are certainly of considerable interest in the study of radioactive
processes, although much work certainly needs to be done in measuring the rele-
vant parameters before his general conclusion is confirmed.

Yet again, though, we find the measurement very much interfering with the
system undergoing decay; indeed, this is the fundamental point in Panov's treat-
ment. As such this class of behaviour falls outside the definition of quantum Zeno
effect discussed earlier in this paper.

It is also worth mentioning that the concept of the first principles treatment of
measurement, although extremely appealing in principle, does not seem to be
workable in practice. As shown in Section 2, a treatment that gives in a satisfactory
way Eq. (2.3), must, as long as one sticks strictly to the Schrödinger equation, also
give the unsatisfactory Eq. (2.7). Little as one may like it, it does appear that we
do not as yet possess a quantum theory for which measurement may, in all cases,
be treated from first principles.

16. CONCLUSIONS

Our conclusions are as follows:

(1) The quantum Zeno effect is to be regarded as a genuine result of quantum
theory, at least of the form of quantum theory we now possess. This leads to the
intriguing possibility that experimental disproof of the effect could cause a recons-
sideration of fundamental aspects of quantum theory (Sections 11 and 13).

(2) In particular, the quantum Zeno effect does not rely on the acceptance of
the projection postulate (Section 13). Rather it is a result of (a) a $t^2$-form of initial
deay, which is the normal quantum case (Section 9), and (b) the use of standard
ideas of quantum measurement theory, in particular the point that many measurements may be considered to be of the first kind (Sections 4 to 7).

(3) The \( r^2 \)-form of initial decay may be considered in many ways analogous to the \( \theta^2 \) spatial behaviour that leads to Bell's inequality. Thus quantum Zeno ideas are very much associated with the mainstream discussion of quantum measurement, not an obscure and perhaps uninteresting backwater (Section 10).

(4) The “paradoxical” nature of the early discussions related to the effect of an external macroscopically separated macroscopic measuring device on an evolving microscopic system. It therefore seems sensible to restrict the use of the term “quantum Zeno effect” to experimental situations of this nature (Section 12).

(5) In particular, then, mere inhibition of a transition of any kind, via a measuring device which interacts directly with the system undergoing the transition, is interesting but scarcely conceptually surprising and would not be regarded as an example of the quantum Zeno effect (according to the above definition) (Section 12).

(6) While total freezing of the decay is clearly the most dramatic prediction of the theory, we certainly do not regard it as playing a crucial role in defining the quantum Zeno effect. It is not even of the essence in discussion of the effect that detectable events occur; the central point is that, since no decay is precisely exponential, the theory predicts influence of a detector on a microscopic evolving system (Section 12).

(7) Equally we do not feel it important that the quantum Zeno effect should be defined only on decaying systems; on the contrary, we would stress the rather general nature of the effect (point (3) above; Section 15).

(8) While the arguments about continuous measurement affect some aspects of the discussion of quantum Zeno, we do not believe that they throw genuine questions on the theoretical prediction itself, which certainly does not require the possibility of continuous observation (Section 14).

(9) Many ingenious and interesting experiments have been performed demonstrating inhibition of evolution by what may be termed measurements. Even apart from their own considerable interest, these experiments certainly contribute to the quantum Zeno debate, by demonstrating the general correctness of the argument concerning \( r^2 \)-evolution and its disturbance. However, the experiments do not fulfill the criteria of point (4) here, and so they do not stand as genuine examples of what we would term the quantum Zeno effect. If that is felt to be too verbal a point, it may be said that these experiments do not exemplify or shed any light on the deepest problems raised by the original writers, nor do they demonstrate that these problems were misconceived (Section 15).

It may be said that virtually all these conclusions relate to matters of considerable dispute in the current literature. The whole area is indeed one of considerable controversy, and it is hoped that the present paper may help to resolve
some of the outstanding issues. Although disagreement may be rife, we may remember what happened to Zeno himself—he was beheaded at Elea in Italy in 435 BC [121]—and we may at least be grateful that none of those presently discussing the modern effect named after him have, at least as yet, suffered such a fate.

REFERENCES

29. R. Omnès, Rev. Mod. Phys. 64 (1992), 339.
34. J. S. Bell, Found. Phys. 12 (1982), 989. Also in Ref. [15, p. 159].
76. J. S. Bell, *Physics (N.Y.)* 1 (1964), 192. Also in Ref. [15, p. 15].
QUANTUM ZENO EFFECT