Foundations of Physics, Vol. 16, No. 6, 1986

Existence of "Free Will" as a Problem of Physics[1]

Asher Peres[2]

Received June 17, 1985

The proof of Bell's inequality is based on the assumption that distant observers can freely and independently choose their experiments. As Bell's inequality is experimentally *violated, it appears that distant physical systems may behave as a single, nonlocal, indivisible entity. This apparent contradiction is resolved. It is shown that the "free will" assumption is, under usual circumstances, an excellent approximation.*

I have set before you life and death, blessing and cursing: therefore choose life....

—*Deuteronomy* XXX, 19

1. BELL'S THEOREM

Consider two observers equipped with various measuring apparatuses. These observers are very distant from each other, so that *no communication* between them is possible while they are doing their experiments. For example, they may perform these experiments at times and places bearing a space-like relationship, as in Fig. 1. Nevertheless, the observers are able, at a later stage, to *compare* their results by sending *messages* to one another. Moreover, they may prearrange their experiments so that they observe *correlated* physical systems, such as particles emitted by a single source, somewhere in the common past of *both* observers (in the intersection of their past light-cones). We now make the following assumptions[1,2]:

(i) The first observer has the *choice* of measuring a dynamical variable A, with possible results $a = \pm 1$, *or* a dynamical variable

[1] Dedicated to Prof. John A. Wheeler on the occasion of his 75th birthday.

[2] Department of Physics, Technion-Israel Institute of Technology, 32 000 Haifa, Israel.

0015-9018/86/0600-0573$05.00/0

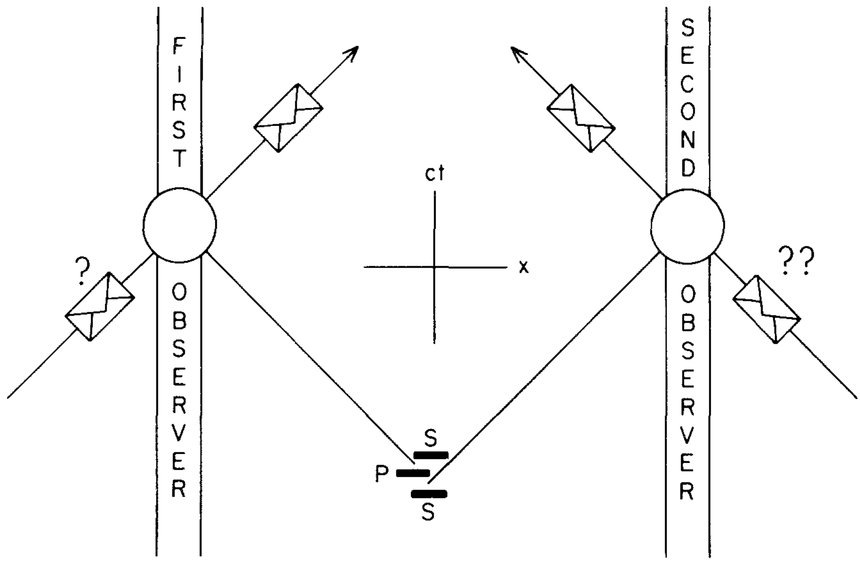

Fig. 1. Space-time diagram showing the world tubes of two observers (assumed mutually at rest), the trajectories of two photons emitted by a single atom (SPS cascade), and messages emitted by the observers in order to compare their results. The observers may select their experiments by "free will" or by means of random messages received from disconnected parts of the Universe (as in Fig. 3, below).

C, with possible results $c = \pm 1$. Likewise, the second observer has the *choice* of measuring a dynamical variable B, with possible results $b = \pm 1$, *or* a dynamical variable D, with possible results $d = \pm 1$.

(ii) Each one of these possible outcomes (a, b, c, and d) is *unpredictable* by the observer, but nevertheless *deterministic.*

(iii) Any outcome obtained by *one* of the observers does not depend on the *choice* of the experiment which is performed (was performed, will be performed) by the *other* observer.

Let us carefully examine these three assumptions:

(i) Each observer *can select* one of two possible *macroscopic procedures,* such as orienting a Stern–Gerlach magnet along the x direction *or* the y direction. These procedures are *mutually exclusive,* but each one is possible and is *freely* chosen by the observer. If we do not want to involve human free will, these procedures can be selected by automated instruments involving random elements. For example, we may have two telescopes pointing at opposite sides of the sky, and counting whether an odd or even number of photons arrive in a predetermined time interval.

Each one of these possible *macroscopic procedures* is called, euphemistically, the "measurement of a dynamical variable," (such as spin, polarization, etc.). These terms should *not* be understood in their *classical meanings*. Nowhere in this paper is it assumed that there exist dynamical variables having numerical values which are originally unknown to the observers, and then are revealed to them by means of measurements.[3] It is only assumed that some macroscopic procedures—called "measurements"—have various macroscopically distinct outcomes, which are conventionally labeled as $a = \pm 1$, etc. An example is shown in Fig. 2.

(ii) The second assumption, which I call *cryptodeterminism*, must be scrutinized very carefully. Consider a classical pinball machine. The trajectory of each ball is perfectly deterministic. Yet, it is utterly unpredictable. This system is *chaotic*[4]: tiny changes in the initial data, or tiny external perturbations, cause huge changes in the final outcome. Only statistical averages can be reliably computed. The same situation prevails in the classical kinetic theory of gases. In principle, all motions are perfectly deterministic. However, because of incomplete knowledge of all initial data, only statistical predictions can be made.

Likewise, the outcomes of quantum measurements are deemed "random" and only their statistical averages are predictable. Are they *really* random? This point of view is supported by many authors. For example, Landé[5] asserts that "the random distribution is a physical reality, and the determinism which only looks random is a purely academic construction." Then, Landé adds the following explanation: "A distribution of effects

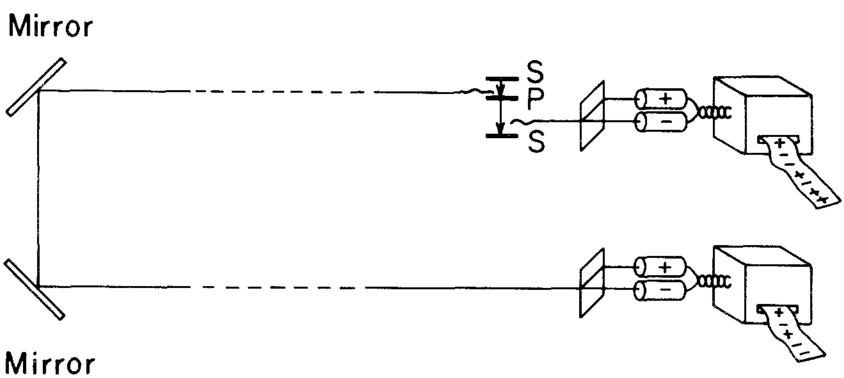

Fig. 2. Two photons with opposite directions, originating in a SPS cascade, must have opposite linear polarizations. In this experimental setup, mirrors reflect one of the photons so that the two analyzers and four detectors can be located next to each other. The recording devices must then yield opposite records.

satisfying the laws of error theory requires, just from the determinist's viewpoint, a corresponding distribution of causes at an earlier time and from there back to a still earlier time. A program of giving a strictly deterministic theory of statistically distributed events leads nowhere." Why nowhere? In the worst case, it may lead in a direction which Landé might, perhaps, have found hazardous. This *is* indeed the direction in which I intend to lead the reader of this essay.

Consider the experiment shown in Fig. 2, whereby one observes the polarizations of photon pairs originating in a SPS cascade. (Another possibility is a $2S-1S$ transition[6] which is a second-order two-photon decay process.) Each recording device yields a sequence of $+1$ and -1 which *appears* perfectly random: This sequence would pass every known test for randomness.[7] Yet, if one of the two sequences is known, the other one can be immediately guessed—it is the opposite of the first one. In particular, if the delay in observing one photon is long enough, one can predict with certainty the outcome of a measurement of its polarization from the result of a measurement already performed on the other photon. It thus appears that the outcome of such a polarization measurement is not at all random. It is perfectly determined, although the whole sequence may *appear* random if we lack *relevant information.*

This suggest the possible existence of *additional variables* (the so-called "hidden variables") which would be the "relevant information" allowing to one predict with certainty the outcome of any experiment. If we have no knowledge at all about these additional variables, we can still compute *averages* over their possible values. These averages ought to coincide with those obtained from quantum theory. Bell's celebrated paper[1] was formulated in a language appropriate to a hidden variable theory (without specific details of the theory). Here, I shall involve no theory at all and stay on strictly phenomenological grounds.

One may feel somewhat uneasy about assumption (ii) because it seems to contradict the spirit, if not the letter, of assumption (i). Observers, after all, are made of ordinary matter and are subject to the ordinary laws of physics.[2] Thus, why should observers be endowed with "free will" while the observed objects behave deterministically? Can't an observer be observed too? These difficult questions may be avoided by using, instead of living observers, some automated devices such as telescopes pointing toward opposite sides of the sky, as explained above. Assumption (i) may then be replaced by:

(i′) The *choices* of the experiments (whether to measure A or C, and whether to measure B or D) can be performed in a way involving *no a priori correlations.*

In other words, *there are neither physical laws nor initial conditions* imposing that if one measurement is, say, A (rather than C) then the other measurement must be B, or must be D. It seems quite obvious that there is no such physical law, which is valid for *any* initial conditions. On the other hand, the hypothesis that we are not subject to special initial conditions, imposing a unique solution for each pair of measurements, is a matter of faith. One could also imagine that there is a worldwide conspiracy whereby everything is predetermined.

(iii) This third assumption is called *locality*, or *separability*. It should be clearly understood that the *result* obtained by each observer may well be correlated to the *result* obtained by the other one, if the particles which they test were correlated in the past (see, e.g., Fig. 2). However, it is reasonable to assume[8] that if the first observer measures A and obtains $a = \pm 1$, say, this result cannot depend on whether the second observer decides (has decided, will decide) to measure B rather than D.

It should be noted that assumption (iii) is *logically meaningless* unless we accept assumption (ii). Indeed, when we formulate (iii), we rely on counterfactual reasoning. We consider the results of two *possible but mutually exclusive* measurements.[9] This has a meaning only if these results are definite, even though only one of them can actually be known.

Having explained all these premises, it is easy now to prove Bell's theorem. Consider a sequence of pairs of measurements, as discussed above. For the jth pair, the outcomes are a_j, b_j, c_j, and d_j. These numbers are unknown, but they can be only ± 1, so that

$$a_j b_j + a_j d_j + c_j b_j - c_j d_j \equiv \pm 2 \tag{1}$$

Taking an average over all values of j (the serial number j effectively plays the same role as Bell's hidden variables), we obtain

$$|\langle ab \rangle + \langle ad \rangle + \langle cb \rangle - \langle cd \rangle| \leqslant 2 \tag{2}$$

Now, although all the numbers a_j, b_j, c_j, and d_j cannot be simultaneously known, *average values* such as $\langle ab \rangle$, namely the correlation of A and B, *can* be measured experimentally. There have been many such experiments, the most famous of which, due to Aspect *et al.*,[10] is schematically shown in Fig. 3. The experimental result clearly *violates* inequality (2). Therefore, at least one of the assumptions (i′) or (ii) or (iii) must be wrong.

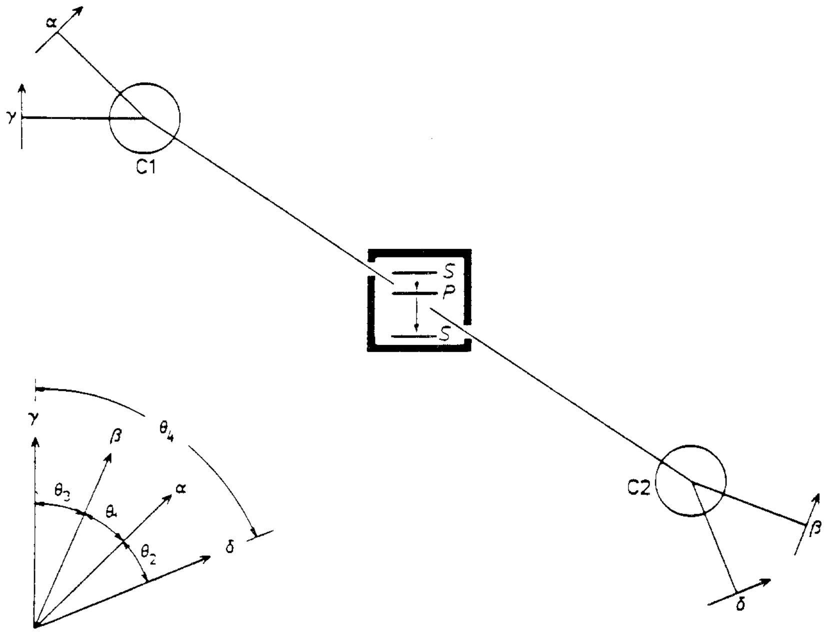

Fig. 3. A theorist's view of Aspect's experiment. Two photons with opposite polarizations pass through random optical commutators C1 and C2, toward polarization analyzers oriented along the directions α, β, γ, and δ. The diagram in the left corner shows the set of angles (three times $\pi/8$) giving the maximum violation of Bell's inequality.

2. ABANDON SEPARABILITY?

One may be reluctant to abandon (i) or (i′). As stated by Bell,[11] "separate parts of the world would be deeply and conspiratorially entangled, and our apparent free will would be entangled with them."

One is even more reluctant to abandon (ii) for which some experimental support is suggested, if not formally proved, by setups such as the one described in Fig. 2. Anyway, as explained above, if one repeals (ii), then (iii) becomes meaningless.

On the other hand, there is neither direct nor indirect experimental support for (iii). The only thing which we know for sure is that distant actions cannot have an immediate (or retroactive!) influence on results of *classical* measurements, or on quantum measurements which happen to be quasi-classical because they are predictable with 100% certainty, or on predictable quantum *average* values such as $\langle ab \rangle$, etc. None of these conflict with inequality (2). In particular, classical correlations always satisfy this inequality.[12]

We are thus led to abandon locality and consider the *pair* of photons in Aspect's experiment as a *single, nonlocal, indivisible entity.*[3,13] This may appear strange at first, but this is not essentially different from considering a *single* photon as a nonlocal, indivisible object: This is indeed how we explain that a single photon can pass through two widely separated slits and interfere with itself. A single photon may even originate in two different lasers.[14] (Actually, the situation is even more curious. In the Aspect experiment, the photons are distinguishable because they have different energies, and *almost* everything happens as if there were two distinct particles with reasonably well-defined positions and momenta at each instant. Only the *polarizations* of the photons are inseparably entangled.)

We have now reconciled all our knowledge with the violation of Eqs. (1) and (2). There are *no* physical quantities such as a_j: One must distinguish a_{jB} from a_{jD} (namely, the values obtained for a_j if the second observer measures B or D, respectively). One may even estimate how often $a_{jB} \neq a_{jD}$: this must happen in more than 41% of cases.[15]

However, once we accept that, doubts about (i) or (i′) appear again, as new Hydra's heads! Why can't our two observers too be considered as a single, nonlocal, indivisible entity? Their past histories are undoubtedly correlated, since they agreed to collaborate in some joint experiment. Why do we believe that each observer has a free will, although he or she consists of atoms subject to deterministic (or cryptodeterministic) laws?

Consider again the situation illustrated in Fig. 2. Each one of the records appears "random" (as if the corresponding apparatus had "free will") in the *absence of information* on the other record. That information is essential, because the two photons have *correlated histories.* They are, as already explained, a single, indivisible, nonlocal object. If we ignore that correlation and make no use of the corresponding information, then *every* outcome looks random. Likewise, when people are considered individually, and all their past interactions with other people or things are ignored, they appear to behave randomly—to have "free will."

Is our apparent free will limited by some unknown correlations? Let us again look at Fig. 2. Randomness was not completely eliminated by correlating the two sequences of results. While we can predict the second sequence after looking at the first one, we cannot predict the first one. Some randomness remains. Could we add information and reduce this randomness again? To do this, we would need to trace back the history of the atom which emitted the SPS cascade. If it was emitted by a coherent process, it might be possible to correlate these two photons with other observable phenomena. On the other hand, if the excitation process was thermal, then further correlations are *practically* lost. Turning now our attention to considerably more complex systems, such as human beings, it is obvious that any EPR[8] correlations must be immediately destroyed in a maze of irreversible processes. *Each one* of us thus behaves unpredictably, as if endowed with free will. This is why words such as "each one" can legitimately be used when we describe the behavior of people.

It *is* an excellent approximation to use pronouns such as "he" and "she" in everyday's life. Lawyers who prepare contracts *can* consider the various signatories as distinct entities. Indeed, these signatories are very complicated systems, which cannot be described completely. We know that actually they are only parts of an interconnected universe, but in our *approximate* description, they *appear* to be distinct and isolated. They behave "randomly." They have free will.

In our daily work as physicists, we are compelled to use only *incomplete* information on the world, because we cannot know everything. The methodology which we use in physics is the following. We divide the world in three parts, which may be called the *system*, the *observer* (or the observing apparatus), and the *rest of the world* (that is, most of it!) which is then considered as irrelevant. We assume that the system under observation is perfectly isolated from everything else, except from the observer testing it, if and when it is tested. This makes things *appear* simple. *This* is what makes physics an exact science. For example, we can compute the properties of the hydrogen atom to umpteen decimal places, because when we do these calculations there is only one hydrogen atom in our conceptual world.

We work as if the world could be dissected into small independent pieces. This is an illusion, because, as Bell's theorem shows, the entire world is interdependent. But there is no other way of working. The practical question with which we are faced is the following: given a finite amount of information, what are the possible outcomes of an ill-defined experiment which we prepare? The answer must be probabilistic, by the very nature of the problem. One may only speculate whether, in a complete description of the whole Universe, including our brains, determinism would be restored.

I don't know the answer to this question. Only God can know that, if He or She exists. However, since this is an essay on science, and not on theology, I shall now turn my attention to a more practical problem.

3. INFORMATION TRANSFER

Assuming that measurements are instantaneous (that is, very brief), can we use these nonlocal properties to transfer information *instantaneously* between distant observers? If we could, Einstein's theory of relativity would be in jeopardy.

There have been some attempts in this direction[16,17] and also some attempts to prove that this is impossible.[18-20] Why did the question arise? After all, nonlocal correlations do exist in classical physics. A classical analogue of the Aspect experiment can easily be imagined as follows: A bomb, initially at rest, explodes into two fragments carrying opposite angular momenta. Two observers measure the projections of these angular momenta along arbitrary directions and find them correlated.[12] Obviously, there is nothing in this scenario allowing one observer to send a signal to the other. Both observers only *receive* information from a common source.

How does this differ from the EPR effect? The point is that the angular momenta of the bomb fragments are believed to have objective values, even before they have been measured. On the other hand, the second photon in Fig. 2 *acquires* a definite polarization as a result of the "collapse of the wavefunction" resulting from the measurement of the polarization of the first photon. If this collapse is naively considered as a real physical phenomenon—which is nonsense[3,13,21]—there is here some new feature which was absent in classical physics.

Can it be used to convey information? For example, can the second observer guess what was the orientation of the apparatus used by the first one? Any proofs purporting to show that this is impossible, by using the Schrödinger equation (or other features of quantum theory), are not acceptable, because the "collapse" itself blatantly violates Schrödinger's equation. Thus, no proof of nonexistence has been proved to exist, as yet.

To prove impossibility, we must *postulate* (in agreement with quantum theory, but without relying on it explicitly) that

(iv) no measurement of a physical system can give us information about its state prior to the measurement, unless some information was already available beforehand.[22,23]

For example, if we know that a photon is certainly polarized along the x direction or the y direction, but we do not know which one of the two, we can ascertain this point by a "nondemolition measurement" that leaves the initial state unchanged. However, if we know *nothing* about the initial state, there is no way to discover it by experimenting on a *single* photon.[24] The most we could learn from such a measurement is that the initial state was not in the orthogonal complement of the final one. Postulate (iv) must be satisfied by any theory claiming that a physical system has a well-defined state, which can change abruptly when a measurement is performed. Otherwise—that is, if (iv) is not valid—it would be possible to convey information instantaneously by measuring correlated systems.

There is one more point in need of clarification. I have not yet properly defined what is information. What is a message? Intuitively, it is a relationship between a *decision* and an *observation*. When I write a letter, I can decide what to write—this is my "free will." But then, my friend who receives the letter cannot decide what to read. He, or she, can only passively observe the contents of the letter. (Notice that my friend and I are tacitly assumed to be distinct entities!) I have thereby defined a message—a relationship between a decision and an observation—but then there is the next hurdle: What is a decision? What is an observation? These obviously are physical processes, but how do we distinguish between them by some objective criterion, without using pronouns such as "I" or "he" or "she"?

My belief is that the distinction between decisions and observations is quantitative, rather than qualitative. A decision is influenced by too many factors to be predictable. It is like the evolution of a chaotic system which, although deterministic, is not computable.[4] Decisions thus *appear* completely random and this is why we claim that we have free will. An observation too is controlled by many factors, but in a way which is conspicuously deterministic, or at least follows simple probability rules. The distinction between past and future, cause and effect, thus seems reducible to the interdependence of events involving chaotic or regular dynamics.[25,26]

How can these qualitative notions be recast into a quantitative, dynamical theory? Consider a model consisting of two physical systems, with dynamical variables collectively called R and C, respectively. These two systems are connected by a chain of Hookean oscillators—a "field"—having, in the long wavelength limit, a well-defined transition time T. Assume that the internal dynamics of R and C are regular and chaotic, respectively. Then, if the Hookean oscillators are at rest at some time $t = t_0$, or have only random (uncorrelated) motions at that time, one obtains, for $t > t_0$, non-vanishing correlations of the type $\langle R(t+T)\, C(t)\rangle$, and *only* of that type. The chaotic system can influence the regular one, with a delay T, but there are *no easily analyzable* effects propagating from the regular system (the "receiver") toward the chaotic one (the "emitter").

Likewise, for $t < t_0$, one has nonvanishing correlations $\langle R(t-T)\, C(t)\rangle$. In that case, the quiescent state assumed at $t = t_0$ (all the field variables at rest or, at least, uncorrelated) results from the *advanced* potential of the emitter. The problem therefore is time-symmetric[27]—it could not be otherwise. The arrow of time cannot result from dynamics alone, if each elementary step is reversible. It necessitates *two* separate assumptions: One is that there is a privileged time t_0 at which the field is quiescent. The other is that the field connects two very different systems, one of which is chaotic and the other regular, acting as the emitter and receiver, respectively.

ACKNOWLEDGMENTS

It is a pleasure to dedicate this essay to John A. Wheeler, with whom these ideas were discussed during my visits to the University of Texas. However, Professor Wheeler bears no responsibility for any errors, misconceptions, or other fantasies in this paper.

Some parts of this article were written in Copenhagen. There could not be a more appropriate location for that. I am very grateful to NORDITA for hospitality.

I am also grateful for the hospitality of the Center for Philosophy and History of Science at Boston University and especially to Prof. Abner Shimony for many helpful comments.

Work done at Technion was supported by the Gerard Swope Fund, the New York Metropolitan Research Fund, and the Fund for Encouragement of Research.

REFERENCES

1. J. S. Bell, *Physics* **1**, 195 (1964).
2. A. Peres and W. H. Zurek, *Am. J. Phys.* **50**, 807 (1982).
3. A. Peres, *Found. Phys.* **14**, 1131 (1984).
4. G. Iooss, R. H. G. Helleman, and R. Stora, eds., *Chaotic Behavior of Deterministic Systems* (North-Holland, Amsterdam, 1983).
5. A. Landé, *Foundations of Quantum Theory* (Yale University Press, New Haven, 1955), p. 4.
6. W. Perrie, A. J. Duncan, H. J. Beyer, and H. Kleinpoppen, *Phys. Rev. Lett.* **54**, 1790 (1985).
7. F. James, *Rep. Prog. Phys.* **43**, 1145 (1980).
8. A. Einstein, B. Podolsky, and N. Rosen, *Phys. Rev.* **47**, 777 (1935).
9. N. Bohr, *Phys. Rev.* **48**, 696 (1935).
10. A. Aspect, J. Dalibard, and G. Roger, *Phys. Rev. Lett.* **49**, 1804 (1982).
11. J. S. Bell, *J. Phys. C* **2**, 41 (1980).
12. A. Peres, *Am. J. Phys.* **46**, 745 (1978).
13. A. Peres, *Am. J. Phys.* **52**, 644 (1984).
14. R. L. Pfleegor and L. Mandel, *Phys. Rev.* **159**, 1084 (1967).
15. S. J. Feingold and A. Peres, *J. Phys. A* **13**, 3187 (1980).
16. N. Cufaro-Petroni, A. Garuccio, F. Selleri, and J. P. Vigier, *C. R. Acad. Sci. Paris B* **290**, 11 (1980).
17. N. Herbert, *Found. Phys.* **12**, 1171 (1982).
18. G. C. Ghirardi, A. Rimini, and T. Weber, *Lett. Nuovo Cimento* **27**, 293 (1980).
19. D. N. Page, *Phys. Lett. A* **91**, 57 (1982).
20. A. Shimony, in *Proc. International Symposium on Foundations of Quantum Mechanics* (Phys. Soc. Japan, Tokyo, 1984).
21. A. Peres, "When is a quantum measurement?", *Am. J. Phys.* **54** (1986), in press.

22. D. Finkelstein, in *Paradigms and Paradoxes*, R. C. Colodny, ed. (Univ. Pittsburgh Press, 1971), Vol. V; reprinted in *Logico-Algebraic Approach to Quantum Mechanics*, C. A. Hooker, ed. (Reidel, Dordrecht, 1975), Vol. II, pp. 141–160.
23. H. Paul, *Am. J. Phys.* **53**, 318 (1985).
24. A. Peres, *Am. J. Phys.* **43**, 1015 (1975).
25. A. Peres and L. S. Schulman, *Int. J. Theor. Phys.* **6**, 377 (1972).
26. L. Gatlin, *Int. J. Theor. Phys.* **19**, 25 (1980).
27. J. A. Wheeler and R. P. Feynman, *Rev. Mod. Phys.* **21**, 425 (1949).