How far do EPR-Bell experiments constrain physical collapse theories?

2007 J. Phys. A: Math. Theor. 40 3141

(http://iopscience.iop.org/1751-8121/40/12/S15)

IOP PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND THEORETICAL

J. Phys. A: Math. Theor. **40** (2007) 3141–3149 doi:10.1088/1751-8113/40/12/S15

How far do EPR-Bell experiments constrain physical collapse theories?

A J Leggett

Department of Physics, University of Illinois at Urbana-Champaign, USA
and
Institute for Quantum Computing, University of Waterloo, Ontario, Canada

Received 4 August 2006, in final form 8 January 2007
Published 7 March 2007
Online at stacks.iop.org/JPhysA/40/3141

Abstract
A class of theories alternative to standard quantum mechanics, including that of Ghirardi *et al* (‘GRWP’), postulates that when a quantum superposition becomes amplified to the point that the superposed states reach some level of ‘macroscopic distinctness’, then some non-quantum-mechanical principle comes into play and realizes one or other of the two macroscopic outcomes. Without specializing to any particular theory of this class, I ask how far such ‘macrorealistic’ theories are *generically* constrained, if one insists that the physical reduction process should respect Einstein locality, by the results of existing EPR-Bell experiments. I conclude that provided one does not demand that the prescription for reduction respects Lorentz invariance, at least some theories of this type, while in principle inevitably making some predictions that conflict with those of standard quantum mechanics, are not refuted by any existing experiment.

PACS numbers: 03.65.Ta, 04.50.+h

Of Professor Giancarlo Ghirardi’s many contributions to fundamental physics, probably the best known is the celebrated Ghirardi–Rimini–Weber model of spontaneous wavefunction collapse [1], subsequently combined [2, 3] with the ideas of Pearle to form what is usually called the ‘GRWP’ (or ‘CSL’) scenario. The principal motivation for the development of this model, as of similar scenarios such as Penrose’s model of collapse induced by gravitational effects [4], is of course to avoid the realization, or as it is more commonly known, ‘measurement’ paradox which arises if the formalism of quantum mechanics is taken to constitute a complete description of the physical world, and which I briefly review below. The GRWP model, and others in the same general class, postulate that the standard formalism of quantum mechanics is *not* the whole truth about the physical universe but needs to be supplemented with other explicitly non-quantum-mechanical assumptions; as a result, the realization of a particular macroscopic outcome is not an arbitrary process materializing out of thin air, but takes place for concrete and identifiable physical reasons.

There are many good discussions in the literature of why we should take scenarios such as the GRWP one seriously; my own best attempt is given in [5], and I just recapitulate here the barest essentials of the argument of that reference. Let us suppose, for the sake of the argument, that the formalism of quantum mechanics indeed constitutes, in principle, a complete description of the physical universe. At the level of single electrons or atoms, most interpretations of the formalism hold that when in the wavefunction describing an ensemble there occur nonzero amplitudes for two alternative physical processes or trajectories (for example, in a standard Young's slits experiment, passage through the upper or the lower slit in a screen), then it is not correct to say that each individual member of the ensemble realizes *either* one alternative *or* the other. If we now extrapolate the formalism to the level of everyday experience, its nature changes not one bit; therefore, neither can nor should our interpretation of its meaning, and thus we are forced (for example) to the conclusion that Schroedinger's unfortunate cat (more precisely, any individual cat drawn from the relevant ensemble) cannot be said to be definitely dead or alive before her state is inspected by a human agent. The common misconception that this problem is somehow resolved by the phenomenon of decoherence is discussed and refuted in [5] (and in many other places in the literature!), and I will not repeat the argument here.

I would argue strongly that, at least to the extent that we refrain from really major revisions of basic concepts such as (e.g.) those concerning the 'arrow of time' in physics, there are only two possible ways out of this dilemma. One is to maintain the belief that the formalism of quantum mechanics is indeed in principle a complete description of the physical universe, but to regard it as nothing more than a formal calculus whose only function is to correctly predict the probabilities of *directly observed* events at the level of everyday life. In this view [6], which is in some sense a logical extension of Bohr's version of the Copenhagen interpretation and which I call the 'extreme statistical' interpretation, the symbols occurring in the formalism correspond to nothing whatever in the 'real world' (whatever that phrase might mean). This point of view, which seems to be increasingly embraced, at least implicitly, by the younger generation of quantum physicists, has no obvious internal inconsistencies, and it seems to me conceivable, if improbable, that it may turn out to be right (in the sense of remaining unrefuted by experiment for the indefinite future), and that our descendants will eventually get as used to it as we are to the world-view of classical physics. (If this should turn out to be the case, then I suspect that the arguments which currently rage between the advocates of the Bohmian, 'many-worlds', modal, consistent histories . . . and other interpretations of the formalism will eventually come to be seen as what I believe they really are, namely mere verbal window-dressing on a deep and fundamental mystery). The only alternative hypothesis which does not involve a really radical (hence almost certainly unforeseeable) breakdown in our current paradigm is that we shall find that at some level between that of the atom and that of our own direct experience quantum mechanics breaks down and needs to be supplemented by some theory of the general nature of the GRWP model.

In this paper I shall not commit myself to the GRWP or any other specific scenario, but will consider a general class of models of 'realization' which is characterized by the following basic postulate, which I call 'macrorealism':

Whenever a macroscopic system has available to it a choice of two (or more) macroscopically distinct states, then it is at all times (apart from short transit times which can be allowed for in the analysis) definitely either in one state or in the other. (I omit the subsidiary assumptions that are necessary to allow the class of macrorealistic theories to be tested experimentally; on this, see [7].)

Needless to say, the physical content of the postulate of macrorealism, and the extent to which one regards it as currently tested or testable, depends critically on one's definition of the notion of ‘macroscopically distinct.’ However, I think it is safe to say that (if one excludes various experiments where the interpretation of the raw data is subject to serious controversy) there is no remotely reasonable definition which would have allowed it to have been tested, even circumstantially, prior to the year 2000. Indeed, even today there is to my knowledge no existing experiment which explicitly refutes realism at any level beyond that of a few atomic-level entities.

However, over the last six years there have been a number of experiments in different areas of physics which have shown that if the raw data are interpreted in quantum-mechanical terms (for the sense of this qualification see [5]), then we can conclude that there exist in nature not just mixtures but quantum superpositions of physical states that (a) possess values of some extensive variable (e.g., the magnetic moment) which differ by an amount of order 10^m times a typical ‘atomic’ value (see below for the number m), and (b) correspond to of order 10^n microscopic particles (electrons or nucleons) behaving ‘appreciably differently’ in the two branches of the superposition. These experiments range from extensions [8] of the original Young’s slits experiment to complex molecules ($m \sim n \sim$ 3–4) to experiments [9] on rf SQUIDs, where m is certainly 9–10 and n may, depending on one’s precise definition of ‘behaving differently’, be 6–7 or 9–10. Since the experiments that would be necessary for an explicit test of macrorealism are quite similar in overall structure to these existing ones (see [7]), it would seem a reasonable hypothesis that if and when they can be performed they will likewise agree with the predictions of quantum mechanics and thus definitively refute macrorealism at this level. However, the level in question is clearly still some distance from that of the everyday world, so that there is plenty of room for some non-quantum-mechanical ‘collapse’ mechanism to inject itself between them. (Incidentally, in view of the content of the rest of this paper, it is worth noting that while the (order-of-magnitude) definitions of m and n are not explicitly Lorentz invariant, they will be agreed upon by any observers not moving with ultrarelativistic velocities with respect to the system(s) in question.)

The question I would like to address in the rest of this paper is: To what extent are such mechanisms *generically* ruled out or constrained by existing experiments in a prima facie different area, namely those relating to Bell’s theorem? The ensuing discussion parallels up to a point that in a recent paper by Kent [10], though the emphasis is somewhat different. Also, while I will be discussing explicitly the EPR-Bell setup, it is clear that some of the issues raised below occur equally in the context of ‘nonlocal Schroedinger cats’ (that is, situations in which a superposition of states of a single microscopic particle is amplified so as to produce, notionally, a superposition of macroscopic events occurring at spacelike separated locations).

For general discussions of the conditions necessary to establish Bell’s theorem (or the CHSH generalization of it) see e.g. [11] or [12] or in more detail [13]. For the purposes of the discussion, I shall assume (a) that the well-known ‘loopholes’ in existing experiments associated with factors such as imperfect detector efficiency, lack of true randomness in the choice of setting, etc, are irrelevant (so that if and when an experiment that is completely loophole free in these respects can be carried out, its results will agree with the predictions of quantum mechanics), (b) that there is no ‘conspiracy’ (see e.g. [13], point (5.2)) and (c) that our usual assumptions concerning the ‘arrow of time’ continue to hold (so that, for example, no information can be transmitted ‘backwards in time’ from the detection stations to the source). Let us further assume that our macrorealistic theory is such that when we are talking about the behaviour of a single spatially isolated macroscopic system in a Schroedinger’s-cat type arrangement, the quantum-mechanical description in terms of a superposition remains valid up to some level, defined for example by the condition that one of the parameters m, n defined above, or some similar parameter, reaches a critical threshold; thereafter, our postulated non-quantum-mechanical collapse (realization) mechanism kicks in and gradually or abruptly realizes a unique macroscopic state. The threshold value in question could be anywhere in the range between the lower limit already set, if it is for *m* or *n*, by existing experiments (at least in SQUIDs, see above) to one corresponding to our direct human experience. For brevity I will refer below to the realization process as the ‘collapse’, without any intended metaphysical implications.

Consider now a standard ‘Bell-EPR’ experimental arrangement, with the photons emitted ‘back-to-back’ in a spin-parity 0+ state and the two mutually distant ‘stations’ labelled S_1 and S_2; until further note I will assume that S_1 and S_2 are equidistant from the source in the rest frame of the latter. Each station consists, schematically, of a polarizer whose transmission axis can be set in an arbitrary direction a (for S_1) or b (for S_2), followed by a detector, let us say for definiteness a standard photomultiplier, whose function is to convert the arrival of a photon into a macroscopic output pulse which is then recorded in some appropriately macroscopic device such as a computer memory. For simplicity I shall assume that any photon that fails to pass the polarizer is absorbed in it, and will count the polarizer as part of the macroscopic ‘measuring device’, which I will label M_1 for station 1 and M_2 for station 2. Thus, (e.g.) the state $|N_1\rangle$ defined below is one in which, while the photomultiplier component of M_1 has failed to fire, the corresponding polarizer is appropriately excited (note that whether or not the photomultiplier is triggered, the photon itself is absorbed and need not be considered further). Also, for the moment (only!) I will treat the complete detection device (‘measuring apparatus’) as physically isolated and initially in a pure state; it will be clear that the latter assumption can be relaxed without affecting the ensuing argument (but the former one is less trivial, see below). We will initially work in a definite Lorentz frame (e.g., that of the source), and suppose, as above, that the spatial distance from the source to the two detection stations is the same. Incidentally, it should be noted that the language of ‘photons’ is used throughout this paper mostly for convenience, without any implication that the basic structure of nature is ‘particle-like’; I believe it would be possible to reformulate everything in the language of quantum fields, but in the present context there seems minimal point in doing so.

Let us now imagine a single photon pair emitted from the source, and suppose that on the run in question the polarizers are set in directions a, b respectively. Consider a stage at which each of the photons, assuming it has passed the relevant polarizer, has entered the anode of the associated photo-multiplier and triggered the latter, but the amplification is at a sufficiently early stage that the collapse has not been triggered, so that a quantum-mechanical description is still valid. Denote the quantum state of measurement apparatus M_i ($i = 1, 2$) in which detector i has fired by $|Y_i\rangle$ and the state, orthogonal to $|Y_i\rangle$, in which it has not fired (but the polarizer has been excited) by $|N_i\rangle$; in the following it is implicit that the states Y_i (at least) are functions of time, as the amplification process progresses.

Then, with our oversimplified ansatz, the joint wavefunction of M_1 and M_2 is

$$\psi(1,2) = 2^{-1/2} \cdot \{(\cos(\theta_{ab}) \cdot (|Y_1\rangle|Y_2\rangle + |N_1\rangle|N_2\rangle) + \sin(\theta_{ab}) \cdot (|Y_1|N_2\rangle + |N_1|Y_2\rangle)\}, \qquad (1)$$

where θ_{ab} is the angle between the polarizer settings. The wavefunction (1) is, needless to say, entangled, and were we to apply the standard measurement axiom at this stage we should, of course, obtain the usual quantum-mechanical predictions for the correlations, which when applied to a set of ensembles of such runs with different choices of a and b would violate the CHSH inequality. However, let us for the moment stay with a single run (i.e. a single experiment with a single pair of settings) and focus on the possibility of engineering, at a slightly later stage, the required physical collapses (one for each measuring device!), and in particular on the constraints set on such a process by the requirements of Einstein locality.

Let us first dispose of two irrelevant points. The first has to do with the concept of macroscopic counterfactual definiteness (MCD) as (implicitly) introduced by Stapp [14]. In the general context of the conceptual implications of the experimental violation of the CHSH inequality this is an important issue, but for our present discussion it is irrelevant, since we are concerned only with the actual behaviour of the system on a particular run, not with its hypothetical behaviour in an unperformed experiment. To put it another way, if we can find a collapse mechanism which is satisfactory in all other respects, the fact that it must violate MCD will be neither more nor less of a worry than it is for standard quantum mechanics (on which point see e.g. [15]).

The second irrelevant point is the observation that provided the only measurements we can make on our measuring devices themselves are diagonal in the $|\mathrm{Y}, \mathrm{N}\rangle$ basis, then the quantum superposition is equivalent, *as regards all its experimental predictions*, to a mixture of the four states involved, which not only is, trivially, unentangled but, even more trivially, does not even involve a superposition of macroscopically distinct states of either device separately. However, it is immediately obvious that to attempt to exploit this circumstance to generate a 'collapse' within the undiluted framework of quantum theory would simply be a special case of the generic 'decoherence' fallacy which I have criticized in [5] and elsewhere.

The real problem is that, in accordance with our usual conceptions of Einstein locality, we would naturally expect any collapse mechanism to be *local*. That is, we would expect all the information necessary to, as it were, instruct measuring device 1 which of the macroscopic states, $|\mathrm{Y}\rangle$ or $|\mathrm{N}\rangle$, to adopt to be available at station 1 or within the backward light cone of the detection event; in particular, it should not be necessary to know the setting b of polarizer 2 (and of course vice versa). As is well known, with this 'local' instruction set it is impossible to violate the CHSH inequality, and thus impossible to generate, for all possible a and b, the experimentally observed correlations. (Needless to say, this condition requires us to think about a sequence of runs with different settings, not as up to now a single run.) It should be emphasized that this postulated insensitivity of the instruction set to the value of b, while it may be thought plausible on general physical grounds, is only guaranteed by Einstein locality to the extent that the act of choice of b is outside the backward light cone of S_1 at the time in question.

In the light of the last sentence, one very obvious way around the problem is to delay the collapse to a sufficiently late stage in the amplifying and recording process, so that it is no longer spacelike separated from the choice of the setting b. This possibility has actually been appreciated for some time (see e.g. [16]); it has been explicitly emphasized recently by Kent [10] under the name of the 'collapse locality loophole.' Indeed, until quite recently the necessary delays were not exorbitant. In particular, in the celebrated experiments of Aspect and co-workers [17] the length of time taken by the photomultipliers to work (*not* the temporal resolution!) was sufficiently large that it would probably have been adequate to take the collapse as occurring at the point where a macroscopic current pulse was produced at the photomultiplier cathode; such a trigger for collapse seems at first sight not unreasonable. However, in the last few years both the spatial intervals and the speed of recording involved in Bell-EPR experiments have increased to the point where in order for a resolution of this kind to work the collapse would have to be pushed back much farther than this. Particularly significant in this respect is the experiment of Weihs *et al* [18], in which the actual recording of the outcome at S_1 took place *in situ* before there was time for a signal to arrive from the setting process at S_2, and rather than correlating the signals from S_1 and S_2 immediately they were stored and compared only hours or days later (membership of a particular pair being identified with the help of accurate timing of the events). Let us suppose that we wish to propose a physical collapse hypothesis for this series of experiments: at what point did the collapse take place? The most extreme view would be that it was only when the information from the two stations is actually combined at a single physical location, which in this case could be literally days later. A simple test of this hypothesis (at least for those of us who are not fans of the ‘many worlds’ interpretation of quantum mechanics!) would be to have two human observers inspect the records at S_1 and S_2 respectively and only subsequently compare notes. This was not in fact done [19] in the experiment of [18], nor to my knowledge in any experiment of this type done to date, but it should be relatively straightforward to do, and I suspect very few people would bet that its outcome would be other than that predicted by a straightforward application of the standard quantum measurement axioms to the state (1). At the other extreme, one might postulate that the collapse takes place as soon as enough time has elapsed for a light signal to propagate from the setting event at S_2 to S_1. In the experiment of [18] this time is of the order of a microsecond, so the hypothesis would presumably not be refutable by involving human observers. So at first sight a collapse scenario having this character, while no doubt somewhat bizarre from the point of view of ‘physical common sense’, should be conceptually and experimentally viable.

Alas, any such scenario would appear to run up against a serious problem, which to the best of my knowledge has not been discussed in the existing literature. Namely, up to now we have assumed not only that the two measurement apparatus M_1 and M_2 are each initially in a pure quantum state, but that each is physically isolated from any ‘environment.’ In real life, however, the apparatus M_1 certainly interacts with various parts of the universe in its vicinity, for example with the ambient radiation field. According to standard arguments in the theory of decoherence, this interaction will tend to correlate different macroscopic states of M_1 to states of the environment that are approximately mutually orthogonal. Thus, the description of the ‘universe’ by the quantum state (1) (multiplied by some definite state of the environment) is too naive; in the approximation that we still neglect the fact that the initial states of M_1 and M_2 are unknown in detail (a complication that may be easily seen not to affect the ensuing analysis in any substantial way), it should be replaced by a wavefunction that is schematically of the form

$$\begin{aligned}|\psi\rangle_{\text{univ}} = 2^{-1/2}\cdot\{&\cos(\theta_{\text{ab}})\cdot(|\mathrm{Y}_1\rangle|\mathrm{Y}_2\rangle|\mathrm{E}_1\rangle|\mathrm{E}_2\rangle + |\mathrm{N}_1\rangle|\mathrm{N}_2\rangle|\mathrm{E}_1'\rangle|\mathrm{E}_2'\rangle)\\ &+\sin(\theta_{\text{ab}})\cdot(|\mathrm{Y}_1\rangle|\mathrm{N}_2\rangle|\mathrm{E}_1\rangle|\mathrm{E}_2'\rangle + |\mathrm{N}_1\rangle|\mathrm{Y}_2\rangle|\mathrm{E}_1'\rangle|\mathrm{E}_2\rangle)\},\end{aligned} \tag{2}$$

where $|\mathrm{E}_1\rangle$, $|\mathrm{E}_1'\rangle$ are approximately mutually orthogonal states of the part of the environment close to S_1 (etc). Evidently, the wavefunction (2) represents the fact that the state of the environment has become entangled with that of the measurement devices, so that the former carries away information about the latter. Moreover, in the absence of pathological good fortune, at least some of this information is likely to be carried away from the neighbourhood of S_1 at the speed of light, thereby erasing any hope of physically affecting it by any operation conducted thereafter at S_1, including the proposed collapse. Thus, our proposed mechanism, while it may be able to reproduce the experimentally measured correlations, has the very unpleasant feature that it leaves a part of the wavefunction, namely that carried off by (part of) the environment, as it were freely floating. Suppose for example that the macroscopic state of M_1 and M_2 realized in the particular event in question corresponds to $|\mathrm{N}_1\rangle|\mathrm{N}_2\rangle$; then, in a physical collapse theory which respects Einstein locality, the ‘surviving’ states of the environment include not only $|\mathrm{E}_1'\rangle$ and $|\mathrm{E}_2'\rangle$, which would be harmless, but also $|E_1\rangle$ and $|\mathrm{E}_2\rangle$. Needless to say, were we interpreting the probability amplitudes (wavefunction) as the ‘extreme statistical’ interpretation of quantum mechanics does, as simply an element in a recipe, with no correspondence to anything in the physical world, this state of affairs would not worry us (in fact, we would simply erase the elements $|E_1\rangle$ and $|E_2\rangle$ from our description at this point); however, the whole motivation of collapse theories is precisely to allow the amplitudes to reflect some ‘element of reality’, so we should indeed be worried about the ontological status of these unwanted pieces of the wavefunction. I will call this difficulty the ‘third-party’ problem. (It is worth noting that it may not arise in Kent’s [10] ‘causal quantum theory’, to the extent that in that scenario the ‘collapse’ can apparently be delayed for as long as we like.)

At this point I feel we might as well grasp the nettle and recognize that *any* physical collapse scenario, no matter to how late a stage the collapse is postponed, is liable not only to appear quite bizarre from the point of view of ‘physical common sense’, but also with high probability to make at least some predictions that are at variance with those of standard quantum mechanics augmented by the usual measurement axiom. (In the case of an ‘early’ collapse (e.g., at or below the level of SQUIDs), it is of course precisely the existence of such contradictions that make existing and currently envisaged tests of macrorealism meaningful; what is perhaps surprising is that they cannot be eliminated by postponing the collapse arbitrarily.) So let us proceed as follows: for the moment, we agree as above to work in a definite (approximate) Minkowski frame (e.g., that in which the earth and thus, under normal experimental conditions, the photon source, is stationary). Then consider the following hypothesis, which I emphasize is not intended as the final one we will adopt, but is introduced simply to illustrate the argument: so long as the description of the universe given by standard quantum mechanics does not contain nonzero amplitudes for two or more macroscopically distinct (by some agreed criterion) states of any macroscopic system, it remains exact and no collapse takes place. As soon as such a ‘macroscopic quantum superposition’ of states of any macroscopic system (or mixture thereof, see below) reaches the prescribed level, the collapse process with respect to the state of that system is activated. Any microscopic systems which, in the epoch when the quantum description was valid, were entangled with the macroscopic system(s) whose state is collapsed are left in a mixture of the relevant states (see example below) and continue to be described by quantum mechanics unless and until they themselves interact with a measuring device (etc) and thereby trigger a macroscopic mixture.

To illustrate this (‘toy-model’) scenario, let us suppose that we were (optimistically, as we shall see) to set the stage at which the collapse occurs very early, for example as soon as a pulse is generated from the cathode of a photomultiplier. Then consider (in distinction to what we have been assuming so far) a *very asymmetric* EPR-Bell setup, in which one detector, say M_1, is much closer to the source than the other, and for the moment ignore the ‘third-party’ (environmental) complication discussed above. Then the evolution of the ‘universe’ following the emission of a photon pair by the source would, according to the above proposal, be as follows: at very early times, when neither photon has reached its detector, the description is just by the usual entangled microscopic state of the pair (multiplied by some uninteresting state of the rest of the universe). When photon 1 has triggered M_1 but the superposition of $|\mathrm{Y}_1\rangle$ and $|\mathrm{N}_1\rangle$ has not yet reached the prescribed level for activation of collapse (and photon 2 has not yet reached its detector), the description is still by a pure quantum state in which the state of M_1 is entangled with that of photon 2: explicitly, if the spin-parity of the photon pair was 0+ and the transmission (rejection) axis of M_1 is labelled y(n), then at this stage

$$|\psi\rangle_{\mathrm{univ}} = 2^{-1/2} \cdot (|\mathrm{Y}_1\rangle|\mathrm{y}_2\rangle + |\mathrm{N}_1\rangle|n_2\rangle), \tag{3}$$

where $|\mathrm{y}_2\rangle$ and $|\mathrm{n}_2\rangle$ denote the polarization states of photon 2 along axes y and n, respectively. Next, once the difference between $|\mathrm{Y}_1\rangle$ and $|\mathrm{N}_1\rangle$ has reached the prescribed level, the (non-quantum mechanical) collapse process kicks in, and the state of M_1 becomes definitely *either* $|\mathrm{Y}_1\rangle$ *or* $|\mathrm{N}_1\rangle$, with probability 0.5 each. What of the state of photon 2, a microscopic entity not subject to the collapse process? According to the prescription given above, following the collapse of the state of M_1 photon 2 should be described as a *mixture* of states $|\mathrm{y}\rangle$ and $|\mathrm{n}\rangle$, in this case each with probability 1/2. This has the consequence that when, much later, photon 2 reaches detector 2, it will trigger a mixture of $|\mathrm{y}_2\rangle$ and $|\mathrm{N}_2\rangle$, which on reaching the prescribed level of ‘macroscopic difference’ will be collapsed into one or other of these states with probability $1/2$ each. It is immediately clear that the correlations predicted by such a scenario are in violent contradiction with those predicted by standard quantum mechanics, so that (assuming that in at least some existing EPR-Bell experiments the setup has been sufficiently asymmetric for the above analysis to apply) the specific proposal considered in this paragraph (that with ‘early’ collapse) is experimentally refuted.

What is much less obvious, however, is whether a scenario of this general nature with ‘sufficiently late’ collapse would be viable (in the minimal sense of not being already refuted by existing experiments). We must now postulate that, whether or not the macroscopic system is entangled with another (possibly distant) one, the collapse takes place only at a sufficiently ‘late’ stage that for all existing experiments a time T has elapsed which is sufficient for communication with the distant device. Undoubtedly, to however late a stage the collapse is postponed, it would *in principle* make some predictions that contradict those of standard quantum mechanics. Consider for example the standard (symmetric) EPR-Bell setup. To the extent that we make the unrealistic assumption of complete isolation of the measuring devices from the environment, it is clear (as already discussed, and see [10]) that by deferring the collapse until not only have both devices reached the prescribed level of macroscopic superposition but they are no longer spacelike separated, we can engineer a collapse which, while respecting Einstein locality, nevertheless recovers the quantum correlations. However, once we take the environment into account, then it is clear that a ‘third-party’ photon will behave similarly to photon 2 in the toy example of the last paragraph; following the collapse of the correlated superposition of the states of M_1 and M_2, this third-party photon will be left in a mixture some of whose components would not be there in the standard quantum analysis, and if we at some later time arrange to detect the state of this photon, the correlations predicted by our scenario will certainly differ from those of standard quantum mechanics. However, it seems extremely unlikely that any experiment along the above lines has been actually done (and for various obvious reasons it would be quite difficult to do in practice), so that it seems highly probable that the scenario is currently viable in at least in the most minimal sense.

Of course, this is only the beginning of our problems. The above analysis was explicitly conducted by working in a specific Minkowski frame, which enabled us to specify unambiguously the time at which the collapse process was initiated. Even in this case, unless the criterion for onset of collapse is perfectly discontinuous, there are obvious difficulties in applying the prescription unambiguously (e.g., in the toy model, what happens if at the moment when the criterion is realized in M_1 photon 2 is just entering the anode of M_2?). If we require that the prescription should respect the constraint of Lorentz invariance, things get a lot worse. In the case of the specific collapse mechanism envisaged by the GRWP scenario, it is well known (see e.g., [20]) that there are serious difficulties in generalizing it so as to do this, and there is no obvious reason why more generic mechanisms should be immune to these problems. In particular, it is clear that in the situation considered in the last paragraph, whenever the events of detection of the original pair are spacelike separated, one can by an appropriate Lorentz transformation reverse their time order (and this was in fact done in the elegant experiment reported in [21], if ‘detection’ is defined as the production of the first reasonably macroscopic event). Does this matter? Obviously, it means that the formal description of the collapse processes must be observer dependent. But the important question is, do all observers predict the same correlations?

Even in the unrealistic approximation of neglect of environmental coupling and the associated third-party effects, it appears we have a problem, since by an appropriate Lorentz transformation we can make the time interval between the points at which the criterion is reached in M_1 and M_2 arbitrarily long, so that for any finite value of T the first collapse will take place before the events can be causally connected. It would appear that as a minimum we must be prepared to attach special status to the frame of reference in which the source and/or the measuring devices are at rest, and specify that the prescriptions for collapse are to be applied in that frame only. This manoeuvre should be sufficient to cope with the ‘third-party’ complications mentioned above, at least as regards all existing experiments. (Incidentally, I should be surprised if the argument of this paragraph, apart perhaps from the last point, has not previously appeared in the literature, although I cannot at the time of writing quote a specific reference).

In conclusion, if we demand full Lorentz invariance (in the sense that the prescriptions for collapse should be independent of the Lorentz frame in which they are applied), then a generic physical collapse mechanism would appear prima facie to be unable to allow all observers to predict correlations in agreement with existing experiment. If on the other hand we are content to apply the prescriptions only in a preferred frame and to demand that the model respect Einstein locality in that frame, then it would appear that such a model is currently viable even when the ‘third-party’ effects are taken into account, in the sense that it predicts nothing which contradicts *existing* experimental results. Needless to say, that is not to say that such a scenario is necessarily elegant or in accordance with the ‘common-sense’ intuition of the majority of physicists; but that is an occupational hazard of any such scheme, and the reader will have to decide for him/herself what weight to give such considerations.

Acknowledgments

This work was supported by the National Science Foundation under grant no. NSF-EIA-01-21568. I am grateful to Lucien Hardy for helpful discussions and for drawing my attention to [10], to Adrian Kent for helpful correspondence, and to Grigor Weihs for information concerning the experiments reported in [18]. It is a pleasure to dedicate this paper to Professor Giancarlo Ghirardi on the occasion of his seventieth birthday, and to wish him many more happy years of activity in the foundations of physics.

References

[1] Ghirardi G C, Rimini A and Weber T 1986 *Phys. Rev.* D **34** 470
[2] Pearle P 1989 *Phys. Rev.* A **39** 2277
[3] Ghirardi G C, Pearle P and Rimini A 1990 *Phys. Rev.* A **42** 78
[4] Penrose R 1996 *Gen. Rel. Grav.* **28** 281
[5] Leggett A J 2002 *J. Phys. Cond. Mat.* **14** R415
[6] Ballentine L E 1970 *Rev. Mod. Phys.* **42** 358
[7] Leggett A J and Garg A 1985 *Phys. Rev. Lett.* **54** 857
[8] Arndt M *et al* 2005 *Phys. World* **18** 35
[9] Friedman J R *et al* 2000 *Nature* **406** 43
[10] Kent A 2005 *Phys. Rev.* A **72** 12107
[11] Clauser J F and Shimony A 1978 *Rep. Prog. Phys.* **41** 1881
[12] Leggett A J 1987 *The Problems of Physics* (Oxford: Oxford University Press) pp 161–7
[13] Valdenebro A G 2002 *Eur. J. Phys.* **23** 569
[14] Stapp H P 1971 *Phys. Rev.* D **3** 1303
[15] Peres A 1978 *Am. J. Phys.* **46** 745
[16] Leggett A J 1987 *Foundations of Quantum Theory in the Light of New Technology* ed M Namiki *et al* (Tokyo: Japanese Physical Society) p 287 (Cf also H P Stapp 1980 *Found. Phys.* **10** 767)
[17] Aspect A, Dalibard J and Roger J 1982 *Phys. Rev. Lett.* **49** 1804
[18] Weihs G *et al* 1998 *Phys. Rev. Lett.* **81** 3563
[19] Weihs G 2006 private communication
[20] Pearle P 2005 *Phys. Rev.* A **71** 32101
[21] Zbinden H *et al* 2001 *Phys. Rev.* A **63** 022111