A Hidden Variable Explanation of Aspect's Asymmetry Anomalies

Caroline H Thompson*

ABSTRACT

Some little-known anomalies reported in Alain Aspect's thesis may be evidence that the accepted quantum mechanics (QM) cosine formulae do not, after all, correctly predict the results of his two-channel paired-photon EPR-type experiment reported in 1982. The anomalies (inequalities between, for example, N_{+-} and N_{-+} coincidence rates), occur completely naturally under a local realist (hidden variable) theory. The latter also predicts that the total of the four coincidence rates should vary slightly with relative detector angle. Aspect's statement that it does not may be approximately true for the particular angles he looked at. The phenomena are illustrated using the Chaotic Ball analogy of a previous paper and some simple assumptions regarding probabilities of detection. The settings of the photomultiplier voltages and discriminator thresholds, which can, in real experiments, be different for each of the four detectors, may be very important. It may not be necessary to do more than look carefully at some of Aspect's unpublished data to prove beyond reasonable doubt this experiment, at least, obeys Einstein, Furry and others' ideas on locality and separability. In how many other experiments does QM's denial of the existence of hidden variables preclude it from making *accurate* predictions of responses to changes in instrument settings?

*Department of Computer Science, UW Aberystwyth, Dyfed, SY23 3DB, UK.
e-mail: cat@aber.ac.uk

INTRODUCTION

Alain Aspect in his thesis (Aspect, 1983, p306) mentions that sometimes in his two-channel experiment (Aspect et al., 1982) coincidence rates that should under quantum theory (QM) have been equal were not so. For example he found that N_{+-} tended to differ from N_{-+} by about 2%. This did not appear to matter, as the two were added when he performed a Bell's test and the sum behaved as predicted. A hidden variable (local realist) model, however, can explain the discrepancy fully, once account is taken of the many sources of asymmetry that are present, and which are discussed in the thesis.

Caser (Caser, 1984) was perhaps the first to recognise that asymmetry could be important as a factor in explaining the violation of Bell's inequalities in real experiments. This paper describes how it can also be important in explaining another class of apparently anomalous results. The mathematics behind this is very simple. I shall attempt, by adapting the Chaotic Ball analogy developed earlier (Thompson, 1995a), to make the the principle clear at an intuitive level, and to make testable predictions.

I shall also discuss briefly some of the practical considerations involved, those concerning the possibility that the shape of the photomultiplier "response curve" may vary as one varies the voltage across the instrument itself and, most importantly, varies the setting of the threshold of the "discriminator" – the instrument that converts irregular pulses output by the photomultiplier into standardised ones ready for input to "integrators" and "coincidence monitors".

It would be relatively easy to check the ideas of this paper by repeating an experiment like Aspect's and systematically varying the instrument voltages. QM would always predict (Lepore and Selleri, 1990) that the extreme values of differences (if any) would occur in the parallel and antiparallel positions of the polarisers. A hidden variable theory can predict that extremes will sometimes be in other positions.

AN ASYMMETRIC CHAOTIC BALL EXPERIMENT

The original chaotic ball model (Thompson, 1995a) involved two assistants, Anne and Bob, who looked at a ball that moved in chaotic fashion. On the ball were marked an N and an S, at opposite poles. A prearranged times, the assistants recorded whether they saw an N or an S. It was assumed that they recorded definite results whenever one of the letters was in their field of vision.

Several modifications are needed in order to model asymmetry in a manner that is reasonably close to the behaviour of real paired-particle systems:

1. We need *four* assistants, not two, in recognition of the fact that real experiments have four detectors and *they do in fact all have different characteristics.* Let us call them 'Anne-' and 'Anne+', standing opposite each other, and 'Bob-' and 'Bob+',similarly opposite each other. We define the vector **a** to be the direction from Anne- to Anne+, and likewise **b** for the two Bobs, with the angle between the two vectors being ϕ. Each assistant records just the Ss seen, ignoring the Ns, which need not even be marked.

2. As in one variation mentioned in my original paper, the assistants may all stand at different distances from the ball, so that they have different-sized missing bands.

3. We have to assume that, even when an assistant can see an S clearly, it has only a certain chance of being recorded. To make this idea concrete, let us assume that results are entered electronically into a device, and this device can be adjusted so as to control the proportion of successful entries. In other words, we are assuming fixed "efficiencies", ϵ_{A-}, ϵ_{A+}, ϵ_{B-}, ϵ_{B+}, (using obvious notation) for the electronic devices.

The model thus has very considerable asymmetry, with the possibility of different rules for each of the four cases. We can, *as in real experiments*, adjust our instruments so as to make the relative "singles rates" take any value we please. This assumes, of course, that we are working with basically low-efficiency instruments, as, if any were perfect, we would not be able to increase its detection rate[1]. One difference between the model and real experiments is that we would not expect in practice to be able to vary the efficiencies and the "ranges of visibility" independently. This does not affect the model's validity for illustrating the effect of real asymmetries.

In Aspect's experiments, it should be noted, efficiencies were commonly of the order of 3%. The detection

[1]Is this last statement true? Looking at graphs in Aspect's thesis (p 217), it would appear that reducing the discriminator threshold increases the "detection rate" without limit. This is presumably a matter of counting more and more tiny pulses that represent pure noise.

rates on the two sides of the experiment differed considerably, due to the sensitivity of the optical instruments to wavelength and the fact that the wavelengths of the two sides were different. Aspect made adjustments, though, to equalise the + and − singles rates within each side. This is not clear from his 1982 paper: he seems to have attached little importance to it, but it is stated in his 1983 thesis. The + and − channels were not intrinsically symmetrical, as one involved reflection and one transmission at a surface between two prisms.

The general idea of the model is shown in Fig. 1.

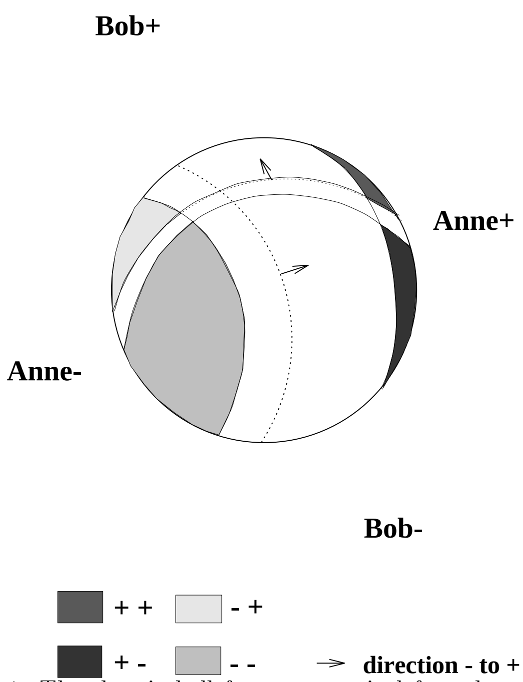


Fig. 1. The chaotic ball for asymmetrical four-channel experiment

The probabilities of *recording* coincidences are obtained by multiplying the appropriate shaded areas by the product of the two efficiencies involved. When all assistants stand at different distances, there is no reason to expect symmetry: the +− region will differ in area from the −+ one, except for special cases when the pairs of assistants stand in particular places. If coincidence probabilities are equal for one particular relative angle, then they cannot remain so for other angles as the proportion of the areas does not remain constant.

It has to be admitted that the analogy is rather far stretched. The reader can, if he so wishes, abandon it and just think in terms of hidden variable spaces. The hidden variable space (the space of possible orientations of the ball) is a real fixed 3-dimensional space. The measuring devices (assistants, together with electronics) define various areas on the space within which each result has a definite probability of occurrence. The probabilities of coincidences are, following the realist "separability" assumption, obtained by multiplying the probabilities of "singles"[2].

MATHEMATICS OF ASYMMETRY

The mathematics behind a slight generalisation of the above idea is as follows:

The experiment is fully defined by four "response functions" that give the probabilities of a result being both observed and recorded for each position of the ball (value of the hidden variable $\boldsymbol{\lambda}$). If we write the functions as $\boldsymbol{p}_{A-}(\boldsymbol{\lambda}, \mathbf{a})$, $\boldsymbol{p}_{A+}(\boldsymbol{\lambda}, \mathbf{a})$ etc., then the singles rates will be proportional, on average, to the probabilities $\boldsymbol{P}_{A-}$, $\boldsymbol{P}_{A+}$ etc. given by expressions such as:

$$\boldsymbol{P}_{A-} = \int \boldsymbol{\rho}(\boldsymbol{\lambda}) \boldsymbol{p}_{A-}(\boldsymbol{\lambda}, \mathbf{a}) \boldsymbol{d\lambda}, \tag{1}$$

$\boldsymbol{\rho}$ being the probability density, assumed uniform over the surface of a sphere and hence equal to $\mathbf{1/4\pi}$.

We assume that we can adjust the response functions so as to make $\boldsymbol{P}_{A-} = \boldsymbol{P}_{A+}$ and $\boldsymbol{P}_{B-} = \boldsymbol{P}_{B+} =$. The mathematical content of this paper consists of the observation that one would not in general expect to find the +− and −+ coincidence rates equal *for all values of* $\boldsymbol{\phi}$ *simultaneously*. These rates are assumed, using realistic notions of separability, factorability, rotational symmetry and "coplanarity" ($\mathbf{a}$ and $\mathbf{b}$ are always in the same plane), to be proportional to the coincidence probabilities, $\boldsymbol{P}_{+-}(\boldsymbol{\phi})$ and $\boldsymbol{P}_{-+}(\boldsymbol{\phi})$, given by expressions such as:

$$\boldsymbol{P}_{+-}(\boldsymbol{\phi}) = \int \boldsymbol{\rho}(\boldsymbol{\lambda}) \boldsymbol{p}_{A+}(\boldsymbol{\lambda}, \mathbf{a}) \boldsymbol{p}_{B-}(\boldsymbol{\lambda}, \mathbf{b}) \boldsymbol{d\lambda}, \tag{2}$$

If the individual response functions were all the same shape, differing at most by multiplicative constants, then

[2] A word of warning: this idea of simply multiplying the singles probabilities will not be *exactly* correct in all experiments. "Non-factorability" can sometimes arise, due to timing variations, unless precautions are taken in relation to the coincidence monitor settings. The deviation from factorability may well be small. It is ignored in the remainder of this paper, but the interested reader can find more in another recent paper (Thompson, 1995b).

it *would* be reasonable to expect $+-$: $-+$ symmetry, but we have no reason to suppose that this is the case.

Now, under QM, all the coincidence probabilities are of the same functional form, and are very nearly proportional to each other. They are all of form $X + Y \cos\phi$, where X and Y are constants[3]. Apart from the fact that these constants have to obey a few constraints, the QM rules make it look possible to adjust them to get at least very close to symmetry. When this is not quite achieved, QM predicts that the difference will be of the same functional form, namely $X + Y \cos\phi$. An important consequence is that maxima and minima will be at $\phi = 0$ and $\phi = \pi$. It would be interesting to know whether or not Aspect's observed discrepancies fitted this picture. It seems likely that they did not, otherwise he would have felt able to explain them. He does not report the positions of his extremes.

WORKED EXAMPLE

The three-dimensional picture is useful intuitively, but when it comes to the arithmetic (and, fortunately, when we look at real experiments involving photon polarisation), a one-dimensional version is more reasonable. We obtain this by constraining the motion of the ball so that the S is always on the equator. 'Anne-', for example, then sees it when it is region R_{A-}, as shown in Fig. 2.

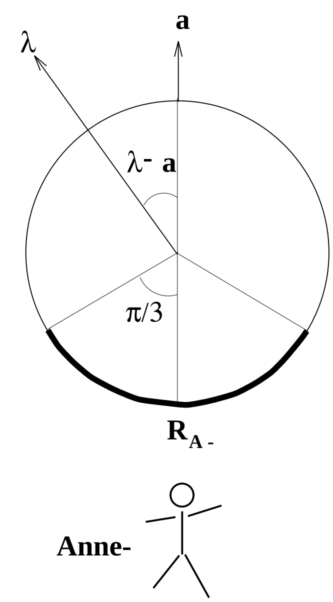
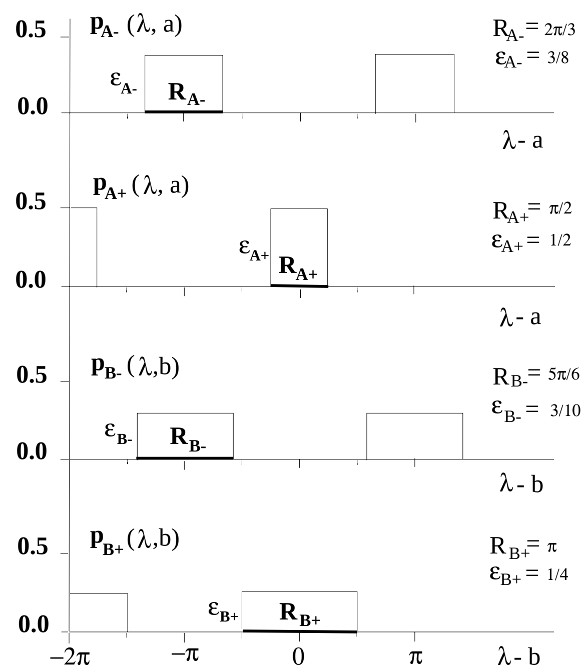


Fig. 2. Anne-'s view of 1-dim version of ball

Let us suppose our response functions are as shown in Fig. 3, where R_{A-}, for example, stands for the angle corresponding to $\mathbf{R_{A-}}$ and the ϵs are efficiencies, as defined earlier. Efficiencies and visibility ranges have been chosen to make the singles rates all equal to $1/8$. (The singles formula (1) reduces to multiplying the area of the rectangle by $\rho = 1/2\pi$).



[3]Actually the formulae involve 2ϕ rather than ϕ for photon experiments: there is a difference of a factor of two between all angles in the ball model and those for photon polarisation.

Fig. 3. Response Functions

The effect of these individual response functions on coincidence rates, obtained by applying equation 2, is shown in Fig. 4, and the difference between $+-$ and $-+$ coincidences by Fig. 5.

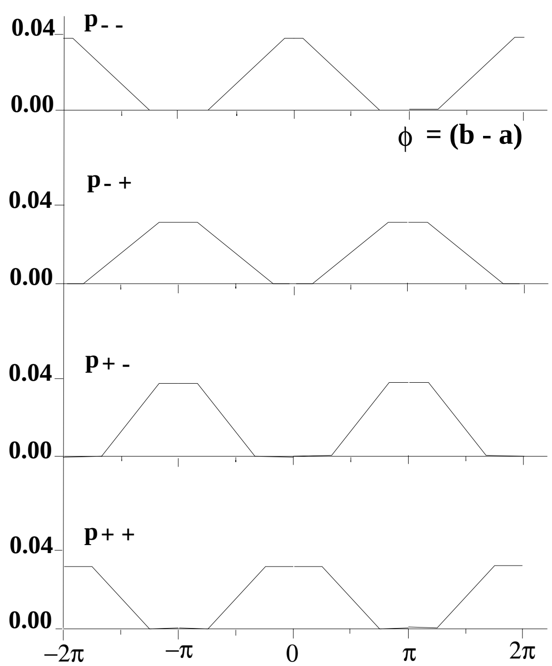


Fig. 4. Coincidence Probabilities

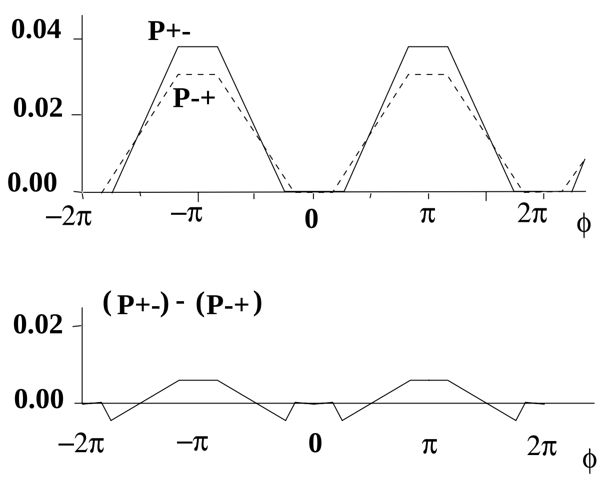


Fig. 5. Predicted Anomaly: P_{+-} does not equal P_{-+}.

The final graph gives an idea of the kind of deviation from the cosine curve predicted by our realist theory: there may be maxima symmetrically placed about either **o** or $\boldsymbol{\pi}$ and minima similarly placed about the opposite value. The values at **o** and $\boldsymbol{\pi}$ are not necessarily themselves either maxima or minima. They can in special circumstances be equal to each other.

Under quantum mechanics, the maximum and minimum can only occur at **o** and $\boldsymbol{\pi}$.

In practice, of course, there is no direct way of knowing the shapes of the individual response functions, so we cannot make definite predictions of the $+-$: $-+$ dependency on $\boldsymbol{\phi}$. We can, however, say that any deviation from a cosine relationship is evidence against QM, and what we would also hope to be able to do is to predict the manner in which the dependency will change as we vary our experimental parameters (in particular, the discriminator threshold). For this, a detailed model such as Marshall's (Marshall, 1988), which allows explicitly for a threshold, might be useful, though we would expect the same qualitative predictions from a wide class of realist models.

REAL EPR EXPERIMENTS

Quantum mechanics embodies the built-in assumption that a "detector" (photomultiplier, for example) is characterised by a single figure, its "quantum efficiency". It is not possible within the theory to discuss the possibility that different inputs might be detected with different efficiencies. It is therefore not possible to conceive of the possibility that the manner in which probabilities of detection vary with amplitude of input signal could be important. Experimenters have felt quite free to adjust the "quantum efficiency" by (in Aspect's case) altering the voltage across the photomultiplier and adjusting the threshold voltage of the discriminator.

Aspect did not feel it necessary to make his $\boldsymbol{A}$ and $\boldsymbol{B}$ sides equal, but he *did* make adjustments to equalise his + and − channels within each side. What effect was this *really* having? At this point, we may get quite different descriptions according to our theoretical predilections, as there seems to be no consensus theory on what photomultipliers *really* do. Do they really detect particles or waves? If particles, then have they all got the same energy? If they are waves, then it is obvious that their amplitude may vary.

I shall restrict my discussion to my own particular predilection, which is highly unorthodox but which I do not believe to be contradicted by any experimental results – and which, indeed, appears to me to be suggested by the totality of Aspect's results.

I see the individual "photon" as a Maxwell-type electromagnetic wave that spreads over the entire receptive surface of the photomultiplier, probably in a manner that is uniform neither in time nor space. The photomultiplier responds with an output pulse, the size and shape (and, incidentally, timing) of which depends not only, as quantum mechanics would have it, on the frequency, but also on the *intensity of the individual signal*. The variations of output pulse size would arise if a single input "photon" causes emission of more than one electron when its "amplitude" is sufficiently great, and/or if high amplitude signals tend to cause emission of high energy electrons. Low intensity signals are lost amongst the noise (the "dark rate signals") produced by what Marshall and Santos's stochastic optics would refer to as the Zero Point Field (ZPF) (Marshall and Santos, 1989). This can be seen simply as all the effectively random electromagnetic signals arising from all the atoms of the rest of the universe. All photomultiplier signals are, in addition, subject to quite a high level of random variation, due to the stochastic nature of the series of electron cascades that builds up the pulse. My theory is that *on average* we have a dependency of pulse size on signal intensity, *even within the "single-photon" level.*

What happens, then, if we alter our discriminator threshold? If we make it very low, we inevitably count as signals some pulses that were really just "dark rate". This, as Aspect recognised, should not matter too much as the signals will not tend to count as coincidences – they will not arrive in synchrony with any pair. If, on the other hand, we increase the threshold, we are in danger of creating or augmenting a missing band. The possibilities for the existence of varying missing bands or, more generally, varying shapes of response functions, is in this description obvious.

TESTING

Firstly, we need to check against Aspect's actual data. It is slightly disturbing that, though Aspect mentions the asymmetry anomaly, he also states that the sum of the four coincidence rates was found *not* to vary with ϕ. Our model predicts small variations. The explanation is likely to be a combination of the fact that variations are indeed very small, together with the fact that Aspect looked only at a few angles, those needed for Bell's tests.

Further testing can be done by modifying any EPR experiment that records both + and − signals, varying the "quantum efficiency" of each detector. Our results may depend to some extent on how we do this. The picture may be slightly different if, for example, we effectively alter the input intensity by defocussing it, as opposed to altering the discriminator setting.

It would seem a high priority also to do experiments to investigate more closely the nature of the photomultiplier response. Given a source of "individual photons" and, say, two polarisers in series, it should be possible to check the belief that has been held ever since 1928 (Lawrence and Beams, 1928), that the size and shape of the output pulse depends only on the frequency of the input. My personal interpretation is that Aspect's experiments contain evidence that it also varies with individual "photon amplitude".

DISCUSSION

We have here a very definite possibility of using a (local) hidden variable theory to make predictions that quantum mechanics cannot be adjusted to match, at least, not without incorporating separability and some equivalent to hidden variables. Without hidden variables, it would seem that there is no way that QM can correctly embody all the information about detector response characteristics that goes into predicting the exact coincidence functions. Do real experiments always produce cosine curves, or is it just that those that do not are, for one reason or another, rejected? It seems to me to be very important that we find out, for if QM really is wrong, we should not be straining our intellects in treating it as anything more than a set of useful algorithms.

The approach of this paper represents an attempt to change the direction of the experimental work in the "quantum theory versus local realism" debate.

Bell's tests represent an area in which QM predicts something that local realist theories do not. Real experiments appear to support QM, yet, not in my view surprisingly, realistic explanations have always been found, so we are no further forward: we have failed to discriminate. What has in fact happened is that Bell's tests have been wrongly applied in attempts to circumvent the need for almost perfect detectors, and the day when detectors of sufficient perfection are found seems still a long way off.

Should not new research be directed towards areas in which realistic theories predict events that QM does not? There have been several such areas suggested already: Philip Pearle (Pearle, 1970) noted that hidden variable models generally predict that the total number of coincidences will be greatest for the parallel and antiparallel settings, least for perpendicular ones; Marshall, Santos and Selleri's model (Marshall et al., 1983) is in agreement with this and they state that the greatest discrepancies between their model and QM will be found in the $\mathbf{0}^o$, $\mathbf{45}^o$ and $\mathbf{90}^o$ settings (which correspond, in photon polarisation experiments, to parallel and perpendicular settings). No experiment has, to my knowledge, ever been conducted with the specific aim of looking into these predictions. The only mention of variation of total coincidences with detector setting that I have found is the one mentioned above, namely Aspect's negative statement, which cannot be taken as general.

The conviction that the real world does contain phenomena that QM cannot "explain" (if, indeed, QM can ever be said to explain anything!), arises from the belief that many people share with Einstein, Podolsky and Rosen (Einstein et al., 1935), Furry (Furry, 1936) and other authorities, that QM is incomplete. It is incompatible with the existence of hidden variables. If, therefore, the latter are a feature of the real world, one would expect them on occasion to reveal their existence. In my opinion, the small deviations from cosine curves that we are discussing are precisely the kind of evidence we should be looking for. In sharp contrast to the Bell's test approach, investigations along these lines should give the best discrimination for imperfect detectors.

REFERENCES

Aspect, A., Grangier, P., and Roger, G., 1982, Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: A new violation of Bell's inequalities: Physical Review Letters, **49**, 91–94.

Aspect, A., 1983, Trois tests expérimentaux des inégalités de Bell par mesure de corrélation de polarisation de photons: Ph.D. thesis, Université de Paris-Sud, Centre D'Orsay.

Caser, S., 1984, Objective local theories and the symmetry between analysers: Physics Letters A, **102**, 152–8.

Einstein, A., Podolsky, B., and Rosen, N., 1935, Can quantum-mechanical description of physical reality be considered complete?: Physical Review, **47**, 777–780.

Furry, W. H., 1936, Note on the quantum-mechanical theory of measurement: Physical Review, **49**, 393–99.

Lawrence, E. O., and Beams, J. W., 1928, The element of time in the photoelectric effect: Physical Review, **32**, 478.

Lepore, V. L., and Selleri, F., 1990, Do performed optical tests disprove local realism?: Foundations of Physics Letters, **3**, 203–220.

Marshall, T. W., and Santos, E., 1989, Stochastic optics: A local realistic analysis of optical tests of Bell inequalities: Physical Review A, **39**, 6271–6283.

Marshall, T. W., Santos, E., and Selleri, F., 1983, Local realism has not been refuted by atomic-cascade experiments: Physics Letters A, **98**, 5–9.

Marshall, T. W., 1988, Stochastic electrodynamics and the Einstein-Podolsky-Rosen argument in: *Quantum Mechanics versus Local Realism, The Einstein-Podolsky-Rosen Paradox* (F. Selleri ed.), Plenum Press, pages 413–432.

Pearle, P., 1970, Hidden-variable example based upon data rejection: Physical Review D, **2**, 1418–25.

Thompson, C. H., 1995a, The chaotic ball: An intuitive analogy for EPR experiments: submitted for publication in Foundations of Physics Letters.

——— 1995b, A non-factorable local realist model for EPR experiments with timing variations: unpublished.