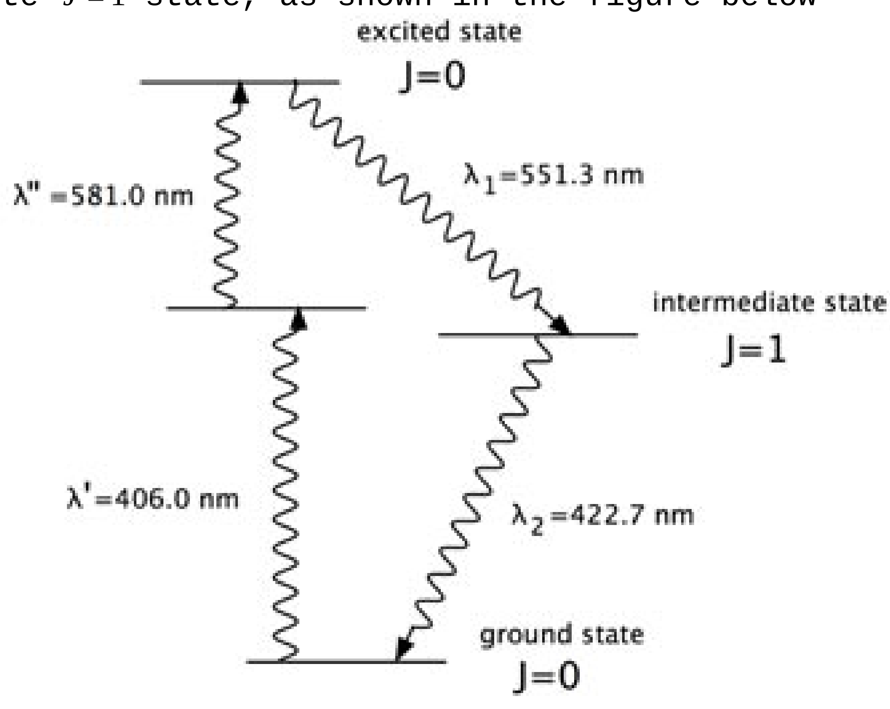
EPR/Bell - The Details

Let us first rethink some quantum mechanical ideas in a context needed for this discussion. This review will hopefully reinforce the ideas you have learned so far.

Single-Photon Interference

All good discussions on quantum mechanics present a long an interesting analysis of the double slit experiment. The crux of the discussion comes when **"the light intensity is reduced sufficiently for photons to be considered as presenting themselves at the entry slit one by one"**. For a long time this point was very contentious, because correlations between two successive photons cannot be ruled out *a priori*. Since 1985, however, the situation has changed. An experiment was performed by Grangier, Roger and Aspect. It was an interference experiment with only a single photon. They used a light source devised for an EPR experiment which guarantees that photons arrive at the entry slit singly. The experiment is difficult to do in practice, but is very simple in principle and it provides an excellent experimental introduction to the concepts of quantum mechanics.

The light source is a beam of calcium atoms, excited by two focused laser beams having wavelengths $\lambda' = 406 nm$ and $\lambda'' = 581 nm$ respectively. Two-photon excitation produces a state having the quantum number $J = 0$ (angular momentum). When it decays, this state emits two monochromatic photons having the wavelengths $\lambda_1 = 551.3 nm$ and $\lambda_2 = 422.7 nm$ respectively, in a cascade of two electronic transitions from the initial $J = 0$ level to the final $J = 0$ state, passing through an intermediate $J = 1$ state, as shown in the figure below



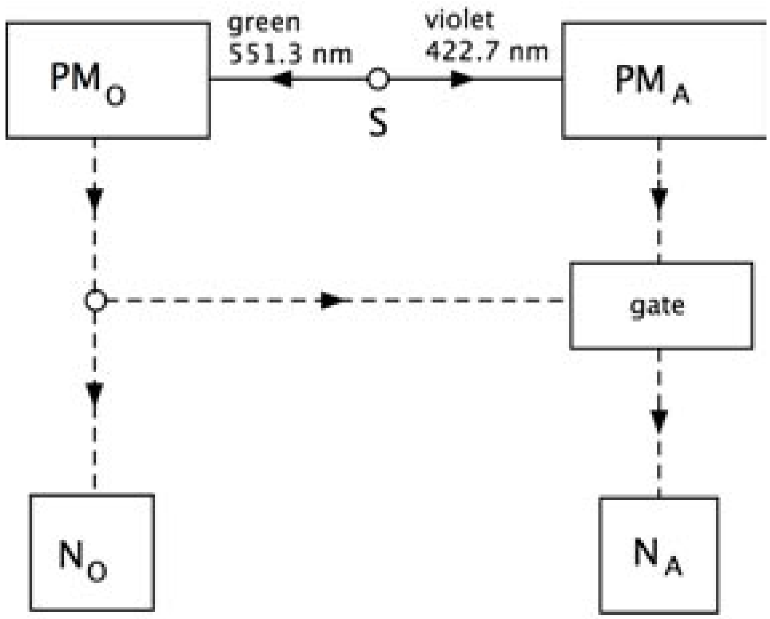
Excitation and decay of the calcium atom

The mean lifetime of the intermediate state is 4.7 ns. To simplify the terminology, we shall call the $\lambda_1 = 551.3 nm$ light green, and the

$\lambda_2 = 422.7nm$ light violet.

Next we describe the experiment, exhibiting its three stages which reveal the complications of the apparatus in progressively greater detail (next three figures).

1. The first stage is a trivial check that the apparatus is working properly; nevertheless it is already very instructive (figure below).



Interference with a single photon (first stage). In this sketch, solid lines are optical paths and dashed lines are electrical connections

On either side of the source S one positions two photomultiplier tubes PM_O and PM_A. These are very sensitive, and can detect the arrival of a single photon. Detection proceeds through photoelectric absorption, followed by amplification which produces an electric signal proportional to the energy of the incident photon. The associated electronic logic circuits can identify the photons absorbed by each detector: the channel PM_O responds only to green light, and the channel PM_A responds only to violet light. The electronic gate is opened (for 9 ns - this is twice the mean lifetime and corresponds to an 85% probability that the photon has been emitted) when green light is detected by PM_O. If, while the gate is open, violet light is emitted by the same atom towards PM_A, then PM_A detects this photon, producing a signal that passes through the gate and is counted in N_A. The counter N_O registers the number of green photons detected by PM_O. It turns out that $N_A << N_O$. As the observation period becomes very long(approximately 5 hours), the ratio N_A / N_O tends to a limit that is characteristic of the apparatus. It represents the probability of detecting a violet photon in PM_A during the 9 ns following the detection of a green photon by PM_O.

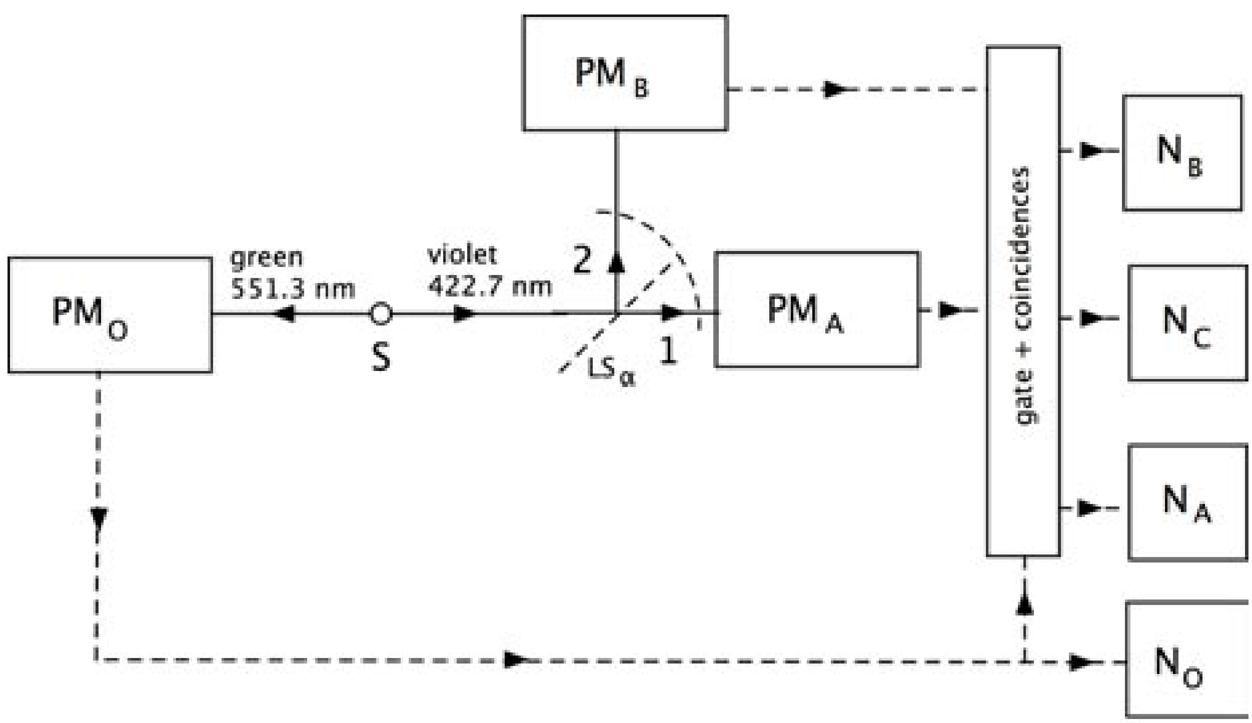
The purpose of this arrangement is to use a green photon in order to open a 9 ns time window, in which to detect a violet photon emitted by the same atom. As we shall see, there is only an extremely small probability of detecting another violet photon emitted by a different atom within the same window.

We will assume that a second observer is in the lab. This observer always feels compelled to present what he thinks are "simple-minded truths" using ordinary words. We will called this second observer Albert. Albert, as we shall see, has a tendency to use, one after another, the three phrases, "I observe", "I conclude", and "I envisage". Consulted about the above experiment, Albert states, with much confidence,

> I observe that the photomultiplier PM_A detects violet light when the source S is on, and that it ceases to detect anything when the source is off. I conclude that the violet light is emitted by S, and that it travelled from S to PM_A.
>
> I observe that energy is transferred between the light and the photomultiplier PM_A always in the same amount, which I will call a quantum.
>
> I envisage the quanta as particles, emitted by the source, propagating freely from S to PM_A, and absorbed by the detector. I shall call these quanta photons.

Albert stops talking at this point.

2. The second stage of the experiment introduces the concept of individual photons (see figure below)



Interference with a single photon (second stage). In this sketch, solid lines are optical paths and dashed lines are electrical connections

Across the path of the violet light one places a half-silvered mirror LS_α, which splits the primary beam into two secondary beams(equal intensity), one transmitted and detected by PM_A, the other reflected and detected by PM_B. As in the first stage, the gate is opened for 9 ns, by PM_O. While it is open, one registers detection by either PM_A (counted as N_A); or by PM_B (counted as N_B); or by both, which we call a coincidence (counted as N_C). The experiment runs for 5 hours again and yields the following results:

(a) The counts N_A and N_B are both of the order of 10^5. By contrast, N_C is much smaller, being equal to 9.

(b) The sequence of counts from PM_A is random in time, as is the sequence of counts from PM_B.

(c) The very low value of N_C show that counts in PM_A and PM_B are mutually exclusive (do not occur at same time).

The experimenters analyze the value of N_C in depth; their reasoning can be outlined as follows:

(a) Suppose two different atoms each emit a violet photon, one being transmitted to PM_A and the other reflected to PM_B, with both arriving during the 9 ns opening of the gate; then the circuitry records a coincidence. In the regime under study, and for a run of 5 hours, quantum theory predicts that the number of coincidences should be $N_C = 9$. The fact that this number is so small means that, in practice, any given single photon is either transmitted or reflected.

(b) If light is considered as a wave, split into two by LS_α and condensed into quanta on reaching PM_A and PM_B, then one would expect the photon counts to be correlated in time, which would entail $N_C >> 9$. Classically speaking this would mean that we cannot have a transmitted wave without a reflected wave.

(c) Experiment yields $N_C = 9$; this quantum result differs from the classical value by 13 standard deviations; hence the discrepancy is very firmly established, and allows us to assert that we are indeed dealing with a source of individual photons.

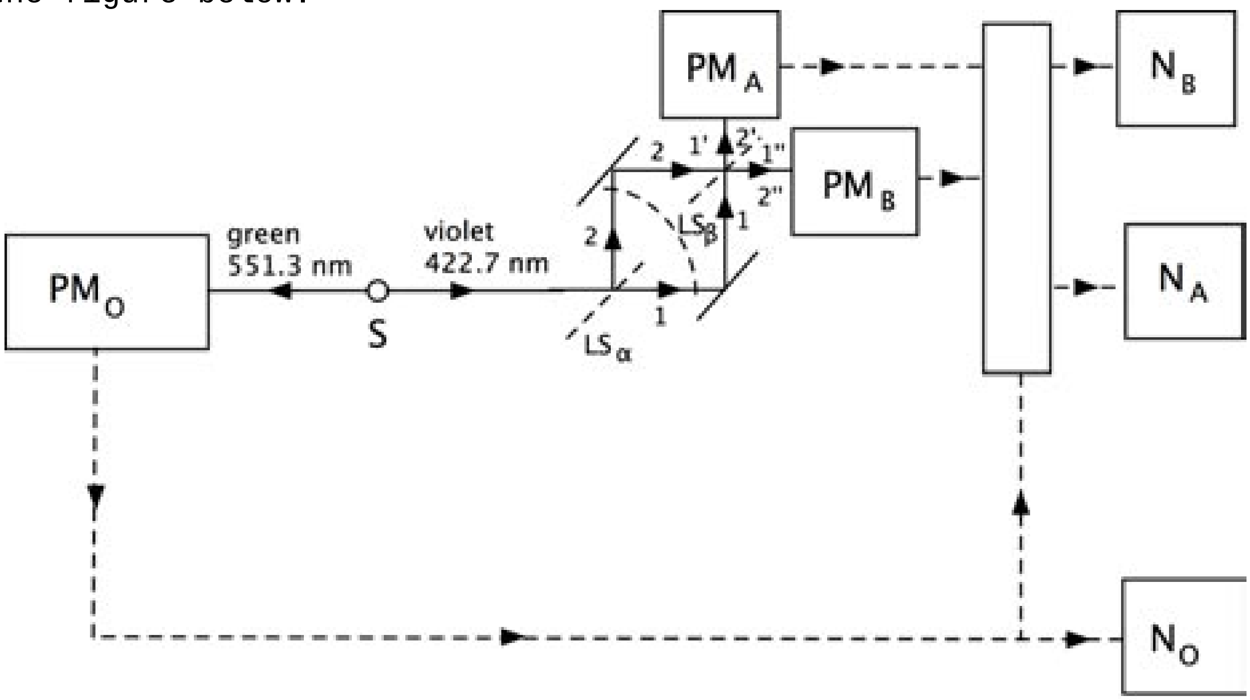
Albert leaves such logical thinking to professionals. Once he notes that N_C is very small, he is quite prepared to treat it as if it were zero. He therefore says

> I observe that light travels from the source to PM_A or to PM_B, because detection ceases when the source is switched off.
>
> I observe the counts N_A and N_B correspond to a game of heads or tails, in that the two possibilities are mutually exclusive, and that the counts are random.
>
> I observe that the optical paths 1 and 2 are

distinguishable, because the experiment allows me to ascertain, for each quantum, whether it has travelled path 1 (detection by PM_A) or path 2 (detection by PM_B).

I envisage that, on arrival at the half-silvered mirror, each photon from the source is directed at random either along path 1 or along path 2; and I assert that it is the nature of photons to play heads or tails.

3. The third stage consists of an interference experiment as shown in the figure below.



Interference with a single photon (third stage). In this sketch, solid lines are optical paths and dashed lines are electrical connections.

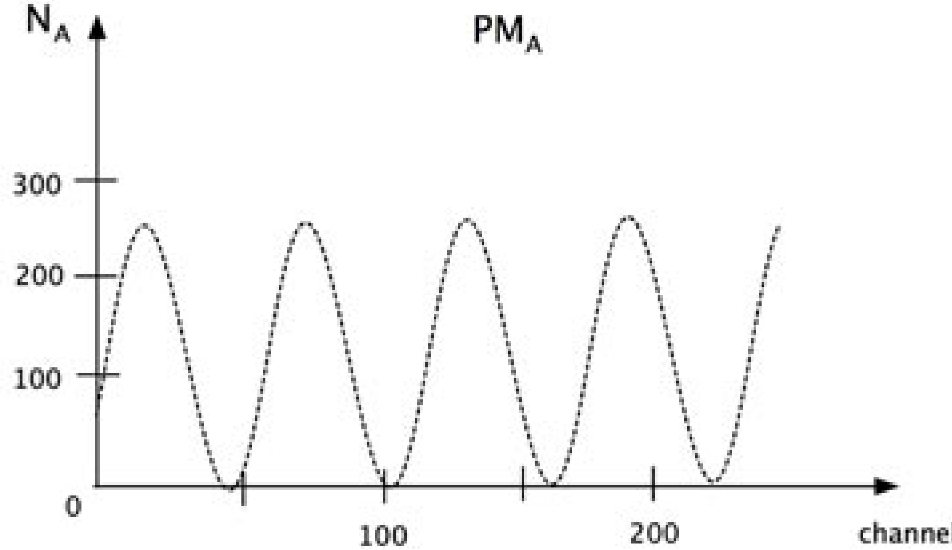
A so-called Mach-Zehnder interferometer is used, allowing one to obtain two interference profiles. The beam of violet light from the source S is split into two by the mirror LS_α. After reflection from two different mirrors, these secondary beams meet on a second half-silvered mirror LS_β. Here, each secondary beam is further split into two; thus one establishes two interference regions, region (1',2') where one places PM_A, and region (1",2") where one places PM_B.

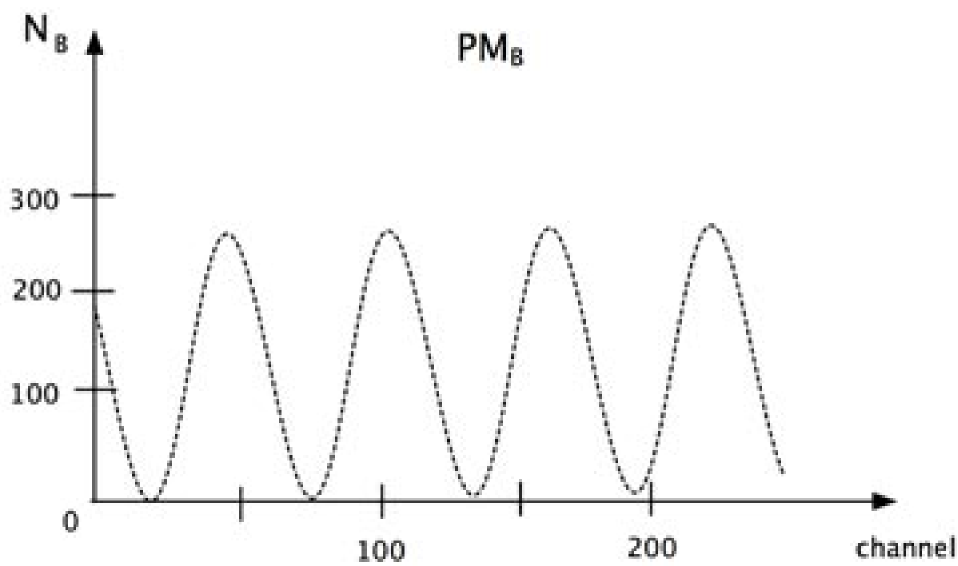
A very high precision piezoelectric system allows one of the mirrors to be displaced so as to vary the path difference between the two arms of the interferometer. In this way one can shift the pattern of interference fringes by regular steps, without moving the detectors PM_A and PM_B; the standard step corresponds to a change of $\lambda/50$ in the difference between the two optical paths.

A sweep, taking 15 sec for each standard step, yields two interference plots corresponding, respectively, to the paths (1',2') and (1",2"); the fringes have good contrast(difference in intensity between maxima and minima), and their visibility

$$\left(N_{A,\max} - N_{A,\min}\right)/\left(N_{A,\max} + N_{A,\min}\right)$$

was measured as 98% as shown in the figure below:





The two interference plots obtained with the Mach-Zehnder interferometer. Note that the maximum counting rates in PM_A correspond to minima in PM_B, indicating a relative displacement of $\lambda/2$ between the two interference patterns.

If we recall that we are reasoning in terms of photons, and that the photons are being processed individually, then we must admit that the interference does not stem from any interaction between successive photons, but that each photon interferes with itself.

What would Albert have to say? He seems exasperated but is still polite. His statements are brief:

> I observe that the optical paths differ in length between LS_α and LS_β, and are then coincident over (1',2') and over (1",2").

In PM_A I observe a process that seems perfectly natural to me, namely

$$light + light \rightarrow light$$

In PM_B I observe a process that I find astounding, namely

$$light + light \rightarrow darkness$$

Such superposition phenomena with light I shall call interference, constructive in PM_A and destructive in PM_B.

In the situation considered before, I envisaged light as consisting of particles called photons, which travelled either along path 1 or along path 2. In the present situation I want to know for each individual photon which path it has travelled; to this end I should like to ask you to close off path 2, since this will ensure that the photons travel by path 1.

Clearly Albert is perturbed. He awaits the new experimental results with some anxiety.

On closing either path, whether 1 or 2, one observes that all interference phenomena disappear. For instance, instead of a very high count N_A and a very low count N_B, we now obtain essentially equal counts from PM_A and PM_B.

Albert is visibly displeased and now very wary. He then continues with his analysis of the experiment:

I observe that in order to produce interference phenomena it is necessary to have two optical paths of different lengths, both open.

Whenever a photon is detected, I note my inability to ascertain whether the light has travelled by path 1 or by path 2, because I have no means for distinguishing between the two cases.

If I were to suppose that photons travel only along 1, then this would imply that path 2 is irrelevant, which is contrary to what I have observed. Similarly, if I were to suppose that photons travel only along 2, then this would imply that path 1 is irrelevant, which is also contrary to my observations.

If I envisage the source S as emitting particles, then I am forced to conclude that each individual photon travels simultaneously along both paths 1 and 2; but this result contradicts the results of the previous experiment (second stage), which compelled me to envisage that every photon chooses, at random, either path 1 or path 2.

I conclude that the notion of particles is unsuited to explaining interference phenomena.

I shall suppose instead that the source emits a wave; this wave splits into two at LS_α, and the two secondary waves travel one along path 1 and the other along path 2. They produce interference by mutual superposition on LS_β constructively in (1',2') and destructively in (1",2"). At the far end of (1',2') or of (1",2") I envisage each of the waves condensing into particles, which are then detected by the photomultipliers (essentially by PM_A since the contrast is 98% means only very few photons are detected by PM_B).

It seems to me that I am beginning to understand the situation. I envisage light as having two complementary forms: depending on the kind of experiment that is being done, it can manifest itself either as a wave, or as a particle, but never as both simultaneously and in the same place. Thus, in the experiment where the path followed by the light cannot be ascertained (third stage), light behaves first like a wave, producing interference phenomena; but it behaves like a particle when, afterwards, it is detected through the photoelectric effect. I conclude that light behaves rather strangely, but nevertheless I have the impression that its behavior can be fully described once one has come to terms with the idea of wave-particle duality.

Albert leaves the room slowly, hesitantly, even reluctantly. He might be impressed by the completeness of all that he has just described or maybe he is worried that more needs to be said.

In fact, something does remain to be said, since the problem of causality remains open. Let us look carefully at the experimental layouts in the second and third stages: we see that they have LS_α in common, and that they differ only beyond some boundary (indicated by the dashed circle downstream from LS_α). We have stated that light behaves like a particle or like a wave depending on whether or not one can ascertain the path it takes through the apparatus; but in the two experiments under consideration, the choice between the alternatives must be decided on LS_α, **before** the light has crossed the crucial boundary, that is, at a stage where nothing can as yet distinguish between the two kinds of apparatus, since they differ only beyond the point of decision. It is as if the light **"chose"** whether to behave like a wave or like a particle before **"knowing"** whether the apparatus it will pass through will elicit interference phenomena or the photoelectric effect. Hence the question of causality is indeed opened up with vengeance.

Albert comes back abruptly. He is disconcerted and wearily says:

Originally I supposed that light would behave like a wave or like a particle, depending on the kind of experiment to which it was being subjected.

I observe that the choice must be made on the half-silvered mirror LS_α, before the light reaches that part of the apparatus where the choice is actually implemented; this would imply that the effect precedes the cause.

I know that both waves and particles obey the principle of causality, that is, that cause precedes effect.

I conclude that light is neither wave nor particle; it behaves neither like waves on the sea, nor like projectiles fired from a gun, nor like any other kind of object that I am familiar with.

I must ask you to forget everything I have said about this experiment, which seems to me to be thoroughly mysterious.

Albert leaves, but quickly returns with a contented smile, and his final statement is not without a touch of malice.

I observe in all cases that the photomultipliers register quanta when I switch on the light source.

I conclude that **"something"** has travelled from the source to the detector. This **"something"** is a quantum object, and I shall continue to call it a photon, even though I know that it is neither a wave nor a particle.

I observe that the photon gives rise to interference when one cannot ascertain which path it follows; and that interference disappears when it is possible to ascertain the path.

For each detector, I observe that the quanta it detects are randomly distributed in time.

If I repeat the experiment several times under identical conditions, then I observe that the photon counts registered by each photomultiplier are reproducible in a statistical sense. For example, suppose that in the first and in the second experiments PM_A registers N_A' and N_A'' respectively; then one can predict that N_A'' has a probability of 0.68 of being in the interval $N_A' \pm (N_A')^{1/2}$.

Thus, these counts enable me to determine experimentally, for any kind of apparatus, the probability that a given detector will detect a quantum, and it is precisely such probabilities that constitute the results of experiments.

I assert that the function of a physical theory is to predict the results of experiments.

What I expect from theoretical physicists is a theory that will enable me to predict, through calculation, the probability that a given detector will detect a photon. This theory will have to take into account the random behavior of the photon, and the absence or presence of interference phenomena depending on whether the paths followed by the light can or cannot be ascertained.

Albert leaves, wishing the physicists well in their future endeavors.

Physicist have indeed worked hard and the much desired theory has indeed come to light, namely, quantum mechanics, as we have seen in our discussions. As we have seen it applies perfectly not only to photons, but equally well to electrons, protons, neutrons, etc; in fact, it applies to all the particles of microscopic physics. For the last 75 years it has worked to the general satisfaction of physicists.

Meanwhile, it has produced two very interesting problems of a philosophical nature.

1. Chance as encountered in quantum mechanics lies in the very nature of the coupling between the quantum object and the experimental apparatus. No longer is it chance as a matter of ignorance or incompetence: it is **chance quintessential and unavoidable**.

2. Quantum objects behave quite differently from the familiar objects of our everyday experience: whenever, for pedagogical reasons, one allows an analogy with macroscopic models like waves or particles, one always fails sooner or later, because the analogy is never more than partially valid. Accordingly, the first duty of a physicist is to force her little grey cells, that is her concepts and her language, into unreserved compliance with quantum mechanics (as we have been attempting to do); eventually this will lead her to view the actual behavior of microsystems as perfectly normal. As a teacher of physics, our duties are if anything more onerous still, because we must convince the younger generations that quantum mechanics is not a branch of mathematics, but an expression of our best present understanding of physics on the smallest scale; and that, like all physical theories, it is predictive.

In this context, let us review the basic formalism of quantum mechanics.

Basic Formalism

We will introduce the elements of quantum mechanics as axioms. Physicists have devised a new mathematical tool. The transition amplitude from initial to final state, and it is this amplitude that enables one to calculate the needed probabilities.

(1) For the experiment where the photon travels from the source S to the detector PM_A (see figure (a) below), we write the transition amplitude from S to PM_A as

$$\langle \text{photon arriving at } PM_A | \text{photon leaving S} \rangle$$

As we know, this is a complex number and it is read from right to left. We write it more symbolically as $\langle f|i\rangle$ which simply means the transition amplitude from initial to final state.

The transition probability from the initial state $|i\rangle$ to the final state $|f\rangle$ is given by

$$|\langle f|i\rangle|^2$$

(2) If the photon emitted by the source can take either of two paths, and if it is in principle possible to ascertain which path it actually does take (figure (b) below) then there are two transition amplitudes:

$$\langle \text{photon arriving at } PM_A | \text{photon leaving S} \rangle$$

$$\langle \text{photon arriving at } PM_B | \text{photon leaving S} \rangle$$

which we symbolize simply as

$$\langle f_1|i\rangle \quad , \quad \langle f_2|i\rangle$$

In this case there are two probabilities:

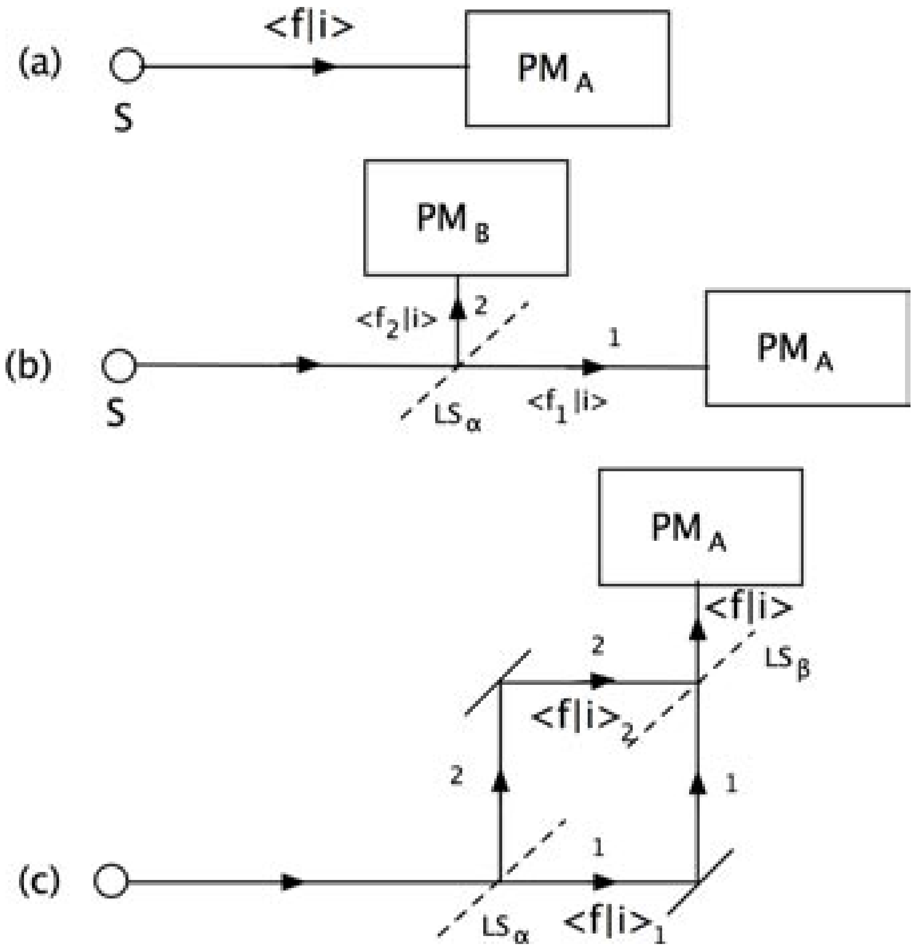
$$\left|\langle f_1|i\rangle\right|^2 \quad , \quad \left|\langle f_2|i\rangle\right|$$

The total probability is their sum:

$$\left|\langle f_1|i\rangle\right|^2 + \left|\langle f_2|i\rangle\right|$$

More generally, we would write

$$\left|\langle f|i\rangle\right|^2 = \sum_k \left|\langle f_k|i\rangle\right|^2$$ where the sum is over all possible paths



Three arrangements sufficient to determine the transition amplitude: (a) a single optical path; (b) two paths, allowing us to ascertain which path has actually been taken; (c) two paths, not allowing us to ascertain which path has actually been taken.

(3) If a photon is emitted by the source S can take either of two paths, but it is impossible to ascertain which path it does take (figure (c) above), then there are again two transition amplitudes:

$$\langle \text{photon arriving at } PM_A | \text{photon leaving S} \rangle_{\text{along path 1}}$$

$$\langle \text{photon arriving at } PM_B | \text{photon leaving S} \rangle_{\text{along path 2}}$$

which we symbolize simply as

$$\langle f|i\rangle_1 \quad , \quad \langle f|i\rangle_2$$

To allow for interference, we assert that in this case it is the amplitudes that must be added; the total amplitude reads

$$\langle f|i\rangle = \langle f|i\rangle_1 + \langle f|i\rangle_2$$

The total probability is:

$$\left|\langle f|i\rangle_1 + \langle f|i\rangle_2\right|^2$$

More generally, we would write

total amplitude: $\langle f|i\rangle = \sum_k \langle f|i\rangle_k$

total probability: $\left|\langle f|i\rangle\right|^2 = \left|\sum_k \langle f|i\rangle_k\right|^2$

where the sums are over all possible paths.

(4) If one wants to analyze the propagation of the light more closely, one can take into account its passage through the half-silvered mirror LS_α, considering this as an intermediate state (figure (b) above). The total amplitude for path 1 is

$$\langle \text{photon arriving at } PM_A | \text{photon leaving S} \rangle$$

However, it results from two successive intermediate amplitudes:

$$\langle \text{photon arriving at } LS_\alpha | \text{photon leaving S} \rangle$$

$$\langle \text{photon arriving at } PM_A | \text{photon leaving } LS_\alpha \rangle$$

Here we consider the total amplitude as the product of the successive intermediate amplitudes; symbolically, labelling the intermediate state as ν, we have

$$\langle f|i\rangle = \langle f|\nu\rangle\langle \nu|i\rangle$$

Finally, consider a system of two mutually independent photons. If photon 1 undergoes a transition from a state i_1 to a state f_1, and photon 2 from a state i_2 to a state f_2, then

$$\langle f_1 f_2 | i_1 i_2\rangle = \langle f_1|i_1\rangle\langle f_2|i_2\rangle$$

The four rules just given suffice to calculate the detection probability in any possible experimental situation. They assume their present form as a result of a long theoretical evolution; but they are best justified **a posteriori**, because in 75 years they have never been found to be wrong. Accordingly, we may consider them as the basic principles governing the observable behavior of all microscopic objects, that is, objects whose action on each other are of order $\hbar$ (Planck's constant). From these principles (they are equivalent to our earlier postulates - just look different because we are using the amplitude instead of the state vector as the fundamental mathematical object in the theory) one can derive all the requisite formalism, that is, all of quantum mechanics.

Quantum mechanics as we have described it earlier and also above, works splendidly, like a well-oiled machine. It, and its basic principles, might therefore be expected to command the assent of every physicist; yet it has evoked, and on occasion continues to evoke, reservations both explicit and implicit. For this there are two reasons:

(1) Quantum mechanics introduces unavoidable chance, meaning that its characteristic randomness is inherent in the microscopic phenomena themselves.

(2) It attributes to microscopic objects properties so unprecedented that we cannot represent them through any macroscopic analogs or models.

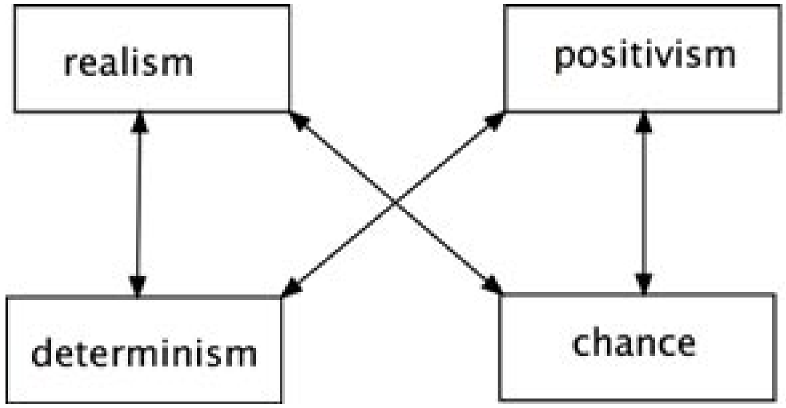
Both features are revolutionary, and it is natural that they should have provoked debate. On the opposite sides of this debate we find two great physicists, Neils Bohr and Albert Einstein, and we will now discuss how the debate evolved from its beginnings in 1927 to its conclusion in 1983 (that is 56 years!).

Inseparable Photons (the EPR Paradox)

Though ornithologists have known about inseparable parrots for a long time, to physicists the existence of inseparable photons has been brought home only during the last two decades, through a beautiful series of experiments by Alain Aspect and his research group at Orsay Laboratory in Paris. The experiments are exemplary, in virtue both of the difficulties they had to overcome and the results achieved, which are exceptionally clear-cut. In fact, the significance of the experiments extends beyond the strict confines of physics, because they provide the touchstone for settling a philosophical debate that has divided physicist for 75 years. The division dates back to the appearance of two mutually contradictory interpretations of quantum mechanics at the Como conference in 1927. To sketch the debate, we start with a brief summary of the philosophy of physics.

The Philosophical Stakes in the Debate

Our summary is best presented diagrammatically as shown in the figure below:



The philosophical elements in a debate between physicists.

(1) For the physicist who is a **realist**, a physical theory reflects the behavior of real objects, whose existence is not brought into question.

(2) For the physicist who is a **positivist**, the purpose of a physical theory is to describe the relations between measurable quantities. The theory does not tell one whether anything characterized by these quantities really exists, nor even whether the question makes sense.

(3) For the physicist who is a **determinist**, exact knowledge of the initial conditions and of the interactions allows the future to be predicted exactly. Determinism is held to be a universal characteristic of natural phenomena, even about those which we know, as yet, little or nothing. In this framework, any recourse to chance merely reflects our own ignorance.

(4) For the physicist who is a probabilist, chance is inherent in the very nature of microscopic phenomena. To her, determinism is a consequence, on the macroscopic level, if the laws of chance operating on the microscopic level; it is appropriate to measurements of mean values of quantities whose relative fluctuations are very weak.

From these four poles, realism, positivism, determinism, and chance, the physicist choose two, one on each axis. Though sometimes the choice is made in full awareness of what it entails, most often it is made subconsciously. In our description of quantum mechanics, we might adopt without reservations, the point of view of the elementary particle physicist. For a start, she believes firmly in the existence of particles, since she spends her time in accelerating, deflecting, focusing, and detecting them. Even though she has never seen or touched them, to her their objective existence is not in any doubt. Next she observes that they impinge on the detectors quite erratically, whence she has no doubts, either, that their behavior is random. Accordingly, the elementary particle experimentalist has chosen realism and chance, most often without realizing that she has made choices at all.

There are other philosophical options that can be adopted with eyes fully open: realism and determinism are the choices of Albert Einstein; positivism and chance are those of Neils Bohr. They are well acquainted and each thinks very highly of the other: which is no bar to their views being incompatible, nor to the two men representing opposite poles of the debate.

From Como to Brussels (1927-30)

On September 26,1927, in Como, Niels Bohr delivered a memorable lecture. His stance is that of an enthusiastic champion of the new quantum mechanics. He puts special weight on the inequalities proved by Heisenberg the year before:

$$\Delta x \Delta p_x \geq \frac{1}{2}\hbar \quad , \quad \Delta t \Delta E \geq \frac{1}{2}\hbar$$

They imply that it is impossible to define exact initial conditions for a microscopic object, which automatically makes it impossible to construct, on the microscopic scale, a deterministic theory patterned on classical mechanics. Only a probabilistic theory is possible, and that theory is quantum mechanics.

Einstein disagrees with this point of view, and his opposition to Bohr's theses becomes public at the Brussels conference in 1930: he adopts the role of a dissenter who knows precisely how to press home the most difficult questions. Deeply shocked by the retreat from determinism, he tries to show via his thought (gedanken) experiments he can contravene the Heisenberg inequalities.

At the cost of several sleepless nights devoted to analyzing the objections of his adversary, Bohr refutes all of Einstein's criticisms, and emerges from the conference as the undoubted winner.

From Brussels to the EPR Paradox (1930-35)

Having lost the argument at Brussels, Einstein tries to define his objections with ever greater precision. Believing as he does that position and momentum exist **objectively and simultaneously**, he considers quantum mechanics to be incomplete and merely provisional. The points of view of the two antagonists at this stage of the debate can be spelled out as follows.

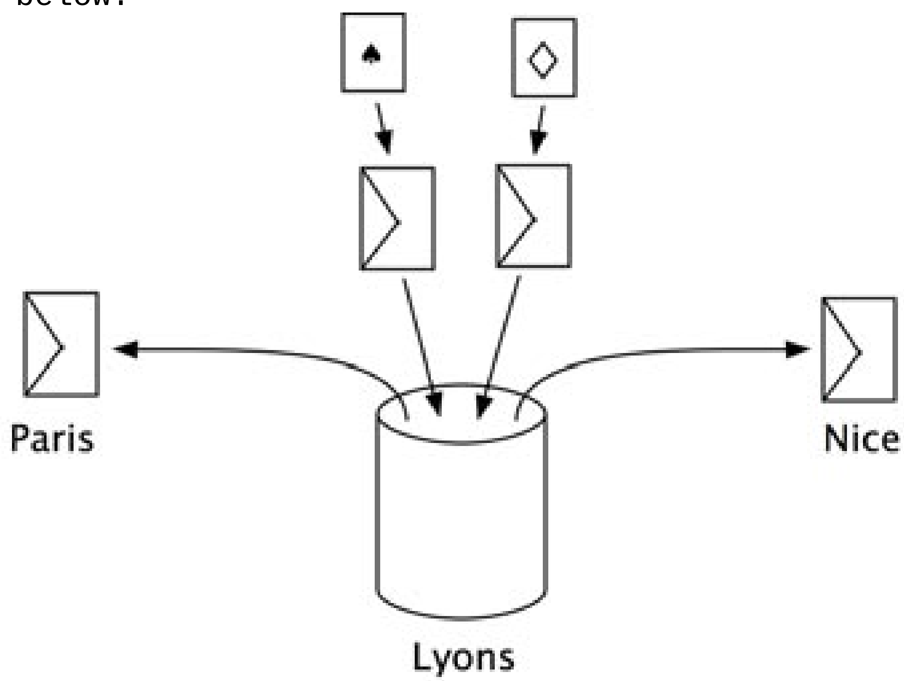
For Einstein, a physical theory must be a deterministic and a complete representation of the objective reality underlying the phenomena. It features known variables that are observable, and others, unknown as yet, called **hidden variables**. Because of our provisional ignorance of the hidden variables, matter at the microscopic level appears to us to behave arbitrarily, and we describe it by means of a theory that is incomplete and probabilistic, namely by quantum mechanics.

For Bohr, a physical theory makes sense only as a set of relations between observable quantities. Quantum mechanics supplies a correct and complete description of the behavior of objects at the microscopic level, which means that the theory itself is likewise complete. The observed behavior is probabilistic, implying that chance is inherent in the nature of the phenomena.

Between chance as a matter of ignorance, as advocated by Einstein, and chance unavoidable, as advocated by Bohr, the debate does not remain merely philosophical. Quite naturally it returns to the plane of physics with the thought experiment proposed by Einstein, Podolsky and Rosen in 1935, which in their view proves that quantum mechanics is indeed incomplete. Their thought experiment is published as a paper in the Physical Review, but it is so important that it reverberates as far as the New York Times. Physicists call the proposal the EPR paradox, after its proponents. It will take fifty years to untangle the question, first in theory and then by experiment. We will not, of course, follow these fifty years blow by blow; instead, we confine attention to three decisive stages reached respectively in 1952, 1964, and 1983. But we start with an illustration that helps one see what the EPR paradox actually is.

An Elementary Introduction to the EPR Paradox

Consider two playing cards, one red(diamond) and one black(spade) as shown below:



Two playing cards help us understand
the stakes in the EPR paradox.

An experimenter in Lyons puts them into separate envelopes which she then seals. She is thus provided with two envelopes looking exactly alike, and she puts both into a container. She shakes the container so as the "shuffle the pack", and the system is ready for the experiment.

At 8:00 two travellers, one from Paris and one from Nice, come to the container (in Lyons), take one envelope each, and then return to Paris and Nice, respectively. At 14:00 they are back at their starting points; each opens her envelope, looks at the card, and telephones to Lyons reporting the color. The experiment is repeated every day for a year, and the observer in Lyons keeps a careful record of the results. At the end of the year the record stands as follows:

1. The reports from Paris are "red" or "black", and the sequence of these reports is random. The situation is exactly the same as in a game of heads or tails, and probability of each outcome is 1/2.

2. The reports from Nice are "red" or "black", and the sequence of these reports is random. Here too probability of each outcome is 1/2.

3. When Paris reports "red", Nice reports "black"; when Paris reports "black", Nice reports "red". One sees that there is perfect(anti) correlation between the report from Paris and the report from Nice.

Accordingly, the experiment we have described displays two features:

(1) It is **unpredictable** and thereby random at the level of individual observations in Paris and Nice.

(2) It is **predictable**, by virtue of the correlation, at the level where one observes the Paris and the Nice results simultaneously.

Einstein and Bohr might have interpreted the correlation as follows.

According to Einstein, the future of the system is decided at 8:00 when the envelopes are chosen, because he believes that the contents of the two envelopes differ. Suppose, for instance, that Paris has (without knowing it) drawn a red card, and Nice the black. The colors so chosen exist in reality, even though we do not know them. The two cards are moved, separately, by the travellers between 8:00 and 14:00, during which time they do not influence each other in any way. The results on opening the envelopes read "red" in Paris and "black" in Nice. Since the choice at 8:00 was made blind, the opposite outcome is equally possible, but the results at 14:00 are always correlated (either red/black or black/red). This correlation at 14:00 is determined by the separation of the colors at 8:00, and we say the theory proposed by Einstein is **realist, deterministic, and separable(or local)**, by virtue of a hidden variable, namely, the color.

According to Bohr, there is a crucial preliminary factor, inherent in the preparation of the system. On shaking the container with the two envelopes, one loses information regarding the colors. Afterwards, one only knows that each envelope contains either a red card (probability 1/2) or a black card (probability 1/2). We will therefore say that a given envelope is in a "brown state", which is a superposition of a red state and of a black state having equal probabilities. At 8:00 the two envelopes are identical: both are in a "brown state", and the future of the system is still undecided. There is no solution until the envelopes are opened at 14:00, since it is only the action of opening them that makes the colors observable. The result is probabilistic. There is a probability 1/2 that in Paris the envelope will be observed to go from the "brown state" to the red, while the envelope in Nice is observed to go from the "brown state" to the black; there is the same probability 1/2 of observing the opposite. But the results of the observations on the two envelopes are always correlated, which means that there is a mutual influence between them, in particular at 14:00; in fact it is better to say that, jointly, they constitute a single and non-separable system, even though one is in Paris and the other is in Nice. Accordingly, the theory proposed by Bohr is **positivist, probabilistic (non-deterministic) and non-separable(non-local)**, interrelating as it does the colors that are actually observed.

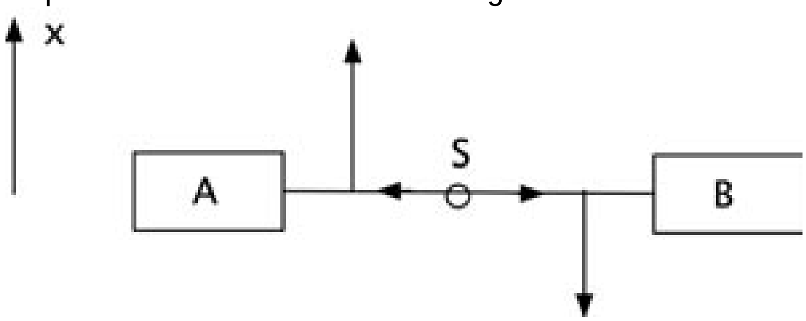
Einstein's view appears to be common sense, while it must be admitted that Bohr's is very startling; however, the point of this macroscopic example is, precisely, to stress how different the quantum view is from the classical.

Proceeding with impeccable logic but from different premises, both theories predict the same experimental results. Can we decide between them? At the level considered here it seems we cannot: for even if the envelopes were opened prematurely while still in Lyons, one would merely obtain the same results at a different time, and without affecting the validity of either interpretation. The solution to the problem must be looked for at the atomic level, by studying the true EPR set-up itself.

The EPR Paradox (1935-52)

Albert Einstein, Boris Podolsky, and Nathan Rosen meant to look for an experiment that could measure, indirectly but simultaneously, two mutually exclusive quantities like position and momentum. Such results would contravene the predictions of quantum mechanics, which allows the measurement of only one such quantity at any one time; that is why the thought experiment is called the EPR paradox.

In 1952, David Bohm showed that the paradox could be set up not only with continuously varying quantities like position and momentum, but also with discrete quantities like spin. This was the first step towards any realistically conceivable experiment. Meanwhile, objectives have evolved, and nowadays it is more usual to talk of the EPR scenario, meaning some sensible experiment capable of discriminating between quantum theory and hidden-variable theories. Such a set-up is sketched in the figure below.



The simplest EPR scenario

A particle with spin 0 decays, at S, into two particles of spin 1/2, which diverge from S in opposite directions. Two Stern-Gerlach type detectors A and B measure the x-components of the spins. Two types of response are possible:

(1) "spin up" at A, "spin down" at B, a result denoted by (+1,-1)

(2) "spin down" at A, "spin up" at B, a result denoted by (-1,+1)

Thus far everyone is agreed, but the interpretation is yet to come.

Einstein reasons that if pairs of particles produced at S elicit different responses (+1,-1) and (-1,+1) from the detector system A,B, then the pairs must have differed already at S, immediately after the decay. It must be possible to represent this difference by a hidden variable λ, which has an objective meaning, and **which governs the future of the system**. After the decay the two particles separate without influencing each other any further, and eventually they trigger the detectors A and B.

Bohr reasons that all the pairs produced at S are identical. Each pair constitutes a non-separable system right up to the time when the photons reach the detectors A and B. At that time we observe the response of the detectors, which is probabilistic, admitting two outcomes (+1,-1) and (-1,+1).

To sum up, Einstein restricts the operation of chance to the instant of decay (at S), whose details we ignore, but which we believe creates pairs whose hidden variables λ are different. By contrast, Bohr believes that chance operates at the instant of detection, and that it is inherent in the very nature of the detection process: this chance is unavoidable. We are still in the realms of thought, and stay there up to 1964.

In 1964, the landscape changes: John Bell, a theorist at CERN, shows that it is possible to distinguish between the two interpretations experimentally. The test applies to the EPR scenario; it is refined by Clauser, Horne, Shimony, and Holt, whence it is called the BCHSH inequality after its five originators.

The BCHSH Inequality (1964)

To set up an EPR scenario, one first needs a source that emits particle pairs. Various experimental possibilities have been explored:

(1) atoms emitting two photons in cascade

(2) electron-positron annihilation emitting two high-energy photons

(3) elastic proton-proton scattering

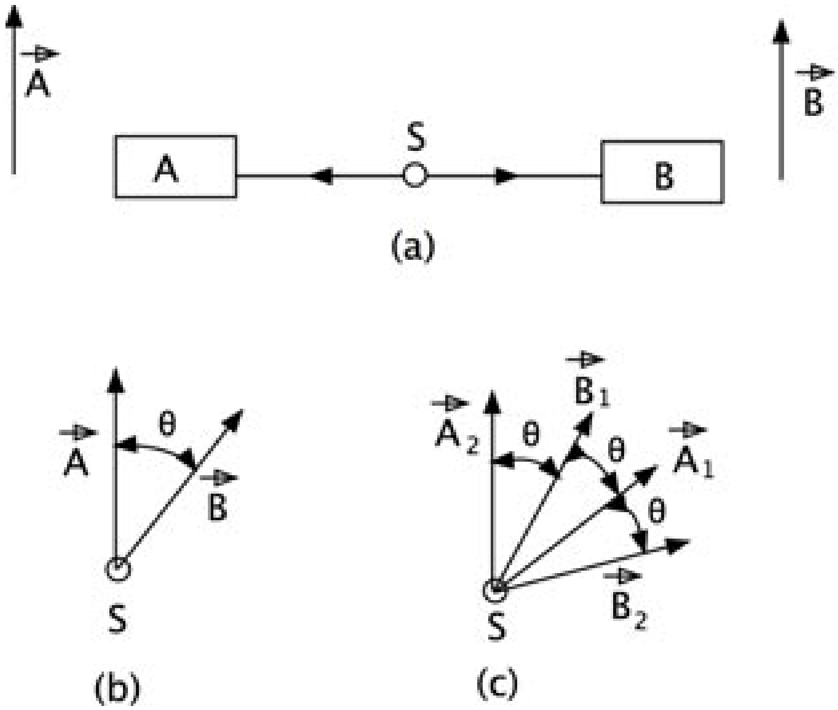
It is solution (1) that has eventually proved the most convenient; it has been exploited by Alain Aspect in particular.

Next one needs detectors whose response can assume one of two values, represented conventionally by +1 and -1. Such a detector might be

(1) for spin 1/2 particles, a Stern-Gerlach apparatus responding to "spin up" or "spin down"

(2) for photons, a polarizer responding to "parallel polarization" or "perpendicular polarization"

Our sketch of the EPR scenario can now be completed as in the figure below.



The most general EPR scenario

(a) views the apparatus perpendicularly to axis, showing the two detectors A and B, with their polarizing directions denoted as $\vec{A}$ and $\vec{B}$

(b) views the apparatus along its axis, and shows that the analyzing directions of the two detectors are not parallel, but inclined to each other at an angle θ

(c) also a view along the axis of the apparatus, and shows the actual settings chosen by Aspect: two orientations are allowed for each detector, $\vec{A}_1$ or $\vec{A}_2$ for one, and $\vec{B}_1$ or $\vec{B}_2$ for the other.

We adopt the following conventions:

(1) $\alpha = \pm 1$ is the response of detector A when oriented along $\vec{A}$

(2) $\beta = \pm 1$ is the response of detector B when oriented along $\vec{B}$

Since each detector has two possible orientations, called 1 and 2, we shall denote their responses as α_1, α_2 and β_1, β_2 respectively.

Now consider the quantity $\langle \gamma \rangle$ defined by

$$\langle \gamma \rangle = \langle \alpha_1 \beta_1 \rangle + \langle \alpha_1 \beta_2 \rangle + \langle \alpha_2 \beta_1 \rangle - \langle \alpha_2 \beta_2 \rangle$$

where the symbol $\langle ... \rangle$ denotes the mean value over very many measured events. We call $\langle \gamma \rangle$ the **correlation function** of the system.

The BCHSH inequality reads

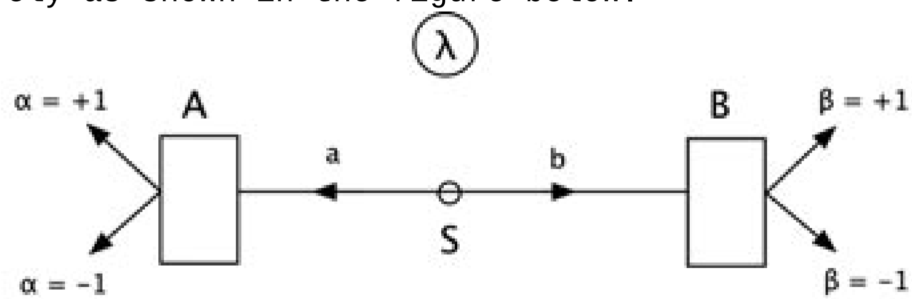
$$-2 \le \langle \gamma \rangle \le 2$$

Its authors have proved that it must be satisfied if mechanics at the microscopic level constitutes a theory that is realist, deterministic, and separable: or in other words if the theory contains a hidden variable. A sketch of the a proof is shown below.

A Proof of Bell's Inequality

A theory that is deterministic and separable:

Suppose that the pair a, b emerging from S can be characterized by a hidden variable λ. The responses of the detectors A, B are $\alpha(\vec{A},\lambda)$ and $\beta(\vec{B},\lambda)$ respectively as shown in the figure below.



The theory is deterministic and separable:

(1) **deterministic**, because the results are determined by the hidden variables plus the settings $\vec{A}$ and $\vec{B}$;

(2) **separable**, because the response of A is independent of the response of B, and vice versa

Since the value of λ is unknown and different for each pair, the responses of A and B seem random. Lacking information about λ, we characterize it by choosing a statistical distribution $\rho(\lambda)$, which then allows us to derive the distribution of the responses $\alpha(\vec{A},\lambda)$ and $\beta(\vec{B},\lambda)$, which can be compared with experiment.

Bell's inequalities have the great virtue that they apply to any hidden variable theory, irrespective of the choice of $\rho(\lambda)$.

Theorem 1. Consider the four numbers $\alpha_1, \alpha_2, \beta_1$, and β_2, each of which can assume only the values 1 or -1. Then the combination

$$\gamma = \alpha_1\beta_1 + \alpha_1\beta_2 + \alpha_2\beta_1 - \alpha_2\beta_2$$

can assume only the values 2 and -2.

To prove the theorem, one constructs a truth table for all 16 possibilities, which shows that 2 and -2 are indeed the only possible values of γ.

α_1	α_2	β_1	β_2	γ
1	1	1	1	2
1	1	1	–1	2
1	1	–1	1	–2
1	1	–1	–1	–2
1	–1	1	1	2
1	–1	1	–1	2
1	–1	–1	1	–2
1	–1	–1	–1	–2
–1	1	1	1	–2
–1	1	1	–1	2
–1	1	–1	1	–2
–1	1	–1	–1	2
–1	–1	1	1	–2
–1	–1	1	–1	–2
–1	–1	–1	1	2
–1	–1	–1	–1	2

Theorem 2. Consider very many sets of four numbers ($\alpha_1, \alpha_2, \beta_1, \beta_2$). The mean value of γ lies in the range [-2,2]. In other words,

$$-2 \le \langle \gamma \rangle \le 2$$

This is obvious, because every value of γ lies in this range, and so therefore must be the mean. The endpoints are included in order to allow for limiting cases.

Note that both theorems are purely mathematical, neither involves any assumptions about physics.

The BCHSH Inequality(or Bell's inequality in the real world)

Within the framework of a theory that is realist, deterministic, and separable, we can describe the photon pair in detail. Realism leads us to believe that polarization is an objective property of each member of the pair, independent of any measurements that may be made later. Determinism leads us to believe that the polarizations are uniquely determined by the decay cascade, and that they are fully specified by the hidden variable λ, which governs the correlation of the polarizations in A and B. Finally, separability leads us to believe that the measurements in A and B do not influence each other, which means in particular that the response of detector a is independent of the orientation of detector B.

Now consider a pair of photons a, b, characterized by a hidden variable λ. The response of the apparatus in its four settings would be as follows:

α_1 and β_1 in the orientation ($\vec{A}_1, \vec{B}_1$)
α_2 and β_2 in the orientation ($\vec{A}_2, \vec{B}_2$)
α_1' and β_2' in the orientation ($\vec{A}_1, \vec{B}_2$)
α_2' and β_1' in the orientation ($\vec{A}_2, \vec{B}_1$)

Recall that the variables α and β can only take on the values 1 and -1.

It is impossible in practice to make four measurements on one and the same pair of photons, because each photon is absorbed in the first measurement made on it; that is why we have spoken conditionally, that is, **of what results would be.** But if we believe that the photon correlations are governed by a theory that is realist, deterministic, and separable, then we are entitled to assume that the responses, of type α or type β, depend on properties that the photons possess before the measurement, so that the responses correspond to some objective reality. In such a framework we can appeal to the principle of separability, which implies, for instance, that detector A would give the same response to the orientations ($\vec{A}_1, \vec{B}_1$) and ($\vec{A}_1, \vec{B}_2$), because the response of A is independent of the orientation of B. Mathematically, this is expressed by the relation

$$\alpha_1 = \alpha_1'$$

Similarly one finds

$$\alpha_2 = \alpha_2' \quad , \quad \beta_1 = \beta_1' \quad , \quad \beta_2 = \beta_2'$$

Thus, we have shown that, for a given pair of photons, all possible responses of the apparatus in its four chosen settings can be specified by means of only four two-valued variables $\alpha_1, \alpha_2, \beta_1$, and β_2. This reduction from eight to four variables depends on the principle of separability. In this way, we are led to a situation covered by Theorem 2, and therefore $-2 \le \langle\gamma\rangle \le 2$.

By making many measurements for each of the four settings we can determine the four mean values $\langle\alpha_1\beta_1\rangle, \langle\alpha_1\beta_2\rangle, \langle\alpha_2\beta_1\rangle, \langle\alpha_2\beta_2\rangle$, and thus the mean value of the correlation

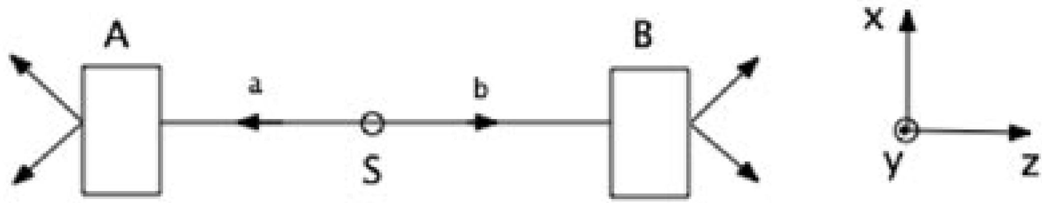
$$\langle\gamma\rangle = \langle\alpha_1\beta_1\rangle + \langle\alpha_1\beta_2\rangle + \langle\alpha_2\beta_1\rangle - \langle\alpha_2\beta_2\rangle$$

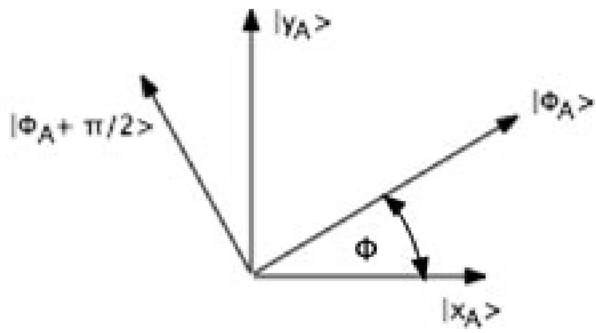
Otherwise, that is according to quantum mechanics (which is positivist, probabilistic, and non-separable), there are cases where the BCHSH inequality is violated. In particular, one can show that for photons in the configuration chosen by Aspect quantum mechanics yields

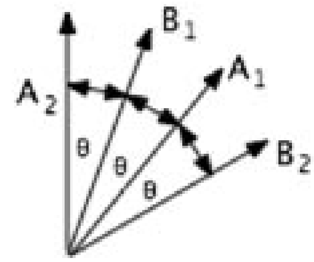
$$\langle\gamma\rangle = 3\cos 2\theta - \cos 6\theta$$

This leads to values well outside the interval [-2,2], for example to $\langle\gamma\rangle = 2\sqrt{2}$ when $\theta = 22.5°$ and to $\langle\gamma\rangle = -2\sqrt{2}$ when $\theta = 67.5°$.

Proof: The laboratory reference frame Oxyz serves to specify the orientations of detectors and polarizers as shown in the figure below:







Before any measurements have been made, the photon pair a, b forms a non-separable entity, represented by the vector

$$|\Phi\rangle = \frac{1}{\sqrt{2}}\left(|x_A, x_B\rangle + |y_A, y_B\rangle\right)$$

The act of measurement corresponds to passage to the ϕ-basis. Hence, we require the transition amplitudes from the two states $|x_A, x_B\rangle$, $|y_A, y_B\rangle$ to the four states $|\phi_A, \phi_B\rangle$, $|\phi_A, \phi_B + \pi/2\rangle$, $|\phi_A + \pi/2, \phi_B\rangle$, $|\phi_A + \pi/2, \phi_B + \pi/2\rangle$.

In the ϕ-basis we have

$$|\Phi\rangle = \frac{1}{\sqrt{2}}\left(\begin{array}{l}\cos(\phi_B - \phi_A)|\phi_A, \phi_B\rangle - \sin(\phi_B - \phi_A)|\phi_A, \phi_B + \pi/2\rangle \\ \qquad + \sin(\phi_B - \phi_A)|\phi_A + \pi/2, \phi_B\rangle + \cos(\phi_B - \phi_A)|\phi_A + \pi/2, \phi_B + \pi/2\rangle\end{array}\right)$$

The square of each amplitude featured here represents the detection probability. For example, the probability if simultaneously detecting photon a polarized at the angle ϕ_A and the photon b polarized at the angle ϕ_B is

$$\left(\frac{1}{\sqrt{2}}\cos(\phi_B - \phi_A)\right)^2 = \frac{1}{2}\cos^2(\phi_B - \phi_A)$$

By convention, we write the responses of detector A to a photon in state $|\phi_A\rangle$ (respectively $|\phi_A + \pi/2\rangle$ as $\alpha = 1$; similarly with β for detector B.

Let us analyze the four possible responses:

(1) $|\phi_A, \phi_B\rangle$ gives $\alpha = 1$, $\beta = 1$ so that $\alpha\beta = 1$; the probability is

$$P_{++} = \frac{1}{2}\cos^2(\phi_B - \phi_A)$$

(2) $|\phi_A, \phi_B + \pi/2\rangle$ gives $\alpha = 1$, $\beta = -1$ so that $\alpha\beta = -1$; the probability is

$$P_{+-} = \frac{1}{2}\sin^2(\phi_B - \phi_A)$$

(3) $|\phi_A + \pi/2, \phi_B\rangle$ gives $\alpha = -1$, $\beta = 1$ so that $\alpha\beta = -1$; the probability is

$$P_{-+} = \frac{1}{2}\sin^2(\phi_B - \phi_A)$$

(4) $|\phi_A + \pi/2, \phi_B + \pi/2\rangle$ gives $\alpha = -1$, $\beta = -1$ so that $\alpha\beta = 1$; the probability is

$$P_{--} = \frac{1}{2}\cos^2(\phi_B - \phi_A)$$

The mean value of $\langle\alpha\beta\rangle_{AB}$ follows immediately as

$$\langle\alpha\beta\rangle_{AB} = P_{++} - P_{+-} - P_{-+} + P_{--} = \cos 2(\phi_B - \phi_A)$$

The settings chosen by Aspect are as shown in the above figure. Corresponding to it we have the four terms

$$\langle\alpha_1\beta_1\rangle = \langle\alpha\beta\rangle_{A_1B_1} = \cos 2(\phi_{B_1} - \phi_{A_1}) = \cos 2\theta$$
$$\langle\alpha_1\beta_2\rangle = \langle\alpha\beta\rangle_{A_1B_2} = \cos 2(\phi_{B_1} - \phi_{A_2}) = \cos 2\theta$$
$$\langle\alpha_2\beta_1\rangle = \langle\alpha\beta\rangle_{A_2B_1} = \cos 2(\phi_{B_2} - \phi_{A_1}) = \cos 2\theta$$
$$\langle\alpha_2\beta_2\rangle = \langle\alpha\beta\rangle_{A_2B_2} = \cos 2(\phi_{B_2} - \phi_{A_2}) = \cos 6\theta$$

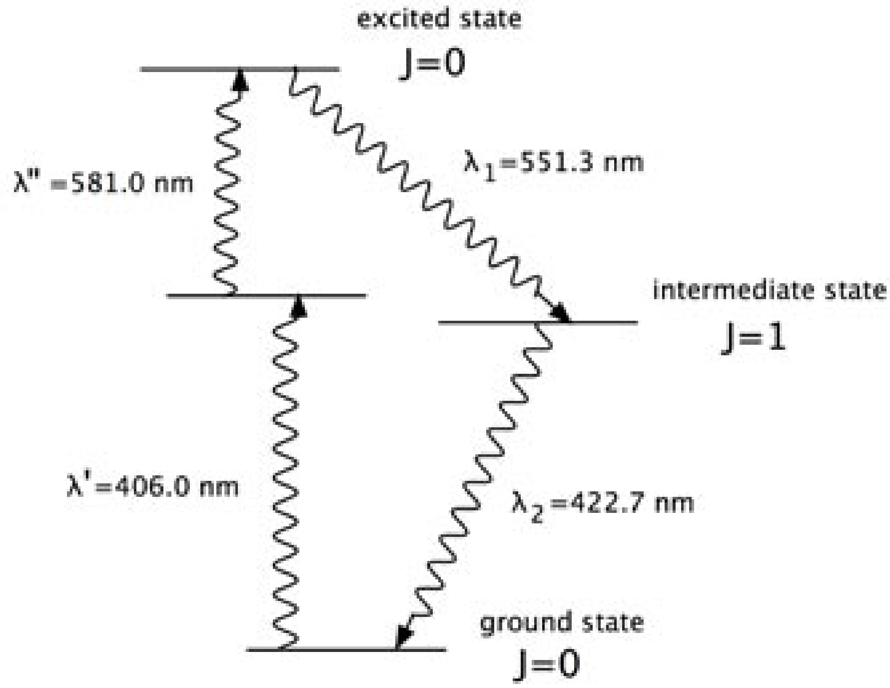
For comparison with Bell's inequality, we introduce the correlation function $\langle\gamma\rangle$:

$$\langle\gamma\rangle = \langle\alpha_1\beta_1\rangle + \langle\alpha_1\beta_2\rangle + \langle\alpha_2\beta_1\rangle - \langle\alpha_2\beta_2\rangle = 3\cos 2\theta - \cos 6\theta$$

Thus, the BCHSH test turns the EPR scenario into an arena for rational confrontation between the two interpretations; it remains only to progress from thought experiments to experiments conducted in the laboratory.

The Beginnings of the Experiment at Orsay (1976)

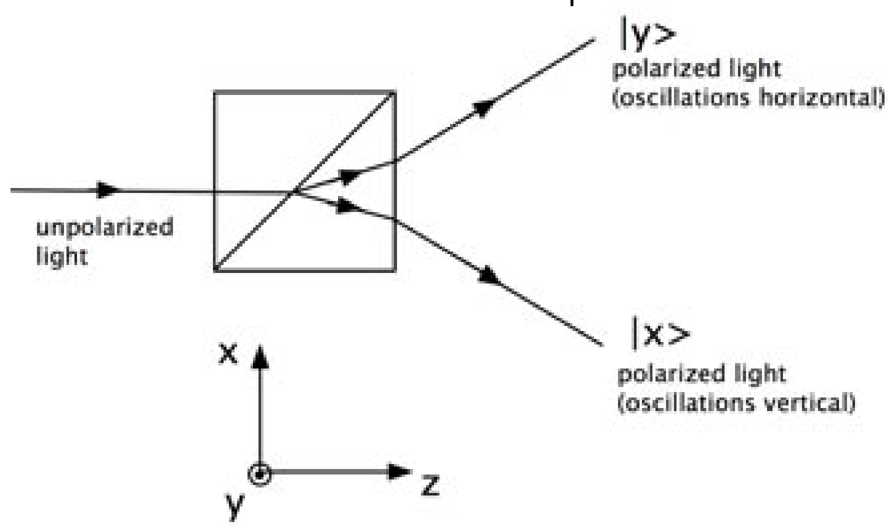
Alain Aspect's experiment studies the correlation between the polarizations of the members of photon pairs emitted by calcium. The light source is a beam of calcium atoms, excited by two focused laser beams having wavelengths $\lambda' = 406nm$ and $\lambda'' = 581nm$ respectively. Two-photon excitation produces a state having the quantum number $J = 0$. When it decays, this state emits two monochromatic photons having the wavelengths $\lambda_1 = 551.3nm$ and $\lambda_2 = 422.7nm$ respectively, in a cascade of two electronic transitions from the initial $J = 0$ level to the final $J = 0$ state, passing through an intermediate $J = 1$ state, as shown in the figure below



Excitation and decay of the calcium atom

The mean lifetime of the intermediate state is 4.7 ns. To simplify the terminology, we shall call the $\lambda_1 = 551.3nm$ light green, and the $\lambda_2 = 422.7nm$ light violet.

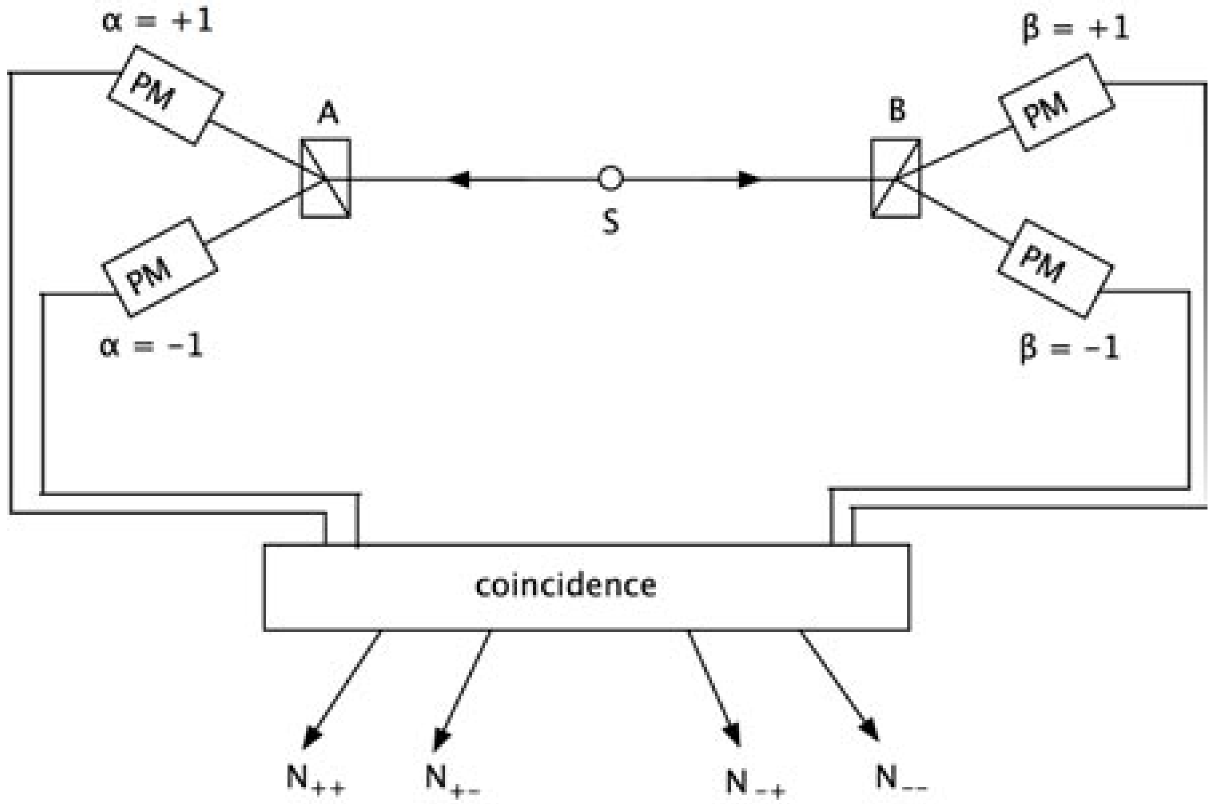
The polarizer, which works like a Wollaston prism shown below



The two-valued response of a Wollaston prism

is made of quartz or of calcite. It splits an incident beam of natural (unpolarized) light into two beams of equal intensity, polarized at $90°$ to each other. If only a single unpolarized photon is incident, it emerges either in the state $|x\rangle$, with probability 1/2, or in the state $|y\rangle$, with probability 1/2. Thus, the response of the system is two-valued.

The photon is detected by the photomultiplier tubes (PM) downstream from the prism. Every electric pulse from these detectors corresponds to the passage of a photon, allowing the photons to be counted. The experimental layout is sketched in the figure below.



Sketch of the first Orsay experiment

It uses a coincidence circuit which registers an event whenever two photons are detected in cascade. In this way four separate counts are recorded simultaneously, over some given period of time. In the EPR scenario envisaged by Bohm, where $\theta = 0$, the only possible responses are (+1,-1) or (-1,+1) (in the situation realized by Aspect, the angle θ is non-zero, and four different responses are possible).

(1) N_{++}, the number of coincidences corresponding to $\alpha = 1$ and $\beta = 1$, that is, to $\alpha\beta = 1$

(2) N_{+-}, the number of coincidences corresponding to $\alpha = 1$ and $\beta = -1$, that is, to $\alpha\beta = -1$

(3) N_{-+}, the number of coincidences corresponding to $\alpha = -1$ and $\beta = 1$, that is, to $\alpha\beta = -1$

(4) N_{--}, the number of coincidences corresponding to $\alpha = -1$ and $\beta = -1$, that is, to $\alpha\beta = 1$

The resolving time of the coincidence circuit is 10 ns, meaning that it reckons two photons as coincident if the they are separated in time by no more than 10 ns. The mean life of the intermediate state of the calcium atom is 4.7 ns. Therefore, after a lapse of 10 ns, that is more than twice the mean lifetime, almost all the atoms have decayed (actually 88%). In other words, the efficiency of the coincidence counter is very high.

The experiment consists in counting, over some given time interval, the four kinds of coincidence: N_{++}, N_{+-}, N_{-+}, and N_{--}. The total number of events is $N = N_{++} + N_{+-} + N_{-+} + N_{--}$.

Accordingly, the different kinds of coincidence have probabilities

$$P_{++} = N_{++}/N \quad \text{corresponding to } \alpha\beta = 1$$
$$P_{+-} = N_{+-}/N \quad \text{corresponding to } \alpha\beta = -1$$
$$P_{-+} = N_{-+}/N \quad \text{corresponding to } \alpha\beta = -1$$
$$P_{--} = N_{--}/N \quad \text{corresponding to } \alpha\beta = 1$$

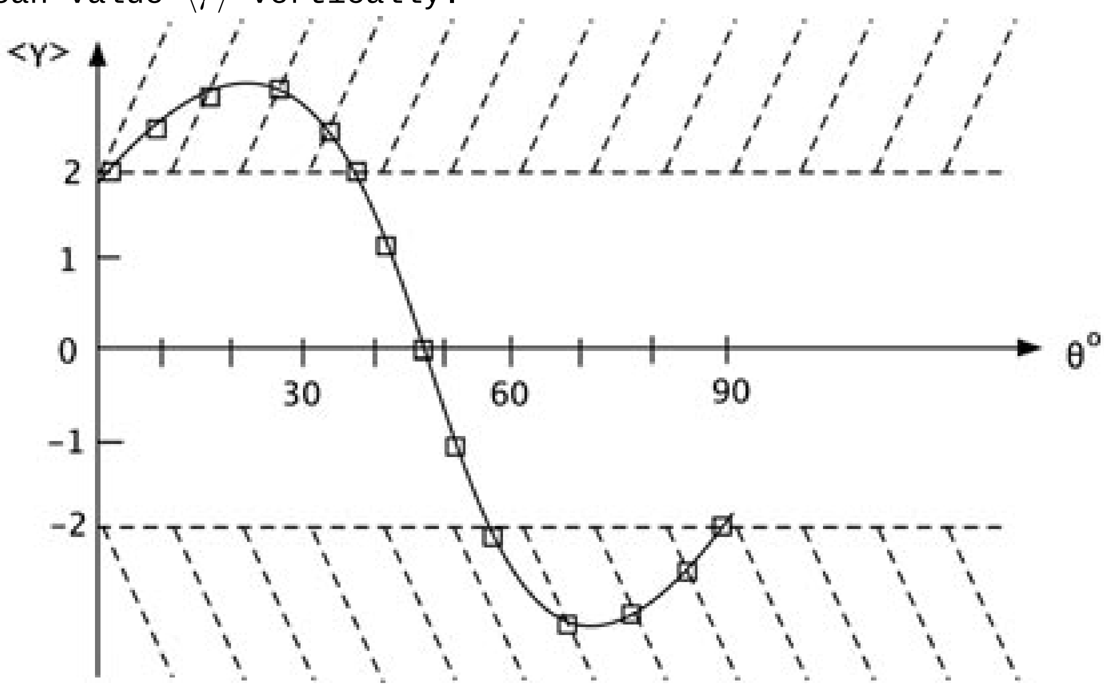
and the measured average of $\alpha\beta$ is

$$\langle\alpha\beta\rangle = \frac{N_{++} - N_{+-} - N_{-+} + N_{--}}{N}$$

Each set of four coincidence counts corresponds to one particular setting of $\vec{A},\vec{B}$, and yields a mean value $\langle\alpha\beta\rangle$. But in order to determine the correlation function $\langle\gamma\rangle$ used in the BCHSH inequality, we need four mean values $\langle\alpha\beta\rangle$. Therefore, we choose, in succession four different settings as shown in figure(c) on page 20; four counting runs then yield the four mean values $\langle\alpha_1\beta_1\rangle, \langle\alpha_1\beta_2\rangle, \langle\alpha_2\beta_1\rangle, \langle\alpha_2\beta_2\rangle$, which then determine the value of $\langle\gamma\rangle$ via $\langle\gamma\rangle = \langle\alpha_1\beta_1\rangle + \langle\alpha_1\beta_2\rangle + \langle\alpha_2\beta_1\rangle - \langle\alpha_2\beta_2\rangle$.

The Results of the First Experiment at Orsay

These results are shown in the figure below. The angle θ which specifies the setting of the polarizers is plotted horizontally, and the mean value $\langle\gamma\rangle$ vertically.



The results of the first Orsay experiment

The correlation function predicted by quantum mechanics reads

$$\langle\gamma\rangle = 3\cos 2\theta - \cos 6\theta$$

It is drawn as the solid curve on the graph(the curve has been corrected for instrumental effects, which explains why its ends are not precisely at 2 and -2). According to the BCHSH inequality

$$-2 \le \langle \gamma \rangle \le 2$$

so that hidden-variable theories exclude the cross-hatched regions of the plane, which correspond to $\langle \gamma \rangle > 2$ or $\langle \gamma \rangle < -2$.

The experimental results from 17 different values of θ are indicated on the figure by squares, where the vertical size of the square gives plus or minus one standard deviation (a measure of the experimental error).

Clearly, there can be no doubt that the BCHSH inequality is violated; many of the experimental points fall outside the interval [-2,2]. At the point where the violation is maximal ($\theta = 22.5°$), one finds

$$\langle \gamma \rangle = 2.70 \pm 0.015$$

which represents a departure of over 40 standard deviations from the extreme value of 2. What is even more convincing is the precision with which the experimental points lie on the curve predicted by quantum mechanics.

Quite evidently, for the EPR scenario one must conclude not only that hidden-variable theories fail, but also that quantum mechanics is positively the right theory for describing the observations.

The Relativistic Test

The EPR experiment just described shows that the measurements in A and B are correlated. What is the origin of the correlations?

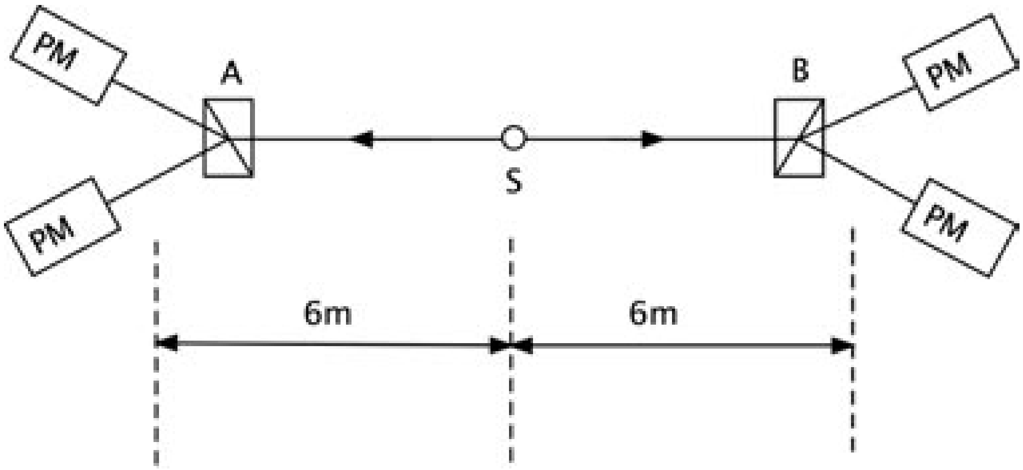
According to quantum theory, before the measurement each particle pair constitutes a single system extending from A to B, whose two parts are non-separable and correlated. This interpretation corresponds to a violation of Bell's inequality.

According to hidden-variables theories, the particle pair is characterized, at the instant of decay, by its hidden variable λ, which determines the correlation between the polarizations measured in A and B. This interpretation satisfies Bell's inequality.

Accordingly, the Orsay experiment supports the quantum interpretation (in terms of the correlation between two parts A and B of a single system).

However, to clinch this conclusion, one must ensure **that no influence is exerted** in the ordinary classical sense through some interaction propagated between the two detectors A and B, that is, no influence which might take effect after the decay at S, and which might be responsible for the correlation actually observed.

Let us therefore examine the Orsay apparatus in more detail as in the figure below.



Einsteinian non-separability

When the detectors at A and B record a coincidence, this means that both have been triggered within a time interval of at most 10 ns, the resolving time of the circuit. Could it happen that, within this interval, A sends to B a signal capable of influencing the response of B? In the most favorable case, such a signal would travel with the speed of light in vacuum, which according to relativity theory is the upper limit on the propagation speed of information, and thereby of energy. To cover the distance AB, which is 12 m in the figure, such a signal would need 40 ns. This is too long by at least 30 ns, and rules out any causal links between A and B in the sense of classical physics. One says that **the interval between A and B is space-like**.

One of the advantages of the Orsay experiment is that it uses a very strong light-source, allowing sufficient distance between the detectors A and B while still preserving reasonable counting rates. By increasing the distance AB step by step, Aspect could check that the correlation persists, even when the interval between A and B becomes space-like. This is the check that guarantees that the two-photon system is non-separable irrespective of the distance AB.

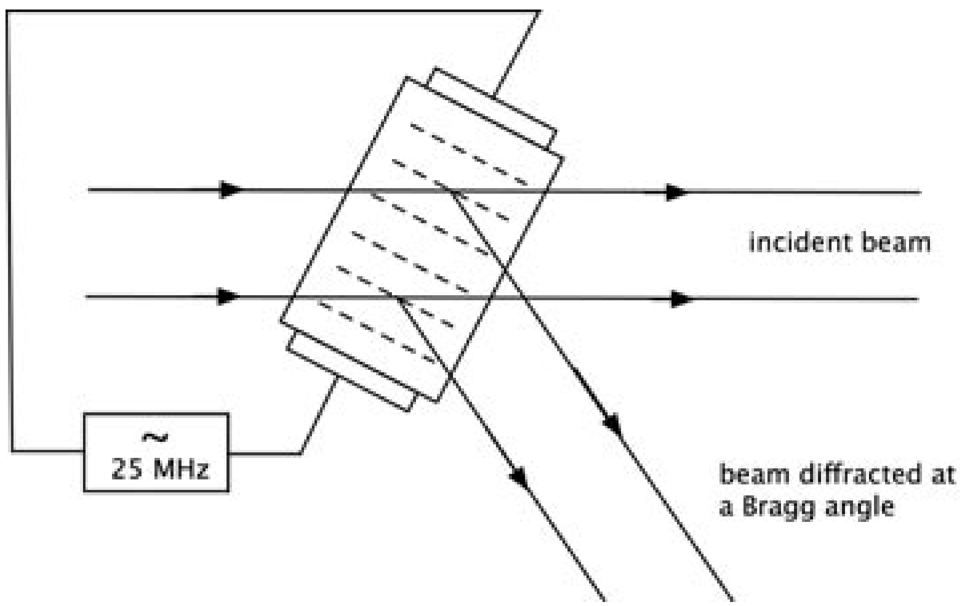
It has become the custom to speak of **the principle of Einsteinian separability** in order to denote the absence of correlations between two events separated by a space-like interval. This is the principle that the Orsay experiment invites us to reconsider, even though our minds, used to the world at the macroscopic level, find it difficult to conceive of two "microscopic" photons 12 m apart as a single indivisible object.

The Final Stage of the Experiment at Orsay (1983)

Though the results of the first Orsay experiment are inarguable and clear-cut, the conclusion they invite is so startling that one should not be surprised at the appearance of a last-ditch objection, which as it happens gave the experimenters a great deal of trouble. In the preceding section we discussed the possible role of interactions between A and B operating after the decay at S, and duly eliminated the objection. But one can also ask whether correlations might be introduced through an interaction operating **before** the decay. We could imagine that the decay itself is preconditioned by the setting of detectors A and B, such influences taking effect through the exchange of signals between the detectors and the source. No such mechanism is known **a priori**, but we do know that, if there is one, then Einsteinian non-separability would cease to be a problem, because the mechanism could come into action long before the decay, removing any reason for expecting a minimum 30 ns delay. Though such a scenario is very unlikely, the objection is a serious one and must be taken into account; to get around it, the experimenter must be able to choose the orientation of the detectors A and B at random after the decay has happened at S. In more picturesque language, we would say that the two photons must leave the source without knowing the orientations of the polarizers A and B. Briefly put, this means that it must be possible to change the detector orientations during the 20 ns transits over SA and SB.

The solution adopted at Orsay employs periodic switching every 10 ns. These changes are governed by two independent oscillators, one for channel A and one for channel B. The oscillators are stabilized, but however good the stabilization it cannot eliminate small random drifts that are different in the two channels, seeing that the oscillators are independent. This ensures that the changes of orientation are random even though the oscillations are periodic, provided the experiment lasts long enough (1 to 3 hours).

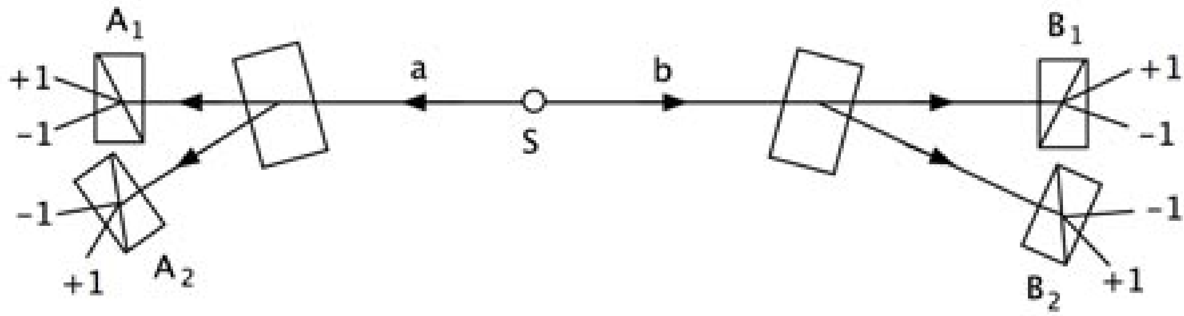
The key element of the second Orsay experiment is the optical switch shown in the figure below.



In a water tank, a system of standing waves is produced by electro-acoustic excitation at a frequency of 25 MHz(corresponds to 10 ns between switchings).

The fluid keeps changing from a state of perfect rest to one of maximum agitation and back again. In the state of rest, the light beam is simply transmitted, In the state of maximum agitation, the fluid arranges itself into a structure of parallel and equidistant plane layers, alternately stationary (nodal planes) or agitated (antinodal planes). Thus, one sets up a lattice of net-like diffracting planes; the diffracted intensity is maximum at the so-called Bragg angles, just as in scattering from a crystal lattice. Here the light beam is deviated through 10^{-2} radians (the angles in the figure are exaggerated for effect). The two numerical values, 25 MHz and 10^{-2} radians, suffice to show the magnitude of the technical achievement. With the acoustic power of 1 watt, the system functions as an ideally efficient switch.

The second Orsay experiment (using optical switches) is sketched in the figure below.



In this set-up, the photons a and b leave S without "knowing" whether they will go, the first to A_1 or A_2, and the second to B_1 or B_2.

The second experiment is less precise than the first, because the light beams must be very highly collimated in order to ensure efficient switching. Nevertheless, its results exhibit an unambiguous violation of Bell's inequality, reaching 5 standard deviations at the peak; moreover the results are entirely compatible with the predictions of quantum mechanics.

The Principle of Non-Separability

Experiment has spoken. Half a century after the Como conference, Bohr's interpretation once again beats Einstein's, in a debate more subtle and also more searching. There were two conflicting theories:

Einstein	Bohr
hidden variables	quantum mechanics
realist	positivist
deterministic	probabilistic
separable	non-separable

The violation of the BCHSH inequality argues for Bohr's interpretation, all the more so as the measured values of $\langle \gamma \rangle$ are in close agreement with the predictions of quantum mechanics.

It remains to ask oneself just why hidden-variable theories do fail. Of the three basic assumptions adopted by such theories, namely realism, determinism, and separability, at least one must be abandoned. In the last resort, it is separability that seems to be the most vulnerable assumption. Indeed, one observes experimentally that the violation of the BCHSH inequality is independent of the distance between the two detectors A and B, even when this distance is 12 m or more. There are still die-hard advocates of determinism, who try to explain non-separability through non-local hidden variables. Such theories, awkward and barely predictive, are typically ad hoc, and fit only a limited number of phenomena. They are weakly placed to defend themselves against interpretations furnished by quantum mechanics, which have the virtues of simplicity, elegance, efficiency, and generality, and which are invariably confirmed by experiment.

The principle of Einsteinian separability asserts that "there are no correlations between two phenomena separated by a space-like interval". In other words, no interaction can propagate faster than light in vacuum. In an EPR scenario this principle must be abandoned, and replaced by a principle asserting non-separability:

> "in a quantum system evolving free of external perturbations, and from well-defined initial conditions, all parts of the system remain correlated, even when the interval between them is space-like"

This assertion reflects the properties of the state vector of a quantum system. For an EPR system, the state vector after the decay of the source reads

$$|\Phi\rangle = \frac{1}{\sqrt{2}}\left(|x_A, x_B\rangle + |y_A, y_B\rangle\right)$$

This expression combines the elements A and B in a non-separable manner, which is what explains the observed correlations. The truth is that all this has been well known ever since the beginnings of quantum mechanics, with the concept of the electron cloud as the most telling illustration. It is for instance hard to imagine separability between the 92 electrons of a uranium atom. What is new is that quantum mechanics, considered hitherto as a microscopic theory applicable on the atomic scale, is now seen to apply to a two-particle system macroscopically, on the scale of meters. The truly original achievement of Aspect's experiment is the demonstration of this fact.

Quantum objects have by no means exhausted their capacity to astonish us by their difference from the properties of the macroscopic objects in our everyday surroundings. In the preceding sections we saw that a photon can interfere with itself and we have shown that two photons 12 m apart constitute but a single object. Thus, it becomes ever more difficult to picture a photon through analogies with rifle bullets, surface waves in water, clouds in the sky, or with any other object of our familiar universe. Such partial analogies fail under attempts to make them more complete, and through their failure we discover new properties pertaining to quantum objects. The only fruitful procedure is to follow the advice of Niels Bohr, namely, to bend one's mind to the new quantum concepts until they become habitual and thereby intuitive. Earlier generations of physicists have had to face similar problems. They had to progress from Aristotle's mechanics to Newton's, and then from Newton's to Einstein's. The same effort is now required of us, at a time favorable in that, by mastering the EPR paradox, quantum mechanics has passed a particularly severe test with flying colors. From this point of view, the principle of non-separability seems as important as the principle of special relativity , and Aspect's experiment plays the same role now that the Michelson-Morley experiment played then.

A Problem with a Solution (Difficult but Rewarding)

Bell's Theorem with Photons

Two photons fly apart from one another, and are in oppositely oriented circularly polarized states. One strikes a polaroid film with axis parallel to the unit vector $\hat{a}$, the other a polaroid with axis parallel to the unit vector $\hat{b}$. Let $P_{++}(\hat{a},\hat{b})$ be the joint probability that both photons are transmitted through their respective polaroids. Similarly, $P_{--}(\hat{a},\hat{b})$ is the probability that both photons are absorbed by their respective polaroids, $P_{+-}(\hat{a},\hat{b})$ is the probability that the photon at the $\hat{a}$ polaroid is transmitted and the other is absorbed, and finally, $P_{-+}(\hat{a},\hat{b})$ is the probability that the photon at the $\hat{a}$ polaroid is absorbed and the other is transmitted.

The classical realist assumption is that these probabilities can be separated:

$$P_{ij}(\hat{a},\hat{b}) = \int d\lambda \rho(\lambda) P_i(\hat{a},\lambda) P_j(\hat{b},\lambda)$$

where i and j take on the values + and −, where λ signifies the so-called hidden variables, and where $\rho(\lambda)$ is a weight function. This equation is called the separable form.

The correlation coefficient is defined by

$$C(\hat{a},\hat{b}) = P_{++}(\hat{a},\hat{b}) + P_{--}(\hat{a},\hat{b}) - P_{+-}(\hat{a},\hat{b}) - P_{-+}(\hat{a},\hat{b})$$

and so we can write

$$C(\hat{a},\hat{b}) = \int d\lambda \rho(\lambda) C(\hat{a},\lambda) C(\hat{b},\lambda)$$

where

$$C(\hat{a},\lambda) = P_+(\hat{a},\lambda) - P_-(\hat{a},\lambda)$$
$$C(\hat{b},\lambda) = P_+(\hat{b},\lambda) - P_-(\hat{b},\lambda)$$

It is required that

(a) $\rho(\lambda) \geq 0$

(b) $\int d\lambda \rho(\lambda) = 1$

(c) $-1 \leq C(\hat{a},\lambda) \leq 1$, $-1 \leq C(\hat{b},\lambda) \leq 1$

The Bell coefficient

$$B = C(\hat{a},\hat{b}) + C(\hat{a},\hat{b}') + C(\hat{a}',\hat{b}) - C(\hat{a}',\hat{b}')$$

combines four different combinations of the polaroid directions.

(1) Show that the above classical realist assumptions imply that $|B| \leq 2$

(2) Show that quantum mechanics predicts that

$$C(\hat{a},\hat{b}) = 2\left(\hat{a}\cdot\hat{b}\right)^2 - 1$$

(3) Show that the maximum value of the Bell coefficient is $2\sqrt{2}$, according to quantum mechanics

(4) Cast the quantum mechanical expression for $C(\hat{a},\hat{b})$ into a separable form. Which of the classical requirements, (a), (b), or (c) above is violated?

Solution

(1) With the separability assumption, we have (from above)

$$C(\hat{a},\hat{b}) = \int d\lambda \rho(\lambda) C(\hat{a},\lambda) C(\hat{b},\lambda)$$

It follows that the Bell coefficient can be written in the form

$$\begin{aligned} B &= C(\hat{a},\hat{b}) + C(\hat{a},\hat{b}') + C(\hat{a}',\hat{b}) - C(\hat{a}',\hat{b}') \\ &= \int d\lambda \rho(\lambda) \Big(C(\hat{a},\lambda)(C(\hat{b},\lambda) + C(\hat{b}',\lambda)) + C(\hat{a}',\lambda)(C(\hat{b},\lambda) - C(\hat{b}',\lambda)) \Big) \end{aligned}$$

Since $|C(\hat{a},\lambda)| \le 1$, $|C(\hat{a}',\lambda)| \le 1$ and $\rho(\lambda) \ge 0$, we have

$$|B| \le \int d\lambda \rho(\lambda) \Big(\left|C(\hat{b},\lambda) + C(\hat{b}',\lambda)\right| + \left|C(\hat{b},\lambda) - C(\hat{b}',\lambda)\right| \Big)$$

Now suppose that for a given λ, $C_M(\lambda)$ is the maximum and $C_m(\lambda)$ is the minimum of $C(\hat{b},\lambda)$ and $C(\hat{b}',\lambda)$, so that $C_M(\lambda) \ge C_m(\lambda)$. Then

$$|B| \le \int d\lambda \rho(\lambda) \Big(\left|C_M(\lambda) + C_m(\lambda)\right| + C_M(\lambda) - C_m(\lambda) \Big)$$

There are two cases to consider.

For the case $C_M(\lambda) \ge 0$, we have $\left|C_M(\lambda) + C_m(\lambda)\right| = C_M(\lambda) + C_m(\lambda)$ so that

$$\begin{aligned} |B| &\le \int d\lambda \rho(\lambda) \Big(C_M(\lambda) + C_m(\lambda) + C_M(\lambda) - C_m(\lambda) \Big) \\ &= 2\int d\lambda \rho(\lambda) C_M(\lambda) \le 2\int d\lambda \rho(\lambda) \left|C_M(\lambda)\right| \le 2\int d\lambda \rho(\lambda) = 2 \end{aligned}$$

For the case $C_M(\lambda) < 0$, we have $\left|C_M(\lambda) + C_m(\lambda)\right| = -C_M(\lambda) - C_m(\lambda)$ so that

$$\begin{aligned} |B| &\le \int d\lambda \rho(\lambda) \Big(-C_M(\lambda) - C_m(\lambda) + C_M(\lambda) - C_m(\lambda) \Big) \\ &= 2\int d\lambda \rho(\lambda)(-C_m(\lambda)) \le 2\int d\lambda \rho(\lambda) \left|C_m(\lambda)\right| \le 2\int d\lambda \rho(\lambda) = 2 \end{aligned}$$

Thus, in all cases $|B| \le 2$.

(2) A photon, traveling in the y direction, might have right- or left-handed circular polarization. The corresponding quantum states are written $|R\rangle$ and $|L\rangle$ respectively. These circular polarization

states can be expressed as coherent superpositions of linearly polarized states in the z and x directions:

$$|R\rangle = \frac{1}{\sqrt{2}}\left(|z\rangle + i|x\rangle\right)$$
$$|L\rangle = \frac{1}{\sqrt{2}}\left(|z\rangle - i|x\rangle\right)$$

Under a rotation of the coordinate axes by an angle θ about the y direction, $|R\rangle \to e^{i\theta}|R\rangle$ and $|L\rangle \to e^{-i\theta}|L\rangle$ or equivalently

$$\begin{pmatrix}|z\rangle \\ |x\rangle\end{pmatrix} \to \begin{pmatrix}|z'\rangle \\ |x'\rangle\end{pmatrix} = \begin{pmatrix}\cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}|z\rangle \\ |x\rangle\end{pmatrix}$$

If each photon is in a state of right-handed circular polarization, we write the corresponding state vector as $|R_1\rangle|R_2\rangle$. However, since the photons are moving in opposite directions, one along the positive, and the other along the negative y axis, it follows that the actual directions in which the electric fields rotate, in time, in the vicinity of the two photons, are opposed to one another. The same hold for the state $|L_1\rangle|L_2\rangle$, corresponding to each photon being in a state of left-handed circular polarization.

The linear combination of these two states,

$$|EPR\rangle = \frac{1}{\sqrt{2}}\left(|R_1\rangle|R_2\rangle + |L_1\rangle|L_2\rangle\right)$$

corresponds to the more general situation in which the photons are in oppositely oriented states of circular polarization, where the sense of this polarization is not specified. We can write this **entangled** or **Einstein-Podolsky-Rosen** state in the form

$$|EPR\rangle = \frac{1}{\sqrt{2}}\left(|z_1\rangle|z_2\rangle - |x_1\rangle|x_2\rangle\right)$$

which is a superposition of states of linear polarization.

Suppose now that a measurement of linear polarization is made on photon 1 in the z direction, and of photon 2 in the z′ direction, that is, the z direction after a rotation of the axes about the y axis. The probability amplitude associated with this measurement on the EPR state is

$$\langle EPR|z_1 z_2'\rangle = \frac{1}{\sqrt{2}}\left(\langle z_1|\langle z_2| - \langle x_1|\langle x_2|\right)\left(|z_1\rangle\left(\cos\theta|z_2\rangle - \sin\theta|x_2\rangle\right)\right)$$
$$= \frac{1}{\sqrt{2}}\cos\theta$$

where we have used $\langle z_1|x_1\rangle = 0$. The probability that photon 1 is found to have linear polarization in the direction z, and photon 2 in the direction z′ is

$$P_{++}(\hat{a},\hat{b}) = \left|\langle EPR|z_1 z_2'\rangle\right|^2 = \frac{1}{2}\cos^2\theta$$

where we have assumed that $\hat{a}$ is in the z direction and $\hat{b}$ is in the z′ direction.

Suppose next that the linear polarization of the linear polarization of photon 1 were measured in the x direction, and that of photon 2 again in the z′ direction. The probability amplitude is

$$\langle EPR|x_1z_2'\rangle = \frac{1}{\sqrt{2}}\Big(\langle z_1|\langle z_2| - \langle x_1|\langle x_2|\Big)\Big(|x_1\rangle\big(\cos\theta|z_2\rangle - \sin\theta|x_2\rangle\big)\Big)$$
$$= \frac{1}{\sqrt{2}}\sin\theta$$

If photon 1 has polarization in the x direction, then it will not be transmitted by a polarizer in the z direction - it will be absorbed. Hence,

$$P_{-+}(\hat{a},\hat{b}) = \left|\langle EPR|x_1z_2'\rangle\right|^2 = \frac{1}{2}\sin^2\theta$$

Similarly,

$$P_{+-}(\hat{a},\hat{b}) = \left|\langle EPR|z_1x_2'\rangle\right|^2 = \frac{1}{2}\sin^2\theta$$

$$P_{--}(\hat{a},\hat{b}) = \left|\langle EPR|x_1x_2'\rangle\right|^2 = \frac{1}{2}\cos^2\theta$$

The correlation coefficient is then

$$C(\hat{a},\hat{b}) = P_{++}(\hat{a},\hat{b}) + P_{--}(\hat{a},\hat{b}) - P_{+-}(\hat{a},\hat{b}) - P_{-+}(\hat{a},\hat{b})$$
$$= \cos^2\theta - \sin^2\theta = 2\cos^2\theta - 1 = \cos 2\theta$$

Since the unit vectors $\hat{a}$ and $\hat{b}$ are at an angle θ with respect to one another, it follows that $\hat{a}\cdot\hat{b} = \cos\theta$, and therefore

$$C(\hat{a},\hat{b}) = 2\cos^2\theta - 1 = 2(\hat{a}\cdot\hat{b})^2 - 1$$

(3) Suppose that the angle between the vectors $\hat{a}'$ and $\hat{a}$ is $x/2$, between $\hat{a}$ and $\hat{b}$ is $y/2$ and between $\hat{b}$ and $\hat{b}'$ is $z/2$. Then the angle between $\hat{a}'$ and $\hat{b}'$ is $(x+y+z)/2$ and according to quantum mechanics, the Bell coefficient has the form

$$B = \cos x + \cos y + \cos z - \cos(x+y+z)$$

This function extrema when

$$\frac{\partial B}{\partial x} = -\sin x + \sin(x+y+z) = 0$$
$$\frac{\partial B}{\partial y} = -\sin y + \sin(x+y+z) = 0$$
$$\frac{\partial B}{\partial z} = -\sin z + \sin(x+y+z) = 0$$

or

$$\sin x = \sin y = \sin z = \sin(x+y+z)$$

This has the solution

$$x = y = z \text{ and } 3x = \pi - x \rightarrow x = \pi/4$$

For this extremum

$$B = 3\cos\frac{\pi}{4} - \cos\frac{3\pi}{4} = \frac{3}{\sqrt{2}} + \frac{1}{\sqrt{2}} = 2\sqrt{2}$$

This is a maximum, since at this point

$$\frac{\partial^2 B}{\partial x^2} = \frac{\partial^2 B}{\partial y^2} = \frac{\partial^2 B}{\partial z^2} = -\cos\frac{\pi}{4} + \cos\frac{3\pi}{4} = -\sqrt{2} < 0$$

(4) Let the vector $\hat{a}$ be at an angle θ_a with respect to some direction in the xz plane, and let $\hat{b}$ be at an angle θ_b with respect to the same direction. Then

$$C(\hat{a},\hat{b}) = \cos 2(\theta_a - \theta_b) = \cos 2\theta_a \cos 2\theta_b + \sin 2\theta_a \sin 2\theta_b$$
$$= \int d\lambda \rho(\lambda) C(\hat{a},\lambda) C(\hat{b},\lambda)$$

and with the assignments

$$\rho(\lambda) = \delta(\lambda+1) + \delta(\lambda-1)$$
$$C(\hat{a},1) = \cos 2\theta_a \quad , \quad C(\hat{a},-1) = \sin 2\theta_a$$
$$C(\hat{b},1) = \cos 2\theta_b \quad , \quad C(\hat{b},-1) = \sin 2\theta_b$$

we then see that

$$\rho(\lambda) \geq 0$$

$$-1 \leq C(\hat{a},\lambda)\, , C(\hat{b},\lambda) \leq 1 \text{ for } \lambda = \pm 1$$

but

$$\int d\lambda \rho(\lambda) = 1 + 1 = 2$$

so that the normalization condition (b) is violated.