A simple proof of Bell's inequality

Lorenzo Maccone

Citation: American Journal of Physics 81, 854 (2013); doi: 10.1119/1.4823600
View online: http://dx.doi.org/10.1119/1.4823600
View Table of Contents: http://scitation.aip.org/content/aapt/journal/ajp/81/11?ver=pdfcov
Published by the American Association of Physics Teachers

Advertisement:
A simple proof of Bell’s inequality

Lorenzo Maccone
Dip. Fisica, INFN Sez. Pavia, Univ. Pavia, via Bassi 6, I-27100 Pavia, Italy

(Received 28 December 2012; accepted 16 September 2013)

Bell’s theorem is a fundamental result in quantum mechanics: it discriminates between quantum mechanics and all theories where probabilities in measurement results arise from the ignorance of pre-existing local properties. We give an extremely simple proof of Bell’s inequality; a single figure suffices. This simplicity may be useful in the unending debate over what exactly the Bell inequality means, because the hypotheses underlying the proof become transparent. It is also a useful didactic tool, as the Bell inequality can be explained in a single intuitive lecture.

I. INTRODUCTION

Einstein had a dream. He believed quantum mechanics was an incomplete description of reality1 and that its completion might explain the troublesome fundamental probabilities of quantum mechanics as emerging from some hidden degrees of freedom—probabilities would arise because of our ignorance of these “hidden variables.” His dream was that probabilities in quantum mechanics might turn out to have the same meaning as probabilities in classical thermodynamics, where they arise from our ignorance of the microscopic degrees of freedom (e.g., the position and velocity of each gas molecule). He wrote, “the statistical quantum theory would, within the framework of future physics, take an approximately analogous position to the statistical mechanics within the framework of classical mechanics.”

A decade after Einstein’s death, John Bell shattered this dream.3–5 Any completion of quantum mechanics with hidden variables would be incompatible with relativistic causality! The essence of Bell’s theorem is that quantum mechanical probabilities cannot arise from the ignorance of local pre-existing variables. In other words, if we want to assign pre-existing (but hidden) properties to explain probabilities in quantum measurements, these properties must be nonlocal. An agent with access to the nonlocal variables could transmit information instantly to a distant location, thus violating relativistic causality and awakening the nastiest temporal paradoxes.6

(We use “local” here in Einstein’s connotation: locality implies that superluminal communication is impossible. In contrast, often quantum mechanics is deemed “nonlocal” in the sense that correlations among properties can propagate instantly, thanks to entanglement.1 This “quantum nonlocality” cannot be used to transfer information instantly, because mere correlations cannot be used in that way. In the remainder of this paper, we will use the terms local and nonlocal in Einstein’s sense, and not in the sense of quantum correlations.)

Modern formulations of quantum mechanics must incorporate Bell’s result at their core. Either they refuse the idea that measurements uncover pre-existing values, or they must make use of nonlocal properties. In the latter case, they must also introduce some censorship mechanism to prevent the use of hidden variables to transmit information. An example of the first formulation is the conventional Copenhagen interpretation of quantum mechanics, which (thanks to complementarity) states that the values of properties arise from the interaction between the quantum system and the measurement apparatus, and so are not pre-existing: “unperformed experiments have no results.”7 An example of the second formulation is the de Broglie-Bohm interpretation of quantum mechanics, which assumes that particle trajectories are hidden variables (they exist independently of position measurements).

Bell’s result is at the core of modern quantum mechanics, as it elucidates the theory’s precarious co-existence with relativistic causality; it has spawned an impressive amount of research. However, it is often ignored in basic quantum mechanics courses because traditional proofs of Bell’s theorem are rather cumbersome and often overburdened by philosophical considerations. Here we give an extremely simple graphical proof of Mermin’s version8,9 of Bell’s theorem. The simplicity of the proof is key to clarifying all of the theorem’s assumptions, the identification of which has generated extensive debate in the literature (e.g., see Ref. 10). Here, we focus on simplifying the proof. We refer the reader who wants to gain an intuition for the quantum part to Refs. 11 and 12, and to Ref. 13 for a proof without probabilities.

II. BELL’S THEOREM AND ITS ASSUMPTIONS

Let us define a “local” theory as one that implies that the outcomes of an experiment on a system are independent of the actions performed on a different system that has no causal connection with the first. (As stated previously, this refers to locality in Einstein’s connotation of the word—the outcomes of the experiment cannot be used to receive information from whoever acts on the second system, if it has no causal connection to the first.) For example, the temperature of my room is independent of whether you choose to wear a purple tie today. Einstein’s relativity provides a stringent condition for causal connections: if two events are outside each others’ light cones there cannot be any causal connection between them.

Let us define as “counterfactual-definite”14,15 a theory whose experiments uncover properties that are pre-existing. In other words, in a counterfactual-definite theory it is meaningful to assign a property to a system (e.g., the position of an electron) independent of whether the measurement of this property is carried out. (Sometime counterfactual definiteness is also called “realism,” but this philosophically laden term can lead to misconceptions.)

Bell’s theorem can be phrased as “quantum mechanics cannot be both local and counterfactual-definite.” A logically equivalent statement is “quantum mechanics is either nonlocal or non-counterfactual-definite.”
To prove this theorem, Bell derived an inequality (involving correlations of measurement results) that is satisfied by all theories that are both local and counterfactual-definite. He then showed that quantum mechanics violates this inequality, and hence cannot be local and counterfactual-definite.

It is important to note that the Bell inequality can also be derived using weaker hypotheses than Einstein locality and counterfactual definiteness. Such a proof is presented in Appendix A, where Einstein locality is relaxed to “Bell locality” and counterfactual definiteness is relaxed to “hidden variable models.” However, from a physical point of view, the big impact of Bell’s theorem is to prove the incompatibility of quantum mechanics with local counterfactual-definiteness, so we will stick to these hypotheses in the main text. (See also Appendix B for a schematic formalization of all these results.)

Two additional hypotheses that underly Bell’s theorem are often left implicit. The first is that our choice of which experiment to perform must be independent of the properties of the object to be measured. This hypothesis is sometimes called “freedom of choice” or “no super-determinism.” If this hypothesis were not true, then, for example, we might falsely conclude that all objects are red when in fact, we are somehow being prevented from choosing to measure the color of non-red objects. The second hypothesis is that future outcomes of an experiment must not influence which apparatus settings were previously chosen. Although apparatus settings will clearly influence the outcomes of experiments, we assume that the reverse does not occur, as a matter of simple causality. This hypothesis is sometimes called “measurement independence.” Both of these hypotheses are usually left implicit because science would be impossible without them.

All experiments performed to date (e.g., see Refs. 17–19) have shown that the Bell inequalities are violated, suggesting that our world cannot be both local and counterfactual-definite. However, it should be noted that no experiment up until now has been able to test the Bell inequalities with complete rigor, because additional assumptions are required to take care of experimental imperfections. These assumptions are all quite reasonable, so that only conspiratorial alternatives to quantum mechanics have yet to be ruled out (where experimental imperfections are fine-tuned to the properties of the objects, so they violate “freedom of choice”). A number of experimental groups are currently pursuing the definitive Bell inequality experiment.

III. PROOF OF BELL’S THEOREM

We consider the version of the Bell inequality proposed by Preskill, following Mermin’s suggestion. Suppose we have two objects that are identical; they have same properties and same values of all properties. Suppose also that the values of these properties are predetermined (counterfactual definiteness) and not generated by their measurement, and that the determination of the property values of one object will not influence any property of the other object (locality).

We will need only three properties, \( A \), \( B \), and \( C \), which can each take the two values 0 and 1. For example, if the objects are coins then \( A = 0 \) might mean that the coin is gold and \( A = 1 \) that the coin is copper (property \( A \): material); \( B = 0 \) means that the coin is shiny and \( B = 1 \) that it is dull (property \( B \): texture); and \( C = 0 \) means that the coin is large and \( C = 1 \) that it is small (property \( C \): size).

Suppose that I do not know any of the property values, because the two coins are a gift in two wrapped boxes. However, I do know that the gift consists of two identical coins. They could be two gold, shiny, small coins \((A = 0, B = 0, C = 1)\), or two copper, shiny, large coins \((1, 0, 0)\), or two copper, dull, large coins \((1, 1, 0)\), and so on. I do know that the property values “exist” (that is, they are counterfactual-definite and predetermined even if I cannot see them directly) and they are local (so acting on one box will not change any property of the coin in the other box). These are quite reasonable assumptions for two coins. My ignorance of the property values is expressed through probabilities that represent either my expectation of finding a value (Bayesian view), or the result of performing many repeated experiments with boxes and coins and averaging over some possibly hidden variable, typically indicated with the letter \( x \) (deterministic view). For example, I might say that the gift bearer will give me two gold coins with a 20% probability (he is usually stingy, but not always).

Bell’s inequality concerns the correlation among measurement outcomes of the property values. Call \( P_{\text{same}}(A, B) \) the probability that the properties \( A \) of the first object and \( B \) of the second have the same value: \( A = B \) and both are 0 (the first coin is gold and the second is shiny) or they are both 1 (the first is copper and the second is dull). For example, \( P_{\text{same}}(A, B) = 1/2 \) would tell me that there is a 50% chance that \( A = B \) (namely, they are both 0 or both 1). Since the two coins have the same counterfactual-definite values, this also implies that there is a 50% chance that I get two gold shiny coins or two copper dull coins. Note that the fact that the two coins are identical (i.e., they possess properties with the same values) means that \( P_{\text{same}}(A, A) = P_{\text{same}}(B, B) = P_{\text{same}}(C, C) = 1 \); if one is gold so is the other, and so on.

Bell’s inequality assumes that three, arbitrary two-valued properties \( A, B, C \) satisfy counterfactual definiteness and locality, and that we have two objects such that \( P_{\text{same}}(X, X) = 1 \) for \( X = A, B, C \) (i.e., the two objects have the same property values). Under these conditions Bell’s inequality states that

\[
P_{\text{same}}(A, B) + P_{\text{same}}(A, C) + P_{\text{same}}(B, C) \geq 1.
\] (1)

The proof of this inequality is given graphically in Fig. 1 and explained in the caption. The inequality says that the sum of the probabilities that the two properties have the same value, if I consider, respectively, \( A \) and \( B \), \( A \) and \( C \), and \( B \) and \( C \), must be larger than 1. This conclusion is also intuitively clear: since the two coins have the same properties, the sum of the probabilities that the coins are gold and shiny, copper and dull, gold and large, copper and small, shiny and large, and dull and small must be greater than 1 because all eight possible three-value combinations have been counted, some more than once. Figure 2 shows how these eight combinations correspond to the different areas in the Venn diagrams of Fig. 1.

This result is true, of course, only if the two objects have the same counterfactual-definite properties and the measurement of one does not affect the outcome of the other. If we lack counterfactual-definite properties we cannot infer that the first coin is shiny only because we measured the second to be shiny, even if we know that the two coins have the
same property values—without counterfactual definiteness we cannot even speak of the first coin’s texture unless we measure it. Moreover, if a measurement of the second coin’s texture can change that of the first coin (nonlocality), we cannot even infer the first coin’s texture from a measurement of the second. Thus, even if we know that the textures of the coins were initially the same, the measurement on the second may change the texture of the first.

To prove Bell’s theorem, we now provide a quantum system that violates the above inequality. Consider two two-level systems (qubits) in the joint entangled state \( |\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2} \) and consider the two-valued properties \( A \), \( B \), and \( C \) defined by the following three sets of eigenstates

\[
A : \begin{cases} 
|a_0\rangle \equiv |0\rangle \\
|a_1\rangle \equiv |1\rangle
\end{cases} \quad B : \begin{cases} 
|b_0\rangle \equiv \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle \\
|b_1\rangle \equiv \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle
\end{cases} \quad C : \begin{cases} 
|c_0\rangle \equiv \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle \\
|c_1\rangle \equiv \frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle.
\end{cases}
\]

(2)

It is easy to check that the states within each pair are orthogonal. It is also easy to check that

\[
|\Phi^+\rangle = \frac{|a_0a_0\rangle + |a_1a_1\rangle}{\sqrt{2}} = \frac{|b_0b_0\rangle + |b_1b_1\rangle}{\sqrt{2}} = \frac{|c_0c_0\rangle + |c_1c_1\rangle}{\sqrt{2}}.
\]

(3)

Fig. 2. Explicit depiction of the property values whose probabilities are represented by the areas of the Venn diagrams in Fig. 1. Each possible set of property values is represented by a triplet of numbers \((A,B,C)\) that indicates the (counterfactual-definite, local) values of the properties \(A\), \(B\), and \(C\) for both objects. Note that in the dotted area \(A\) must be different from both \(B\) and \(C\), so that \(B\) and \(C\) must be equal there (the values of \(B\) and \(C\) are also equal in the intersection between the two smaller circles, but that is irrelevant to the proof).
so that the two qubits have the same values for all properties; that is, \( P_{\text{same}}(A, A) = P_{\text{same}}(B, B) = P_{\text{same}}(C, C) = 1 \), so that a measurement of the same property on both qubits always yields the same outcome, both 0 or both 1.

We are now ready to calculate the left-hand side of Bell’s inequality (1). To calculate any one of the three terms, we write the state \( |\Phi^+\rangle \) in terms of the corresponding eigenstates of the two individual qubits. For example, we can find the value of \( P_{\text{same}}(A, B) \) if we write

\[
|\Phi^+\rangle = \frac{|a_0\rangle(|b_0\rangle + \sqrt{3}|b_1\rangle) + |a_1\rangle(|\sqrt{3}|b_0\rangle - |b_1\rangle)}{2\sqrt{2}},
\]

(4)

and

\[
|\Phi^+\rangle = \frac{(|b_0\rangle + \sqrt{3}|b_1\rangle)(|c_0\rangle + \sqrt{3}|c_1\rangle) - (\sqrt{3}|b_0\rangle - |b_1\rangle)(\sqrt{3}|c_0\rangle - |c_1\rangle)}{4\sqrt{2}},
\]

(6)

respectively. Summarizing, we have found

\[
P_{\text{same}}(A, B) + P_{\text{same}}(A, C) + P_{\text{same}}(B, C) = \frac{3}{4} < 1,
\]

(7)

which violates Bell’s inequality (1).

This proves Bell’s theorem: all theories that are both local and counterfactual-definite must satisfy inequality (1), which is violated by quantum mechanics. Therefore, quantum mechanics cannot be a local counterfactual-definite theory; it must either be non-counterfactual-definite (as in the Copenhagen interpretation) or nonlocal (as in the de Broglie-Bohm interpretation).21

ACKNOWLEDGMENTS

The author acknowledges G. M. D’Ariano, G. Introzzi, P. Perinotti, and W. H. Zurek for interesting discussions and/or useful feedback. In particular, P. Perinotti suggested Appendix A.

APPENDIX A: HIDDEN VARIABLE MODELS

This appendix is addressed only to more advanced readers. In the spirit of the original proof of Bell’s theorem,22 one can relax both the “counterfactual definiteness” and the “Einstein locality” hypotheses somewhat. In fact, instead of supposing that there are some pre-existing properties of the objects (counterfactual definiteness), we can suppose that the properties are not completely pre-determined, but that a hidden variable \( \lambda \) exists and the properties have a probability distribution that is a function of \( \lambda \). The “hidden variable model” hypothesis is weaker than counterfactual definiteness. If the properties are pre-existing, then their probability distribution in \( \lambda \) is trivial: there is a value of \( \lambda \) that determines uniquely the property, e.g., a value \( \lambda_0 \) such that the probability \( P_i(a = 0|A, \lambda_0) = 1 \) and hence \( P_i(a = 1|A, \lambda_0) = 0 \), so it is certain that property \( A \) for object \( i \) has value \( a = 0 \) for \( \lambda = \lambda_0 \),

We can also relax the “Einstein locality” hypothesis by simply requiring that the probability distributions of measurement outcomes factorize (“Bell locality”).4,22,23 Call

\[
P(x, x'|X, X', \lambda) \text{ the probability distribution (due to the hidden variable model) that the measurement of the property } X \text{ on the first object gives result } x \text{ and the measurement of } X' \text{ on the second gives } x' \text{, where } X, X' = A, B, C \text{ denote the three two-valued properties } A, B, \text{ and } C. \text{ By definition, “Bell locality” is the property that the probability distributions of the properties of the two objects factorize, namely}
\]

\[
P(x, x'|X, X', \lambda) = P_1(x|X, \lambda)P_2(x'|X', \lambda).
\]

(A1)

The factorization of the probability means that the probability of seeing some value \( x \) of the property \( X \) for object 1 is independent of which property \( X' \) one chooses to measure and what result \( x' \) one obtains on object 2 (and vice-versa). The “Bell locality” condition (A1) is implied by, and hence weaker than, Einstein locality. In fact, Einstein locality implies that the measurement outcomes at one system cannot be influenced by the choice of which property is measured on a second, distant system. So the probability of the outcomes of the first system \( P_1 \) must be independent of the choice of the measured property of the second system \( X' \), namely \( P_1(x|X, X', \lambda) = P_1(x|X, \lambda) \). The same reasoning applies to the second system, which leads to condition (A1).

Following Ref. 22, we now show that a Bell-local, hidden variable model, together with the requirement that the two systems can have identical property values, implies counterfactual definiteness. This means that we can replace “counterfactual definiteness” with “hidden variable model” in the above proof of Bell’s theorem.

If two objects have the same property values, then \( P_{\text{same}}(X, X) = 1 \) so the probability that a measurement of the same property \( X \) on the two objects gives opposite results (say, \( x = 1 \) and \( x' = 0 \)) is zero. Written as a formula we have

\[
\sum_{\lambda} P(x = 1, x' = 0|X, X, \lambda) p(\lambda) = 0,
\]

(A2)

where the \( \sum_{\lambda} \) emphasizes that we are averaging over the hidden variables (since they are hidden): \( p(\lambda) \) is the probability distribution of the hidden variable \( \lambda \) in the initial (joint) state of the two systems. Note that in Eq. (A2) we are

857
Am. J. Phys., Vol. 81, No. 11, November 2013

Lorenzo Maccone

857
measuring the same property \( X \) on both objects but we are looking for the probability of obtaining opposite results \( x' \neq x \). Using the Bell locality condition (A1), the probability factorizes so Eq. (A2) becomes

\[
\sum_{\lambda} P_1(x = 1 | X, \lambda) P_2(x' = 0 | X, \lambda) p(\lambda) = 0. \tag{A3}
\]

Since \( P_1, P_2, \) and \( p \) are probabilities, they must be positive or zero. Consider the values of \( \lambda \) for which \( p(\lambda) > 0 \). Then the above sum can be zero only if either \( P_1 \) or \( P_2 \) is zero; that is, if \( P_1(x = 1 | X, \lambda) = 0 \) (which implies that \( X \) has the predetermined value \( x = 0 \)) or \( P_2(x' = 0 | X, \lambda) = 0 \) (which means that \( X \) has predetermined value \( x' = 1 \)). We remind the reader that counterfactual definiteness means that \( P_i(x | X, \lambda) \) is either 0 or 1—it is equal to 0 if the property \( X \) of the \( i \)th object does not have the value \( x \), and it is equal to 1 if it does have the value \( x \). We have, hence, shown that Eq. (A3) implies counterfactual definiteness for property \( X \): its value is predetermined for one of the two objects.

Summarizing, if we assume that a Bell-local hidden variable model admits two objects that have the same values of their properties, then we can prove counterfactual definiteness. This means that we can relax the “counterfactual definiteness” and “Einstein locality” hypotheses in the proof of Bell’s theorem, replacing them with the “existence of a hidden variable model” and with “Bell locality” respectively, so that Bell’s theorem takes the meaning that “no Bell-local hidden variable model can describe quantum mechanics.” The hypothesis that two objects can have the same values for the properties is implicit in the fact that such objects exist in quantum mechanics; see Eq. (3). Therefore, if we want to use a hidden variable model to describe quantum mechanics (as in the de Broglie-Bohm interpretation), this model must violate Bell locality. Otherwise, if we want to maintain Bell locality we cannot use a hidden variable model (as in the Copenhagen interpretation).

**APPENDIX B: SUMMARY OF THE HYPOTHESES AND LOGIC OF BELL’S THEOREM**

We have given two different proofs of the Bell inequality, based on different hypotheses. In this appendix we summarize the logic behind the Bell inequality proofs.

The hypotheses we used (rigorously defined above) were:

(A) “Counterfactual Definiteness”

(B) “Einstein locality”

(C) “No super-determinism”

(D) “Measurement independence”

(A’) “Hidden variable model,” implied by (A) and by the fact that systems with the same property values exist (see Appendix A)

(B’) “Bell locality,” implied by (B) (see Appendix A)

In the main text we have proven (Fig. 1) the following theorem:

\[(A) \land (B) \land (C) \land (D) \Rightarrow \text{Bell inequality} \Rightarrow \neg \text{QM} \tag{B1}\]

where by “\( \neg \text{QM} \)” we mean that quantum mechanics (QM) violates the Bell inequality and is, hence, incompatible with it. Using the fact that \( X \land Y \Rightarrow \neg Z \), we can state the above theorem equivalently as

\[\text{QM} \Rightarrow \neg (A) \lor \neg (B) \lor \neg (C) \lor \neg (D). \tag{B2}\]

Since one typically assumes that both (C) and (D) are true, they can be dropped and the theorem can be written more compactly as

\[\text{QM} \Rightarrow \neg (A) \lor \neg (B). \tag{B3}\]

That is (assuming “no super-determinism” and “measurement independence”) quantum mechanics implies that either “counterfactual definiteness” or “Einstein locality” must be dropped. This is the most important legacy of Bell.

We have also seen that the hypotheses (A) and (B) can be weakened somewhat, so that the Bell inequality can also be derived using only (A’) and (B’). That is, we can prove (see Appendix A) that

\[(A') \land (B') \land (C) \land (D) \Rightarrow \text{Bell inequality} \Rightarrow \neg \text{QM}. \tag{B4}\]

In other words (assuming “no super-determinism” and “measurement independence”), quantum mechanics is incompatible with Bell-local hidden variable models.

---

858 Am. J. Phys., Vol. 81, No. 11, November 2013

Lorenzo Maccone 858


Delzenne’s Circle

The Earth Inductor is a flip coil used to measure the magnitude and direction of the earth’s magnetic field. The rotating coil is quickly flipped through 180 degrees, and the output is fed to a ballistic galvanometer. This device, with a long period of oscillation, measures the total charge delivered by the induced EMF; the charge is then proportional to the magnitude of the magnetic field threading the coil. The experiment is done with the coil horizontal (to find the vertical component of the earth’s field) and vertical (to find the horizontal component). The direction and net field can then be readily calculated. The apparatus was devised by Charles Edouard Joseph Delzenne (1776-1866). It is listed at $42.00 in the 1920 Central Scientific Catalogue and resides in the Greenslade Collection. (Notes and photograph by Thomas B. Greenslade, Jr., Kenyon College)