Simplest proof of Bell's inequality - Lorenzo Maccone

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 of what exactly the Bell inequality means, since the hypothesis at the basis of the proof become extremely transparent. It is also a useful didactic tool, as the Bell inequality can be explained in a single intuitive lecture.

1 Introduction

Einstein had a dream. He believed quantum mechanics was an incomplete description of reality [1] 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 refer to 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" [2].

A decade after Einstein's death, John Bell [3, 4] shattered this dream in the worst possible way from Einstein's point of view [4]: 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 non-local. This non-locality is of the worst possible kind: an agent with access to the non-local variables would be able to transmit information instantly to a distant location, thus violating relativistic causality and awakening the nastiest temporal paradoxes [5].

Modern formulations of quantum mechanics must incorporate Bell's result at their core: either they refuse the idea that measurements uncover pre-existing properties, or they must make use of non-local 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 states that the properties arise from the interaction between the quantum system and the measurement apparatus, they are not pre-existing: "unperformed experiments have no results" [6]. An example of the second formulation is the "de Broglie-Bohm interpretation" of quantum mechanics that 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 equilibrium with relativistic causality. It has spawned an impressive amount of research. However, it is often ignored in basic quantum mechanics courses since 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 version [7, 8] of Bell's theorem. The simplicity of the proof is key to clarifying all the theorem's assumptions, the identification of which generated a large debate in the literature (e.g. see [9]).

2 Bell's theorem

Let us define a "local" theory as a one where the outcomes of an experiment on a system are independent of the actions performed on a different system which has no causal connection with the first. For example, the temperature of this room is independent on whether I choose to wear purple socks today. Einstein's relativity provides a stringent condition for causal connections: if two events are outside their respective light cones, there cannot be any causal connection among them.

Let us define a "counterfactual" theory [10, 11] as one whose experiments uncover properties that are pre-existing. In other words, in a counterfactual theory it is meaningful to assign a property to a system (e.g. the position of an electron) independently of whether the measurement of such property is carried out. [Sometime this counterfactual definiteness property is also called "realism", but it is best to avoid such philosophically laden term to avoid misconceptions].

Bell's theorem can be phrased as "quantum mechanics cannot be both local and counterfactual". A logically equivalent way of stating it is "quantum mechanics is either non-local or non-counterfactual".

To prove this theorem, Bell provided an inequality (referring to correlations of measurement results) that is satisfied by all local *and* counterfactual theories. He then showed that quantum mechanics violates this inequality, and hence cannot be local and counterfactual.

All experiments [12] performed to date have shown that Bell inequalities are violated, suggesting that our world cannot be both local and counterfactual. However, it should be noted that no experiment up to now has been able to test Bell inequalities rigorously, 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 (where experimental imperfections are fine-tuned to the properties of the objects [13]) have yet to be ruled out. In the next couple of years the definitive Bell inequality experiment will be performed: many research groups worldwide are actively pursuing it.

If we want to be extremely pedantic in enumerating the hypothesis at the basis of Bell's theorem, we must also request

1. that our choice of which experiment to perform is independent of the properties of the object to be measured (technically, "freedom of choice" or "no super-determinism" [4]): e.g., if we decided to measure the color of red objects only, we would falsely conclude that all objects are red;

2. that future outcomes of the experiment do not influence which apparatus settings were previously chosen [14] (whereas clearly the apparatus settings will influence the outcomes): a trivial causality requirement (technically, "measurement independence").

These two hypothesis are usually left implicit because science would be impossible without them.

3 Proof of Bell's theorem

We use the Bell inequality proposed by Preskill [8], following Mermin's suggestion [7]. Suppose we have two identical objects, namely they have the same properties. Suppose also that these properties are predetermined (counterfactual definiteness) and not generated by their measurement, and that the determination of the properties of one object will not influence any property of the other object (locality).

We will only need three properties A, B, and C that can each take 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 the coin is shiny and $B = 1$ it is dull (property B, texture), and $C = 0$ means the coin is large and $C = 1$ it is small (property C, size).

Suppose I do not know the properties because the two coins are a gift in two wrapped boxes: I only know the gift is two identical coins, but I do not know whether they are two gold, shiny, small coins $(A = 0, B = 0, C = 1)$ or two copper, shiny, large coins $(1, 0, 0)$ or two gold, dull, large coins $(1, 1, 0)$, etc. I do know that the properties "exist" (namely, they are counterfactual and predetermined even if I cannot see them directly) and they are local (namely, acting on one box will not change any property of the coin in the other box: the properties refer separately to each coin). These are quite reasonable assumptions for two coins! My ignorance of the properties is expressed through probabilities that represent either my expectation of finding a property (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 λ [4], that determines the property (frequentist view) [6]. For example, I might say the gift bearer will give me two gold coins with a 20% probability (he is stingy, but not always).

Bell's inequality refers to the correlation among measurement outcomes of the properties: call $P_{same}(A, B)$ the probability that the properties A of the first object and B of the second are the same: A and B are both 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, $P_{same}(A, B) = 1/2$ tells me that with 50% chance $A = B$ (namely they are both 0 or both 1). Since the two coins have equal counterfactual properties, this also implies that with 50% chance I get two gold shiny coins or two copper dull coins. Note that the fact that the two coins have the same properties means that $P_{same}(A,A) = P_{same}(B,B) = P_{same}(C,C) = 1$: if one is made of gold, also the other one will be, or if one is made of copper, also the other one will be, etc.

4 Bell's Inequality [8]

Under the conditions that three arbitrary two-valued properties A, B, C satisfy counterfactual definiteness and locality, and that $P_{same}(X,X) = 1$ for $X = A, B.C$ (i.e. the two objects have same properties), the following inequality among correlations holds,

$$P_{same}(A,B) + P_{same}(A,C) + P_{same}(B,C) \geq 1 \tag{1}$$

namely, a Bell inequality. The proof of such inequality is given graphically in Figure 1 below. The inequality basically says that the sum of the probabilities that the two properties are the same if I consider respectively A and B, A and C, and B and C must be larger than one. This is 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 small, dull and large is greater than one: all the combinations have been counted, possibly more than once. In Figure 2 [15] the events to which the probabilities represented by the Venn diagrams of Figure 1 refer are made explicit. This is true, of course, only if the two objects have same counterfactual properties and the measurement of one does not affect the outcome of the other. If we lack counterfactual 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 properties: 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 the one of the first coin (non-locality) again we cannot infer the first coin's texture from a measurement of the second: even if we know that the initial texture of the coins was the same, the measurement on the second may change such property of the first. The "counterfactual definiteness" hypothesis we used here can be relaxed somewhat, as shown in the appendix.

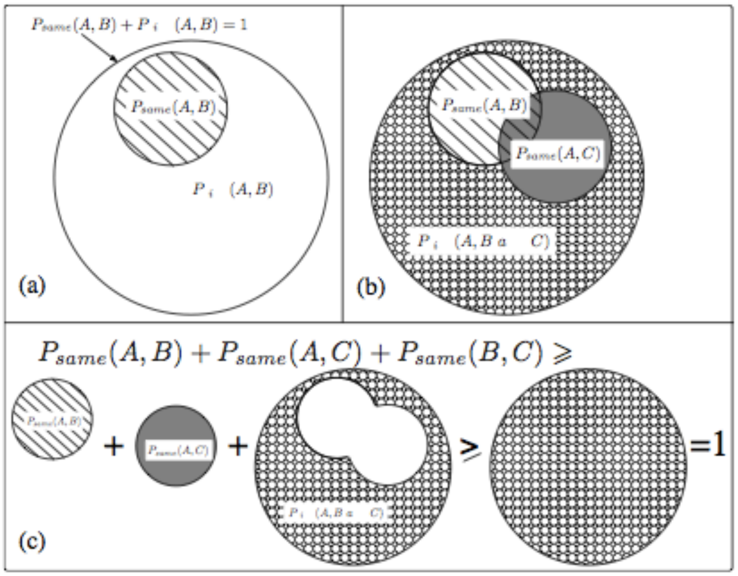
$P_{same}(A,B) + P_{i}\quad(A,B) = 1$
$P_{same}(A,B)$
$P_{i}\quad(A,B)$
(a)
$P_{same}(A,B)$
$P_{same}(A,C)$
$P_{i}\quad(A,B\ a\quad C)$
(b)
$P_{same}(A,B) + P_{same}(A,C) + P_{same}(B,C) \geqslant$
$P_{same}(A,B)$
+
$P_{same}(A,C)$
+
$P_{i}\quad(A,B\ a\quad C)$
$\geqslant$
=1
(c)

Figure 1: Proof of Bell inequality (1) using areas to represent probabilities. (a) The dashed area represents the probability that property A of the first object and B of the second are equal (both 1 or both 0): $P_{same}(A,B)$. The white area represents the probability that they are different: $P_{diff}(A,B)$. The whole circle has area 1 = $P_{same}(A,B) + P_{diff}(A,B)$. (b) The gray area represents the probability that A and C are equal, and the non-gray area represents the probability that A and C are different. If A of the first object is different from both B and C of the second (dotted area), then B and C of the second object must be the same. Hence, the probability that B and C are the same must be larger than (or equal to) the dotted area: since B is the same for the two objects, $P_{same}(B,C)$ must be larger than (or equal to) the dotted area. (c) The quantity $P_{same}(A,B) + P_{same}(A,C) + P_{same}(B,C)$ is hence larger than (or equal to) the sum of the dashed + gray + dotted areas, which is in turn larger than (or equal to) the full circle of area 1: this proves the Bell inequality (1). The reasoning fails if we do not employ counterfactual properties, for example if complementarity prevents us from assigning values to both properties B and C of the second object. It also fails if we employ non-local properties, for example if a measurement of B on an object to find its value changes the value of A of the other object.

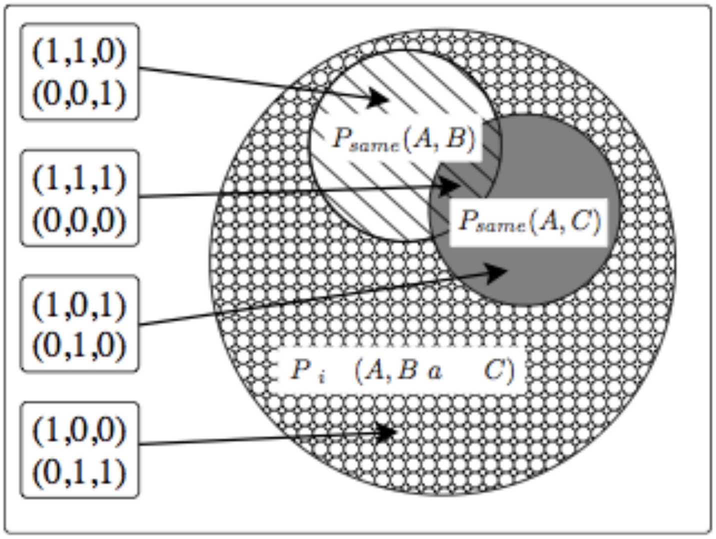

Figure 2: Explicit depiction of the properties whose probabilities are represented by the areas of the Venn diagrams in Figure1. The properties are represented by a triplet of numbers (A, B, C) that indicate the (counterfactual, local) values of the properties A, B, and C for both objects. Note that in the dotted area A must be different from both B and C, so that B and C must be equal there (B and C are equal also in the intersection between the two smaller sets, but that is irrelevant to the proof).

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 obtained by projecting the qubit on the states

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

where it is easy to check that $|b_1\rangle$ is orthogonal to $|b_0\rangle$ and $|c_1\rangle$ is orthogonal to $|c_0\rangle$. It is also easy to check that

$$|\Phi^+\rangle = \frac{|a_0a_0\rangle + |a_1a_1\rangle}{\sqrt{2}} = \frac{|b_0b_0\rangle + |b_1b_1\rangle}{\sqrt{2}} = \frac{|c_0c_0\rangle + |c_1c_1\rangle}{\sqrt{2}} \tag{3}$$

so that the two qubits have the same properties, namely $P_{same}(A, A) =$

$P_{same}(B,B) = P_{same}(C,C) = 1$: the 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 quantity on the left of Bell's inequality (1). Just write the state $|\Phi^+\rangle$ in terms of the eigenstates of the properties A, B, and C. E.g., it is easy to find the value of $P_{same}(A,B)$ if we write

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

In fact, the probability of obtaining zero for both properties is the square modulus of the coefficient of $|a_0\rangle\,|b_0\rangle$, namely, $|1/2\sqrt{2}|^2 = 1/8$, while the probability of obtaining one for both is the square modulus of the coefficient of $|a_1\rangle\,|b_1\rangle$, again $1/8$. Hence, $P_{same}(A,B) = 1/8 + 1/8 = 1/4$. Analogously, we find that $P_{same}(A,C) = 1/4$ and that $P_{same}(B,C) = 1/4$ by expressing the state as

$$|\Phi^+\rangle = \frac{|a_0\rangle\,(|c_0\rangle + \sqrt{3}\,|c_1\rangle) - |a_1\rangle\,(\sqrt{3}\,|c_0\rangle - |c_1\rangle)}{2\sqrt{2}}$$

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

Summarizing, we have found

$$P_{same}(A,B) + P_{same}(A,C) + P_{same}(B,C) = \tfrac{3}{4} < 1 \qquad (4)$$

which violates Bell's inequality (1).

This proves Bell's theorem: all local counterfactual theories must satisfy inequality (1) which is violated by quantum mechanics. Then, quantum mechanics cannot be a local counterfactual theory: it must either be non-counterfactual (as in the Copenhagen interpretation) or non-local (as in the de Broglie-Bohm interpretation).

5 Reference

[1] A. Einstein, B. Podolsky, N. Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev. **47**, 777 (1935).

[2] A. Einstein, in *A. Einstein, Philosopher-Scientist*, ed. by P.A. Schilpp, Library of Living Philosophers, Evanston (1949), pg. 671.

[3] J. S. Bell, *On the Einstein Podolsky Rosen Paradox*, Physics **1**, 195 (1964); Bell J. S., *On the problem of hidden variables in quantum mechanics*, Rev. Mod. Phys. **38**, 447 (1966).

[4] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge Univ. Press, Cambridge, 1987).

[5] "Bell's telephone" in R. Werner, *Quantum Information Theory - an Invitation*, Springer Tracts in Modern Physics 173, 14, (2001), available from http://arxiv.org/pdf/quant-ph/0101061

[6] A. Peres, *Unperformed experiments have no results*, Am. J. Phys. **46**, 745 (1978).

[7] N.D. Mermin, *Bringing home the atomic world: Quantum, mysteries for anybody*, Am. J. Phys. **49**, 940 (1981).

[8] J. Preskill, lecture notes at http://www.theory.caltech.edu/people/preskill/ph229.

[9] G. Auletta, *Foundations and Interpretation of Quantum Mechanics* (World Scientific, Singapore, 2000).

[10] A. Peres, *Existence of "Free will" as a problem of Physics*, Found. Phys. **16**, 573 (1986).

[11] H.P. Stapp, *S-Matrix Interpretation of Quantum Theory*, Phys. Rev. D **3**, 1303 (1971); *Bell's theorem and world process*, Nuovo Cimento **29B**, 270 (1975); *Are superluminal connections necessary?*, Nuovo Cimento **40B**, 191 (1977); *Locality and reality*, Found. Phys. **10**, 767 (1980); W. De Baere, *On Some Consequences of the Breakdown of Counterfactual Definiteness in the Quantum World*, Fortschr. Phys. **46**, 843 (1998).

[12] A. Aspect, P. Grangier, G. Roger, *Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities*, Phys. Rev. Lett. **49**, 91 (1982).

[13] P.M. Pearle, *Hidden-Variable Example Based upon Data Rejection*, Phys. Rev. D **2**, 1418 (1970).

[14] Y. Aharonov, A. Botero, M. Scully, *Locality or non-locality in quantum mechanics: Hidden variables without "spooky action-at-a-distance"*, Z. Natureforsh. **A56**, 5 (2001).

[15] G. Ghirardi, *On a recent proof of nonlocality without inequalities*, Found. Phys. **41**, 1309 (2011), also at arXiv:1101.5252.

6 APPENDIX: Hidden variable models

In the spirit of the original proof of Bell's theorem [4, 15], one can relax the "counterfactual definiteness" hypothesis 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 λ exists and the properties have a probability distribution that is a function of λ. The "hidden variable model" hypothesis is weaker than counterfactual definiteness: if the properties are pre-existing, then their probability distribution in λ is trivial: there is a value of λ that determines uniquely the property, e.g., a value λ_0 such that the probability $P_i(a = 0|A, \lambda_0) = 1$ and hence $P_i(a = 1|A, \lambda_0) = 0$, namely it is certain that property A for object i has the value $a = 0$ for $\lambda = \lambda_0$.

Following [15], we now show that a local, hidden variable model together with the request that the two systems can have identical properties, implies counterfactual definiteness. This means that we can replace "counter-factual definiteness" with "hidden variable model" in the proof of Bell theorem, which, with these relaxed hypothesis states that "no local hidden variable model can rep- resent quantum mechanics".

Call $P(x, x'|X, X', \lambda)$ the probability distribution (due to the hidden variable model) that the measurement of the property X on the first object gives result x and the measurement of X' on the second gives x', where $X, X' = A, B, C$ denote the three two-valued properties A, B, and C. The locality means that the probability distributions of the properties of the two objects factorize, namely $P(x, x'|X, X', \lambda) = P_1(x|X, \lambda)P_2(x'|X', \lambda)$: 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).

If two objects have the same property, then $P_{same}(X,X) = 1$, namely the probability that a measurement of the same property X on the two objects gives opposite results (say, $x = 1$ and $x' = 0$) is null. In formulas,

$$\sum_\lambda P(x = 1, x' = 0|X, X, \lambda)p(\lambda) = 0 \tag{5}$$

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 λ in the initial (joint) state of the two systems. Note that in Eq. (5), we are measuring the same property X on both objects but we are looking for the probability of obtaining opposite results $x' \neq x$. As argued above, locality implies factorization of the probability, namely Eq. (5) becomes

$$\sum_\lambda P_1(x = 1|X, \lambda)P_2(x' = 0|X, \lambda)p(\lambda) = 0 \tag{6}$$

Since P_1, P_2, and p are probabilities, they must be positive. Consider the values of λ for which $p(\lambda) > 0$: the above sum can be *null* only if either P_1 or P_2 is null. Namely if $P_1(x = 1|X, \lambda) = 0$ (which implies that X has the predetermined value $x = 0$) or if $P_2(x' = 0|X, \lambda) = 0$ (which means that X has predetermined value $x' = 1$): we remind 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. (6) implies counterfactual definiteness for property X: its value is predetermined for one of the two objects.

Summarizing, if we assume that a 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" hypothesis in the proof of the Bell theorem, replacing it with the "existence of a hidden variable model", so that the Bell theorem takes the meaning that "no 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)]. Namely, if we want to use a hidden variable model to describe quantum mechanics (as in the de Broglie-Bohm interpretation), such model must be non-local. Otherwise we cannot use a hidden variable model (as in the Copenhagen interpretation).