Why the Fact that Nature Violates Bell Inequalities is So Important

Randy S

Abstract Quantum theory is the foundation for our current understanding of nature. When people first begin to learn quantum theory, they tend to question its necessity, wondering if the same phenomena could be explained using something that feels more like common sense. Such questions are healthy as long as we pay close attention to the experimental facts.

Part of what we might have called *common sense* leads to an inequality called the **CHSH bound**. This article explains how this inequality is derived and then describes a class of experiments in which the CHSH bound is violated. For people whose goal is to understand how nature works, this is one of the most important phenomena ever observed, because it shows in a relatively direct way that part of what we might have called *common sense* cannot be correct.

Contents

1 Preview of the phenomenon **3**

2 Why the phenomenon is important **4**

3 An easy derivation of the CHSH bound **5**

4 A more general derivation of the CHSH bound **6**

5 Derivation of (9) **8**

6 Polarizing filters **9**

7 Polarizing filters applied to single photons **10**

8 Polarizing beamsplitter **11**

9 How to observe violations of the CHSH bound **12**

10 The no-signaling condition **15**

11 Real experiments **16**

12 The obsession with closing loopholes **17**

13 References **18**

14 References in this series **19**

1 Preview of the phenomenon

Physics revolves around what we observe, both through everyday experience and through careful experiments. Some observations are especially important because they challenge assumptions that most other observations don't seem to challenge. This article reviews one of those especially important observations.

An **observable** is something that can be measured. This article is about a phenomenon that involves the simplest type of observable, one whose measurement has only two possible outcomes. This is called a **dichotomic** observable (from the word "dichotomy"). We could label the two possible outcomes "yes" and "no," or "up" and "down," or whatever. To make the math easy, let's label them $+1$ and -1.

Suppose we have four dichotomic observables, A, B, C, D. Suppose that we can measure either A or C, and at the same time we can also measure either B or D, maybe in a different location. If we measure A and B, we get two numbers, both with magnitude $|\pm 1| = 1$. When we multiply these two numbers together, the result is either $+1$ or -1. Let $\langle AB\rangle$ denote the value of this product, averaged over many trials.[1] This average is a number between -1 and $+1$.

In the real world, even if we do our best to prepare the system the same way in every trial and independently of which two observables will be measured, the quantity

$$\Omega(A,B,C,D) \equiv \langle AB\rangle + \langle CB\rangle + \langle CD\rangle - \langle AD\rangle \tag{1}$$

can be as large as $2\sqrt{2} \approx 2.8$. This remains true even after an enormous number of trials, so that statistical errors are negligible.

Sections 9-11 review a class of experiments in which this phenomenon has been demonstrated. Sections 2-5 explain why the phenomenon is important.

[1] We might do our best to prepare the system the same way in every trial, but we still consider the *average* because the measurement outcomes might vary for reasons that are beyond our control.

2 Why the phenomenon is important

The phenomenon described in section 1 is important because it isn't compatible with certain assumptions about the way nature works – assumptions that might seem like common sense. In particular, it violates the **CHSH bound**.[2,3] Section 3 presents an easy derivation of the CHSH bound, and section 4 presents another derivation that assumes even less. The fact that the CHSH bound is *violated* in the real world tells us that the assumptions that go into the CHSH bound, as sensible as they may seem, cannot all be correct.

The derivations in sections 3 and 4 assume that observables have well-defined values whether or not we actually measure them. Quantum theory does *not* assume this. Unlike the assumptions we made above, quantum theory is consistent with the fact that the CHSH bound can be violated in the real world. This article isn't about quantum theory, though. The important message here is simply that the assumptions leading to the CHSH bound are *not* consistent with reality.

[2] The name is derived from the initials of the authors of Clauser *et al* (1969).

[3] The CHSH bound is an example of a **Bell inequality** (Aspect (2002)). Bell inequalities are reviewed in Brunner *et al* (2014).

3 An easy derivation of the CHSH bound

Suppose we have four dichotomic observables, A, B, C, D. For each observable, label its two possible outcomes $+1$ and -1. In any given trial, let a, b, c, d denote the outcomes that would be obtained when A, B, C, D is measured. Each of the numbers a, b, c, d is either $+1$ or -1. The numbers a, b, c, d may vary from one trial to the next, but we're assuming that in any given trial, each of them is either $+1$ or -1. The fact that a, b, c, d all have magnitude 1 implies that the quantity

$$(a + c)b + (c - a)d \tag{2}$$

is always either $+2$ or -2. The individual numbers a, b, c, d may vary from one trial to the next, but this particular combination is always either $+2$ or -2. The average of a bunch of numbers in that range is another number in that range, so if we repeat the experiment many times, the average of the quantity (2) will be between $+2$ and -2. In symbols,

$$-2 \leq \Big\langle (A + C)B + (C - A)D \Big\rangle \leq 2.$$

The average is linear, so we can also write this as

$$-2 \leq \langle AB \rangle + \langle CB \rangle + \langle CD \rangle - \langle AD \rangle \leq 2. \tag{3}$$

This inequality (3) is called the **CHSH bound**.

In the real world, the CHSH bound can be violated (section 1). The arithmetic shown above is correct, so the problem must be in the assumptions. We explicitly assumed that all four observables have definite values in every trial whether or not we measure them, and we implicitly assumed that the act of measurement reveals those values with perfect accuracy. The next section shows that the CHSH bound still holds even if we relax the perfect-accuracy assumption.

4 A more general derivation of the CHSH bound

This section shows that the CHSH bound still holds even if the measurements are not perfectly accurate.[4]

Let S denote everything about the pre-measurement state of the system *except* any information that anticipates which observables we will measure. The outcome of a measurement depends on S, and it may also be affected by the act of measurement itself – maybe because the measurement is not perfectly accurate, or even because the measurement inadvertently changes the value of the thing being measured.

Our procedure for preparing the system (which we repeat at the beginning of each trial) might not produce the same state S every time, because S might involve microscopic details over which we have no control. To accommodate this, let P_S denote the proportion of trials in which our procedure produces the state S. By definition of "proportion," we have

$$\sum_S P_S = 1, \tag{4}$$

where the sum is over all of the states that our procedure might produce.

Even for a given pre-measurement state S, the outcome of a measurement of the observable A may still vary from one trial to the next, because the measurement might not be perfectly accurate and might even change the value of the thing being measured. To accommodate this, let $\langle A\rangle_S$ denote the average of the outcomes of an A-measurement when the pre-measurement state is S. Each individual outcome is either -1 or $+1$, so the average $\langle A\rangle_S$ is a number between -1 and $+1$. Similarly, let $\langle AB\rangle_S$ denote the average of the product of the outcomes of the A and B measurements when the pre-measurement state is S.

Suppose, however, that the act of measuring A or C does not affect the outcome of any measurement of B or D, or conversely. This seems especially reasonable if

[4]This is also derived in section 9.33 in Aspect (2002) and in the text surrounding equation (4) in Brunner *et al* (2014).

the location in which A or C is measured is far away from the location in which B or D is measured at the same time. This implies

$$\langle AB\rangle_S = \langle A\rangle_S\,\langle B\rangle_S, \tag{5}$$

and likewise for the other compatible pairs (CB, CD, and AD). Using this notation, the quantity (1) may be written

$$\Omega(A,B,C,D) = \sum_S \Omega_S(A,B,C,D)\,P_S \tag{6}$$

with

$$\Omega_S(A,B,C,D) \equiv \langle AB\rangle_S + \langle CB\rangle_S + \langle CD\rangle_S - \langle AD\rangle_S. \tag{7}$$

Equation (5) implies

$$\Omega_S(A,B,C,D) = \langle A\rangle_S\langle B\rangle_S + \langle C\rangle_S\langle B\rangle_S + \langle C\rangle_S\langle D\rangle_S - \langle A\rangle_S\langle D\rangle_S. \tag{8}$$

Each of the quantities

$$a \equiv \langle A\rangle_S \qquad b \equiv \langle B\rangle_S \qquad c \equiv \langle C\rangle_S \qquad d \equiv \langle D\rangle_S$$

is a number between -1 and $+1$, and section 5 shows that this implies

$$-2 \leq (a+c)b + (c-a)d \leq 2. \tag{9}$$

Use this in (8) to get

$$-2 \leq \Omega_S(A,B,C,D) \leq 2. \tag{10}$$

Use this together with equations (6) and (4) to get the CHSH bound (3).

The assumptions used in this derivation are sometimes called **Bell locality**,[5] maybe modulo some philosophical nuances that are beside the point here. The point is that *any* derivation of the CHSH bound involves one or more assumptions that must be incorrect, because the CHSH bound can be violated in the real world.

[5] https://plato.stanford.edu/entries/bell-theorem

5 Derivation of (9)

This section derives a lemma that was used in the derivation in section 4. The lemma says that if a, b, c, d are four numbers, each between -1 and $+1$, then the inequality (9) holds. We already deduced this in the case where the four numbers are each *equal* to either -1 or $+1$. Here, we will see that the inequality still holds if the four numbers are anywhere in the interval between -1 and $+1$.

To derive this, we will start with two relatively obvious inequalities. The inequality

$$|x+y| \le |x| + |y| \tag{11}$$

obviously holds for any two real numbers x, y, and the inequality

$$|w| + |z| \le \max\Big(|w+z|,\, |w-z|\Big) \tag{12}$$

obviously holds for any two real numbers w, z. Now define

$$x \equiv (a+c)b \qquad\qquad y \equiv (c-a)d$$

and

$$w \equiv a+c \qquad\qquad z \equiv c-a.$$

With these definitions, the general inequality (11) combined with the conditions $|b| \le 1$ and $|d| \le 1$ implies

$$\Big|x+y\Big| \le |w| + |z|, \tag{13}$$

and the general inequality (12) combined with the conditions $|a| \le 1$ and $|c| \le 1$ implies

$$|w| + |z| \le 2. \tag{14}$$

Combine the inequalities (13) and (14) to get

$$\Big|(a+c)b + (c-a)d\Big| \le 2,$$

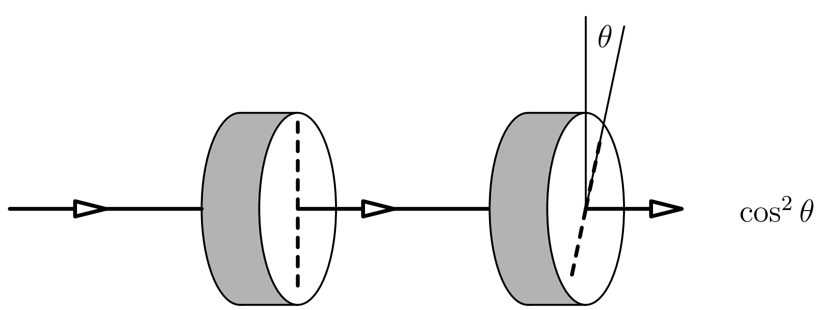
which clearly implies (9). This completes the derivation.

6 Polarizing filters

The phenomenon previewd in section 1 involves dichotomic observables. A measurement of such an observable has only two possible outcomes. One example of a dichotomic observable is the polarization of a single photon. This section introduces the concept of polarization, and sections 7-8 introduce some basic facts about single-photon polarization measurements. These will be used in section 9 to explain how violations of the CHSH bound can be observed.

Polarizing sunglasses are transparent to light that has one linear polarization and are opaque to light that has the orthogonal linear polarization.[6] You can verify this by removing one of the lenses from a pair of polarizing sunglasses and holding it in front of the other lens. The amount of light that passes through the second lens depends on how it is rotated with respect to the first lens.

In more detail: Each lens is a **polarizing filter**. Light emerging from such a filter is linearly polarized. When two filters are applied in sequence, the intensity of the light emerging from the second filter depends on how it is oriented relative to the first filter. Ideally, when the angle between the two filters is θ, the second filter reduces the intensity of the light by a factor of $\cos^2\theta$, which is a number between 0 and 1. This is illustrated here, with each filter drawn as a cylinder, and using a dashed line to indicate the orientation of the polarization that passes through:



[6] https://www.polarization.com/water/water.html

7 Polarizing filters applied to single photons

Now consider what happens when the intensity of the incoming light is reduced so much that individual photons pass through the system one-at-a-time. Empirically, a photon that passes through the second filter has the same energy as it did before it entered the second filter. Either the whole photon passes through, or it none of it passes through.[7]

- If $\theta = 0$, then *all* of the photons that pass through the first filter will also pass through the second filter.[8]
- If $\theta = \pi/2$, then *none* of the photons pass through the second filter: all of the photons are blocked.[8]
- For a generic angle θ, some of the photons that pass through the first filter also pass through the second filter, and some do not.

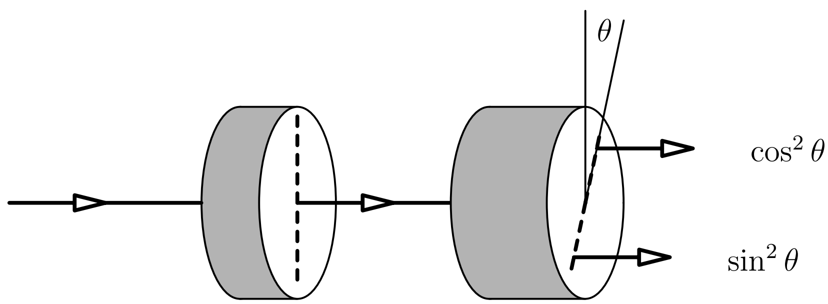
For a generic angle θ, nobody knows how to predict which of the individual photons will pass through the second filter, but we can predict the average number of photons that pass through the second filter. On average, the number of photons that pass through the second filter is $\cos^2\theta$ times the number that entered it, where θ is the angle defined in the preceding diagram. We don't know how to predict the fate of an individual photon, but do know the distribution.

[7]The fact that low-intensity light is detected in the form of discrete bursts of energy (called "photons") was one of the first discoveries that led to the development of quantum theory. The experiment described in section 9 takes advantage of this fundamental property of light.

[8] This is an idealization, assuming perfect filters. This article uses such idealizations to simplify the discussion. Eliminating the idealizations would not change the conclusions in any conceptually significant way.

8 Polarizing beamsplitter

If we don't detect a photon at the output of the second filter, then how do we know that a photon *entered* the second filter at all? To address this, we can replace the second filter by a **polarizing beamsplitter**. A polarizing beamsplitter is transparent to all of the light that enters it,[9] but it separates the incoming light into two distinct outgoing beams. The relative intensities of the outgoing beams depend on the angle θ. This is illustrated below, using a thin cylinder for the filter and a thicker cylinder for the beamsplitter:[10]



Now consider what happens when photons pass through this arrangement one-at-a-time. Downstream, single-photon detectors (not shown in the diagram) count the number of photons emerging from each output. Ideally, ignoring background noise and inefficiencies, each photon is detected either by one detector or by the other – never by both, and never by neither. For a generic value of θ, nobody knows how to predict which photon will be detected at which output, but we can still predict the averages. As illustrated in the diagram, the average numbers are $\cos^2\theta$ and $\sin^2\theta$, respectively, times the number of photons that enter the beamsplitter. The identity $\cos^2\theta+\sin^2\theta=1$ says that none of the photons are blocked: every photon that enters the beamsplitter is detected at one of the two outputs.[11]

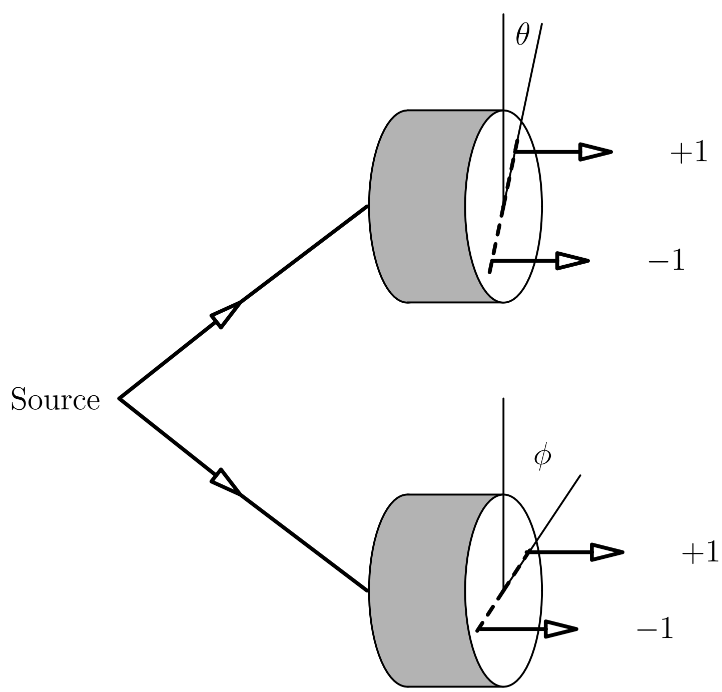
[9] This is an idealization.

[10] A real birefringent crystal doesn't look like this. I'm drawing it this way so I can indicate it's orientation more clearly.

[11] Again, this is an idealization.

9 How to observe violations of the CHSH bound

This section describes a type of experiment that has been done many times by many different researchers.[12] For simplicity, the description presented here will be schematic. The experiment produces paired photons in a special kind of state, using a process called parametric down-conversion.[13] The two photons in each pair travel to two different stations, as illustrated here:



The two stations may be far away from each other. At each station, that photon passes through a polarizing beamsplitter with single-photon detectors at each output. The orientations of the two polarizing beamsplitters are indicated in the diagram by the angles θ and ϕ. The integers $+1$ and -1 are used to label the two

[12]Some examples are cited in section 11.

[13]This happens naturally when a laser of the right frequency is sent through the right kind of crystal (Kwiat *et al* (1995), Couteau (2018), Dehlinger and Mitchell (2002a), Dehlinger and Mitchell (2002b))

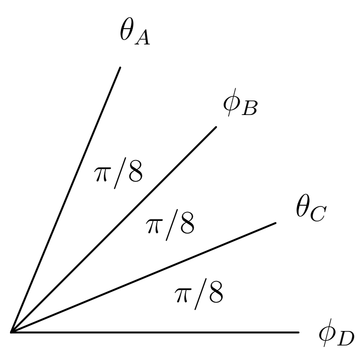
outputs of each polarizing beamsplitter. Each time a pair of photons is detected, one at each station, there are four possible combinations of outcomes:

Outcome at one station	Outcome at other station
$+1$	$+1$
-1	-1
$+1$	-1
-1	$+1$

The distribution of outcomes is shown in this table:

Outcomes	Relative frequency
$+1$ and $+1$	$\frac{1}{2}\cos^2(\theta-\phi)$
-1 and -1	$\frac{1}{2}\cos^2(\theta-\phi)$
$+1$ and -1	$\frac{1}{2}\sin^2(\theta-\phi)$
-1 and $+1$	$\frac{1}{2}\sin^2(\theta-\phi)$

This distribtion violates the CHSH bound (3). To confirm this, consider two different choices for θ, denoted θ_A and θ_C. In any given trial, we may use either one of these settings for θ, but not both. Also consider two different choices for ϕ, denoted ϕ_B and ϕ_D. In any given trial, we may use either one of these settings for ϕ, but not both. Choose these four angles so that they are related as shown here:



The same arrangement may be described in equations like this:

$$\theta_A - \phi_B = \pi/8 \qquad \phi_B - \theta_C = \pi/8$$
$$\theta_C - \phi_D = \pi/8 \qquad \theta_A - \phi_D = 3\pi/8.$$

For a particular pair of settings, say θ_A and ϕ_B, the product of the outcomes is either $+1$ or -1. The experiment generates many photon-pairs, one pair at a time, using the same process each time. The results are used to compute $\langle AB\rangle$, the average value of the product of outcomes when the settings are θ_A and ϕ_B, and similarly for the other pairs of settings. By doing this for different combinations of the settings, we get four averages: $\langle AB\rangle$, $\langle CB\rangle$, $\langle CD\rangle$, and $\langle AD\rangle$. The distribution tabulated on the previous page implies

$$\begin{aligned}
\langle AB\rangle &= \cos^2(\theta_A - \phi_B) - \sin^2(\theta_A - \phi_B) \\
\langle CB\rangle &= \cos^2(\theta_C - \phi_B) - \sin^2(\theta_C - \phi_B) \\
\langle CD\rangle &= \cos^2(\theta_C - \phi_D) - \sin^2(\theta_C - \phi_D) \\
\langle AD\rangle &= \cos^2(\theta_A - \phi_D) - \sin^2(\theta_A - \phi_D).
\end{aligned}$$

For the special angles that were specified above, the identities

$$\cos^2\left(\frac{\pi}{8}\right) - \sin^2\left(\frac{\pi}{8}\right) = \frac{1}{\sqrt{2}} \qquad \cos^2\left(\frac{3\pi}{8}\right) - \sin^2\left(\frac{3\pi}{8}\right) = \frac{-1}{\sqrt{2}}$$

imply

$$\boxed{\langle AB\rangle + \langle CB\rangle + \langle CD\rangle - \langle AD\rangle = 2\sqrt{2}\,,} \tag{15}$$

which violates the CHSH bound, as claimed.[14]

This description is idealized, of course. The imperfections of real experiments prevent the result from being exactly $2\sqrt{2}$, but well-designed experiments come close. Section 11 cites some real examples.

[14] Section 9.4.2 in Aspect (2002) shows that this is the maximum possible violation for the distribution shown on the previous page.

10 The no-signaling condition

The experimental results described in section 9 violate the CHSH bound, but they still satisfy the **no-signaling** condition, which says that we cannot transmit information faster than the speed of light.[15] To see that this distribution satisfies the no-signaling condition, use the identity

$$\frac{1}{2}\cos^2(\theta-\phi)+\frac{1}{2}\sin^2(\theta-\phi)=\frac{1}{2}$$

to see that the distribution of outcomes of a polarization-measurement of the first photon is independent of the angle that is used when measuring the polarization of the second photon, and conversely. In other words, we cannot infer the angle of the second beamsplitter just from the distribution of outcomes at the first beamsplitter, or conversely. The phenomenon described in section 9 is profound, but it cannot be used to achieve faster-than-light communication.

[15]Relativistic quantum field theory includes a postulate called **microlocality**, or **Einstein causality**, or **causal locality**, or sometimes just plain **causality** or **locality**. (References are listed in article 21916.) Don't confuse this with Bell locality (see section 4), a condition which is *not* satisfied in quantum field theory or in the real world. The microcausality postulate enforces the no-signaling condition for realistic observables, but it doesn't quite achieve that goal for some not-so-realistic observables (article 41818).

11 Real experiments

Section 9 described a typical experiment whose results violate the CHSH bound, but that description was idealized. Section 9.9.4 in Aspect (2002) reviews one of the first experiments, reporting 2.697 ± 0.015 for the measured magnitude of (1). Here are some other examples of reported results:

Measured magnitude of (1)	Reference
2.6900 ± 0.0066	Kwiat *et al* (1995)
2.7007 ± 0.0029	Kwiat *et al* (1999)
2.73 ± 0.02	Weihs *et al* (1998)
2.25 ± 0.03	Rowe *et al* (2001)
2.307 ± 0.035	Dehlinger and Mitchell (2002a)
2.56 ± 0.04	Salart *et al* (2008)
2.0732 ± 0.0003	Ansmann *et al* (2009)

All of these results violate the CHSH bound by a large margin, and some of them approach the maximum value allowed by quantum theory, which is[16] $2\sqrt{2} \approx 2.8$.

[16]This is **Tsirelson's bound** (also spelled **Cirelson's bound**). Derivations of this bound are shown in Cirelson (1980), Cabello (2002), and section 2 of Maldacena (2015).

12 The obsession with closing loopholes

Compared to experiments that demonstrate other phenomena, a disproportionately large amount of effort that has gone into closing "loopholes" in experiments like these. Examples:

- The experiment reported in Weihs *et al* (1998) used "sufficient physical distance between the measurment stations, ... ultra-fast and random setting of the analyzers, and ... completely independent data registration."
- The experiment reported in Salart *et al* (2008) used receiving stations separated by 18 kilometers, and the photon detections at each station triggered the displacement of a macroscopic mass in less than the time it would take for light to travel from one station to the other.
- The experiments reported in Rowe *et al* (2001) and Ansmann *et al* (2009) were designed to eliminate the "detection loophole."

Regarding the extensive efforts to close loopholes in such experiments, page 95 in Clauser (2002) says this:

> ...it is logically possible that in such an experiment the two different particle detectors, located on opposite sides of one's laboratory, are conspiratorially communicating with each other, with the specific motive of defeating the experimental test. ...if all experimentalists were similarly paranoid, then experimental physics, in general, would seem to be a pointless endeavor.

Whether or not one thinks that closing loopholes in these experiments is worth any extraordinary effort, the fact that such extraordinary efforts have been made is a testament to the phenomenon's importance. It is important because it rules out a broad class of non-quantum theories.

13 References

(Open-access items include links.)

Ansmann *et al* (2009) "Violation of Bell's inequality in Josephson phase qubits" *Nature* **461**: 504-506

Aspect (2002) "Bell's theorem: The naive view of an experimentalist" Pages 119-154 in *Quantum [Un]speakables: From Bell to Quantum Information*, edited by Bertlmann and Zeilinger (Springer)

Brunner *et al* (2014) "Bell nonlocality" *Rev. Mod. Phys.* **86**: 419, https://arxiv.org/abs/1303.2849

Cabello (2002) "Violating Bell's inequality beyond Cirel'son's bound" *Phys. Rev. Lett.* **88**: 060403, https://arxiv.org/abs/quant-ph/0108084

Cirelson (1980) "Quantum generalizations of Bell's inequality" *Letters in Mathematical Physics* **4**: 93-100, http://www.tau.ac.il/~tsirel/download/qbell80.html

Clauser *et al* (1969) "Proposed experiment to test local hidden-variable theories" *Physical Review Letters* **23**: 880-884

Clauser (2002) "Early history of Bell's theorem" Pages 61-98 in *Quantum [Un]speakables: From Bell to Quantum Information*, edited by Bertlmann and Zeilinger (Springer)

Couteau (2018) "Spontaneous parametric down-conversion" *Contemporary Physics* **59(3)**: 291-304, https://arxiv.org/abs/1809.00127

Dehlinger and Mitchell (2002a) "Entangled photons, nonlocality and Bell inequalities in the undergraduate laboratory" https://arxiv.org/abs/quant-ph/0205171

Dehlinger and Mitchell (2002b) "Entangled photon apparatus for the undergraduate laboratory" https://arxiv.org/abs/quant-ph/0205172

Kwiat *et al* (1995) "New high-intensity source of polarization-entangled photon pairs" https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.75.4337

Kwiat *et al* (1999) "Ultra-bright source of polarization-entangled photons" *Physical Review A* **60**: 773-776, https://arxiv.org/abs/quant-ph/9810003

Maldacena (2015) "A model with cosmological Bell inequalities" https://arxiv.org/abs/1508.01082

Rowe *et al* (2001) "Experimental violation of a Bell's inequality with efficient detection" *Nature* **409**: 791-794

Salart *et al* (2008) "Space-like Separation in a Bell Test assuming Gravitationally Induced Collapses" *Phys. Rev. Lett.* **100**: 220404, https://arxiv.org/abs/0803.2425

Weihs *et al* (1998) "Violation of Bell's inequality under strict Einstein locality conditions" *Phys. Rev. Lett.* **81**: 5039-5043, https://arxiv.org/abs/quant-ph/9810080

14 References in this series

Article **21916** (https://cphysics.org/article/21916):
"Local Observables in Quantum Field Theory"

Article **41818** (https://cphysics.org/article/41818):
"Microcausality, Sorkin's Paradox, and (Un)measurable Observables"