EPR and Bell’s Theorem -- The Ongoing Debate

Mike Ulrey   Bellevue College   May 1, 2018
Why is there still so much activity (papers, conferences, debates) surrounding a paper written in 1964, and going back to ideas from the 1930’s?

https://scholar.google.com/citations?user=Bd9KX0oAAAAJ&hl=en #d=gsc_md_hist&p=&u=

The Vaxjo conferences


Books: “Beyond Quantum”, Andrei Khrennikov,


Andrei Khrennikov is Professor of Mathematics at the Department of Mathematics at Linnaeus University. Andrei is also director of the research group International Center for Mathematical Modeling (ICMM) and organizer of some 20 conferences in the field of quantum theory at Linnaeus University.
One quantum physicist’s opinion --

Christopher Fuchs:
Professor of Physics, University of Massachusetts Boston, 2015–
Research Fellow, Max Planck Institute for Quantum Optics, Garching, Germany, 2014
Senior Researcher, Perimeter Institute for Theoretical Physics, Waterloo, Canada, 2007–2013
Member of Technical Staff, Bell Laboratories, Alcatel-Lucent, Murray Hill, New Jersey, 2000–2007

From “Quantum Mechanics as Quantum Information, Mostly”, Christopher A. Fuchs, Bell Labs, 2002]

“In this paper, I try to cause some good-natured trouble. The issue is, when will we ever stop burdening
the taxpayer with conferences devoted to the quantum foundations? The suspicion is expressed that no
end will be in sight until a means is found to reduce quantum theory to two or three statements of crisp
physical (rather than abstract, axiomatic) significance.”

“The task is not to make sense of the quantum axioms by heaping more structure, more definitions,
more science fiction imagery on top of them, but to throw them away wholesale and start afresh. We
should be relentless in asking ourselves: From what deep physical principles might we derive this
exquisite mathematical structure? Those principles should be crisp; they should be compelling. They
should stir the soul.”

“When I was in junior high school, I sat down with Martin Gardner’s book Relativity for the Million
and came away with an understanding of the subject that sustains me today: The concepts were
strange, but they were clear enough that I could get a grasp on them knowing little more mathematics
than simple arithmetic. One should expect no less for a proper foundation to quantum theory. Until we
can explain quantum theory’s essence to a junior-high-school or high-school student and have them
walk away with a deep, lasting memory, we will have not understood a thing about the quantum foundations.”
Mike’s dream concerning Bell’s ‘Theorem’

To understand the “mapping” between the natural language (words like “locality” and “realism” and “hidden variables” and “counterfactual definiteness”) and the actual physical and mathematical assumptions that Bell made (and others that later extended or generalized the ‘theorem’).

So many questions:
- Do the mathematical assumptions “faithfully” represent the corresponding words? (whatever that means)
- Are the mathematical assumptions and subsequent argument relevant to EPR?
- Are the mathematical assumptions and subsequent argument responsive to Einstein’s concerns?
- Or, is Einstein vindicated in his conviction that QM is incomplete and there needs to be a deeper “sub-quantum” theory that subsumes the existing framework? (sort of like GR subsumes Newton)

This presentation will be focus on Fine’s Theorem (1982), which brings into question whether Bell’s ‘Theorem’ has stupendous physical consequences, or is merely a minor probabilistic result.

If I’m lucky, also cause a little good-natured trouble.
Outline

- Review of EPR
- Review of Bell’s Theorem
- Deep Dive: Bell’s inequality, CHSH, and Fine’s Theorem
- The search for a deeper explanation -- the “sub-quantum” world
Review of EPR
What is the Einstein-Podolsky-Rosen challenge?

-- EPR (Einstein-Podolsky-Rosen paper 1935

“Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?”

The EPR argument: “QM (Predictions of entanglement) + Locality + Reality → QM is incomplete”

In other words, IF quantum theory is true, and IF you can’t transfer information faster than light, and IF you adhere to a commonsense view of reality, THEN there is a deeper, heretofore undiscovered, sub-quantum explanation that needs to be found.
EPR argument in brief

(Based on Sec’s 6.5 & 6.6 of “Quantum Processes, Systems, and Information”, Schumacher and Westmoreland)

EPR Criterion of Reality: If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of reality corresponding to that quantity.

1. What QM can and cannot tell us: For a single qubit*, if the qubit has a definite value for $X$, then its value for $Z$ must be indeterminate, and vice-versa. No quantum state of either qubit in the picture above can be an eigenstate of both the $X$ and $Z$ observables. In particular, Bob’s qubit possesses a physical quantity that is not an element of reality.

1.1. *‘Qubit’ refers to a system represented as a Hilbert space of dimension 2. Quantum analog to a classical ‘bit’.

2. Now suppose Alice’s and Bob’s qubits are part of an “entangled” pair of qubits (for example in a so-called singlet state).

3. Alice can definitely predict (probability 1) either $X$ or $Z$ on Bob’s qubit (but not both at the same time) by performing one measurement or the other on her qubit.

4. We assume Alice’s choice of which observable to measure does not in any way disturb Bob’s qubit. For example, Alice and Bob are separated by a great distance (e.g., earth to Alpha Centauri -- 4.3 LY).

5. By the Criterion of Reality, and (3) and (4), the values for Bob’s $X$ and $Z$ observables must exist as elements of reality before any measurement by Bob.

6. Therefore: Items (1) and (5) are in contradiction. One or the other must go. We choose to keep (5), hence (1) must go. Since we don’t wish to throw out the baby with the bathwater, this only means that quantum theory is somehow incomplete.
Review of Bell’s Theorem
Bell’s inequality and theorem

John Stewart Bell Irish physicist. On leave from CERN in 1964, spending time at Stanford, University of Wisconsin, and Brandeis, wrote “On the Einstein-Podolsky-Rosen Paradox”.

Bell derived an inequality that any local hidden variable theory must satisfy, he claims. He then showed that this inequality is inconsistent with the predictions of QM. He thus turned the EPR argument on its head.

Thesaurus for the correspondences between EPR language and Bell language:

<table>
<thead>
<tr>
<th>EPR</th>
<th>Bell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effects can't travel faster than light</td>
<td>Locality</td>
</tr>
<tr>
<td>Something is missing from current understanding of QM</td>
<td>Hidden variables</td>
</tr>
</tbody>
</table>

The EPR statement: “Entanglement (QM) + Locality \(\implies\) Hidden variables”

Bell’s counter statement: “Hidden variables + Locality \(\implies\) Some predictions inconsistent with QM”

Roughly speaking, this statement is what is known as “Bell’s Theorem”.

Bell’s inequality is indisputable (in the right context). Bell’s theorem is still controversial even today.

Note: EPR had no problem with QM predictions, only with its descriptive power.

Bell’s conclusion: EPR is wrong.

The proof is by contradiction: Bell derives a simple probabilistic inequality based on the hidden variable and locality assumptions. This inequality is contradicted by certain QM theoretical predictions. Later, several experiments seem to confirm this state of affairs (starting in the 1970’s but especially with Aspect et al in 1982).
Derivation of the Bell inequality

(Based on Sec 6.6 of "Quantum Processes, Systems, and Information", Schumacher and Westmoreland)

Imagine Alice and Bob each receive one of two entangled particles in a total spin 0 state, as in the EPR thought experiment (as modified later by David Bohm and others, called EPR-B). Note: The diagram above illustrates a photon spin version of the experiment, not a pair of spin 0 particles, but at the level we are interested in the following description is good enough. See Appendix 5 for more.

- Alice can measure spin components $A_1$ or $A_2$.
- Bob can measure spin components $B_1$ or $B_2$.
- Thus there are four possible joint measurements $(A_1, B_1)$, $(A_1, B_2)$, $(A_2, B_1)$, $(A_2, B_2)$.
- Choose units to make the outcomes $\pm 1$.

The CHSH form of the Bell inequality

- By the hidden variables assumption, $A_1$, $A_2$, $B_1$, $B_2$ all have values every time we do the experiment, even though we only find out some of the values. (counterfactual definiteness)
- By the locality assumption, the value of Alice’s measurement of $A_1$ does not depend on whether Bob measures $B_1$ or $B_2$, for example.
- As a result of these assumptions, we conclude that $B_1 + B_2$ can take on values -2, 0, or +2 and $B_1 - B_2$ can take on values 0, ±2, or 0, respectively
- This allows us to form the observable $S = A_1(B_1 - B_2) + A_2(B_1 + B_2)$ and make certain conclusions that follow below:
- Since $A_i$, $B_i = \pm 1$ for $i = 1, 2$, we have that $S = \pm 2$
- Thus the expected value $\langle S \rangle$ of the random variable $S$ is constrained by $-2 \leq \langle S \rangle \leq 2$,
- Hence, by expanding the expression for $S$ and using the linearity of expectation*, we get:
- $-2 \leq \langle A_1 B_1 \rangle - \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_2 B_2 \rangle \leq +2$, where $\langle A_i B_j \rangle$ denotes the expected value of the product of $A_i$ and $B_j$. 

* Linearity of expectation: $\langle aX + bY \rangle = a\langle X \rangle + b\langle Y \rangle$ for random variables $X$, $Y$ and constants $a$, $b$.
*Note 1: This step is the subject of much controversy. First, the \( A_i \)'s and \( B_j \)'s are random variables conditioned on Alice’s and Bob’s measurement choices. Second, only one of these pairs can be measured on any given trial of the experiment. Hence invoking “linearity of expectation” is highly suspect! Is the “first” \( A_1 \) the same as the “second” \( A_1 \)? It’s possible we should be considering each expectation with respect to different “context-dependent” measures.

*Note 2: In any setting where we can assume that \((A_1, B_1, A_2, B_2)\) have an overall joint probability distribution that is consistent with the given correlations, then we can invoke linearity of expectation.

This is called a CHSH inequality, a specific example of a Bell-type inequality.

Neither Alice nor Bob can determine the value of \( S \) in any single experiment, but they can (jointly) measure any of the products \( A_i B_j \).

By repeating the experiment many times, statistical averages can be computed for each of the terms in the inequality, providing estimates of the expectations.

This produces an experimentally testable statement based on the assumptions of hidden variables and locality.
Deep Dive: Bell’s inequality, CHSH, and Fine’s Theorem
Bell’s Theorem in a nutshell (CHSH formulation):

Two models and some test functions

Bell's Theorem in a nutshell (CHSH formulation):

Two models and some test functions
The space of outcomes for \((A_1, A_2, B_1, B_2)\) is \(S_4 = \{-1, 1\}^4\), consisting of sixteen 4-tuples of -1 and +1’s.

Bell model: Assume that there exists a pmf 
\(q = \{q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_{11}, q_{12}, q_{13}, q_{14}, q_{15}, q_{16}\}\) on \(S_4\) that is consistent with the specified (spin) correlations \(C_{11} = \langle A_1 B_1 \rangle, C_{12} = \langle A_1 B_2 \rangle, C_{21} = \langle A_2 B_1 \rangle, C_{22} = \langle A_2 B_2 \rangle\).

By “pmf” (probability mass function), we mean \(q_k \geq 0\) all \(k\) and \(\sum_k q_k = 1\). Since
\[
\begin{pmatrix}
S_1 \\
S_2 \\
S_3 \\
S_4
\end{pmatrix} =
\begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1
\end{pmatrix}
\begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22}
\end{pmatrix} =
\begin{pmatrix}
2q_1 - 2q_2 - 2q_3 - 2q_4 - 2q_5 - 2q_6 + 2q_7 + 2q_8 + 2q_9 + 2q_{10} - : \\
2q_1 + 2q_2 - 2q_3 - 2q_4 - 2q_5 + 2q_6 - 2q_7 + 2q_8 + 2q_9 - 2q_{10} + : \\
2q_1 - 2q_2 + 2q_3 - 2q_4 + 2q_5 + 2q_6 - 2q_7 - 2q_8 - 2q_9 - 2q_{10} + : \\
2q_1 + 2q_2 - 2q_3 - 2q_4 + 2q_5 - 2q_6 + 2q_7 - 2q_8 - 2q_9 + 2q_{10} - :
\end{pmatrix}.
\]

It is obvious that \(-2 \leq \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix} \leq 2\) by the assumption that the \(q_k\)’s form a pmf.

These 8 inequalities (4 pairs) are called “CHSH inequalities” after Clauser, Horne, Shimony, and Holt.

\textbf{QM Model:} On the other hand, the QM model says the correlations are given by

\[
\begin{pmatrix}
Q_{11} \\
Q_{12} \\
Q_{21} \\
Q_{22}
\end{pmatrix} =
\begin{pmatrix}
< A_1 B_1 > \\
< A_1 B_2 > \\
< A_2 B_1 > \\
< A_2 B_2 >
\end{pmatrix} =
\begin{pmatrix}
\cos 2 \theta_1 \\
\cos 2 \theta_2 \\
\cos 2 \theta_3 \\
\cos 2 \theta_4
\end{pmatrix}.
\]

where the angles \(\theta_k\) are the differences between Alice’s and Bob’s (independently) selected measurement angles.

Note that the angles \(\theta_k\) are not independent -- for example, as labeled in this diagram, \(\theta_1 = \theta_2 + \theta_3 + \theta_4\).

Now consider the test function \(S_1 = -Q_{11} + Q_{12} + Q_{21} + Q_{22}\), for example.

Let the angles be \(\theta_1 = 67.5^\circ\), \(\theta_2 = 22.5^\circ\), \(\theta_3 = 22.5^\circ\), \(\theta_4 = 22.5^\circ\).

Then

\[
S_1 = -Q_{11} + Q_{12} + Q_{21} + Q_{22} = -\cos(2 \times 67.5^\circ) + \cos(2 \times 22.5^\circ) + \cos(2 \times 22.5^\circ) + \cos(2 \times 22.5^\circ) = 2 \sqrt{2} > 2
\]
contradicting the Bell (CHSH) inequality
\[ S_1 = -C_{11} + C_{12} + C_{21} + C_{22} = 2q_1 - 2q_2 + 2q_3 - 2q_4 - 2q_5 - 2q_6 + 2q_7 + 2q_8 + 2q_9 + 2q_{10} - 2q_{11} - 2q_{12} - 2q_{13} + 2q_{14} - 2q_{15} + 2q_{16} \leq 2 \]

Check --
\( (-\cos[2\theta_1] + \cos[2\theta_2] + \cos[2\theta_3] + \cos[2\theta_4]) / . \)
\( \{\theta_1 \rightarrow 67.5 \text{ Degree}, \theta_2 \rightarrow 22.5 \text{ Degree}, \theta_3 \rightarrow 22.5 \text{ Degree}, \theta_4 \rightarrow 22.5 \text{ Degree}\} \)
% ≤ 2

\( 2.82843 \)
\( \text{False} \)
Bell test experiments

Alain Aspect et al (1982)

Alain Aspect has performed numerous beautiful experiments, proving conclusively that our universe violates Bell’s Inequalities big time. And quantum mechanics explains the effects quite nicely.

A source $S$ produces entangled pairs of photons with spin and routes one member of each pair to the “A” wing of the experimental apparatus and the other member of the pair to the “B” wing. These “wings” are separated by some distance, across the lab, across campus, or a few kilometers.

Each photon passes through a polarization analyzer.

There are four possible joint measurements corresponding to angle pairs $(a, b)$, $(a', b')$, $(a', b)$, and $(a, b')$. Angles $a$ or $a'$ are selectable in the “A” wing and angles $b$ or $b'$ are selectable in the “B” wing.

The outputs in each wing are $\pm 1$. These then go to a “Coincidence Monitor” which records a +1 if the two outputs are the same (i.e., ++ or --), and a -1 if they are different (i.e., + - or - +)

The following table shows a possible fragment of the results of a large number $N$ of trials. Note that each row has only one entry. The other entries are not 0, they are just not there, since only one pair can be measured each time.
### Outcomes

<table>
<thead>
<tr>
<th>Trial</th>
<th>Detector settings</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(a,b)</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>(a',b')</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>(a,b)</td>
<td>(a',b')</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sums</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(N₁⁺,N₁⁻)</td>
<td>(N₂⁺,N₂⁻)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correlation estimates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ĉ₁₁=(N₁⁺−N₁⁻)/(N₁⁺+N₁⁻)</td>
<td>Ĉ₁₂=(N₂⁺−N₂⁻)/(N₂⁺+N₂⁻)</td>
</tr>
<tr>
<td>Ĉ₂₁=(N₃⁺−N₃⁻)/(N₃⁺+N₃⁻)</td>
<td>Ĉ₂₂=(N₄⁺−N₄⁻)/(N₄⁺+N₄⁻)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test statistic</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ș₁ = Ĉ₁₁+Ĉ₁₂+Ĉ₂₁+Ĉ₂₂</td>
<td></td>
</tr>
</tbody>
</table>

At the end of the experiment, the data is collected and the correlations are computed together with test functions such as Ș₁, Ș₂, Ș₃, Ș₄ (only Ș₁ is shown in the table as an example).

The results of the test functions are compared to the Bell model and QM model predictions. For example, is |Ș₁| > 2? If so, then this is an indication that the QM model is right and one or more hypotheses from the Bell model must be rejected.
Fine’s Theorem

Arthur Fine:
University of Washington
Professor Emeritus of Philosophy, Adjunct Professor of History, Adjunct Professor of Physics
PhD University of Chicago 1963
His research concentrates on foundations of quantum physics and interpretive issues relating to the development of the natural and social sciences.

Recall the definition of the test functions:

\[
\begin{pmatrix}
S_1 \\
S_2 \\
S_3 \\
S_4
\end{pmatrix} \equiv \begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1
\end{pmatrix}
\]

\[
\begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22}
\end{pmatrix} \equiv \begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1
\end{pmatrix}
\]

\[
\begin{pmatrix}
< A_1 B_1 > \\
< A_1 B_2 > \\
< A_2 B_1 > \\
< A_2 B_2 >
\end{pmatrix}
\]

(1)

and the 8 CHSH inequalities

\[
-2 \leq \begin{pmatrix}
S_1 \\
S_2 \\
S_3 \\
S_4
\end{pmatrix} \leq 2
\]

(2)

Then consider the following proposition:

There exists a joint distribution (pmf) for \((A_1, A_2, B_1, B_2)\) that is consistent with the specified correlations \(C_{13}, C_{14}, C_{23}, \text{ and } C_{24} \iff \text{the 8 CHSH inequalities (2) hold.}\)

(3)

Necessity \((\Longrightarrow)\) is easy and is in fact what Clauser, Horne, Shimony, and Holt showed in their paper (1969?, 1978?, 1982?). Sufficiency \((\Longleftarrow)\) is harder and is what A. Fine showed (1982).
Proof of Necessity for proposition (3) (easy)

Existence of joint distribution (pmf) that is consistent with specified correlations \(\implies\) all 8 CHSH inequalities hold

The space of outcomes for \((\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}_1, \mathcal{B}_2)\) is \((-1, 1)^4\), consisting of sixteen 4-tuples of -1 and +1’s.

Assume that there exists a pmf \(q = \{q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_{11}, q_{12}, q_{13}, q_{14}, q_{15}, q_{16}\}\) on \((-1, 1)^4\) that is consistent with the specified (spin) correlations

\[ C_{11} = \langle A_1 B_1 \rangle, C_{12} = \langle A_1 B_2 \rangle, C_{21} = \langle A_2 B_1 \rangle, C_{22} = \langle A_2 B_2 \rangle. \]

<table>
<thead>
<tr>
<th>(A_1)</th>
<th>(A_2)</th>
<th>(B_1)</th>
<th>(B_2)</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>(q_1)</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>(q_2)</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>(q_3)</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>(q_4)</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>(q_5)</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>(q_6)</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>(q_7)</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(q_8)</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>(q_9)</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>(q_{10})</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>(q_{11})</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>(q_{12})</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>(q_{13})</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>(q_{14})</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(q_{15})</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>(q_{16})</td>
</tr>
</tbody>
</table>

The four (spin) correlations are then given by (see Appendix 2)

\[
\begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22}
\end{pmatrix} = \begin{pmatrix}
q_1 + q_2 - q_3 - q_4 + q_5 + q_6 - q_7 - q_8 - q_9 - q_{10} + q_{11} + q_{12} - q_{13} - q_{14} + q_{15} + q_{16} \\
q_1 - q_2 + q_3 - q_4 + q_5 - q_6 + q_7 - q_8 - q_9 + q_{10} - q_{11} + q_{12} - q_{13} + q_{14} - q_{15} + q_{16} \\
q_1 + q_2 - q_3 - q_4 - q_5 + q_6 + q_7 + q_8 + q_9 + q_{10} - q_{11} - q_{12} - q_{13} + q_{14} + q_{15} + q_{16} \\
q_1 - q_2 + q_3 - q_4 - q_5 - q_6 - q_7 + q_8 - q_9 - q_{10} + q_{11} - q_{12} - q_{13} + q_{14} - q_{15} + q_{16}
\end{pmatrix},
\]

and therefore the four S-test functions of these correlations are given by

\[
\begin{pmatrix}
S_1 \\
S_2 \\
S_3 \\
S_4
\end{pmatrix} = \begin{pmatrix}
q_1 + q_2 - q_3 - q_4 + q_5 + q_6 - q_7 - q_8 - q_9 - q_{10} + q_{11} + q_{12} - q_{13} - q_{14} + q_{15} + q_{16} \\
q_1 - q_2 + q_3 - q_4 + q_5 - q_6 + q_7 - q_8 - q_9 + q_{10} - q_{11} + q_{12} - q_{13} + q_{14} - q_{15} + q_{16} \\
q_1 + q_2 - q_3 - q_4 - q_5 + q_6 + q_7 + q_8 + q_9 + q_{10} - q_{11} - q_{12} - q_{13} + q_{14} + q_{15} + q_{16} \\
q_1 - q_2 + q_3 - q_4 - q_5 - q_6 - q_7 + q_8 - q_9 - q_{10} + q_{11} - q_{12} - q_{13} + q_{14} - q_{15} + q_{16}
\end{pmatrix}.\]
\[
\begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22}
\end{pmatrix} = 
\begin{pmatrix}
2q_1 - 2q_2 + 2q_3 - 2q_4 - 2q_5 - 2q_6 + 2q_7 + 2q_8 + 2q_9 + 2q_{10} - 2q_{11} - 2q_{12} \\
2q_1 + 2q_2 - 2q_3 - 2q_4 - 2q_5 + 2q_6 - 2q_7 + 2q_8 + 2q_9 - 2q_{10} + 2q_{11} - 2q_{12} \\
-2q_1 - 2q_2 + 2q_3 - 2q_4 + 2q_5 + 2q_6 - 2q_7 - 2q_8 - 2q_9 - 2q_{10} + 2q_{11} + 2q_{12} \\
2q_1 + 2q_2 - 2q_3 - 2q_4 + 2q_5 - 2q_6 + 2q_7 - 2q_8 - 2q_9 + 2q_{10} - 2q_{11} - 2q_{12}
\end{pmatrix}
\]

It is obvious that \(-2 \leq \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{pmatrix} \leq 2\) by the assumption that the \(q_k\)'s form a pmf. In other words all 8 CHSH inequalities hold.
Proof of sufficiency for proposition (3) (harder)

Fine’s Theorem: All 8 CHSH inequalities hold \(\implies\) Existence of joint distribution (pmf) that is consistent with the specified correlations

Assume that all 8 CHSH inequalities hold, that is,

\[
-2 \leq \begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1 \\
\end{pmatrix}
\begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22} \\
\end{pmatrix}
= \begin{pmatrix}
-C_{11} + C_{12} + C_{21} + C_{22} \\
C_{11} - C_{12} + C_{21} + C_{22} \\
C_{11} + C_{12} - C_{21} + C_{22} \\
C_{11} + C_{12} + C_{21} - C_{22} \\
\end{pmatrix} \leq 2
\]

Then what we must show is that for any such

\[
\begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22} \\
\end{pmatrix}
\]

there exists a pmf

\[
\{q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_{11}, q_{12}, q_{13}, q_{14}, q_{15}, q_{16}\}
\]

such that

\[
\begin{pmatrix}
q_1 + q_2 - q_3 - q_4 + q_5 + q_6 - q_7 - q_8 - q_9 - q_{10} + q_{11} + q_{12} - q_{13} - q_{14} + q_{15} + q_{16} \\
q_1 - q_2 + q_3 - q_4 + q_5 - q_6 + q_7 - q_8 - q_9 + q_{10} - q_{11} + q_{12} - q_{13} + q_{14} - q_{15} + q_{16} \\
q_1 + q_2 - q_3 - q_4 - q_5 + q_6 + q_7 + q_8 + q_9 + q_{10} - q_{11} - q_{12} - q_{13} - q_{14} + q_{15} + q_{16} \\
q_1 - q_2 + q_3 - q_4 - q_5 + q_6 - q_7 + q_8 + q_9 - q_{10} + q_{11} - q_{12} - q_{13} + q_{14} - q_{15} + q_{16} \\
\end{pmatrix}
= \begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22} \\
\end{pmatrix}
\]

For notational convenience, we will temporarily adopt the following notation for the correlation 4-vectors, i.e.

\[
\begin{pmatrix}
w \\
x \\
y \\
z \\
\end{pmatrix}
= \begin{pmatrix}
C_{11} \\
C_{12} \\
C_{21} \\
C_{22} \\
\end{pmatrix}
\]

Let

\[
I = \{ (w, x, y, z) \in \mathbb{R}^4 \mid -2 \leq \begin{pmatrix}
- w + x + y + z \\
w - x + y + z \\
w + x - y + z \\
w + x + y - z \\
\end{pmatrix} \leq 2 \text{ and } -1 \leq \begin{pmatrix}
w \\
x \\
y \\
z \\
\end{pmatrix} \leq 1 \}.
\]
1. Pick an arbitrary point \( p = (w, x, y, z) \in I \).

2. Claim \( p \) can be written as a convex combination of the rows of \( V = \begin{pmatrix} w & w & 1 & 1 \\ w & -1 & -w & 1 \\ w & 1 & w & 1 \\ w & -w & 1 & 1 \\ w & w & -1 & -1 \\ w & 1 & -w & -1 \\ w & -1 & w & -1 \\ w & w & 1 & -1 \\ w & 1 & -1 & -w \\ w & 1 & 1 & w \\ w & -1 & 1 & -w \\ w & -1 & 1 & w \end{pmatrix} \).

That is, there exists a pmf \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_{12}) \) s.t \( p = (w, x, y, z) = \alpha V \).

Note: When \( w = \pm 1 \) there will only be 4 (distinct) rows, i.e.,

\[
V = \begin{pmatrix} -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix}
\]

when \( w = -1 \) and

\[
V = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \end{pmatrix}
\]

when \( w = +1 \), but we still have \( p = (w, x, y, z) = \alpha V \), but for some pmf \( \alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4) \).
Motivation for the claim

For any fixed $-1 \leq w \leq 1$, define $I_w = \{(x, y, z) \in R^3 \mid -2 \leq \begin{pmatrix} -w + x + y + z \\ w - x + y + z \\ w + x - y + z \\ w + x + y - z \end{pmatrix} \leq 2 \text{ and } -1 \leq \begin{pmatrix} x \\ y \\ z \end{pmatrix} \leq 1 \}.$

Now let’s visualize the sets $I_w$:

Same plot but we explicitly show the coordinates of the 12 (or 4) extreme points of each convex set $I_w$.

Now use the following theorem by Minkowski:
Minkowski’s Theorem (1911): Every compact, convex set in $\mathbb{R}^n$ is the convex hull of its extreme points.

Hermann Minkowski (22 June 1864 – 12 January 1909) was a German mathematician and professor at Königsberg, Zürich and Göttingen. He created and developed the geometry of numbers and used geometrical methods to solve problems in number theory, mathematical physics, and the theory of relativity.

From “Linear Programming” by Cooper and Steinberg:
1. “Any point in a closed, strictly bounded convex set S, with a finite number of extreme points, can be written as a convex combination of extreme points”. (p 50)
2. Here $p$ is an “extreme point” in S if it is NOT POSSIBLE to write $p = \lambda p_0 + (1 - \lambda) p_1$ for two other points $p_0$ and $p_1$ in S and where $0 < \lambda < 1$.
3. Theorem 4-3 (p 57) “The objective function of a linear programming problem assumes its maximum at an extreme point of the convex set of feasible solutions.

Minkowski is all we need, but here is a one generalization and one related result:

Krein-Milman Theorem (1940): Every compact convex subset of a locally convex Hausdorff topological vector space is the closed convex hull of its extreme points.

Mark Grigorievich Krein (3 April 1907 – 17 October 1989) was a Soviet Jewish mathematician, one of the major figures of the Soviet school of functional analysis. He is known for works in operator theory (in close connection with concrete problems coming from mathematical physics), the problem of moments, classical analysis and representation theory.

David Pinhusovich Milman (15 January 1912, Chichelnik near Vinnitsia – 12 July 1982, Tel Aviv) was a Soviet and later Israeli mathematician specializing in functional analysis.[1] He was one of the major figures of the Soviet school of functional analysis. In the 70s he emigrated to Israel and was on the
faculty of Tel Aviv University.

Carathéodory's Theorem (1907): If a point $x \in \mathbb{R}^d$ lies in the convex hull of a set $P$, then $x$ can be written as the convex combination of at most $d + 1$ points in $P$

Constantin Carathéodory (13 September 1873 – 2 February 1950) was a Greek mathematician who spent most of his professional career in Germany. He made significant contributions to the theory of functions of a real variable, the calculus of variations, and measure theory.
Strategy (outline of proof) continued

3. For the $i$th row of $V$, find a pmf $q_i = \{q_{i1}, q_{i2}, q_{i3}, q_{i4}, q_{i5}, q_{i6}, q_{i7}, q_{i8}, q_{i9}, q_{i10}, q_{i11}, q_{i12}, q_{i13}, q_{i14}, q_{i15}, q_{i16}\}$ such that the correlations reproduce that row. For example, for the first row of $V$ find $q_{1,k}$'s such that

\[
q_{1,1} + q_{1,2} - q_{1,3} - q_{1,4} + q_{1,5} + q_{1,6} - q_{1,7} - q_{1,8} - q_{1,9} - q_{1,10} + q_{1,11} + q_{1,12} - q_{1,13} - q_{1,14} + q_{1,15} + q_{1,16} \\
q_{1,1} - q_{1,2} + q_{1,3} - q_{1,4} + q_{1,5} - q_{1,6} + q_{1,7} - q_{1,8} - q_{1,9} + q_{1,10} - q_{1,11} + q_{1,12} - q_{1,13} + q_{1,14} - q_{1,15} + q_{1,16} \\
q_{1,1} + q_{1,2} - q_{1,3} - q_{1,4} - q_{1,5} - q_{1,6} + q_{1,7} + q_{1,8} + q_{1,9} + q_{1,10} - q_{1,11} - q_{1,12} - q_{1,13} - q_{1,14} + q_{1,15} + q_{1,16} \\
q_{1,1} - q_{1,2} + q_{1,3} - q_{1,4} - q_{1,5} + q_{1,6} - q_{1,7} + q_{1,8} + q_{1,9} - q_{1,10} + q_{1,11} - q_{1,12} - q_{1,13} + q_{1,14} - q_{1,15} + q_{1,16}
\]

\[
\begin{pmatrix}
  w \\
  w \\
  1 \\
  1
\end{pmatrix}
\]

Similarly for the rest of the rows.

Hint: One solution as follows. The $i$th row of the following matrix $Q$ is a 16-tuple pmf $q_i$ that reproduces the correlations in the $i$th row of $V$ for $i = 1, \ldots, 12$ (or $i = 1, \ldots, 4$).

\[
Q = \begin{pmatrix}
  \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

for $-1 < w < 1$

and for the two special end point cases:

\[
Q = \begin{pmatrix}
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

for $w = -1$, and
\[ Q = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
\end{pmatrix} \] for \( w = 1 \).

**4.** Let \( q = \alpha Q \). Claim \( q \) is the pmf that reproduces the correlations that comprise the original arbitrary point \( p \).

That is, we have the following correspondence:

*The convex combination of the extreme points that yields the given 4-tuple of correlations is also the convex combination of (extreme point) pmf's that yields the overall pmf on the outcome space that reproduces the given correlations!* More simply put --

\[ p = \alpha V \iff q = \alpha Q. \quad (4) \]
Visualization of this correspondence

Here are plots of a random point and the corresponding convex region of allowable correlations, together with a table of the correlations, random point coefficients, and the required pmf q that yields these correlations. Confirms that these correlations recover the original random point.

Manipulate[Module[{αCoeff, extPts, randomPt, randPtPlot, extPts3D, qExt, qq, corr, subs},
  αCoeff = If[ww ≠ 1 && ww ≠ -1, sumOfRandomReals[0, 1, 12, 1][[1]], sumOfRandomReals[0, 1, 4, 1][[1]]];
  extPts = If[ww ≠ 1 && ww ≠ -1, ptsw4 /. {w -> ww},
    If[ww == -1, extPtsWMinus1Subs1, extPtsWPlus1Subs2]];
  extPts3D = Rest[#] & /@ extPts;
  randomPt = αCoeff.extPts;
  randPtPlot = Graphics3D[{Purple, PointSize[Large], Point[Rest[randomPt]]}];
  qExt = If[ww ≠ 1 && ww ≠ -1, q16Subs1 /. {w -> ww}, If[ww == -1, qWMinus1Subs1, qWPlus1Subs2]];
  qq = αCoeff.qExt;
  corr = correlations /. Table[qk → qq[[k]], {k, 1, Length[qq]}];
  {Show[ConvexHullMesh[extPts3D],
    MeshCellStyle -> {{0, All} → {PointSize[Medium], Red}, {2, All} → {Opacity[0.2]}},
    ViewPoint -> view, AxesLabel -> {"x", "y", "z"}, PlotTheme -> "Detailed"], randPtPlot},
  Grid[{{"Random point p = ", "Correlations \n based on pmf q = ", "Are they equal?",
    "Random Point \n Coefficients α", "pmf q"}, {randomPt // MatrixForm, corr // MatrixForm,
    randomPt == corr, αCoeff // MatrixForm, qq // MatrixForm} // MatrixForm}, Frame -> All]]],
  {ww, 0, "Slice at w"}, {-1, 1}, {{view, {1.3, -2.4, 2}, "Viewpoint"},
  {{1.3, -2.4, 2}, {0, -2, 0}, {0, -2, 2}, {0, -2, -2}, {-2, -2, 0}, {2, -2, 0}, {0, 0, 2}}}]
Slice at \( w \) with Viewpoint \( \{1.3, -2.4, 2\} \)

Random point \( p \):
\[
\begin{pmatrix}
0. \\
-0.0824364 \\
-0.0465162 \\
0.0412725
\end{pmatrix}
\]

Correlations based on pmf \( q \):
\[
\begin{pmatrix}
0. \\
-0.0824364 \\
-0.0465162 \\
0.0412725
\end{pmatrix}
\]

Are they equal? True

Random Point Coefficients:
\[
\begin{pmatrix}
0.158322 \\
0.0738934 \\
0.0643056 \\
0.13376 \\
0.12843 \\
0.0621439 \\
0.0866291 \\
0.111806 \\
0.0542809 \\
0.0118927 \\
0.0512355 \\
0.0633016
\end{pmatrix}
\]

pmf \( q \):
\[
\begin{pmatrix}
0.11726 \\
0.124835 \\
0. \\
0. \\
0.0310719 \\
0.0369467 \\
0. \\
0. \\
0.141725 \\
0.0929211 \\
0.0985307 \\
0.0913553 \\
0.13918 \\
0.126173 \\
0. \\
0.
\end{pmatrix}
\]
Results related to Fine’s Theorem

This result from probability and logic was already presaged by George Boole in 1862 (hence the name ‘Boole-Bell inequality’ is often given to Bell’s inequality)

From Wikipedia: George Boole (/bəʊl/; 2 November 1815 – 8 December 1864) was a largely self-taught English mathematician, philosopher and logician, most of whose short career was spent as the first professor of mathematics at Queen’s College, Cork in Ireland. Aside from his academic work, he was also known to be a precursor to the feminism movement. He worked in the fields of differential equations and algebraic logic, and is best known as the author of The Laws of Thought (1854) which contains Boolean algebra. Boolean logic is credited with laying the foundations for the information age.[3] Boole maintained that:

No general method for the solution of questions in the theory of probabilities can be established which does not explicitly recognise, not only the special numerical bases of the science, but also those universal laws of thought which are the basis of all reasoning, and which, whatever they may be as to their essence, are at least mathematical as to their form.[4]

More complete and general versions of the result were worked out by J Bass (1955) and N. N. Vorobev (1959 in Russian and 1962 in English), before Bell’s paper on EPR in 1964. At the time these were purely probability papers, hence the reason they didn’t immediately affect the quantum physicists’ thinking.


Nikolai Nikolayevich Vorobyov (sometimes spelled Vorobev or Vorob’ev or Vorobedy or Vorobiiev) (Russian: Николай Николаевич Воробьёв, 18 September 1925, Leningrad — July 14, 1995) was a Soviet and Russian mathematician, an expert in the field of abstract algebra, mathematical logic and probability theory, the founder of the Soviet school of game theory. He is an author of two textbooks, three monographs, a large number of mathematical articles and a number of popular science books. He supervised over 30 kandidat and doktor dissertations.[1]
From Russian Academy of Sciences website.


Couldn’t find a picture or biography for J Bass.
Because it brings into question whether or not Bell proved what he claimed.

Bell posited a common hidden variable (or variables) $\lambda$ (or $\Lambda$) on which all random variables depend, thereby *implicitly assuming* that all relevant probabilities and expectations could be referred back to a single Kolmogorovian probability space (i.e., a joint distribution (pmf) on $(A_1, A_2, B_1, B_2)$).

In other words, his argument reduces to a purely mathematical demonstration concerning probabilities and expectations.

See papers by Fine [5], Hess et al [8], and others.

But wait -- does the “hidden variables” assumption somehow *imply* the existence of an overall joint distribution?
More why it is important

Some excerpts from the war between Hess, Philipp et al and Mermin, Leggett, Peres, et al

Karl Hess (born June 20, 1945 in Trumau, Austria) is the Swanlund Professor Emeritus in the Department of Electrical and Computer Engineering at the University of Illinois at Urbana–Champaign (UIUC).[1][2] He helped to establish the Beckman Institute for Advanced Science and Technology at UIUC.

Nathaniel David Mermin (/ˈmɜrmɪn/; born 1935) is a solid-state physicist at Cornell University best known for the eponymous Mermin–Wagner theorem, his application of the term "boojum" to superfluidity, his textbook with Neil Ashcroft on solid-state physics, and for contributions to the foundations of quantum mechanics and quantum information science.

From *The Bell Theorem as a Special Case of a Theorem of Bass*, Karl Hess and Walter Philipp

\[ \text{(p 2)} \text{“We reconsider here these concepts and show that violations of the Bell inequalities have a purely mathematical reason, in particular that the Bell inequalities represent a special case of theorems given earlier by Bass [10], Vorob’ev [11], [12] and Schell [13]. These theorems permit us to deduce that, for all the possible Bell inequalities to be valid, it is a necessary and sufficient condition that the random variables involved in their proof are defined on one common probability space. Beyond this we show that the requirement of the use of one common probability space does not follow from the requirements of objective local spaces and vice versa. In fact, we show that there exist objective local random variables that can not be defined on a common probability space and therefore do not need to obey the Bell inequalities.”} \]

From *Breakdown of Bell’s Theorem for incompatible measurements*, Karl Hess
(p 5) “As we will show below, Bell’s mathematical formulation (BMF) [2] and his parameter $\lambda$ neither prove the existence of nor the necessity to use three or more functions on one common probability space to describe EPR experiments.”

(p 5) “Before we show all of this we emphasize that no actual set of experimental results (such as those of Aspect [18]) can relate directly to all those higher order probability distributions because the corresponding EPR experiments (with two settings on each side) are incompatible and mutually exclusive at a given measurement time or for a given entangled pair. Thus Bell and followers did not and could not possibly have derived the existence of some of the higher order ($\geq 3$) probability distributions by any direct appeal to results of actual experiments. The experimental evidence and the quantum theory for entangled pairs relate only to pair correlations and thus to pairs of random variables and not to the joint probabilities that three or more random variables assume certain values.”
So where does that leave us?

- Is there a “classical” (local realist) explanation of the quantum correlations?
- Was Einstein right that QM is incomplete?
  - From conclusion of Breakdown of Bell’s Theorem for incompatible measurements, Karl Hess
    - “The combination of Bell inequalities and the results of Aspect type experiments does therefore not disprove the EPR hypothesis of the existence of elements of reality plus Einstein locality.”
- What ARE the consequences of Bell’s (and others’) assumptions of locality and hidden variables?
  - From Irrelevance of Bell’s Theorem for experiments involving correlations in space and time: a specific loophole-free computer-example, Hans De Raedt, Kristel Michielsen, Karl Hess
    “A CFD*-compliant model of the EPRB** experiment that incorporates post-selection by a time window can violate the inequality $|S| \leq 2$ but cannot violate Eq. (9)**. Furthermore, with the proper choice of model parameters, this model reproduces the results of the quantum theoretical description of the EPRB experiment in terms of the singlet state. Therefore, we have demonstrated that in the case of the EPRB experiment, CFD does not separate nor distinguish classical from quantum physics. The CFD-compliant model, which may be viewed as having physical time involved in the post-selection process as a hidden variable, provides a counter example to the dogma that CFD implies a Bell-type inequality.”

*CFD = “counter-factual definiteness” and **EPRB = “Einstein-Podolsky-Rosen-Bohm”

**$E(a_1,a_2) - E(a_1,a_2') + E(a_1',a_2) + E(a_1',a_2') \leq 4 - 2\delta$. [9]
The no-communication theorem

- **No-communication theorem**: From Wikipedia -- [https://en.wikipedia.org/wiki/No-communication_theorem](https://en.wikipedia.org/wiki/No-communication_theorem)

  “In physics, the no-communication theorem is a no-go theorem from quantum information theory which states that, during measurement of an entangled quantum state, it is not possible for one observer, by making a measurement of a subsystem of the total state, to communicate information to another observer. The theorem is important because, in quantum mechanics, quantum entanglement is an effect by which certain widely separated events can be correlated in ways that suggest the possibility of instantaneous communication. The no-communication theorem gives conditions under which such transfer of information between two observers is impossible. These results can be applied to understand the so-called paradoxes in quantum mechanics, such as the EPR paradox, or violations of local realism obtained in tests of Bell’s theorem. In these experiments, the no-communication theorem shows that failure of local realism does not lead to what could be referred to as “spooky communication at a distance” (in analogy with Einstein’s labeling of quantum entanglement as “spooky action at a distance”).”

- And isn’t communication of *information* (at a speed faster than light) what Einstein explicitly denied? That is, Einstein locality is not the same as Bell locality, so it seems to me there is no “paradox”.

- Doesn’t this vindicate Einstein’s point of view concerning locality? In other words, given the choice: choose 2 out of 3: QM or locality or realism, we can choose QM and realism, and drop locality in the Bell sense only, thus keeping QM and realism AND Einstein locality. Need to think about this -- is it Bell’s formulation of “realism” or “locality” that can be safely dropped??
The search for a deeper explanation -- the “sub-quantum” world
“When I speak with somebody and get to know his interpretation, I understand immediately it is wrong. The main problem is that I do not know whether my own interpretation is right.” (Theo Nieuwenhuizen)

This is the standard problem of participants of Vaxjo conferences.

Einstein’s Dreams: It is well known (Chapter 1) that Einstein did not believe in irreducible randomness and completeness of QM. He dreamed of a better, so to say “prequantum,” model [79]:

1. Reduction of quantum randomness to classical.
2. Renaissance of causal description.
3. Instead of particles, classical fields will provide a complete description of reality, reality of fields.

Einstein’s Dreams 1 and 3 (but not Dream 2) came true in PCSFT*, a version of CSM** in which fields play the role of particles. In particular, composite systems can be described by vector random fields, i.e., by the Cartesian product of state spaces of subsystems, but not the tensor product. The basic postulate of PCSFT can be formulated in the following way [162]: A quantum particle is the symbolic representation of a “prequantum” classical field fluctuating on the space–time scale which is essentially finer than the space–time scale of present measurements.

*PCSFT = Pre-Quantum Classical Statistical Field Theory; **CSM = Classical Statistical Mechanics

Main message: QM is a version of classical signal theory. It describes noisy and temporary and spatially singular signals.

In CSM a composite system S = (S 1, S 2) is mathematically described by the Cartesian product of state spaces of its parts S 1 and S 2. In QM it is described by the tensor product. The majority of researchers working in quantum foundations, especially quantum information theory, consider this difference in the mathematical representations as crucial. In particular, the entanglement which is a consequence of the tensor space representation is treated as a totally nonclassical phenomenon. However, Einstein was sure that the EPR states encrypt classical correlations generated by a common preparation. In PCSFT Einstein’s dream of the entanglement as a classical correlation will come true.

Space is a huge random wave; quantum systems are spikes on this wave; they are correlated via this space-wave. Thus quantum correlations have two contributions: (1) initial preparation; (2) coupling via the vacuum field. The picture is purely classical. In this model the background field is the source of additional correlations. It seems that this background field is an additional (purely classical) computational resource increasing the power of quantum algorithms. According to PCSFT, quantum computers really have additional computational resource given by the background field.
The Heisenberg’s uncertainty relation is the very heart of QM. In fact, this relation was the main source of the Copenhagen interpretation of QM. Bohr proposed his principle of complementarity after a number of conversations with Heisenberg who was advocating his uncertainty principle. Roughly speaking, the fathers of Copenhagen QM (Bohr, Heisenberg, Pauli, . . . ) were not interested in various no-go theorems (such as, e.g., von Neumann theorem).

For them, it was completely clear that the Heisenberg’s uncertainty relation is the first and final no-go theorem, see Stanford Encyclopedia of Philosophy (article: Uncertainty Principle): ‘One striking aspect of the difference between classical and quantum physics is that whereas classical mechanics presupposes that exact simultaneous values can be assigned to all physical quantities, quantum mechanics denies this possibility, the prime example being the position and momentum of a particle. According to quantum mechanics, the more precisely the position (momentum) of a particle is given, the less precisely can one say what its momentum (position) is.’

In PCSFT approach to the Heisenberg’s uncertainty relation, the presence of the random background (vacuum fluctuations) will play a crucial role. Roughly speaking, all quantum mysteries arise from the ignorance to the presence of these fluctuations. To be more precise, from the ignorance of the fact that QM is a special mathematical formalism elaborated to cancel the contribution of vacuum fluctuations. Although this is a fruitful strategy for dealing with experimental statistical data (because otherwise we would operate with noisy data) one cannot simply forget about these fluctuations. For example, the right interpretation of the Heisenberg’s uncertainty relation can be obtained only by taking vacuum fluctuations into account. The crucial point is that the “quantum dispersion” is not really the statistical dispersion, but it gets rise from ignoring the dispersion of the random background.
Appendices
Appendix 1: Comparing Bell and QM visually

3D slices of correlations plots: $w = -1, -1/2, 0, 1/2, \text{ and } 1$

Row 1: Bell regions: $\{(x, y, z) \in R^3 \mid -2 \leq \begin{pmatrix} -w + x + y + z \\ w - x + y + z \\ w + x - y + z \\ w + x + y - z \end{pmatrix} \leq 2 \text{ and } -1 \leq \begin{pmatrix} x \\ y \\ z \end{pmatrix} \leq 1 \}$.

Row 2: QM contours: $\{(\cos(2(\theta_1 - \theta_3 - \theta_4)), \cos(2\theta_3), \cos(2\theta_4)) \mid \theta_1 = \frac{1}{2} \arccos(w), -90^\circ \leq \theta_3, \theta_4 \leq 90^\circ \}$.

Row 3: Row 1 and Row 2 combined.
Bell and QM test function $S_1$ compared as a function of
measurement angle differences $\theta_3$ and $\theta_4$

The following plot is a visual illustration of Bell’s ‘theorem’ (at least the ‘mathy’ part of it). The green square represents the highest “Bell-achievable” value of 2. The yellow wavy contour represents the QM $S_1$ test function $S_1 = -\cos 2\theta_1 + \cos 2\theta_2 + \cos 2\theta_3 + \cos 2\theta_4$, visualized as a 3D slice with the following parameterization:

$$\begin{align*}
\theta_1 &= \theta_2 + \theta_3 + \theta_4, \\
\theta_3, \theta_4 &\leq 90^\circ
\end{align*}$$

We fix $\theta_2 = 22.5^\circ$ and we have the linear relation $\theta_1 = \theta_2 + \theta_3 + \theta_4$. This leaves just two free parameters. We choose $\theta_3$ and $\theta_4$, which represent the $x$- and $y$-axes. The $z$-axis represents the value of the $S_1$ test function at the corresponding $\theta_3$ and $\theta_4$ values.

Note that any point on the yellow QM contour above the green square represents a violation of the Bell inequality. The maximum possible violation is shown as a purple dot at the highest point $(22.5^\circ, 22.5^\circ, 2\sqrt{2})$.

The black curve of the intersection represents the angle combinations for which BOTH the Bell $S_1$ and the QM $S_1$ are equal to 2.
Appendix 2: Calculation of the Bell model correlations

Overall joint distribution:

<table>
<thead>
<tr>
<th>A₁</th>
<th>A₂</th>
<th>B₁</th>
<th>B₂</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>q₁</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>q₂</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>q₃</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>q₄</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>q₅</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>q₆</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>q₇</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>q₈</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>q₉</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>q₁₀</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>q₁₁</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>q₁₂</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>q₁₃</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>q₁₄</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>q₁₅</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>q₁₆</td>
</tr>
</tbody>
</table>

Marginal distributions of the four relevant pairs:

Finally, distributions of the pair products, together with their expectations, that is the correlations, in the last row of each table:
<table>
<thead>
<tr>
<th>$A_1B_1$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$q_1$</td>
</tr>
<tr>
<td>1</td>
<td>$q_2$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_3$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_4$</td>
</tr>
<tr>
<td>1</td>
<td>$q_5$</td>
</tr>
<tr>
<td>1</td>
<td>$q_6$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_7$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_8$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_9$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_{10}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{11}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{12}$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_{13}$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_{14}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{15}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{16}$</td>
</tr>
<tr>
<td>Correlation</td>
<td>$q_1 + q_2 - q_3 - q_4 + q_5 + q_6 - q_7 - q_8 - q_9 - q_{10} + q_{11} + q_{12} - q_{13} - q_{14} + q_{15} + q_{16}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$A_1B_2$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$q_1$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_2$</td>
</tr>
<tr>
<td>1</td>
<td>$q_3$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_4$</td>
</tr>
<tr>
<td>1</td>
<td>$q_5$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_6$</td>
</tr>
<tr>
<td>1</td>
<td>$q_7$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_8$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_9$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{10}$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_{11}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{12}$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_{13}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{14}$</td>
</tr>
<tr>
<td>-1</td>
<td>$q_{15}$</td>
</tr>
<tr>
<td>1</td>
<td>$q_{16}$</td>
</tr>
<tr>
<td>Correlation</td>
<td>$q_1 - q_2 + q_3 - q_4 + q_5 - q_6 + q_7 - q_8 - q_9 + q_{10} - q_{11} + q_{12} - q_{13} + q_{14} - q_{15} + q_{16}$</td>
</tr>
</tbody>
</table>
### Appendix 2: Visualization of the pmf’s corresponding to extreme points

Here we show the coordinates of the (12 or 4) extreme points (only) for each value of $w$ and check that the computed pmf’s $q = \{q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}, q_{11}, q_{12}, q_{13}, q_{14}, q_{15}, q_{16}\}$ actually reproduce the specified correlations.
Out[
200
]=

Slice at viewpoint \(\{1.3, \ -2.4, 2\}\)

![Diagram](image)

<table>
<thead>
<tr>
<th>Extreme points</th>
<th>Correlations based on (q_k)'s</th>
<th>Are they equal?</th>
<th>(q) matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\begin{pmatrix} 0 &amp; 0 &amp; 1 &amp; 1 \ -1 &amp; 0 &amp; 1 \ 0 &amp; 1 &amp; 0 &amp; 1 \ 0 &amp; 0 &amp; -1 &amp; 1 \ 0 &amp; 0 &amp; -1 &amp; -1 \ 0 &amp; 1 &amp; 0 &amp; -1 \ -1 &amp; 0 &amp; -1 \ 0 &amp; 0 &amp; 1 &amp; -1 \ 0 &amp; 1 &amp; -1 &amp; 0 \ 0 &amp; 1 &amp; 1 &amp; 0 \ 0 &amp; -1 &amp; 1 &amp; 0 \ 0 &amp; -1 &amp; -1 &amp; 0 \end{pmatrix})</td>
<td>(\begin{pmatrix} 0 &amp; 0 &amp; 1 &amp; 1 \ -1 &amp; 0 &amp; 1 \ 0 &amp; 1 &amp; 0 &amp; 1 \ 0 &amp; 0 &amp; -1 &amp; 1 \ 0 &amp; 0 &amp; -1 &amp; -1 \ 0 &amp; 1 &amp; 0 &amp; -1 \ -1 &amp; 0 &amp; -1 \ 0 &amp; 0 &amp; 1 &amp; -1 \ 0 &amp; 1 &amp; -1 &amp; 0 \ 0 &amp; 1 &amp; 1 &amp; 0 \ 0 &amp; -1 &amp; 1 &amp; 0 \ 0 &amp; -1 &amp; -1 &amp; 0 \end{pmatrix})</td>
<td>True</td>
<td>(\begin{pmatrix} 1 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 \ \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 \ \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 \ 0 &amp; 0 &amp; 0 &amp; 0 &amp; 0 &amp; \frac{1}{2} &amp; 0 &amp; 0 &amp; 0 &amp; 0 \end{pmatrix})</td>
</tr>
</tbody>
</table>
Appendix 3: Perils of 3D view of 4D world

Consider the QM point of maximum violation for the test function $S_2$
\[
\left( \cos 2 \times 22.5^\circ, \cos 2 \times 67.5^\circ, \cos 2 \times 22.5^\circ, \cos 2 \times 22.5^\circ \right) = \left( 1/2 \sqrt{2}, -1/2 \sqrt{2}, 1/2 \sqrt{2}, 1/2 \sqrt{2} \right),
\]
making
\[
S_2 = 1/2 \sqrt{2} - (-1/2 \sqrt{2}) + 1/2 \sqrt{2} + 1/2 \sqrt{2} = 2 \sqrt{2}.
\]
Now consider the position of this point relative to some Bell and QM “slices”.

First, let’s compare apples and oranges, that is, Bell and QM “slices” both with $w = 1/2 \sqrt{2}$ . We also show the QM extreme point, or at least the final three coordinates $(x, y, z)$, namely $(−1/2 \sqrt{2}, 1/2 \sqrt{2}, 1/2 \sqrt{2})$, shown as a purple dot. Note that it is exactly on the QM surface, but outside the Bell region.

Now let’s “cheat” and change the Bell region “slice” to $w = -1$. Thus we get a tetrahedron, and it “swallows up” the QM extreme point. See next plot.

Doesn’t this mean that Bell’s model can achieve even the most extreme values, namely, $S_2 = 2 \sqrt{2}$ ? No it does not.

Remember that each $(x, y, z)$ point that we can visualize can be associated with an infinite number of $w$ values (except for the corner cases, where there may be only one $w$ value), as long as all of the test functions of the FULL $(w, x, y, z)$ are not more than 2 in absolute value. So even though it LOOKS like the tetrahedron contains a QM extreme point, it really doesn’t because the $w$ value associated with the $(x, y, z)$ point is $w = -1$, hence $S_2 = -1 - (-1/2 \sqrt{2}) + 1/2 \sqrt{2} + 1/2 \sqrt{2} = 3/2 \sqrt{2} - 1 \approx 1.12132 < 2.$
comparoPlot2

Bell region with \( w = 1 \) and QM contour with \( w = \cos(45^\circ) = 1/\sqrt{2} \)

Note extreme QM point (purple dot) is inside the tetrahedron
Appendix 4: Test functions computed on vertices, edges, and faces

Note: In the interest of keeping notational encumbrances to a minimum, in the following I often take the liberty of identifying a 4D-vector \((w, x, y, z)\) with its 3D counterpart \((x, y, z)\) when discussing vertices, edges, and faces of a 3D object. This does not cause any real problems in this context since every 3D polyhedron is a “slice” through a portion of the 4D polyhedron with \(w\) held fixed. Therefore when I take convex linear combinations of various 4D vectors as if they were 3D vectors, it all comes out in the end since the initial \(w\) remains undisturbed, i.e., \(\sum_k \alpha_k w = w\) whenever the \(\alpha_k\)'s form a pmf.

Define the matrix of test functions:

\[
\begin{pmatrix}
-1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 \\
1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1
\end{pmatrix}
\]

Vertices

The special cases of \(w = \pm 1\):

\[
\begin{pmatrix}
-1 & -1 & -1 & -1 \\
1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1
\end{pmatrix}, \quad \begin{pmatrix}
2 & 2 & -2 & 2 \\
2 & -2 & -2 & -2 \\
-2 & 2 & -2 & -2
\end{pmatrix}
\]

Summary:

For the end cases of \(w = \pm 1\), it is clear that ALL FOUR test functions equal exactly \(\pm 2\) at EVERY vertex of the corresponding tetrahedron. See the second matrix in each of the last two results above -- all \(\pm 2\)'s.

As for the general case \((-1 < w < 1)\), it is more complicated and each test function is only exactly equal to \(\pm 2\) for some of the vertices.

The general case with \(-1 < w < 1\):

The last matrix in the following output represents ALL of the test functions of ALL of the rows of \(V\). In particular, the entry in the \(i\)th row and the \(j\)th column is the \(i\)th test function evaluated at the \(j\)th row of \(V\). Notice that all of the entries are between \(-2\) and \(+2\), since \(-1 \leq w \leq 1\).
By inspection of the second matrix above with \( w \)'s in it, we see
Row 1 (first test function): Exactly +2 for vertices (columns) 1,3,10 and exactly -2 for vertices (columns) 5,7,12

Row 2 (second test function): Exactly +2 for vertices (columns) 1,2,11 and exactly -2 for vertices (columns) 5,6,9

Row 3 (third test function): Exactly +2 for vertices (columns) 3,4,9 and exactly -2 for vertices (columns) 7,8,11

Row 4 (fourth test function): Exactly +2 for vertices (columns) 6,8,10 and exactly -2 for vertices (columns) 2,4,12

Faces

Let's do the two simple cases first: \( w = \pm 1 \). These are the tetrahedrons of course, so there are four faces in each case. In each case, the four faces are obtained simply by deleting exactly one vertex, and the remaining three form one of the faces.
\[\{\alpha, \beta, \gamma\} \cup \{\text{facesWMinus1}\} \cup \{\text{facesWPlus1}\} \]

**Summary:**
For \(w = -1\), the only entries in the first matrix above GUARANTEED to be exactly -2 or +2 are

- (1,4) -2
- (2,1) +2
- (3,3) -2
- (4,2) -2

For \(w = +1\), the only entries in the second matrix above GUARANTEED to be exactly -2 or +2 are

- (1,3) +2
- (2,2) +2
- (3,4) +2
- (4,1) -2

**Edges**

The edges can be obtained by taking all pairs of points (six of them).

\[\text{edgesWMinus1} = \text{Subsets}[\text{ptsWMinus1}, \{2\}];\]
\[\text{edgesWPlus1} = \text{Subsets}[\text{ptsWPlus1}, \{2\}];\]

Now take arbitrary points on each edge (with coefficients \(\alpha, \beta\)) and compute all four test functions in each case. Obviously we get values between -2 and +2 in all cases, but only twice in each row (corresponding to a different test function) is there EXACTLY either a -2 or a +2 (unless of course the values of \(\alpha, \beta\) are such to make the combination equal to -2 or +2).
\( \{\alpha, \beta\}.\# & /@ \text{edgesWMinus1} \)
\( \text{sBell}.\# & /@ \% // \text{MatrixForm} \)
\( \{\alpha, \beta\}.\# & /@ \text{edgesWPlus1} \)
\( \text{sBell}.\# & /@ \% // \text{MatrixForm} \)

\( \text{out} \{1\} \)
\[ \{\{\alpha - \beta, -\alpha + \beta, \alpha + \beta, \alpha - \beta\}, \{\alpha - \beta, -\alpha + \beta, \alpha + \beta, \alpha - \beta\}\} \]
\( \text{out} \{2\} \)
\[ \{\{\alpha + \beta, \alpha - \beta, \alpha - \beta, \alpha + \beta\}, \{\alpha + \beta, \alpha - \beta, \alpha - \beta, \alpha + \beta\}\} \]

\( \text{Bell Theorem Update 4 No Initialization.nb} \)

**Summary:**
For \( w = -1 \), the only entries in the first matrix GUARANTEED to be exactly -2 or +2 are
- (1,3) and (1,4) -2
- (2,3) and (2,4) -2
- (3,1) +2 and (3,3) -2
- (4,2) and (4,4) -2
- (5,1) +2 and (5,2) -2
- (6,2) and (6,3) -2

For \( w = +1 \), the only entries in the second matrix GUARANTEED to be exactly -2 or +2 are
- (1,2) and (1,3) +2
- (2,3) and (2,4) +2
- (3,2) and (3,4) +2
- (4,1) -2 and (4,3) +2
- (5,1) -2 and (5,2) +2
- (6,1) -2 and (6,4) +2
Appèndix 5: Correlation confusion

In the literature, the QM correlations $C_{11}$, $C_{12}$, $C_{21}$, $C_{22}$ may appear different in different contexts, depending on the specific physics of the Bell experiment. There are two traditional experiments, one involving spin-1/2 particles and one involving photon polarization. The following table illustrates the differences in the underlying probabilities as a function of certain specified measurement angles, with the subsequent consequences for the correlations and the test function. The only test function illustrated in this table is $S_2$.

In this paper, we have used the photon polarization values.

**Table of comparisons between elements of photon polarization analysis and spin-1/2 entangled particle analysis**

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$P_{+}(a,b)$</th>
<th>$P_{-}(a,b)$</th>
<th>$P_{-}(a,b)$</th>
<th>Correlation plot</th>
<th>Correlation $S_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Photon Polarization</strong></td>
<td>$\frac{1}{2}\cos^2[(a,b)]$</td>
<td>$\frac{1}{2}\sin^2[(a,b)]$</td>
<td>$\frac{1}{2}\sin^2[(a,b)]$</td>
<td>$\frac{1}{2}\cos^2[(a,b)]$</td>
<td>$\cos[2,(a,b)]$</td>
</tr>
<tr>
<td><strong>Spin-1/2 singlet state</strong></td>
<td>$\frac{1}{2}\sin^2[(a,b)/2]$</td>
<td>$\frac{1}{2}\cos^2[(a,b)/2]$</td>
<td>$\frac{1}{2}\cos^2[(a,b)/2]$</td>
<td>$\frac{1}{2}\sin^2[(a,b)/2]$</td>
<td>$-\cos[(a,b)]$</td>
</tr>
</tbody>
</table>
Appendix 6: On covariance and correlations

Note: In the following we assume all random variables have finite first and second moments (at least).

**Covariance:** \( \text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \)

**Correlation coefficient:** \( \rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \).

Note that \( \text{Cov}(X, Y) = \mathbb{E}[XY] - \mu_X \mu_Y \). Thus \( \text{Cov}(X, Y) = 0 \) if \( X \) and \( Y \) are independent, because \( \mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y] = \mu_X \mu_Y \) in this case. The converse is not true in general.

What is the significance of covariance?

**Note:** In this section, “large” means “large in absolute value”.

Covariance can be thought of as a measure of “coherence” between random variables. That is,

- \( \text{Cov}(X, Y) \) tends to be large and positive when both \( X - \mu_X \) and \( Y - \mu_Y \) tend to be large and positive together OR large and negative together, on average,
- \( \text{Cov}(X, Y) \) tends to be large and negative when \( X - \mu_X \) and \( Y - \mu_Y \) tend to be large and and oppositely positive or negative together, on average,
- \( \text{Cov}(X, Y) \) tends to be close to zero when the products \( (X - \mu_X)(Y - \mu_Y) \) tend to be positive or negative approximately equally often, on average. For example, when \( X - \mu_X \) is large and positive, \( Y - \mu_Y \) is positive or negative equally often, on average.

**Correlation coefficient**

Show that the correlation coefficient \( \rho_{XY} \) satisfies \( |\rho_{XY}| \leq 1 \). Expressed another way,

\[ |\text{Cov}(X, Y)| \leq \sigma_X \sigma_Y, \text{ or } \sigma_{XY} \leq \sigma_X \sigma_Y \text{ where } \sigma_{XY} = |\text{Cov}(X, Y)|. \]

Suppose \( X \) and \( Y \) are r.v.’s with mean 0 and standard deviation 1. Then

\[ 0 \leq \mathbb{E}[(X + Y)^2] = \mathbb{E}[X^2] + 2 \mathbb{E}[XY] + \mathbb{E}[Y^2] = \sigma_X^2 + 2 \mathbb{E}[XY] + \sigma_Y^2 = 2(1 + \mathbb{E}[XY]) \implies \mathbb{E}[XY] \geq -1, \text{ and} \]

\[ 0 \leq \mathbb{E}[(X - Y)^2] = \mathbb{E}[X^2] - 2 \mathbb{E}[XY] + \mathbb{E}[Y^2] = \sigma_X^2 - 2 \mathbb{E}[XY] + \sigma_Y^2 = 2(1 - \mathbb{E}[XY]) \implies \mathbb{E}[XY] \leq +1. \]

Then for any two r.v.’s \( X \) with mean \( \mu_X \) and standard deviation \( \sigma_X \) and \( Y \) with mean \( \mu_Y \) and standard deviation \( \sigma_Y \), we have that \( U = \frac{X - \mu_X}{\sigma_X} \) and \( V = \frac{Y - \mu_Y}{\sigma_Y} \) have mean 0 and standard deviation 1, so that

\[ |\mathbb{E}[UV]| \leq 1 \implies |\mathbb{E}[\frac{X - \mu_X}{\sigma_X} \frac{Y - \mu_Y}{\sigma_Y}]| \leq 1 \implies |\frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}| \leq 1 \implies |\frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}| \leq 1 \implies |\rho_{XY}| \leq 1. \]

The correlation coefficient is convenient because, compared to the covariance,

- It is dimensionless.
- It is insensitive to scale, since product of the the standard deviations in the denominator scale with the covariance in the numerator.
- It is confined to the interval \([-1,1]\), with an intuitive interpretation of “size”.

Relation between \( \mathbb{E}[UV] \) and \( \mathbb{E}[XY] \) when \( U = (X - \mu_X) / \sigma_X \) and \( V = (Y - \mu_Y) / \sigma_Y \).
Now \( E[U \ V] = E[\frac{X-\mu_X}{\sigma_X} \ \frac{Y-\mu_Y}{\sigma_Y}] = E[(X-\mu_X) (Y-\mu_Y)] / (\sigma_X \ \sigma_Y) = (E[XY] - \mu_X \mu_Y) / (\sigma_X \ \sigma_Y) \), so

We have \( E[XY] = \sigma_X \ \sigma_Y E[U \ V] + \mu_X \mu_Y \) so as \( E[U \ V] \) goes from -1 to +1, \( E[XY] \) goes from \( \mu_X \mu_Y - \sigma_X \ \sigma_Y \) to \( \mu_X \mu_Y + \sigma_X \ \sigma_Y \).

Covariance is invariant to translation and scales with product of scaling coefficients of \( X \) and \( Y \)

\[
\text{Cov}(a \ X + b, \ c \ Y + d) = E[(a \ X + b - a \mu_X - b) (c \ Y + d - c \mu_Y - d)] = a \ c \text{Cov}(X, Y)
\]

Linear relationship: If \( Y \) is a linear function of \( X \): \( Y = c \ X + d \), then \( \rho_{XY} = \pm 1 \)

In particular, if \( Y = c \ X + d \), with \( c > 0 \), we have

\[
\rho_{c \ X, X} = \frac{c \ \sigma_Y}{\sigma_X} \frac{c \ \sigma_X}{\sigma_X} = \rho_{c \ X, X} = 1.
\]

If \( c < 0 \), we have

\[
\rho_{c \ X, X} = \frac{c \ \sigma_Y}{\sigma_X} \frac{(-c) \ \sigma_X}{\sigma_X} = -1.
\]

Converse: If \( \rho_{XY} = \pm 1 \), then \( Y \) is a linear function of \( X \): \( Y = c \ X + d \) for some real \( c, \ d \)

As a hint at a possible proof, notice that

\[
\text{Cov}(X, Y) = \text{Cov}(X, \frac{\sigma_Y}{\sigma_X} X + d) = \text{Cov}(X, \frac{\sigma_Y}{\sigma_X} X) = \frac{\sigma_Y}{\sigma_X} \text{Cov}(X, X) = \frac{\sigma_Y \ \sigma_X}{\sigma_X} = \sigma_X \ \sigma_Y = \rho_{XY} \ \sigma_X \ \sigma_Y \text{ when} \ \rho_{XY} = 1.
\]

Assume \( \rho_{XY} = +1 \). We want to show \( P(Y = c \ X + d) = 1 \) for some real \( c > 0, \ d \). Let \( c = \frac{\sigma_Y}{\sigma_X} \) and \( d \) be any real number.

Now define (for any real number \( d \)):

\[
Z = Y - \frac{\sigma_Y}{\sigma_X} X - d
\]

(5)

Then

\[
\mu_Z = \mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X - d
\]

(6)

and
\[
\sigma^2_Z = E[(Y - \frac{\sigma_Y}{\sigma_X} X - d - \mu_Y + \frac{\sigma_Y}{\sigma_X} \mu_X + d)^2] = E[(Y - \frac{\sigma_Y}{\sigma_X} (X - \mu_X) - \mu_Y)^2] = E[(Y - \mu_Y) - (X - \mu_X)]
\]

\[
= E[(Y - \mu_Y)^2] + \left(\frac{\sigma_Y}{\sigma_X}\right)^2 E[(X - \mu_X)^2] - 2 \frac{\sigma_Y}{\sigma_X} E(Y - \mu_Y)(X - \mu_X)
\]

\[
= \sigma_Y^2 + \left(\frac{\sigma_Y}{\sigma_X}\right)^2 \sigma_X^2 - 2 \frac{\sigma_Y}{\sigma_X} \text{Cov}(X, Y)
\]

\[
= 2 \sigma_Y^2 - 2 \frac{\sigma_Y}{\sigma_X} \sigma_X \sigma_Y
\]

\[
= 2 \sigma_Y^2 - 2 \sigma_Y^2 = 0
\]

In summary
\[
\sigma^2_Z = \sigma_Y^2 - \frac{\sigma_Y}{\sigma_X} X - d = 0 \tag{7}
\]

hence
\[
P\left(Y = \frac{\sigma_Y}{\sigma_X} X + d\right) = 1 \tag{8}
\]

The case for \(\rho_{XY} = -1\) is similar with the appropriate sign change(s).

**Special case for r.v.'s that take values \pm 1 only -- why is \(-1 \leq \text{Cov}(X, Y) \leq 1\) without the normalizing denominator \(\sigma_X \sigma_Y\)?**

Now suppose that \(X\) is a r.v. with \(P(X = 1) = p = 1 - P(X = -1)\). Then \(E X = 2p - 1\) and \(\sigma_X = 2\sqrt{p(1-p)}\). Thus we have
\[
|E X| \leq 1 \text{ and } 0 \leq \sigma_X \leq 1.
\]

See the plot below. Note this is an ellipse since we can re-arrange \(y = 2\sqrt{1-x}\) into \(\left(\frac{x-1}{2}\right)^2 + \left(\frac{y}{1}\right)^2 = 1\)

which is an ellipse with center at \((1/2, 0)\) and \(a = 1/2, \ b = 1\).

Now the spin variables \(A_1, A_2, B_1, B_2\) and their products \(A_i B_j\) all fall into this category, that is, taking only the two values -1 and +1. Hence the their expectations \(2p - 1\) for some \(0 \leq p \leq 1\) are all between -1 and +1. In other words, whatever the structure is that determines the pmf's for the \(A_1, A_2, B_1, B_2\)
r.v.’s, the products $A_i B_j$ can only take values $\pm 1$, so of course the expectation will be between -1 and +1 inclusive. 
(Note we didn’t really use the facts about the standard deviations here.)

Recall:

$$\text{Cov}(A, B) = E[(A - \mu_A)(B - \mu_B)] = E[AB] - \mu_A \mu_B$$

or $E[AB] = \text{Cov}(A, B) + \mu_A \mu_B$

Set up the necessary notation.

```math
\text{Set up the basic outcome spaces and associated probabilities.}
```

```math
\text{In[227]:=} outcomes = Tuples[{-1, 1}, 2]
\text{probs = Array[q # &, 4]}
\text{outcomesAndProbs = Table[outcomes[[k]], probs[[k]]], \{k, 1, Length[outcomes]\}}
\text{cond1 = And @@ (![\#] > 0 & /@ probs)}
\text{cond2 = Total[probs] >= 1}
```

```math
\text{Out[227]= {\{-1, -1\}, \{-1, 1\}, \{1, -1\}, \{1, 1\}}} 
\text{Out[228]= \{q_1, q_2, q_3, q_4\}} 
\text{Out[229]= \{\{-1, -1\}, q_1\}, \{\{-1, 1\}, q_2\}, \{(1, -1), q_3\}, \{(1, 1), q_4\}} 
\text{Out[230]= q_1 >= 0 && q_2 >= 0 && q_3 >= 0 && q_4 >= 0} 
\text{Out[231]= q_1 + q_2 + q_3 + q_4 = 1} 
```

```math
\text{Out[229]= {{{1}, q_1}, {{1}, q_2}, {{1}, q_3}, {{1}, q_4}}} 
\text{Out[230]= {{{1}, q_1}, {{1}, q_2}, {{1}, q_3}, {{1}, q_4}}} 
\text{Out[231]= {{{1}, q_1}, {{1}, q_2}, {{1}, q_3}, {{1}, q_4}}} 
\text{Out[232]= {{{1}, q_1}, {-1, q_2}, {{1}, q_3}, {{1}, q_4}}} 
\text{Out[233]= {{{1}, q_1}, {-1, q_2}, {-1, q_3}, {1, q_4}}} 
\text{Out[234]= \{\{1, q_1\}, \{-1, q_2\}, \{1, q_3\}, \{1, q_4\}\}} 
\text{Out[235]= \{\{1, q_1\}, \{-1, q_2\}, \{-1, q_3\}, \{1, q_4\}\}} 
```

```math
\text{Compute marginal expectations and variances and one covariance.}
```
Max and Min for Cov(A,B). ±1 as expected.

Maxes and Mins for σ₂ₐ and σ₂₂ₐ. Why are the Min's = 1?? They should be 0. FindMaximum and FindMinimum only find local maxes and mins starting from a system-chosen point. I had to give it a little clue to find one of the global minimums of ρ.

ListPlots of the three parameters {σ₂ₐ, σ₂₂ₐ, ρₐₐ} for many randomly chosen joint probabilities {q₁, q₂, q₃, q₄} for the outcomes in [-1, 1]^4. Note the variances are in [0,1] and the correlation coefficients are in [-1,1], as expected.
Chop[FindMaximum[[covAB/Sqrt[sigmaSquaredA sigmaSquaredB], cond1 && cond2], probs], 10^-7]

Chop[FindMinimum[[covAB/Sqrt[sigmaSquaredA sigmaSquaredB], cond1 && cond2], probs], 10^-7]

{1., {q1 -> 0.5, q2 -> 0, q3 -> 0, q4 -> 0.5}}

{-1., {q1 -> 0, q2 -> 0.5, q3 -> 0.5, q4 -> 0}}

So finally, to answer my original question: How can $E[AB]$ and $\rho_{AB} = \frac{\text{Cov}(A,B)}{\sigma_A \sigma_B}$ simultaneously be measures of a meaningful “correlation” between -1 and 1?

$E[AB] = \sigma_A \sigma_B \rho_{AB} + \mu_A \mu_B = \text{Cov}(A, B) + \mu_A \mu_B$

Are they equal at the degenerate pmf’s? YES.

$\{\text{expectAB, covAB + expectA expectB} / . # & @\}

Table[qk -> Permutations[{1, 0, 0, 0}] ([[j, k]], [j, 1, 4], [k, 1, 4])]

{1, 1, -1, -1, -1, 1, 1}

Are they equal at some random pmf’s? YES.

$\text{randQ1 = sumOfRandomReals[0, 1, 4, 10]}$

$\text{tabQ1 = Table[qk -> randQ1[[j, k]], [j, 1, Length[randQ1]], [k, 1, 4]];}

\{\text{expectAB, covAB + expectA expectB} / . # & @\}

$\text{tabCompare = \{expectAB, covAB/Sqrt[sigmaSquaredA sigmaSquaredB]\} / . # &@\}

Now let’s compare $E[AB]$ and $\rho_{AB}$ at random points. They are not equal but all between -1 and +1 as expected.

$\text{randQ2 = sumOfRandomReals[0, 1, 4, 1000];}

\text{tabQ2 = Table[qk -> randQ2[[j, k]], [j, 1, Length[randQ2]], [k, 1, 4]];}

\text{tabCompare = \{expectAB, covAB/Sqrt[sigmaSquaredA sigmaSquaredB]\} / . # & @}\$

How about a ListPlot comparison...these correlation statistics seemed to be correlated! Hmmm...sort of a regression line...interesting....
Appendix 7: More on Fine’s Theorem: From Hess and Philipp [8]

“Consider Bell’s original proof [4]. Here Bell assumes that all random variables A,B,C are in turn functions of a single random variable Λ. Then it is clear that A,B,C are defined on one common probability space and therefore the inequalities can not be violated by the pair expectation values as explained above. It is clear that no Λ can exist that leads to a violation of the inequalities for purely mathematical reasons as already found by Bass much earlier. Bell’s physical justification is wanting because he attempts to show that the inequalities follow from the fact that Λ does not depend on the settings a,b,... In fact, it does not matter on what Λ depends as long as the resulting A,B and C are random variables defined on one probability space. We will discuss this in more detail below.

Other well known proofs [5] invoke “counterfactual” reasoning of the following kind: If, for example, A is measured given a certain information that we denote by λ (a value that Λ assumes in a given experiment) and that is carried by the correlated spin pair, then one could have measured with another setting, say b and the same λ. As we have explained in more detail previously [18], it is permissible to ask the question of what would have been obtained if the measurement had been performed with a different setting. It is also permissible to hypothesize the existence of an element of reality related to that different setting if that different setting had been chosen. However, to assume then, as is always done in Bell type proofs, that all these possible different measurement results are actually contained in the data set of actual outcomes of the idealized experiment is arbitrary and against all the rules of modelling and simulation especially for the particular case of the Aspect-type experiment and all other known EPR experiments [18].

Naturally, we do not have to pay for all items on a restaurant’s menu just because we could have chosen them. We call this latter assumption the extended counterfactual assumption (ECA). ECA is equivalent to the assumption that A,B and C only depend on one random variable Λ and is therefore an assumption, not a proof. As a consequence, ECA implies that A,B and C are defined on one common probability space. In view of the Bass-Vorob’ev theorem it leads to a contradiction from the outset irrespective and independent of any physical considerations.”
Appendix 8. Derivation of the Bell inequality

Imagine Alice and Bob each receive one of two entangled particles in a total spin 0 state, as in the EPR thought experiment.

Alice can measure spin components \( A_1 \) or \( A_2 \).

Bob can measure spin components \( B_1 \) or \( B_2 \).

Thus there are four possible joint measurements \((A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2)\).

Let \( P(X = Y) \) denote the probability that the \( X \) and \( Y \) measurements give the same result, either both \( \pm \hbar/2 \).

By the hidden variables assumption, \( A_1, A_2, B_1, B_2 \) all have values every time we do the experiment, even though we only find out some of the values.

By the locality assumption, the value of Alice’s measurement of \( A_1 \) does not depend on whether Bob measures \( B_1 \) or \( B_2 \), for example.

Note that \((A_1 = B_1) \land (A_2 = B_2) \land (B_1 = A_2) \implies (A_1 = B_2)\), so 
\[ P(A_1 = B_1) \land (A_2 = B_2) \land (B_1 = A_2) \leq P(A_1 = B_2) , \]
and thus (by one of de Morgan’s laws and simple probability), \( P(A_1 \neq B_1) \lor (A_2 \neq B_2) \lor (B_1 \neq A_2) \geq P(A_1 \neq B_2) = 1 - P(A_1 = B_2) \)

Assume: \( P(A_1 = B_1) = P(A_2 = B_2) = 0.85 \) and \( P(B_1 = A_2) = 1 \). Then
\[ P(A_1 = B_2) \geq 1 - P(A_1 \neq B_1) \lor (A_2 \neq B_2) \lor (B_1 \neq A_2) \geq 1 - P(A_1 \neq B_1) - P(A_2 \neq B_2) - P(B_1 \neq A_2) = 1 - 0.15 - 0.15 - 0 = .70 \]

But QM predicts that if we create two spins in a total spin 0 state, then the probability of agreement between two spin measurements, where \( \alpha \) is the angle between the measurement axes, is given by (Note: \textit{We will derive this relationship in general on the next few slides}): 

<table>
<thead>
<tr>
<th>( \alpha )</th>
<th>0</th>
<th>45°</th>
<th>90°</th>
<th>135°</th>
<th>180°</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of agreement</td>
<td>0</td>
<td>0.15</td>
<td>0.5</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

We choose 4 spin axes so that \( \angle A_1 B_1 = 135^\circ, \angle A_2 B_1 = 180^\circ, \angle A_2 B_2 = 135^\circ \) and \( \angle A_1 B_2 = 90^\circ \)

Thus our assumptions are satisfied: \( P(A_1 = B_1) = P(A_2 = B_2) = 0.85 \) and \( P(B_1 = A_2) = 1 \), so we get \( P(A_1 = B_2) \geq .70 \)

But as seen in the table above for a 90° angle, QM predicts that \( P(A_1 = B_2) = .50 < .70 \), hence the Bell inequality is violated!
Appendix 9: The CHSH form of the Bell inequality


(Based on Sec 6.6 of "Quantum Processes, Systems, and Information", Schumacher and Westmoreland)

Assume (more or less in line with the EPR point of view)

- **Hidden variables**: The results of any measurement on any individual system are pre-determined. Any probabilities we may use only reflect our ignorance of the (hidden) definite values.

- ** Locality**: Alice’s choice of measurement does not affect the outcomes of Bob’s measurements, and *vice versa*.

  - The statement “The value of $B_1$ is +1” means “If Bob were to measure $B_1$, then the result +1 would be obtained”, regardless of whether Alice chooses to measure $A_1$ or $A_2$. (Note Alice’s *outcome* may affect Bob’s probabilities, but this does not allow Alice to send a message to Bob (also see the “no-communication” theorem)).

  - This is where the hidden variables and locality assumptions come in.

  - Hidden variables: All four measurements $A_1$, $A_2$, $B_1$, $B_2$ have definite values that are predetermined, that is, exist before any measurement is made.

  - Locality: Alice’s choice of measurement does not affect Bob’s outcome, and vice-versa.

  - As a result of these assumptions, we conclude that $B_1 + B_2$ can only take on values -2, 0, or +2 and $B_1 - B_2$ can take on values 0, ±2, or 0, respectively.

- This allows us to form the observable $S = A_1(B_1 - B_2) + A_2(B_1 + B_2)$ and make certain conclusions that follow below:

  - Since $A_i, B_i = ±1$ for $i = 1, 2$, we have that $S = ±2$

  - Thus the expected value $\langle S \rangle$ of the random variable $S$ is constrained by $-2 \leq \langle S \rangle \leq 2$,

  - Hence, by expanding the expression for $S$ and using the linearity of expectation*, we get:

    - $-2 \leq \langle A_1 B_1 \rangle - \langle A_1 B_2 \rangle + \langle A_2 B_1 \rangle + \langle A_2 B_2 \rangle \leq +2$, where $\langle A_i B_j \rangle$ denotes the expected value of the product of $A_i$ and $B_j$. 


*Linearity of expectation: If $X$ and $Y$ are random variables, then $\langle aX + bY \rangle = a \langle X \rangle + b \langle Y \rangle$, for any constants $a$ and $b$. 

---

**Bell Theorem Update 4 No Initialization.nb**
*Note 1:* This step is the subject of much controversy. First, the $A_i$'s and $B_j$'s are random variables *conditioned* on Alice's and Bob's measurement choices. Second, only one of these pairs can be measured on any given trial of the experiment. Hence invoking “linearity of expectation” is highly suspect! Is the “first” $A_i$ the same as the “second” $A_i$? It’s possible we should be considering each expectation with respect to different “context-dependent” measures.

*Note 2:* In any setting where we can assume that $(A_1, B_1, A_2, B_2)$ have an *overall joint* probability distribution that is *consistent* with the given correlations, then we can invoke linearity of expectation.

This is called a CHSH inequality, a specific example of a Bell-type inequality.

Neither Alice nor Bob can determine the value of $Q$ in any single experiment, but they can (jointly) measure any of the products $A_i B_j$.

By repeating the experiment many times, statistical averages can be computed for each of the terms in the inequality, providing estimates of the expectations.

This produces an experimentally testable statement based on the assumptions of hidden variables and locality.

### Table: Outcomes and Detector Settings

<table>
<thead>
<tr>
<th>Trial</th>
<th>$(a,b)$</th>
<th>$(a,b')$</th>
<th>$(a',b)$</th>
<th>$(a',b')$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>-1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sums</td>
<td>$(N_1, N_1)$</td>
<td>$(N_2, N_2)$</td>
<td>$(N_3, N_3)$</td>
<td>$(N_4, N_4)$</td>
</tr>
</tbody>
</table>

### Correlation estimates

$\hat{C}_{11} = \frac{(N_1 - N_1)}{(N_1 + N_1)}$,  
$\hat{C}_{12} = \frac{(N_2 - N_2)}{(N_2 + N_2)}$,  
$\hat{C}_{21} = \frac{(N_3 - N_3)}{(N_3 + N_3)}$,  
$\hat{C}_{22} = \frac{(N_4 - N_4)}{(N_4 + N_4)}$

### Test statistic

$\hat{S}_1 = -\hat{C}_{11} + \hat{C}_{12} + \hat{C}_{21} + \hat{C}_{22}$
Appendix 10: Other facts about convex sets

1. The intersection of any number of closed or open half spaces (i.e. one side or another of a hyper-plane) in $R^n$ is a convex set. Because
   1a. Open or closed half-spaces are convex.
   1b. Intersection of convex sets is convex.
2. Indeed any closed convex set is the intersection of all half-spaces that contain it.
Appendix 11: Locality

Some thoughts based on the YouTube video
https://www.youtube.com/watch?v=ZuvK-od647c

Namely, that although Alice’s and Bob’s “wings” of the experiment are described from the “God’s eye point of view” so it makes it LOOK like Bob’s outcomes are immediately correlated with Alice’s, it is of course not possible for anyone that is part of the experiment to do an immediate comparison. The outcomes of Alice and Bob look random to each of them, and it would become evident only later if indeed they were able to get together and compare results (if indeed they could do this). That is, they would see they had the “same randomness”, or at least connected randomness. This is related to the fact that they cannot communicate information (in the Shannon sense) faster than the speed of light. Doesn’t this vindicate Einstein’s point of view concerning locality? In other words, given the choice: choose 2 out of 3: QM or locality or realism, we can choose QM and realism, and drop locality in the Bell sense only, thus keeping QM and realism AND Einstein locality.

DeBroglie-Bohm Theory: Deterministic and explicitly non-local: From Wikipedia

The de Broglie–Bohm theory, also known as the pilot wave theory, Bohmian mechanics, Bohm’s interpretation, and the causal interpretation, is an interpretation of quantum mechanics. In addition to a wavefunction on the space of all possible configurations, it also postulates an actual configuration that exists even when unobserved. The evolution over time of the configuration (that is, the positions of all particles or the configuration of all fields) is defined by the wave function by a guiding equation. The evolution of the wave function over time is given by the Schrödinger equation. The theory is named after Louis de Broglie (1892–1987) and David Bohm (1917–1992).

The theory is deterministic[1] and explicitly nonlocal: the velocity of any one particle depends on the value of the guiding equation, which depends on the configuration of the system given by its wavefunction; the latter depends on the boundary conditions of the system, which, in principle, may be the entire universe.

Pilot-wave theory is explicitly nonlocal, which is in ostensible conflict with special relativity. Various extensions of "Bohm-like" mechanics exist that attempt to resolve this problem. Bohm himself in 1953 presented an extension of the theory satisfying the Dirac equation for a single particle. However, this was not extensible to the many-particle case because it used an absolute time.[17]

A renewed interest in constructing Lorentz-invariant extensions of Bohmian theory arose in the 1990s; see Bohm and Hiley: The Undivided Universe, and,[18][19] and references therein. Another approach is given in the work of Dürr et al.,[20] in which they use Bohm–Dirac models and a Lorentz-invariant foliation of space-time.

Thus, Dürr et al. (1999) showed that it is possible to formally restore Lorentz invariance for the Bohm–
Dirac theory by introducing additional structure. This approach still requires a foliation of space-time. While this is in conflict with the standard interpretation of relativity, the preferred foliation, if unobservable, does not lead to any empirical conflicts with relativity. In 2013, Dürr et al. suggested that the required foliation could be covariantly determined by the wavefunction.[21]

The relation between nonlocality and preferred foliation can be better understood as follows. In de Broglie–Bohm theory, nonlocality manifests as the fact that the velocity and acceleration of one particle depends on the instantaneous positions of all other particles. On the other hand, in the theory of relativity the concept of instantaneousness does not have an invariant meaning. Thus, to define particle trajectories, one needs an additional rule that defines which space-time points should be considered instantaneous. The simplest way to achieve this is to introduce a preferred foliation of space-time by hand, such that each hypersurface of the foliation defines a hypersurface of equal time.

Initially, it had been considered impossible to set out a description of photon trajectories in the de Broglie–Bohm theory in view of the difficulties of describing bosons relativistically.[22] In 1996, Partha Ghose had presented a relativistic quantum-mechanical description of spin-0 and spin-1 bosons starting from the Duffin–Kemmer–Petiau equation, setting out Bohmian trajectories for massive bosons and for massless bosons (and therefore photons).[22] In 2001, Jean-Pierre Vigier emphasized the importance of deriving a well-defined description of light in terms of particle trajectories in the framework of either the Bohmian mechanics or the Nelson stochastic mechanics.[23] The same year, Ghose worked out Bohmian photon trajectories for specific cases.[24] Subsequent weak-measurement experiments yielded trajectories that coincide with the predicted trajectories.[25][26]

Chris Dewdney and G. Horton have proposed a relativistically covariant, wave-functional formulation of Bohm's quantum field theory[27][28] and have extended it to a form that allows the inclusion of gravity.[29]

Nikolič has proposed a Lorentz-covariant formulation of the Bohmian interpretation of many-particle wavefunctions.[30] He has developed a generalized relativistic-invariant probabilistic interpretation of quantum theory,[31][32][33] in which \(|\psi|^2\) is no longer a probability density in space, but a probability density in space-time. He uses this generalized probabilistic interpretation to formulate a relativistic-covariant version of de Broglie–Bohm theory without introducing a preferred foliation of space-time. His work also covers the extension of the Bohmian interpretation to a quantization of fields and strings.[34]

See also: Quantum potential § Relativistic and field-theoretic extensions

Roderick I. Sutherland at the University in Sydney has a Lagrangian formalism for the pilot wave and its beables. It draws on Yakir Aharonov’s retrocasual weak measurements to explain many-particle entanglement in a special relativistic way without the need for configuration space. The basic idea was already published by Costa de Beauregard in the 1950s and is also used by John Cramer in his transactional interpretation except the beables that exist between the von Neumann strong projection operator measurements. Sutherland’s Lagrangian includes two-way action-reaction between pilot wave and
beables. Therefore, it is a post-quantum non-statistical theory with final boundary conditions that violate the no-signal theorems of quantum theory. Just as special relativity is a limiting case of general relativity when the spacetime curvature vanishes, so, too is statistical no-entanglement signaling quantum theory with the Born rule a limiting case of the post-quantum action-reaction Lagrangian when the reaction is set to zero and the final boundary condition is integrated out.[35]
References

1. On the Einstein, Podolsky, Rosen Paradox, J. S. Bell, Department of Physics, University of Wisconsin, Madison, Wisconsin, Physics Vol. 1, No. 3, pp. 195-290, 1964


9. Bell's theorem: Critique of proofs with and without inequalities, Karl Hess and Walter Philipp

10. The digital computer as a metaphor for the perfect laboratory experiment: Loophole-free Bell experiments, Hans DeRaedt, Kristel Michielsenb, KarlHessc, Computer Physics Communications, August 2016.


21. Methods and Applications of Linear Programming, Cooper and Steinberg, Saunders Publishing, 1974

22. Extreme Points and the Krein-Milman Theorem, Simon textbook, Chapter 8 (see PDF file in Math papers, under Linear Programming.


24. 5-cell, https://en.wikipedia.org/wiki/5-cell


