Entanglement, Information, Causality

Gennaro Auletta[1],[a]

[1] *University of Cassino, Italy*
Department Letters and Philosophy

Abstract. The paper is divided in two parts. In the first one a summary of the main issues about quantum non–locality is provided. In the second part, the connections with information and causality are considered. In particular, it is shown that a principle of information causality implies that hyper–correlations among experimental settings are not possible but only correlations among possible outcomes. Since a setting is for measuring a particular observable and the eigenbasis of this observable can be considered a code, this means that information codification is a local procedure.

1 EPR

According to EPR, the *correctness* of a theory consists in the degree of agreement between its conclusions and human experience—the objective reality, while its *completeness* is defined as [10]: A theory is complete if every element of objective reality has a counterpart in it. The aim of the EPR article is to show the *incompleteness* of quantum mechanics in the sense of its inability to give a satisfactory explanation of entities which are considered fundamental—in a word, it is a 'disproof' and not a positive proof. Indeed, theories can be disproved by experience and (even thought) experiments.

The core of the argument is constituted by the *Separability principle*, which we can express as follows: Two dynamically independent systems cannot influence each other. The separability principle consists in the assumption that any form of interdependency among physical systems is of dynamical and causal type. Therefore, it is important to carefully distinguish the problem of relativistic locality—i.e., the existence of bounds in the transmission of signals and physical effects—from that of separability, which concerns only the impossibility of a correlation between separated systems in the case in which there are *no dynamical and causal connections*. Part of the EPR argument is that, in the absence of physical interactions, the systems are also separated.

EPR state a sufficient condition for the reality of observables, which can be formulated as follows:

> If, without in any way disturbing a system, we can predict with certainty the value of a physical quantity, then, independently of our measurement procedure, there exists an element of the physical reality corresponding to this physical quantity.

The words "without in any way disturbing a system" tells us that the systems are considered as dynamically independent. The aim of EPR is to show that, assuming separability and the sufficient condition

---

[a]e-mail: gennaro.auletta@gmail.com

of reality, quantum mechanics is not complete: in logical terms, for quantum mechanics the following statement holds:

$$[(\text{Suff. Cond. Reality}) \wedge (\text{Separability})] \implies \neg\text{Completeness}, \tag{1}$$

where $\wedge$, $\neg$, and the arrow are the logical symbols for conjunction (AND), negation and implication, respectively.

The argument of EPR is structured as follows. From (i) the definition of completeness, (ii) the principles of physical reality and separability, and (iii) the fact that, according to quantum mechanics, two non–commuting observables cannot simultaneously have definite values, it follows that the following two statements are incompatible:

- The statement r that the quantum mechanical description of reality given by the wave function is not complete and

- The statement s that when the operators describing two physical quantities do not commute, the two quantities cannot have simultaneous reality.

In formal terms,

$$r \mathbin{\succ\mkern-5mu\prec} s, \tag{2}$$

where the symbol $\succ\mkern-5mu\prec$ means a XOR. The meaning of the statement (2) is the following: if it is possible to show that two non–commuting observables have in fact simultaneous reality, we can logically conclude that quantum mechanics cannot be a complete description of reality (from the falsity of s we infer the truth of r).

Let us consider a one-dimensional system $\mathcal{S}$ made of two subsystems $\mathcal{S}_1$ and $\mathcal{S}_2$ interacting during the time interval between t_1 and t_2, with momenta in the position representations:

$$\hat{p}_x^{(1)} = -\imath\hbar\frac{\partial}{\partial x_1} \quad \text{and} \quad \hat{p}_x^{(2)} = -\imath\hbar\frac{\partial}{\partial x_2}, \tag{3}$$

with momentum eigenfunctions

$$\langle p_1 \,|\, \varphi \rangle = \varphi_p(x_1) \quad \text{and} \quad \langle p_2 \,|\, \psi \rangle = \psi_p(x_2) \,, \tag{4}$$

respectively. The vectors $|\varphi\rangle$ and $|\psi\rangle$ describe the states of the particles 1 and 2, respectively. The eigenfunctions in the position representation are

$$\varphi_p(x_1) = \frac{1}{\sqrt{2\pi}}e^{\frac{\imath}{\hbar}px_1} \quad \text{and} \quad \psi_p(x_2) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\imath}{\hbar}(x_2-x_0)p} \,, \tag{5}$$

respectively, where x_0 is a fixed position (constant) and the eigenfunction $\varphi_p(x_1)$ corresponds to eigenvalue $+p$ whilst $\psi_p(x_2)$ corresponds to the eigenvalue $-p$ of the second particle's momentum (in other words, the two particles are moving away from each other with the same direction into opposite senses).

Therefore, the compound system is described by the wave function

$$\Psi(x_1, x_2) = \int\limits_{-\infty}^{+\infty} dp\,\psi_{-p}(x_2)\varphi_p(x_1). \tag{6}$$

Now, I summarize the scheme of the first thought experiment [2, Chap. 16] [4, Chap. 10]:

(a) We locally measure the momentum on particle 1: let us assume that we find an eigenvalue p'.

(b) Therefore, the state (6) reduces to

$$\psi_{-p'}(x_2)\varphi_{p'}(x_1). \tag{7}$$

(c) Then, it is evident that particle 2 must be in state $\psi_{-p'}$ and this result can be predicted with absolute certainty.

(d) However, we were able to formulate such a prediction without disturbing particle 2 (assumption of separability).

(e) Then, as a consequence of (c) and (d) and of the sufficient condition of reality, $\hat{p}_x^{(2)}$ is an element of reality.

Note that steps (a)–(c) are purely quantum mechanical. Only steps (d)–(e) are connected to the specific EPR argument.

However, if we had chosen to consider another observable of particle 1, say $\hat{x}_1$, whose eigenfunctions are $\varphi_x(x_1)$ (whereas $\psi_x(x_2)$ are the eigenfunctions of the observable $\hat{x}_2$ of particle 2), then we would have written the state Ψ of the compound system as

$$\Psi(x_1, x_2) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{+\infty} dx \psi_x(x_2)\varphi_x(x_1). \tag{8}$$

Let us now repeat the previous procedure for the position measurement.

(a') We locally measure the position on particle 1 and find the eigenvalue x'.

(b') Now it is clear that the state (8) reduces to

$$\psi_{x'}(x_2)\varphi_{x'}(x_1). \tag{9}$$

(c') Then, it is evident that the particle 2 must be in the state $\psi_{x'}$ and this result can be predicted with absolute certainty.

(d') However, we have not disturbed particle 2 (assumption of separability).

(e') Then, as a consequence of (c') and (d') and of the sufficient condition of reality, $\hat{x}^{(2)}$ is an element of reality.

Conclusions (e) and (e') look incompatible on the basis of the fact that position and momentum observables of particle 2 do not commute: going back to Propositions r and s [Eq. (2)], EPR have in this way shown that, assuming that r (the quantum mechanical description of reality is not complete) is false, s is proved to be false as well since both $\hat{p}_x^{(2)}$ and $\hat{x}^{(2)}$ have simultaneous reality. Then, the previous assumption must be rejected, and r must be true. Therefore, according to the EPR argument, quantum mechanics cannot be considered as a complete theory and the wave functions (6) and (8) cannot be considered as complete descriptions of the state of the particles.

2 Bohm's reformulation

The argument as formulate in the original EPR paper is difficult tom test. However, a great step was provided by David Bohm. Consider now two particles with spin $\frac{1}{2}$ that are in a state in which the total spin is zero, that is, they are in a singlet state [8]. They can be produced by a single atom radioactive decay. After a time t_0 the two particles begin to separate and at time t_1 they no longer interact. On the hypothesis that they are not disturbed, the law of angular momentum conservation guarantees that they remain in a singlet state. Considering the projection of the spin along the z–direction, the singlet state may be written in the form

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} \left(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_2 \right), \tag{10}$$

where the subscripts 1 and 2 refer to the particles. This implies that, if a measurement of the spin component along the z direction of particle 1 leads to a result $+1/2$, that of particle 2 along the same direction must give the value $-1/2$, and vice versa. This means that $|\Psi_0\rangle$ is an eigenket of the z component of the spin observables $\hat{\sigma}_{1z}\hat{\sigma}_{2z}$ of the two systems.

Entanglement is a property of the state that is independent of the basis used. In order to see this rotational invariance, let us write it in terms of the z–component eigenvectors as

$$|\Psi_0\rangle = \frac{1}{\sqrt{2}} \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}_1 \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}_2 - \begin{pmatrix} 0 \\ 1 \end{pmatrix}_1 \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}_2 \right]. \tag{11}$$

Then, $|\Psi_0\rangle$ turns out to be also an eigenvector of $\hat{\sigma}_{1x}\hat{\sigma}_{2x}$ and $\hat{\sigma}_{1y}\hat{\sigma}_{2y}$. For example, let us consider the y orientation. First, let us expand the z–eigenkets into the y–eigenkets:

$$|\uparrow\rangle = \frac{\sqrt{2}}{2} \left(|\uparrow\rangle_y + |\downarrow\rangle_y \right), \tag{12a}$$

$$|\downarrow\rangle = -\frac{\iota\sqrt{2}}{2} \left(|\uparrow\rangle_y - |\downarrow\rangle_y \right). \tag{12b}$$

Then, we can write the singlet state (10) in the y expansion:

$$\begin{aligned}
\frac{1}{\sqrt{2}} \left(|\uparrow\rangle_1 \otimes |\downarrow\rangle_2 - |\downarrow\rangle_1 \otimes |\uparrow\rangle_2 \right) &= \frac{1}{\sqrt{2}} \left[-\frac{\iota}{2} \left(|\uparrow\rangle_y + |\downarrow\rangle_y \right)_1 \otimes \left(|\uparrow\rangle_y - |\downarrow\rangle_y \right)_2 \right. \\
&\quad \left. + \frac{\iota}{2} \left(|\uparrow\rangle_y - |\downarrow\rangle_y \right)_1 \otimes \left(|\uparrow\rangle_y + |\downarrow\rangle_y \right)_2 \right] \\
&= \frac{\iota}{\sqrt{2}} \left[\left(|\uparrow\rangle_y \right)_1 \otimes \left(|\downarrow\rangle_y \right)_2 - \left(|\downarrow\rangle_y \right)_1 \otimes \left(|\uparrow\rangle_y \right)_2 \right].
\end{aligned} \tag{13}$$

Consequently, we have

$$\left(\hat{\sigma}_{1y}\hat{\sigma}_{2y} \right) |\Psi_0\rangle = \frac{\iota}{\sqrt{2}} \left(\hat{\sigma}_{1y}\hat{\sigma}_{2y} \right) \left[\left(|\uparrow\rangle_y \right)_1 \otimes \left(|\downarrow\rangle_y \right)_2 - \left(|\downarrow\rangle_y \right)_1 \otimes \left(|\uparrow\rangle_y \right)_2 \right], \tag{14}$$

which, by making use Pauli matrices and of a reformulation of expression (11) in the y basis, implies

$$\begin{aligned}
\left(\hat{\sigma}_{1y}\hat{\sigma}_{2y} \right) |\Psi_0\rangle &= \frac{\iota}{2\sqrt{2}} \begin{bmatrix} 0 & -\iota \\ \iota & 0 \end{bmatrix}_1 \begin{bmatrix} 0 & -\iota \\ \iota & 0 \end{bmatrix}_2 \\
&\quad \times \left[\begin{pmatrix} 1 \\ \iota \end{pmatrix}_1 \otimes \begin{pmatrix} 1 \\ -\iota \end{pmatrix}_2 - \begin{pmatrix} 1 \\ -\iota \end{pmatrix}_1 \otimes \begin{pmatrix} 1 \\ \iota \end{pmatrix}_2 \right] \\
&= \frac{\iota}{2\sqrt{2}} \left[\begin{pmatrix} 1 \\ \iota \end{pmatrix}_1 \otimes \begin{pmatrix} -1 \\ \iota \end{pmatrix}_2 - \begin{pmatrix} -1 \\ \iota \end{pmatrix}_1 \otimes \begin{pmatrix} 1 \\ \iota \end{pmatrix}_2 \right] \\
&= \frac{\iota}{\sqrt{2}} \left[\left(|\uparrow\rangle_y \right)_1 \otimes \left(-|\downarrow\rangle_y \right)_2 - \left(-|\downarrow\rangle_y \right)_1 \otimes \left(|\uparrow\rangle_y \right)_2 \right].
\end{aligned} \tag{15}$$

Now, let us back–substitute this expression into the z expansion:

$$\begin{aligned}
\frac{\iota}{\sqrt{2}}\left[\left(|\!\uparrow\rangle_y\right)_1\left(-|\!\downarrow\rangle_y\right)_2 - \left(-|\!\downarrow\rangle_y\right)_1\left(|\!\uparrow\rangle_y\right)_2\right] &= -\frac{\iota}{\sqrt{2}}\frac{1}{2}\left[(|\!\uparrow\rangle + \iota|\!\downarrow\rangle)_1\left(-|\!\uparrow\rangle + \iota|\!\downarrow\rangle\right)_2\right. \\
&\quad \left. -(-|\!\uparrow\rangle + \iota|\!\downarrow\rangle)_1\left(|\!\uparrow\rangle + \iota|\!\downarrow\rangle\right)_2\right] \\
&= \frac{\iota}{\sqrt{2}}\left(\iota|\!\uparrow\rangle_1|\!\downarrow\rangle_2 - \iota|\!\downarrow\rangle_1|\!\uparrow\rangle_2\right) \\
&= -\frac{1}{\sqrt{2}}\left(|\!\uparrow\rangle_1|\!\downarrow\rangle_2 - |\!\downarrow\rangle_1|\!\uparrow\rangle_2\right) \\
&= -|\Psi_0\rangle,
\end{aligned} \tag{16}$$

where I have dropped the symbol $\otimes$ of the sake of simplicity.

3 Bell Theorem

Bell assumed the existence of a hidden parameter λ such that, given λ, the function $A_{\mathbf{a}}$ describing the results obtained by measuring with a device A the spin of the first particle along a chosen direction $\mathbf{a}$ (i.e., the observable $\hat{\sigma}_1\cdot\mathbf{a}$), depends only on λ and on $\mathbf{a}$ [6]. Similarly, the function $B_{\mathbf{b}}$ describing the results when measuring with a device B the spin of the second particle along a chosen direction $\mathbf{b}$ (i.e., $\hat{\sigma}_2\cdot\mathbf{b}$), depends only on $\mathbf{b}$ and λ. The separability principle denies that there can be a form of interdependence between two systems if they do not dynamically interact (factorization rule):

$$A_{\mathbf{a}}B_{\mathbf{b}} = A_{\mathbf{a}}(\lambda)B_{\mathbf{b}}(\lambda), \tag{17}$$

where therefore $A_{\mathbf{a}}$ and $B_{\mathbf{b}}$ represent two deterministic functions of the hidden parameter. Eq. (17) expresses the fact that the probability distributions for the two particles are mutually independent. I assume that the result of each measurement can be either +1 (representing spin up) or -1 (representing spin down), that is,

$$A_{\mathbf{a}}(\lambda) = \pm 1, \quad B_{\mathbf{b}}(\lambda) = \pm 1. \tag{18}$$

Following Eq. (17), if $\wp(\lambda)$ denotes the probability distribution of the hidden parameter λ, then the expectation value of the product of the two components $\hat{\sigma}_1\cdot\mathbf{a}$ and $\hat{\sigma}_2\cdot\mathbf{b}$ is

$$\langle(\hat{\sigma}_1\cdot\mathbf{a})(\hat{\sigma}_2\cdot\mathbf{b})\rangle = \int_{\Lambda}\wp(\lambda)A_{\mathbf{a}}(\lambda)B_{\mathbf{b}}(\lambda)d\lambda, \tag{19}$$

where Λ represents the set of all possible values of λ.

In the present context, $A_{\mathbf{a}}(\lambda)$ and $B_{\mathbf{b}}(\lambda)$ are functions defining the possible measurement results or the eigenvalues of the measured observables. Since we do not know the values of the hidden parameters λ, we must integrate over all the possible values $\lambda \in \Lambda$. Because $\wp(\lambda)$ is supposed to be a normalized probability distribution, we have

$$\int_{\Lambda}\wp(\lambda)d\lambda = 1, \tag{20}$$

and, given the values (18), we also have

$$-1 \le \langle\mathbf{a}, \mathbf{b}\rangle \le +1, \tag{21}$$

where I have rewritten the expression $\langle(\hat{\sigma}_1 \cdot \mathbf{a})\,(\hat{\sigma}_2 \cdot \mathbf{b})\rangle$ in the simplified form $\langle \mathbf{a}, \mathbf{b}\rangle$. Our aim is to compare the prediction of a deterministic HV theory as expressed by Eq. (19) with the quantum mechanical expectation value, which for the singlet state $|\Psi_0\rangle$ [Eq. (10)] is given by

$$\langle \mathbf{a}, \mathbf{b}\rangle_{\Psi_0} = \langle \Psi_0 |(\hat{\sigma}_1 \cdot \mathbf{a})\,(\hat{\sigma}_2 \cdot \mathbf{b})| \Psi_0 \rangle = -\mathbf{a} \cdot \mathbf{b}. \tag{22}$$

This result can be derived when considering the previous products between observables and vectors as sum of Cartesian components

$$\hat{\sigma}_1 \cdot \mathbf{a} \;=\; a_x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_1 + a_y \begin{bmatrix} 0 & -\iota \\ \iota & 0 \end{bmatrix}_1 + a_z \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_1, \tag{23a}$$

$$\hat{\sigma}_2 \cdot \mathbf{b} \;=\; b_x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_2 + b_y \begin{bmatrix} 0 & -\iota \\ \iota & 0 \end{bmatrix}_2 + b_z \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}_2, \tag{23b}$$

The expectation value on the singlet state (11) of these two products gives 9 terms, of which the first three have the form

$$\begin{aligned}
\langle \Psi_0 |a_x b_x \hat{\sigma}_{1x} \hat{\sigma}_{2x}| \Psi_0\rangle \;&=\; \langle \Psi_0| \frac{a_x b_x}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}_2 \\
&\quad \times \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}_1 \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}_2 - \begin{pmatrix} 0 \\ 1 \end{pmatrix}_1 \otimes \begin{pmatrix} 1 \\ 0 \end{pmatrix}_2 \right] \\
&=\; \langle \Psi_0 |-a_x b_x| \Psi_0\rangle = -a_x b_x.
\end{aligned} \tag{24}$$

Indeed, similar calculations show that we also have

$$\left\langle \Psi_0 \left| a_y b_y \hat{\sigma}_{1y} \hat{\sigma}_{2y} \right| \Psi_0 \right\rangle = -a_y b_y \quad \text{and} \quad \langle \Psi_0 |a_z b_z \hat{\sigma}_{1z} \hat{\sigma}_{2z}| \Psi_0\rangle = -a_z b_z. \tag{25}$$

The remaining six cross terms are instead all zero, so that we may finally conclude that

$$\langle \Psi_0 |(\hat{\sigma}_1 \cdot \mathbf{a})\,(\hat{\sigma}_2 \cdot \mathbf{b})| \Psi_0\rangle \;=\; -(a_x b_x + a_y b_y + a_z b_z) = -\mathbf{a} \cdot \mathbf{b}. \tag{26}$$

When the two orientations $\mathbf{a}$ and $\mathbf{b}$ are parallel, quantum mechanical calculations [see Eq. (15)] show that

$$\langle \mathbf{a}, \mathbf{a}\rangle_{\Psi_0} = -1, \tag{27}$$

as it should be since there is a perfect *anticorrelation* (spin–up versus spin–down) between the results of the two measurements.

Since the value given by Eq. (27) for perfect anticorrelation is an experimental fact, also a HV theory must satisfy this requirement. On the other hand, $\langle \mathbf{a}, \mathbf{a}\rangle = -1$ holds if and only if we also have

$$A_{\mathbf{a}}(\lambda) = -B_{\mathbf{a}}(\lambda), \tag{28}$$

for any direction $\mathbf{a}$. In this case, Eq. (19) reaches the minimum value [see also Eq. (21)]. Under this assumption, we can drop any reference to the B device and rewrite Eq. (19) as

$$\langle \mathbf{a}, \mathbf{b}\rangle = -\int d\lambda \wp(\lambda) A_{\mathbf{a}}(\lambda) A_{\mathbf{b}}(\lambda). \tag{29}$$

Now we consider two alternative orientations, say $\mathbf{b}$ and $\mathbf{c}$, of the spin measurement of particle 2:

$$\begin{aligned}
\langle \mathbf{a}, \mathbf{b}\rangle - \langle \mathbf{a}, \mathbf{c}\rangle \;&=\; -\int d\lambda \wp(\lambda)[A_{\mathbf{a}}(\lambda) A_{\mathbf{b}}(\lambda) - A_{\mathbf{a}}(\lambda) A_{\mathbf{c}}(\lambda)] \\
&=\; \int d\lambda \wp(\lambda) A_{\mathbf{a}}(\lambda) A_{\mathbf{b}}(\lambda)[A_{\mathbf{b}}(\lambda) A_{\mathbf{c}}(\lambda) - 1],
\end{aligned} \tag{30}$$

because of the property (18) and since, for any orientation $\mathbf{n}$, we have $[A_\mathbf{n}(\lambda)]^2 = 1$, which implies

$$A_\mathbf{a}(\lambda)A_\mathbf{b}(\lambda)A_\mathbf{b}(\lambda)A_\mathbf{c}(\lambda) = A_\mathbf{a}(\lambda)A_\mathbf{c}(\lambda). \tag{31}$$

Then, from Eq. (30) we may prove the inequality

$$|\langle \mathbf{a}, \mathbf{b}\rangle - \langle \mathbf{a}, \mathbf{c}\rangle| \leq \int d\lambda \wp(\lambda)[1 - A_\mathbf{b}(\lambda)A_\mathbf{c}(\lambda)]. \tag{32}$$

This result is obtained when one considers that for any integrable function $f(x)$, we have

$$\left| \int dx f(x) \right| \leq \int dx |f(x)|, \tag{33}$$

and, given again the property (18), we also have

$$|A_\mathbf{b}(\lambda)A_\mathbf{c}(\lambda) - 1| = 1 - A_\mathbf{b}(\lambda)A_\mathbf{c}(\lambda). \tag{34}$$

Therefore, given the property (20) we finally obtain

$$|\langle \mathbf{a}, \mathbf{b}\rangle - \langle \mathbf{a}, \mathbf{c}\rangle| \leq 1 + \langle \mathbf{b}, \mathbf{c}\rangle, \tag{35}$$

where

$$\langle \mathbf{b}, \mathbf{c}\rangle = -\int d\lambda \wp(\lambda)A_\mathbf{b}(\lambda)A_\mathbf{c}(\lambda). \tag{36}$$

A reformulation of the Bell inequality (35) is the so-called CHSH inequality, a widely used form,

$$\left| \langle \mathbf{a}, \mathbf{b}\rangle + \langle \mathbf{a}, \mathbf{b}'\rangle + \langle \mathbf{a}', \mathbf{b}\rangle - \langle \mathbf{a}', \mathbf{b}'\rangle \right| \leq 2, \tag{37}$$

where $\mathbf{a}'$ is a setting alternative to $\mathbf{a}$ as well as $\mathbf{b}'$ to $\mathbf{b}$. We may associate to this inequality the following Bell operator:

$$\begin{aligned}
\hat{\mathcal{B}} &= \hat{\sigma}_1 \cdot \mathbf{a}\,(\hat{\sigma}_2 \cdot \mathbf{b} + \hat{\sigma}_2 \cdot \mathbf{b}') + \hat{\sigma}_1 \cdot \mathbf{a}'\,(\hat{\sigma}_2 \cdot \mathbf{b} - \hat{\sigma}_2 \cdot \mathbf{b}') \\
&= (\hat{\sigma}_1 \cdot \mathbf{a})\,(\hat{\sigma}_2 \cdot \mathbf{b}) + (\hat{\sigma}_1 \cdot \mathbf{a})\,(\hat{\sigma}_2 \cdot \mathbf{b}') \\
&\quad + (\hat{\sigma}_1 \cdot \mathbf{a}')\,(\hat{\sigma}_2 \cdot \mathbf{b}) + (\hat{\sigma}_1 \cdot \mathbf{a}')\,(\hat{\sigma}_2 \cdot \mathbf{b}'),
\end{aligned} \tag{38}$$

which will play a crucial role later on. I recall indeed that e.g. $\langle \mathbf{a}, \mathbf{b}\rangle$ is a shorthand for $\langle(\hat{\sigma}_1 \cdot \mathbf{a})\,(\hat{\sigma}_2 \cdot \mathbf{b})\rangle$, which allows us to write

$$\left| \langle \hat{\mathcal{B}} \rangle \right| \leq 2. \tag{39}$$

4 Experiments and Loopholes

Tests of the Bell theorem already started in the mid of 1970s. However, several loopholes were discovered that affected these early experiments and could be dealt with step by step. The *first loophole* we consider is the locality loophole. In all experiments, one should consider the possibility that the result of a measurement obtained by using a certain polarizer direction depend on the orientation of the other polarizer. This problem was overcome by Aspect's team [1] as outlined in Fig. 1.

Another difficulty (*second loophole*) concerns the angular correlation: Because of the cosine–squared angular correlation of the directions of the photons emitted in an atomic cascade, an inherent polarization decorrelation is present. Hence the very polarization correlation which could result in a
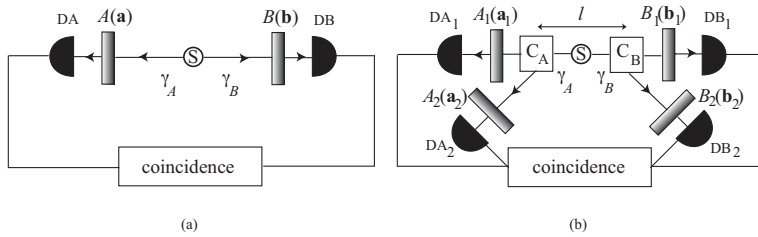
Figure 1. One should consider the possibility that the results obtained using a certain polarization direction could depend on the other polarizer. (a) Friedman–Clauser experiment: The correlated photons γ_A, γ_B coming from the source S impinge upon the linear polarizers A, B oriented in directions $\mathbf{a}$, $\mathbf{b}$, respectively. (b) Experiment proposed by Aspect: The optical commutator C_A directs the photon γ_A either towards polarizer A_1 with orientation $\mathbf{a}_1$ or to polarizer A_2 with orientation $\mathbf{a}_2$. Similarly for C_B for B_1 and B_2. The two commutators work independently (the time intervals between two commutations are taken to be stochastic). The four joint detection rates are monitored and the orientations $\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2$ are not changed for the whole experiment. l is the separation between the switches.
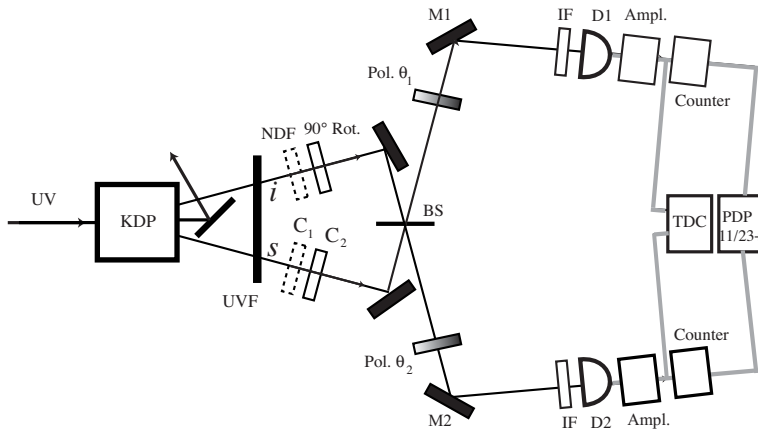

Figure 2. Because of the cosine–squared angular correlation of the directions of the photons emitted in an atomic cascade, an inherent polarization decorrelation is present. Outline of the Alley–Shih and Ou–Mandel's experiment. Light from the 351.1–nm line of an argon–ion laser falls on a non–linear crystal of potassium dihydrogen phosphate (KDP), where down-converted photons of wavelength of about 702 nm are produced. Down–conversion can be tuned in order that linearly polarized signal and idler photons emerge at angles of about $\pm 2°$ relative to the ultraviolet (UV) pump beam with the electric vector in the plane of the diagram. The *idler* (i) photons pass through a 90° polarization rotator, while the *signal* (s) photons traverse a compensating glass plate C_1 producing an equal time delay. The two photons are then directed from opposite sides towards a beam splitter (BS). The input to the BS consists of an x–polarized s–photon and of a rotated y–polarized i–photon. The light beams emerging from BS, consisting of a mixing of i-photons and s-photons, pass through linear polarizers set at adjustable angles θ_1 and θ_2, through similar interference filters (IF) and finally fall on two photodetectors D_1 and D_2. The photoelectric pulses from D_1 and D_2 are amplified and shaped and fed to the start and stop inputs of a time-to-digital converter (TDC) under computer control which functions as a coincidence counter.

violation of one of the Bell inequalities is reduced for non–collinear photons. The problem can be overcome by using SPDC sources instead of atomic cascade ones [14]. Pairs of photons resulting from SPDC can have an angular correlation of better than 1 mrad, although in general they need not be collinear. The set up is shown in Fig. 2.

A further issue (*third loophole*) is represented by the detection loophole. In fact, we may raise the question of how high the detection efficiencies must be for the experimental confirmation of the
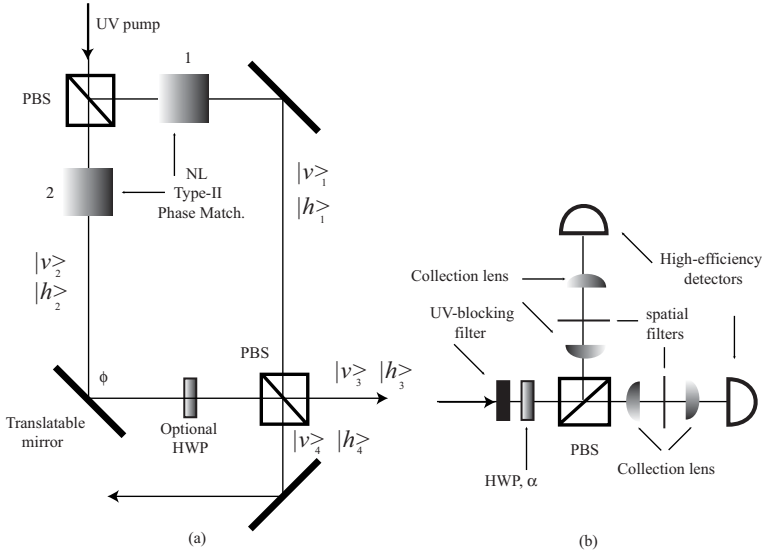
Figure 3. The question is how high the detection efficiency must be for the experimental confirmation of the quantum predictions. In Aspect's experiment the required is 83%. With SPDC experiments, we are obliged to discard part of the counts (when both photons are in the same channel). Proposed experiment for solving the detection loophole. A possible solution is to directly produce a pair of photons in a singlet–kind state avoiding in this way any post–selection.

(a) An ultraviolet pump photon may be spontaneously down-converted in either of two nonlinear crystals, producing a pair of collinear orthogonally polarized photons at half the frequency (type–II phase matching). The outputs are directed toward a second PBS. When the outputs of both crystals are combined with an appropriately relative phase ϕ, a true singlet- or triplet-like state may be produced. By using a half-wave plate (HWP) to effectively exchange the polarizations of photons originating in crystal 2, one overcomes several problems arising from nonideal phase matching. An additional mirror is used to direct the photons into opposite direction towards separated analyzers.

(b) A typical analyzer, including an HWP to rotate by θ the polarization component selected by the analyzing BS, and precision spatial filters to select only conjugate pairs of photons. In an advanced version of the experiment, the HWP could be replaced by an ultrafast polarization rotator (such as Pockels or Kerr cells) to close also the locality loophole.

quantum theoretical predictions. The problem with SPDC–type experiments is that, even with high detection efficiency, one must discard part of the counts, since we are obliged to discard all events where both photons are in the same channel, and one could rise the question whether this selection might represent a bias. Even though this is a remote possibility, in order to exclude any ambiguity a more refined solution is required [12]. A possibility is to directly produce a pair of photons in singlet–type state, thus avoiding any post–selection. One of the first proposals for doing this is shown and summarized in Fig. 3. By means of this apparatus it is possible to produce output photons in the state

$$|\Psi\rangle \simeq |v\rangle_3 |h\rangle_4 + e^{\iota\phi} (|h\rangle_3 |v\rangle_4) \ . \tag{40}$$

5 Non–Locality and Information

I have a general remark. Given any quantum system described by the density matrix $\hat{\rho}$, its von Neumann entropy is [11]

$$S(\hat{\rho}) = -\text{Tr}(\hat{\rho} \ln \hat{\rho}) \ . \tag{41}$$

In fact, the density matrix can be seen as the operator which carries maximal information about the state of the system. If we consider an orthonormal basis $\{|b_k\rangle\}$ of eigenvectors of the density operator

$\hat{\rho}$ for a system $\mathcal{S}$ such that

$$\hat{\rho}\,|b_k\rangle = r_k\,|b_k\rangle \;, \tag{42}$$

where the r_k's are the eigenvalues of $\hat{\rho}$, we may rewrite Eq. (41) as

$$S(\hat{\rho}) = -\sum_j r_j \ln r_j \;, \tag{43}$$

The eigenvectors $|b_k\rangle$ are the possible outcomes of a measurement when we choose to measure the observable of which they are eigenvectors. Let us define the entanglement between systems [5] 1 and 2 as

$$\mathrm{E}(1,2) \quad = \quad S(1,2) - S(1) - S(2) \;, \tag{44}$$

where $S(1,2)$ is the joint (total) entropy of systems 1 and 2, and

$$S(1) = S(\hat{\varrho}_1) \quad \text{and} \quad S(2) = S(\hat{\varrho}_2) \tag{45}$$

are the entropies calculated on the reduced density matrices of the subsystems 1 and 2, respectively, relative to $\hat{\rho}_{12}$. This reflects the fact that entanglement is a quantum form of mutual information: Two entangled systems are correlated because they share an amount of information that is not foreseen classically: indeed the possible outcomes are interdependent.

Are there specific quantum mechanical bounds on information acquisition? Is the bound found with inequality (35) a necessity or are there more rigorous bounds? And if they are, what is their meaning? Let us take advantage of the CHSH inequality (37). Since each of the terms in Eq. (37) lies between -1 and $+1$ [Eq. (21)], the natural upper bound for the entire expression is $+4$. This is precisely the case if we demand that the probabilities satisfy only the causal communication constraint [16], i.e., that they do not violate relativistic locality (what is called non–signaling requirement). In this case, we have

$$\left|\langle \mathbf{a}, \mathbf{b}\rangle + \langle \mathbf{a}, \mathbf{b}'\rangle + \langle \mathbf{a}', \mathbf{b}\rangle - \langle \mathbf{a}', \mathbf{b}'\rangle\right| \le 4 \quad \text{or} \quad \left|\langle \hat{\mathcal{B}}\rangle\right| \le 4 \;. \tag{46}$$

Indeed, the non–signaling requirement is that the operations one can perform locally here are not influenced by the operations one performs elsewhere, which implies in particular that the probability to obtain a certain outcome (say 1) when choosing the direction $\mathbf{a}$ is independent from the outcomes (either $+1$ or -1) when elsewhere one choses a direction $\mathbf{b}$ or $\mathbf{b}'$, that is,

$$\wp_{a,b}(1,1) + \wp_{a,b}(1,-1) = \wp_{a,b'}(1,1) + \wp_{a,b'}(1,-1). \tag{47}$$

Similar considerations hold for any direction. If we consider only this requirement, we are allowed to build the set of probabilities

$$\wp_{a,b}(1,1) = \wp_{a,b}(-1,-1) = \frac{1}{2}\;, \qquad \wp_{a,b'}(1,1) = \wp_{a,b'}(-1,-1) = \frac{1}{2}$$

$$\wp_{a',b}(1,1) = \wp_{a',b}(-1,-1) = \frac{1}{2}\;, \qquad \wp_{a',b'}(1,-1) = \wp_{a',b'}(-1,1) = \frac{1}{2}, \tag{48}$$

while all other probabilities are zero and where I remark that only the $\wp_{a',b'}$ probabilities show anti–correlation.

All the four different expectation values in inequality (46) can be formulated as the following one:

$$\langle \mathbf{a}, \mathbf{b}\rangle \quad = \quad \wp_{a,b}(1,1) + \wp_{a,b}(-1,-1) - \wp_{a,b}(1,-1) - \wp_{a,b}(-1,1), \tag{49}$$

where the negative sign of the latter two probabilities is due to the fact that both represent anticorrelations. However, this expectation value in the paramount case in which all probabilities are equal mirrors the separability condition (17), i.e., the absence of correlations ($\langle \mathbf{a}, \mathbf{b} \rangle = 0$) between the two systems. Due to the above assumptions, however, the latter two probabilities are zero so that the whole expression is reduced to

$$\langle \mathbf{a}, \mathbf{b} \rangle = \wp_{a,b}(1, 1) + \wp_{a,b}(-1, -1), \tag{50}$$

and similarly for the other three correlations. In this way, taking into account the probabilities (48), the upper bound 4 of inequality (46) is easily obtained.

Let $\hat{O}^a, \hat{O}^{a'}, \hat{O}^b, \hat{O}^{b'}$ be arbitrary Hermitian operators on a Hilbert space $\mathcal{H}$ satisfying the condition $[\hat{O}^a, \hat{O}^b] = 0$ and so on for the other couples $(a, b'), (a', b), (a', b')$ [17]. Moreover, each operator has eigenvalues 1 and -1. We define a generalization of the Bell operator (38), since we are no longer considering the spin observable only,

$$\hat{B} = \hat{O}^a \hat{O}^b + \hat{O}^{a'} \hat{O}^b + \hat{O}^a \hat{O}^{b'} - \hat{O}^{a'} \hat{O}^{b'}. \tag{51}$$

From the previous assumptions, it follows that the square of each operator is equal to the identity, which implies

$$
\begin{aligned}
2\sqrt{2} - \hat{B} &= \frac{1}{\sqrt{2}} \left[(\hat{O}^a)^2 + (\hat{O}^{a'})^2 + (\hat{O}^b)^2 + (\hat{O}^{b'})^2 \right] - \hat{B} \\
&= \frac{1}{\sqrt{2}} \left[\left(\hat{O}^a - \frac{\hat{O}^b + \hat{O}^{b'}}{\sqrt{2}} \right)^2 + \left(\hat{O}^{a'} - \frac{\hat{O}^b - \hat{O}^{b'}}{\sqrt{2}} \right)^2 \right] = \hat{A}.
\end{aligned}
\tag{52}
$$

Now, we wish to prove that $\hat{B}$ can be expanded as above and that:

$$\left| \langle \hat{B} \rangle \right| \le 2\sqrt{2} \,. \tag{53}$$

Let us expand the operator $\hat{A}$ as

$$
\begin{aligned}
&\frac{1}{\sqrt{2}} \left[(\hat{O}^a)^2 + \frac{\left(\hat{O}^b + \hat{O}^{b'} \right)^2}{2} - 2 \frac{\hat{O}^a \left(\hat{O}^b + \hat{O}^{b'} \right)}{\sqrt{2}} + (\hat{O}^{a'})^2 + \frac{\left(\hat{O}^b - \hat{O}^{b'} \right)^2}{2} - 2 \frac{\hat{O}^{a'} \left(\hat{O}^b - \hat{O}^{b'} \right)}{\sqrt{2}} \right] \\
&= \frac{1}{\sqrt{2}} \cdot \frac{1}{2\sqrt{2}} \left[2\sqrt{2}(\hat{O}^a)^2 + \sqrt{2}(\hat{O}^b)^2 + \sqrt{2}(\hat{O}^{b'})^2 + 2\sqrt{2}\hat{O}^b \hat{O}^{b'} - 4\hat{O}^a \hat{O}^b - 4\hat{O}^a \hat{O}^{b'} \right. \\
&\qquad \left. +2\sqrt{2}(\hat{O}^{a'})^2 + \sqrt{2}(\hat{O}^b)^2 + \sqrt{2}(\hat{O}^{b'})^2 - 2\sqrt{2}\hat{O}^b \hat{O}^{b'} - 4\hat{O}^{a'} \hat{O}^b + 4\hat{O}^{a'} \hat{O}^{b'} \right] \\
&= \frac{1}{4} \left[2\sqrt{2}(\hat{O}^a)^2 + \sqrt{2}(\hat{O}^b)^2 + \sqrt{2}(\hat{O}^{b'})^2 + 2\sqrt{2}(\hat{O}^{a'})^2 + \sqrt{2}(\hat{O}^b)^2 + \sqrt{2}(\hat{O}^{b'})^2 \right. \\
&\qquad \left. +2\sqrt{2}\hat{O}^b \hat{O}^{b'} - 2\sqrt{2}\hat{O}^b \hat{O}^{b'} - 4\hat{O}^a \hat{O}^b - 4\hat{O}^a \hat{O}^{b'} - 4\hat{O}^{a'} \hat{O}^b + 4\hat{O}^{a'} \hat{O}^{b'} \right] \\
&= \frac{1}{4} \left[2\sqrt{2} \left((\hat{O}^a)^2 + (\hat{O}^b)^2 + (\hat{O}^{a'})^2 + (\hat{O}^{b'})^2 \right) - 4 \left(\hat{O}^a \hat{O}^b + \hat{O}^{a'} \hat{O}^b + \hat{O}^a \hat{O}^{b'} - \hat{O}^{a'} \hat{O}^{b'} \right) \right] \\
&= \frac{1}{4} \left(8\sqrt{2}\hat{I} - 4\hat{B} \right).
\end{aligned}
\tag{54}
$$

This proves the expansion of $\hat{B}$. Now, consider that the sum or the difference between Hermitian operators is itself a Hermitian operator, which shows that the following two operatorial expressions are Hermitian:

$$\frac{\hat{O}^b + \hat{O}^{b'}}{\sqrt{2}} \quad \text{and} \quad \frac{\hat{O}^b - \hat{O}^{b'}}{\sqrt{2}} \,. \tag{55}$$

The last step implies that also

$$\hat{O}^a - \frac{\hat{O}^b + \hat{O}^{b'}}{\sqrt{2}} \quad \text{and} \quad \hat{O}^{a'} - \frac{\hat{O}^b - \hat{O}^{b'}}{\sqrt{2}} \tag{56}$$

are. Therefore, the operator

$$\hat{A} = \frac{1}{\sqrt{2}} \left[\left(\hat{O}^a - \frac{\hat{O}^b + \hat{O}^{b'}}{\sqrt{2}} \right)^2 + \left(\hat{O}^{a'} - \frac{\hat{O}^b - \hat{O}^{b'}}{\sqrt{2}} \right)^2 \right], \tag{57}$$

which consists of the sum of squares of Hermitian operators, has clearly an expectation value

$$\left\langle \hat{A} \right\rangle \geq 0. \tag{58}$$

Since by taking the mean value on both sides of Eq. (53) we have

$$\left\langle \hat{A} \right\rangle = 2\sqrt{2} - \left\langle \hat{B} \right\rangle, \tag{59}$$

this leads to the conclusion:

$$\left\langle \hat{B} \right\rangle \leq 2\sqrt{2}. \tag{60}$$

A similar argument leads to

$$\left\langle \hat{B} \right\rangle \geq -2\sqrt{2}, \tag{61}$$

which finally implies

$$\left| \left\langle \hat{B} \right\rangle \right| \leq 2\sqrt{2}. \tag{62}$$

The importance of Tsirelson result lies in the fact that it proves that quantum mechanics *does not* fill the entire gap between the bounds set by Eqs. (37) and (46). The former inequality sets bounds (i.e., 2) for classical separable theories whilst quantum mechanics satisfy the bound $2\sqrt{2}$, which is still stricter than the bound (i.e., 4) imposed by inequality (46). In other words, quantum mechanics certainly allows for correlations that are not allowed by classical HV theories. However, there is a wide spectrum of "hyper–correlations" that do not contradict causal communication constraints (they satisfy the bound imposed by inequality (46)) but are nevertheless not allowed by quantum mechanics (since they do not satisfy inequality (62)). Therefore, we need still to clarify the relations between these different bounds.

To examine this point, let us reformulate the CHSH inequality (37) as an equality with the maximal bound attainable, i.e., B=2, which could be rewritten as [13]

$$\frac{1}{2}B - 1 = 0, \tag{63}$$

where B expresses again the Bell operator written in terms of a numerical parameter B. However we are interested in more general cases than those allowed by the classical separability requirement. In those case, instead of putting a 0 on the right-hand side we write another numerical parameter, D, as follows:

$$D = \frac{1}{2}B - 1. \tag{64}$$

Moreover, we like to write the parameter B as a combination of correlations C_{jk} (where $j, k = a, b, a', b'$) expressed in informational terms, that is, with $j, k = 1, 0$. We would also like to express the

possible outputs when Alice and Bob measure in informational term 1,0. In other words, instead of speaking of polarization directions $\mathbf{a}$, $\mathbf{a}'$, $\mathbf{b}$, and $\mathbf{b}'$, or of observables $\hat{O}^a$, $\hat{O}^{a'}$, $\hat{O}^b$, $\hat{O}^{b'}$, we would like to introduce generic inputs $a, b = 0$ and $a', b' = 1$. Moreover, instead to have possible results $-1, 1$, we like to introduce information outputs 0,1. With these assumptions, we rewrite the correlation $\langle \mathbf{a}, \mathbf{b} \rangle$ for the non–signaling case as expressed in the formula (50) as the sum of two conditional probabilities:

$$C_{00} = \wp(11|00) + \wp(00|00), \tag{65a}$$

where what follows the vertical lines are the inputs and what precedes the vertical line the outputs. Similarly, we have

$$C_{10} = \wp(11|10) + \wp(00|10), \quad C_{01} = \wp(11|01) + \wp(00|01), \quad C_{11} = \wp(10|11) + \wp(01|11),$$

where the first equality is a reformulation of the correlation $\langle \mathbf{a}', \mathbf{b} \rangle$, the second of the correlation $\langle \mathbf{a}, \mathbf{b}' \rangle$, and the latter of the correlation $\langle \mathbf{a}', \mathbf{b}' \rangle$, and again I remark that only the latter one is an anticorrelation.

Then we can write:

$$
\begin{aligned}
D &= \frac{1}{2}B - 1 \\
&= \frac{1}{2}\left(C_{00} + C_{01} + C_{10} - C_{11}\right) - 1.
\end{aligned}
\tag{66}
$$

Let us now introduce a simplification. In the case in which

$$C_{00} = C_{01} = C_{10} = -C_{11} \geq 0, \tag{67}$$

we can write

$$C = C_{00} = C_{01} = C_{10} = -C_{11}, \tag{68}$$

which implies B= 4C (or C = B/4) that allows us to make the parameter D dependent on C and to rewrite the expression (66) as

$$D(C) = \frac{1}{2} \cdot 4C - 1 \quad = \quad 2C - 1. \tag{69}$$

It is easy to see that

- When C = 1 we also have D = 1, which implies that $B_{ns} = 4$, which is precisely the non–signaling case (when only the causal requirement in the transmission of signals is observed).
- Instead, we have the classical separability $B_c = 2$ when C = 1/2 and D = 0.
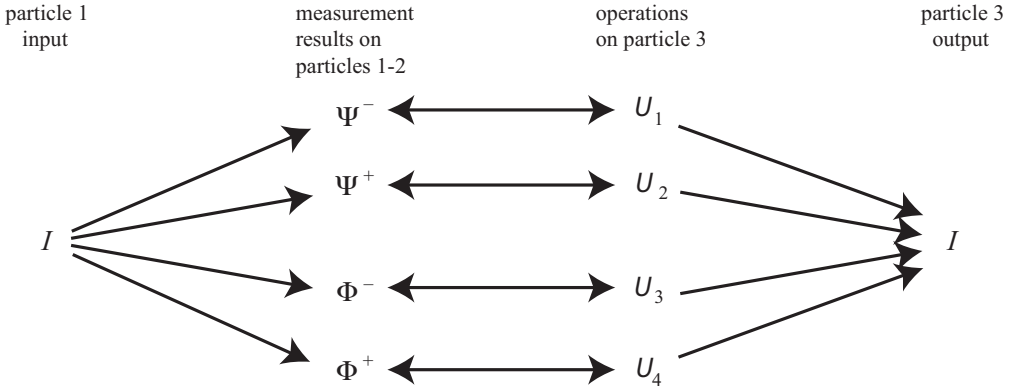- In the quantum case, we have $B_q = 2\sqrt{2}$ when $C = 1/\sqrt{2}$ and $D = \sqrt{2} - 1$.

To have a concrete model, let us briefly consider how teleportation works [7]. The eigenbasis of the Bell operator is given by

$$\left|\Psi^-\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(\left|\uparrow\right\rangle_1 \left|\downarrow\right\rangle_2 - \left|\downarrow\right\rangle_1 \left|\uparrow\right\rangle_2\right), \tag{70}$$

$$\left|\Psi^+\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(\left|\uparrow\right\rangle_1 \left|\downarrow\right\rangle_2 + \left|\downarrow\right\rangle_1 \left|\uparrow\right\rangle_2\right), \tag{71}$$

$$\left|\Phi^-\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(\left|\uparrow\right\rangle_1 \left|\uparrow\right\rangle_2 - \left|\downarrow\right\rangle_1 \left|\downarrow\right\rangle_2\right), \tag{72}$$

$$\left|\Phi^+\right\rangle_{12} = \frac{1}{\sqrt{2}}\left(\left|\uparrow\right\rangle_1 \left|\uparrow\right\rangle_2 + \left|\downarrow\right\rangle_1 \left|\downarrow\right\rangle_2\right). \tag{73}$$

Figure 4. Scheme of teleportation: the fact that each measurement result is mapped in a certain way to the input information I allows through the ebit that Alice classical instructs Bob about the kind of operation to be performed. It is a code.

The state of the three particles can be described as

$$|\Psi\rangle_{123} = \frac{1}{2} \Big[\big|\Psi^-\big\rangle_{12} (-c\,|\uparrow\rangle_3 - c'\,|\downarrow\rangle_3) + \big|\Psi^+\big\rangle_{12} (-c\,|\uparrow\rangle_3 + c'\,|\downarrow\rangle_3)$$
$$+ \big|\Phi^-\big\rangle_{12} (c\,|\downarrow\rangle_3 + c'\,|\uparrow\rangle_3) + \big|\Phi^+\big\rangle_{12} (c\,|\downarrow\rangle_3 - c'\,|\uparrow\rangle_3) \Big]. \tag{74}$$

It suffices an unitary operation (a mechanical instruction) to recover the information of Particle 1 on 3 once the Bell operator has been measured:

$$\hat{U}_1 = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}; \quad \hat{U}_2 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \hat{U}_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad \hat{U}_4 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{75}$$

We can assume an information causality principle which (in a teleportation protocol) relates to the amount of information that Bob can gain about a data set belonging to Alice, the contents of which are completely unknown to him [15]. Using all his local resources (which may be correlated with her resources) and allowing classical communication from Alice to Bob, the amount of information that the latter can recover is bounded by the information volume (n) of the communication. Namely, if Alice classically communicates n bits to Bob, the total information obtainable by Bob cannot be greater than n. Consider the easiest case in which a two-bit information has been classically transmitted. Then, Bob can at most recover this information (the instruction to perform a particular unitary operation out of four on his particle) and not the whole set of potential information from which Alice has selected the message she has sent. Then, in this simple case, the acquired information must be bound as

$$I \leq 2, \tag{76}$$

since $\lg 4 = 2$. However, if the sole causal (non–signaling) requirement would rule this information exchange, Bob would indeed recover 4 bits (the whole code mapping four outcomes into four operations), thus violating the information causality principle. Therefore, the amazing result that was found by Zukowski *et al.* is that a hypothetical theory which fulfill the requirements of causality but exceeds the Tsirelson bound, also violate the principle of information causality.

Let $\hat{O}_1$ and $\hat{O}_2$ be two observables on subsystems $\mathcal{S}_1$ and $\mathcal{S}_2$ of a system $\mathcal{S}$, respectively, and $\wp(o_a, \mathbf{a}; o_b, \mathbf{b})$ be the probability that the results of a measurement of $\hat{O}_1$ on $\mathcal{S}_1$ and $\hat{O}_2$ on $\mathcal{S}_2$ yield o_a and o_b when certain settings of the measurement apparata are $\mathbf{a}$ and $\mathbf{b}$, respectively. Since this assumption is of absolute generality, we not need to consider the specific spin model previously introduced.

According to Eberhard, the probability distribution of $\hat{O}_1$ (or $\hat{O}_2$), independently of the measurement operations on $\hat{O}_2$ (or $\hat{O}_1$), obtained by integrating or summing the probabilities $\wp(o_a, \mathbf{a}; o_b, \mathbf{b})$ over the possible outcomes o_b (or o_a), needs to be independent of the other setting $\mathbf{b}$ (or $\mathbf{a}$), that is, the two probabilities must depend on local settings only [9]:

$$\sum_{o_b} \wp(o_a, \mathbf{a}; o_b, \mathbf{b}) = \wp(o_a, \mathbf{a}) \; ; \quad \sum_{o_a} \wp(o_a, \mathbf{a}; o_b, \mathbf{b}) = \wp(o_b, \mathbf{b}). \tag{77}$$

According to Eberhard, if this requirement were violated, we would have a *causal non–local interdependence* between the two subsystems. Actually, a violation of the above requirement does not necessarily imply a non–local causal interconnection because there could still be a form of interdependence but satisfying the non–signaling requirement.

In order to prove the theorem, let

$$\hat{P}_{o_a, \mathbf{a}} = |o_a, \mathbf{a}\rangle \langle o_a, \mathbf{a}| \quad \text{and} \quad \hat{P}_{o_b, \mathbf{b}} = |o_b, \mathbf{b}\rangle \langle o_b, \mathbf{b}| \tag{78}$$

be the projectors on the state $|o_a, \mathbf{a}\rangle$ of subsystem $\mathcal{S}_1$ when the setting is $\mathbf{a}$ and the outcome $|o_a\rangle$, and on the state $|o_b, \mathbf{b}\rangle$ of subsystem $\mathcal{S}_2$ when the setting is $\mathbf{b}$ and the outcome $|o_b\rangle$, respectively, and $\hat{\rho}$ a density matrix which represents the compound state of $\mathcal{S} = \mathcal{S}_1 + \mathcal{S}_2$. The probability $\wp(o_a, \mathbf{a})$ that, by measuring the observable $\hat{O}_1$ on $\mathcal{S}_1$, we obtain the outcome $|o_a\rangle$ (or the eigenvalue o_a), is

$$\wp(o_a, \mathbf{a}) = \text{Tr}\left[\hat{P}_{o_a, \mathbf{a}} \hat{\rho}\right]. \tag{79}$$

After a measurement of $\hat{O}_1$ when the setting is $\mathbf{a}$ with result o_a we obtain the transformation

$$\hat{\rho} \mapsto \hat{\rho}' = \frac{\hat{P}_{o_a, \mathbf{a}} \hat{\rho} \hat{P}_{o_a, \mathbf{a}}}{\wp(o_a, \mathbf{a})}. \tag{80}$$

If we perform a second measurement on the second subsystem, the conditional probability of obtaining $|o_b\rangle$ (or o_b) by measuring $\hat{O}_2$ when the setting is $\mathbf{b}$, is given by

$$\wp'(o_b, \mathbf{b}|o_a, \mathbf{a}) = \text{Tr}\left[\hat{P}_{o_b, \mathbf{b}} \hat{\rho}'\right] = \frac{\text{Tr}\left[\hat{P}_{o_b, \mathbf{b}} \hat{P}_{o_a, \mathbf{a}} \hat{\rho} \hat{P}_{o_a, \mathbf{a}}\right]}{\wp(o_a, \mathbf{a})}. \tag{81}$$

For any events A and B, the classical probability calculus tells us that the probability of their joint occurrence can be expressed as

$$\wp(A, B) = \wp(A)\wp(B|A). \tag{82}$$

Therefore, the joint probability of obtaining the two results o_a and o_b given the settings $\mathbf{a}$ and $\mathbf{b}$, is given by combining Eqs. (79) and (81):

$$\begin{aligned} \wp(o_a, \mathbf{a}; o_b, \mathbf{b}) &= \wp(o_a, \mathbf{a})\wp'(o_b, \mathbf{b}|o_a, \mathbf{a}) \\ &= \wp(o_a, \mathbf{a})\frac{\text{Tr}\left[\hat{P}_{o_b, \mathbf{b}} \hat{P}_{o_a, \mathbf{a}} \hat{\rho} \hat{P}_{o_a, \mathbf{a}}\right]}{\wp(o_a, \mathbf{a})} \\ &= \text{Tr}\left[\hat{P}_{o_b, \mathbf{b}} \hat{P}_{o_a, \mathbf{a}} \hat{\rho} \hat{P}_{o_a, \mathbf{a}}\right]. \end{aligned} \tag{83}$$

Given these assumptions, we can obtain the following result that is in accordance with Eqs. (77):

$$\sum_{o_a} \wp(o_a, \mathbf{a}; o_b, \mathbf{b}) = \text{Tr} \sum_{o_a} \left[\hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \hat{P}_{o_a,\mathbf{a}} \right] = \text{Tr} \left[\hat{P}_{o_b,\mathbf{b}} \hat{\rho} \right] = \wp(o_b, \mathbf{b}). \tag{84}$$

To derive this result, first note that

$$\sum_{o_a} \text{Tr} \left[\hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \hat{P}_{o_a,\mathbf{a}} \right] = \text{Tr} \sum_{o_a} \left[\hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \hat{P}_{o_a,\mathbf{a}} \right]. \tag{85}$$

Moreover, I have made use of the cyclic properties of the trace, i.e., given any three arbitrary observables, we have

$$\text{Tr} \left[\hat{O}_1 \hat{O}_2 \hat{O}_3 \right] = \text{Tr} \left[\hat{O}_3 \hat{O}_1 \hat{O}_2 \right] = \text{Tr} \left[\hat{O}_2 \hat{O}_3 \hat{O}_1 \right]. \tag{86}$$

This property implies that

$$\text{Tr} \sum_{o_a} \left[\hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \hat{P}_{o_a,\mathbf{a}} \right] = \text{Tr} \sum_{o_a} \left[\hat{P}_{o_a,\mathbf{a}} \hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \right]. \tag{87}$$

Moreover, $\hat{P}_{o_a,\mathbf{a}}$ and $\hat{P}_{o_b,\mathbf{b}}$ commute because they pertain to different subsystems, and therefore we have

$$\text{Tr} \sum_{o_a} \left[\hat{P}_{o_a,\mathbf{a}} \hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \right] = \text{Tr} \sum_{o_a} \left[\hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \right]. \tag{88}$$

However, any orthogonal set of projectors $\{\hat{P}_{o_a,\mathbf{a}}\}$ satisfies the two properties $\hat{P}_{o_a,\mathbf{a}}^2 = \hat{P}_{o_a,\mathbf{a}}$ and $\sum_{o_a} \hat{P}_{o_a,\mathbf{a}} = \hat{I}$, from which we finally obtain

$$\text{Tr} \sum_{o_a} \left[\hat{P}_{o_b,\mathbf{b}} \hat{P}_{o_a,\mathbf{a}} \hat{P}_{o_a,\mathbf{a}} \hat{\rho} \right] = \text{Tr} \left[\hat{P}_{o_b,\mathbf{b}} \hat{\rho} \right]. \tag{89}$$

We may proceed in a similar way starting from the conditional probability $\wp'(o_a, \mathbf{a} | o_b, \mathbf{b})$ in order to derive the second equality (77).

What would happen in a world in which the quantum bound is violated but the locality (non–signaling) requirement is satisfied [3]? Let us now reformulate the quantum–mechanical Eqs. (77) in analogy with Eq. (47) as

$$\wp_{a,b}(1,1) + \wp_{a,b}(1,-1) = p_a(1) \quad \text{and} \quad \wp_{a,b}(1,1) + \wp_{a,b}(-1,1) = \wp_b(1), \tag{90}$$

and similarly for the other outcomes. This clearly shows that quantum mechanics requires a full independence of the settings (here expressed by the orientation $\mathbf{a}$), which need to be local operations performed in complete separation from other operations that could be performed elsewhere. Quantum correlations are indeed interdependencies of possible outcomes and not of settings. In other words, a violation of the quantum mechanical bound (and of the information causality principle) would imply that there are correlations between settings. If we consider the abstract forms (65) in which I have written the correlations entering in the CHSH inequality, we see that they are expressed in terms of pure conditional probabilities of the form $\wp(11|00)$ Following the customary approach in quantum mechanics (and our physical experience) we have naturally interpreted probabilities of this form as expressing e.g. the probability that both Alice and Bob get the output 1 given that they have both chosen the setting 0. In fact, if Bob knows which was the setting of Alice (whether 0 or 1) he is able to infer which was her outcome. On this procedure is indeed based quantum cryptography.

However, nothing forbids to interpret such a probability as telling us that, in a Bayesian inversion, Bob is able to predict that Alice has chosen the setting 1 once that he knows that he and Alice have obtained the outcome 0. This is still allowed by the non–signaling condition (47). As a matter of fact, given Eqs. (65), Bob is able to predict any setting of Alice if he knows her outcome:

- If Bob choses the setting 0, obtains the outcome 0 and knows that A has also obtained the outcome 0, he knows that she has chosen the setting 0;

- If Bob choses the setting 1, obtains the outcome 0 and knows that A has also obtained the outcome 0, he knows that she has chosen the setting 1;

- If Bob choses the setting 0, obtains the outcome 0 and knows that A has obtained the outcome 1, he knows that she has chosen the setting 0;

- If Bob choses the setting 1, obtains the outcome 0 and knows that A has obtained the outcome 1, he knows that she has chosen the setting 1;

- If Bob choses the setting 0, obtains the outcome 1 and knows that A has obtained the outcome 0, he knows that she has chosen the setting 0;

- If Bob choses the setting 1, obtains the outcome 1 and knows that A has obtained the outcome 0, he knows that she has chosen the setting 1;

- If Bob choses the setting 0, obtains the outcome 1 and knows that A has also obtained the outcome 1, he knows that she has chosen the setting 1;

- If Bob choses the setting 1, obtains the outcome 1 and knows that A has also obtained the outcome 1, he knows that she has chosen the setting 0.

6 Conclusion

In a world in which settings (and not only outcomes) are shared and so there would be kinds of non–local settings, this would imply that also information codification is shared. Indeed, it can be shown that information codification deals with the choice of a basis which in a measurement context is the choice of a setting. In other words, in a world showing hyper–correlations based on the sharing of settings, information codification would be no longer a local procedure.

The fact that this is forbidden justifies quantum information as a general theory of information since

- It satisfies and saturates the bounds that are imposed by the principle of information causality, and in so doing

- It also sets specific constraints on both the possible interdependencies and the possible interactions (also causal interconnections) in our universe.

The most general conclusion is that information can be defined as correlation among possible outcomes or events, as we have seen above for the mutual–information expression of entanglement.

References

[1] Aspect, A. , Dalibard, J., and Roger, G., Physical Review Letters **49**, 1804–1807 (1982).

[2] Auletta, G., Fortunato, M., and Parisi, G., *Quantum Mechanics* (Cambridge University Press, Cambridge, 2009).

[3] Auletta, G., Journal of Modern Physics **2**, 958–61 (2011).

[4] Auletta, G. and Wang, S.–Y., *Quantum Mechanics for Thinkers* (PanStanford Pub., Peking, 2013).

[5] Barnett, S. M. and Phoenix, S. J. D., Physical Review **A40**, 2404–2409 (1989).

[6] Bell, J. S., Physics **1**, 195–200 (1964).

[7] Bennett, C. H., Brassard, G., Crepeau, C., Jozsa, R., Peres, A., and Wootters, W. K., Physical Review Letters **70**, 1895–1899 (1993).

[8]   Bohm, D., *Quantum Theory* (Prentice-Hall, New York, 1951).

[9]   Nuovo Cimento **46B**, 392–419 (1978).

[10] Einstein, A., Podolsky, B., and Rosen N., Physical Review **47**, 777-780 (1935).

[11] Fano, U., Review of Modern Physics **29**, 74–93 (1957).

[12] Kwiat, P. G., Eberhard, P. H., Steinberg, A. M., and Chiao, R. Y., Physical Review **A49**, 3209–20 (1994).

[13] Masanes, L., Acin, A., and Gisin, N., Physical Review **A73**, 012112-1–9 (2006).

[14] Ou, Z. Y. and Mandel, L., Physical Review Letters **61**, 50–53 (1988).

[15] Pawłowski, M., Paterek, T., Kaszlikowski, D., Scarani, V., Winter, A., Zukowski, M. Z., Nature **461**, 1101–1104 (2009).

[16] Popescu, S. and Rohrlich, D., Foundations of Physics **24**, 379–85 (1994).

[17] Tsirelson, B. S., Letters in Mathematical Physics **4**, 93–100 (1980).