Mathematical and physical meaning of the Bell inequalities

Emilio Santos

Departamento de Física, Universidad de Cantabria, Santander, Spain

E-mail: santose@unican.es

Received 26 March 2016, revised 19 May 2016
Accepted for publication 24 May 2016
Published 29 June 2016

Abstract
It is shown that the Bell inequalities are closely related to the triangle inequalities involving distance functions amongst pairs of random variables with values \{0, 1\}. A hidden variables model may be defined as a mapping between a set of quantum projection operators and a set of random variables. The model is noncontextual if there is a joint probability distribution. The Bell inequalities are necessary conditions for its existence. The inequalities are most relevant when measurements are performed at space-like separation, thus showing a conflict between quantum mechanics and local realism (Bell’s theorem). The relations of the Bell inequalities with contextuality, the Kochen–Specker theorem, and quantum entanglement are briefly discussed.

Keywords: Bell inequalities, triangle inequalities in probability, contextuality, locality, entanglement

1. Introduction
More than 50 years have elapsed since John Bell derived his celebrated inequalities [1, 2]. These inequalities are fulfilled by all classical theories but are violated by quantum predictions for some quantum states. Thus they prove the impossibility of supplementing quantum mechanics with local hidden variables, a result known as Bell’s theorem. This is popularly stated saying that local realism is incompatible with quantum mechanics. Bell’s work had a relevant impact on the foundations of quantum mechanics. In particular, it has produced a renewed interest in the study of quantum entanglement. It has also led to increasingly refined experiments and it has been a stimulus for the development of quantum information theory. In the last year Bell’s theorem has received new attention due to the fact that loophole-free tests of the inequalities have been performed for the first time [3–5].

In a wide sense the name ‘Bell inequality’ is applied to any inequality that is a necessary condition for the existence of local realistic (or local hidden variables models (HVM)). For a
recent review see [6]. In this paper I will be concerned with a different, although related, definition of Bell inequalities, namely quadrilateral inequalities involving distances defined in (classical) probability theory for sets of dichotomic random variables. The fulfillement, or not, of the inequalities for similar distances defined for expectation values of projection operators in a Hilbert space allows us to know whether some subsets of operators may be represented, or not, in terms of (classical) probability theory. This connection is stressed in our derivation of the inequalities in section 2.

Actually, those properties are related to the distributivity of classical logic. Indeed, as Birkhoff and von Neumann pointed out [7], the essential difference between classical and quantum physics appears already at the level of the logic (or propositional calculus), which is distributive in the classical case but not in the quantum case. This relation of the Bell inequalities with the distributivity of the logic has been discussed since long ago [8–13].

The physical implications of the Bell inequalities will be briefly discussed in section 3. However, I shall touch but slightly on the experimental tests of the inequalities and not at all on the applications to quantum information theory. The main purpose of the section will be to comment on the relevance of the inequalities for the interpretation of quantum mechanics, in particular the question whether a realistic interpretation is possible. There are a lot of papers and many books concerned with this question, that will not be commented here (see, e.g. [14–18]). The subject is still controversial [19].

2. Mathematical content

From the mathematical point of view the Bell inequalities are necessary conditions for the existence of a joint probability distribution associated to a set of projection operators (projectors) and a density operator on a Hilbert space. A relevant result is that, if all projectors of the set commute with each other, then a joint probability distribution exists and all Bell inequalities hold true for any density operator. Although this section deals with mathematical properties, and therefore it is independent of any physical theory, I will use the widely known language of quantum mechanics. In particular in this section I shall name ‘observable’ any self-adjoint operator in the Hilbert space.

2.1. Probability distribution of an observable

For any pair \( \{ \hat{A}, \hat{\rho} \} \) of an observable and a density operator it is possible to define the probability density, \( f(a) \), of a random variable, \( a \), associated with the observable. In fact the standard rule about the expectation value, \( \langle \hat{A}^n \rangle \), of the \( n \)th power of an observable \( \hat{A} \) in the state given by the density operator \( \hat{\rho} \) allows getting the characteristic function, \( C(\zeta) \), of the associated probability distribution, that is

\[
\langle \hat{A}^n \rangle = \text{Tr}[\hat{A}^n \hat{\rho}] \Rightarrow C(\zeta) = \text{Tr}[\exp(i\zeta \hat{A}) \hat{\rho}],
\]

where \( \text{Tr}(\hat{x}) \) means the trace of the operator \( \hat{x} \). The probability density is obtained by Fourier transform, that is

\[
f(a) = \frac{1}{2\pi} \int \frac{d\zeta}{2\pi} \exp(-i\zeta a)C(\zeta),
\]

\[
= \frac{1}{2\pi} \int \frac{d\zeta}{2\pi} \exp(-i\zeta a)\text{Tr}[\exp(i\zeta \hat{A}) \hat{\rho}].
\]

**Example 1.** Observable \( \hat{A} \) having discrete nondegenerate spectrum and \( \hat{\rho} \) being a pure state, that is

2
Assuming that the eigenfunctions \{\ket{\psi_j}\} form an orthonormal basis we have

\[
\exp(i\zeta \hat{A}) = \sum_k \left( \frac{i\zeta \sum_j a_j \ket{\psi_j}\bra{\psi_j}}{k!} \right)^k,
\]

whence, taking equation (1) into account, the probability distribution is

\[
f(a) = \frac{1}{2\pi} \sum_j \int d\zeta \exp(-i\zeta a + i\zeta a_j) \langle \phi | \psi_j \rangle \langle \psi_j | \phi \rangle = \sum_j \delta(a - a_j) p_j, \quad p_j \equiv |\langle \phi | \psi_j \rangle|^2.
\]

This means that the distribution is discrete and it associates the probability \(|\langle \phi | \psi_j \rangle|^2\) to the eigenvalue \(a_j\).

**Example 2.** \(\hat{A}\) is a projector and \(\hat{\rho}\) arbitrary (not necessarily representing a pure state). \(\hat{A}\) fulfills

\[
\hat{A} = \hat{A}^2 \Rightarrow \hat{A}^n = \hat{A},
\]

for any natural number \(n\). Then we have

\[
f(a) = \frac{1}{2\pi} \int d\zeta \exp(-\zeta a) \text{Tr}[\hat{\rho} \exp(i\zeta \hat{A})],
\]

\[
= \frac{1}{2\pi} \int d\zeta \exp(-\zeta a + \zeta a_j) \text{Tr}(\hat{\rho} \hat{A}) = \delta(a - 1) \text{Tr}(\hat{\rho} \hat{A}),
\]

which means that \(\text{Tr}(\hat{\rho} \hat{A}) = p(a)\) is the probability that the projector takes the value 1.

### 2.2. Joint distribution for several observables

Generalizing equation (1) we might define the joint probability distribution of two observables, \(f(a, b)\), as follows

\[
f(a, b) = \frac{1}{4\pi^2} \int d\chi \int d\zeta \exp(-i \zeta a - i\chi b) \text{Tr}[\exp(i\zeta \hat{A} + i\chi \hat{B}) \hat{\rho}] = \frac{1}{4\pi^2} \int d\chi \int d\zeta \exp(-i \zeta a - i\chi b) \text{Tr}[\exp(i\zeta \hat{A} + i\chi \hat{B}) \hat{\rho}]. \tag{2}
\]

If the observables \(\hat{A}\) and \(\hat{B}\) commute with each other, the function \(f(a, b)\) is indeed the desired joint distribution and it has the properties of a classical joint distribution, as shown in the following. In fact in this case there is a complete set of orthonormal simultaneous eigenvectors of the two observables, that we will label \{\ket{\phi_j}\} (assuming for simplicity that the set is discrete). Thus we may use the resolution of the identity operator \(\hat{I}\) in terms of these eigenvectors in order to define the trace, i.e.

\[
\hat{I} = \sum_j \ket{\phi_j}\bra{\phi_j}, \quad \text{Tr}\hat{M} = \sum_j \langle \phi_j | \hat{M} | \phi_j \rangle.
\]
Hence equation (2) becomes, after some algebra
\[ f(a, b) = \sum_j P_j \rho_{aj}(a) \rho_{bj}(b), \quad P_j = \langle \phi_j | \hat{P} | \phi_j \rangle, \] (3)
where \( \rho_{aj}(a) \) is the probability distribution of \( a \) in the state \( j \), as given by equation (1).
Equation (3) has the same form as a classical joint probability distribution written in terms of the probability distributions of the individual variables \( a \) and \( b \). In particular if the sum in \( j \) contains only one term, the distributions \( \rho_{aj}(a) \) and \( \rho_{bj}(b) \) are uncorrelated.

In sharp contrast if \( \hat{A} \) and \( \hat{B} \) do not commute the function \( f(a, b) \) defined by equation (2) may not be semidefinite positive. For instance in a system consisting of a single particle (without spin) in one dimension, the position and momentum observables, \( \hat{x} \) and \( \hat{p} \), do not commute and equation (2) leads to
\[
W(x, p) = \frac{1}{4\pi^2} \int d\xi \int d\chi \exp(-i\chi x - i\xi p) \text{Tr}\{\exp(i\xi \hat{x} + i\chi \hat{p})\hat{\rho}\},
\]
which is the Wigner function associated to the state \( \hat{\rho} \). As is well known the Wigner function is not always positive. We might try other choices for the ordering of the operators \( \hat{x} \) and \( \hat{p} \), but none fully satisfactory has been found. Thus the question arises, is it always possible to find a joint (positive semidefinite) probability distribution for several observables and a given density operator? The answer is affirmative if all observables commute, but negative in general. In order to prove that assertion let us start studying joint probability distributions in the mathematical theory of probability.

2.3. Joint distribution for several dichotomic random variables

What I will present in the following is well known for any mathematical measure, in particular a probability distribution, defined on a set. In mathematics a measure on a set is a mapping of subsets on nonnegative numbers, with some topological restrictions on the subsets if the set is not finite. But for our purposes it is enough to study probability distributions on a finite set \( \{a_j, j = 1, 2, \ldots n\} \) of random variables having values \( \{0, 1\} \).

The question that I will try to answer is the following. Given the probabilities \( p_j(a_j) \) that \( a_j = 1 \), and the probabilities \( p_k(a_k) \) that \( a_k = a_l = 1 \) for all \( j, k \), we want to know whether these probabilities are the marginals of some joint probability distribution for all the variables. A well known result of probability theory is the following lemma: a necessary condition for the existence of a joint probability distribution is that all triangle inequalities of the form
\[ d(a_j, a_k) + d(a_k, a_l) \geq d(a_j, a_l) \] (4)
hold true. The ‘distance function’ \( d(a_j, a_k) \) is defined by
\[ d(a_j, a_k) = p(a_j, \bar{a}_k) + p(\bar{a}_j, a_k) = p_1(a_j) + p_1(a_k) - 2p_2(a_j, a_k), \] (5)
where \( \bar{a}_j = 1 - a_j \) is the variable that takes the value 1(0) when \( a_j \) takes the value 0(1). The distance function fulfils
\[ 0 \leq d(a_j, a_k) \leq 1, \quad d(a_j, a_j) = 0, \quad d(a_j, \bar{a}_j) = 1. \] (6)
We see that the distance is zero when the two variables are maximally correlated (they might be considered the same variable) and it is unity if they are maximally anticorrelated. The proof of the lemma is easy. We consider three random variables \( \{a, b, c\} \) with values \( \{0, 1\} \) and use a Venn diagram where the random variables are represented by circles and the areas are proportional to the probabilities. Thus it is easy to check that
\[ p(ab) + p(ab) + p(ac) + p(ac) \geq p(bc) + p(bc). \]  

(7)

In algebraic form the derivation is as follows:

\[ p(ab) = p(abc) + p(ab\tilde{c}), \quad p(ab) = p(\tilde{a}bc) + p(ab\tilde{c}), \]
\[ p(ac) = p(ac\tilde{b}) + p(ac\tilde{b}), \quad p(ac) = p(acb) + p(ac\tilde{b}). \]

Adding the four equations gives

\[ p(ab) + p(ab) + p(ac) + p(ac) = 2p(ab\tilde{c}) + 2p(ac\tilde{b}) + p(bc) + p(bc), \]

whence equation (7) follows if the first two terms on the right are omitted. Hence a triangle inequality similar to equation (4) follows.

In the following I show that the triangle inequalities equation (4) are closely related to the Bell inequalities [1].

2.4. Random variables representations

Let us consider a finite set, \( P \equiv \{\hat{a}_j\} \), of projectors (with eigenvalues 1 and 0) and a density operator \( \hat{\rho} \). A set \( R \equiv \{\hat{a}_j\} \) of random variables with values \{0, 1\} is here defined as a ‘random variables representation’ (RVR) of the pair \( \{P, \hat{\rho}\} \) if there is an one to one mapping \( P \rightarrow R \) such that, for any subset of \( P \) involving only projectors that commute with each other, the following equality holds true

\[ \text{Tr}[\hat{\rho}(\hat{a}_j\hat{a}_k \ldots \hat{a}_i)] = p(a_ja_k \ldots a_i), \]  

(8)

where \( p(a_j, a_k \ldots a_i) \) is the probability that all variables \{\( a_j, a_k \ldots a_i \)\} take the value 1. We assume that both sets \( P \) and \( R \) are complete in the sense that if \( a_j \in P \) also \( \hat{a}_j \in P \), \( \hat{I} \) being the identity operator. Hence if \( a_j \in R \) also \( \hat{a}_j \in R \).

The quantities on the right side of equation (8) have all the properties of probabilities. In fact they are positive or zero and the marginals are consistent with the mapping \( P \rightarrow R \), that is

\[ p(a_k \ldots a_i) = p(a_ja_k \ldots a_i) + p(\hat{a}_ja_k \ldots a_i), \]
\[ = \text{Tr}[\hat{\rho}(\hat{a}_j\hat{a}_k \ldots \hat{a}_i)] + \text{Tr}[\hat{\rho}(\hat{I} - \hat{a}_j)\hat{a}_k \ldots \hat{a}_i]), \]  

(9)

assuming that the operators \{\( \hat{a}_j, \hat{a}_k \ldots \hat{a}_i \)\} commute pairwise (as was stated in equation (8)). In conclusion a ‘RVR’ exists for any pair \( \{P, \hat{\rho}\} \).

**Example.** Let us consider four projectors \( \{\hat{a}_1, \hat{b}_1, \hat{a}_2, \hat{b}_2\} \) such that \( \hat{a}_j \) commutes with \( \hat{b}_k \) for any \( j, k \in \{1, 2\} \), but neither \( \hat{a}_1 \) commutes with \( \hat{a}_2 \) nor \( \hat{b}_1 \) with \( \hat{b}_2 \). Thus equation (8) allow defining 24 probabilities, namely

\[ p(a_1), p(b_k), p(\hat{a}_j), p(\hat{b}_k), p(a_jb_k), p(a_j\hat{a}_k), p(a_j\hat{b}_k), p(a_j\hat{b}_k). \]  

(10)

These probabilities, in addition to being positive, fulfil the conditions equation (9) for the marginals. For instance we should have, for any \( j \) and \( k \)

\[ p(a_j) = p(a_jb_k) + p(a_j\hat{b}_k). \]

But there are many probabilities that are not defined by equation (8), namely those involving two or more noncommuting operators. For instance the following probabilities cannot be derived that way

\[ p(a_1a_2), p(b_1b_2), p(a_1a_2b_k), p(b_1b_2a_j), p(a_1a_2b_1b_2). \]  

(11)
The probabilities equation (10) define a RVR, while the quantities equation (11) do not enter in that RVR or even they do not exist with all required conditions (e.g. being positive and fulfil relations similar to equation (9)).

There are two different kinds of RVR. A RVR of the pair \( \{ P, \hat{\rho} \} \) is ‘complete’ if there is a joint probability distribution for all random variables of \( R \) such that the probabilities given by equation (8) are marginals of the joint distribution. It is incomplete if such a joint probability does not exist.

In the above example a complete RVR would require the existence of the 16 probabilities

\[
p(a_1, a_2, b_1, b_2), p(a_1, a_2, b_1, \overline{b}_2), \ldots, p(\overline{a}_1, \overline{a}_2, \overline{b}_1, \overline{b}_2),
\]

whence all other probabilities could be obtained as marginals. However some of these could not be derived from equation (8), for instance this would be the case for the probabilities equation (11) in our example. On the other hand it may be that probabilities equation (12) do not exist with the condition that all probabilities derivable from equation (8) are marginals of these equation (12). In this case the RVR will not be complete. The complete and incomplete RVR correspond to the so-called noncontextual and contextual hidden variables theories in quantum mechanics, respectively, as will be clarified in the next section.

It is not difficult to generalize RVR to observables not necessarily projectors. Actually any set of observables may be written (or accurately approximated) in terms of projectors.

2.5. Quadrilateral inequalities as tests of completeness

An interesting task is to find conditions that allow to discover whether the RVR of a pair \( \{ P, \hat{\rho} \} \) is complete. In view of the subsection 1.3 a necessary condition would be that all triangle inequalities involving probabilities defined by equation (8) are fulfilled. However this criterion is useless. In fact for any three commuting projectors the triangle inequality holds true either if the RVR is complete or not. In fact if the operators commute they are compatible and quantum mechanics predicts the existence of a joint probability distribution. Hence the triangle inequality follows as commented in section 2.3. On the other hand if two projectors do not commute their distance function cannot be defined because the joint probability is not given by equation (8). In this cases one of the sides of the triangle inequality is unknown.

The solution to the problem is to consider quadrilateral inequalities as follows. If the RVR is complete, then a joint probability is defined for any three random variables of the model. As a consequence, choosing any four variables \( \{ a_1, a_2, b_1, b_2 \} \) the following two triangle inequalities hold true

\[
d(a_1, b_1) \leq d(a_1, a_2) + d(a_2, b_1),
\]

\[
d(a_1, a_2) \leq d(a_1, b_2) + d(b_2, a_2),
\]

although it may be that some of the distances cannot be obtained via equation (8) (but they might be obtained as marginals of the assumed joint probability distribution as the RVR is complete). The addition of these two inequalities gives the quadrilateral inequality

\[
d(a_1, b_1) \leq d(a_1, b_2) + d(b_2, a_2) + d(a_2, b_1),
\]

which is useful because all four distances may be obtained via equation (8). As a consequence a necessary condition for a RVR to be complete is that all quadrilateral inequalities like equation (13) hold true. The inequalities are not trivial if some of the projectors associated with the variables do not commute. For instance, in our example above the projectors \( \hat{a}_1 \) and \( \hat{a}_2 \) do not commute.

The quadrilateral inequalities are closely related to the most typical Bell inequalities [1, 2]. In fact, taking equation (5) into account, (13) may be written in the form
which is an inequality derived by Clauser and Horne [20] in 1974 from different (but equivalent) assumptions as those used here. Most popular is the inequality derived in 1969 by Clauser, Horne, Shimony and Holt (CHSH) [21] which is acually equivalent to equation (14).

In fact, let us consider dichotomic variables \( \{A_j, B_k\} \) with values \{-1, 1\}, related to the variables \( \{a_j, b_k\} \), with values \{0, 1\}, by means of

\[
a_j = \frac{1}{2} (A_j + 1), \quad b_k = \frac{1}{2} (B_k + 1).
\]

Now for a random variable \( a \) with values \{0, 1\} the probability \( p(a) \) of taking the value 1 is identical to the expectation value \( \langle a \rangle \). Thus the inequality equation (14) may be rewritten

\[
\langle a_i \rangle + \langle b_i \rangle \geq \langle a_i b_i \rangle + \langle a_2 b_1 \rangle + \langle a_1 b_2 \rangle - \langle a_2 b_2 \rangle,
\]

which, taking equation (15) into account, leads to

\[
\frac{1}{2} \langle A_1 + 1 \rangle + \frac{1}{2} \langle B_1 + 1 \rangle \geq \frac{1}{4} \langle (A_1 + 1)(B_1 + 1) \rangle + \frac{1}{4} \langle (A_2 + 1)(B_1 + 1) \rangle
\]

\[ + \frac{1}{4} \langle (A_1 + 1)(B_2 + 1) \rangle - \frac{1}{4} \langle (A_2 + 1)(B_2 + 1) \rangle.
\]

Hence simple algebra gives

\[
2 \geq \langle A_1 B_1 \rangle + \langle A_2 B_1 \rangle + \langle A_1 B_2 \rangle - \langle A_2 B_2 \rangle,
\]

which is the CHSH inequality [21].

I shall point out that in papers dealing with the CHSH inequality it is frequent to call ‘correlation’ the expectation value of a product of observables like \( \langle AB \rangle \). The name does not agree with the standard one in mathematical statistics, where the correlation between two random variables, \( A \) and \( B \), is usually defined by the dimensionless quantity

\[
\text{Corr}(A, B) \equiv \frac{\langle AB \rangle - \langle A \rangle \langle B \rangle}{\sqrt{\langle A^2 \rangle - \langle A \rangle^2} \sqrt{\langle B^2 \rangle - \langle B \rangle^2}}.
\]

An inequality similar to CHSH with the correlation equation (17) substituted for the expectation of the product, \( \langle AB \rangle \), may be violated by classical (hidden variables) models. This fact has lead some authors to misunderstand, and criticize, Bell’s work.

### 3. Physical implications

The physical meaning of the Bell inequalities has been widely discussed and a detailed survey is beyond the scope of this paper. Here I will give a summary of the fundamental ideas. The interested reader may consult the review by Brunner et al [6]. At a difference with the previous (mathematical) section, here it is necessary to distinguish between an ‘observable’ of a physical system, say \( A \), and the ‘operator’ associated to it in the Hilbert space formalism of quantum mechanics, say \( \hat{A} \).

#### 3.1. Observables and elements of reality

**3.1.1. From Einstein, Podolsky and Rosen (EPR) to hidden variables.** The starting point of Bell’s work was the celebrated article by EPR [22]. Indeed the title of the pioneer Bell publication made reference to that paper [1]. EPR article begins with what may be seen as a definition of *epistemological realism*, that is the conditions for a realistic interpretation of
Any serious consideration of a physical theory must take into account the distinction between the objective reality, which is independent of any theory, and the physical concepts with which the theory operates. These concepts are intended to correspond with the objective reality, and by means of these concepts we picture this reality to ourselves [22]. Then the authors present a state of a pair of particles entangled in positions and momenta. They stress the quantum prediction that, even if the particles are well separated, a measurement of position (momentum) on the first particle allows an instantaneous knowledge of the position (momentum) of the second one. Thus we should attach to the second particle different quantum states without in any way making a physical action on it. The authors conclude that either there is an action at a distance (nonlocality) or quantum mechanics associates two different quantum wavefunctions to the same actual state of the second particle. Thus the EPR paper proves that quantum mechanics is either nonlocal or incomplete, and the authors considered the second alternative more plausible. Bell showed that if quantum mechanics is incomplete (and it admits hidden variables) it is also incompatible with local realism (Bell’s theorem). Therefore the question of completeness is irrelevant for the question of compatibility. In a soon reply to EPR, Bohr supported completeness and a kind of wholeness of quantum mechanics that implies nonlocality [23]. Thus Bell’s theorem is currently interpreted as a vindication of Bohr against Einstein.

Many physicists do not accept EPR realism, they believe that ‘pictures of the objective reality’ are irrelevant for physics where only the agreement of the theoretical predictions with the empirical evidence is required. However other physicists search for a clear picture of the reality, which is not offered by the quantum formalism. A standard approach to a picture is the introduction of the so-called ‘hidden variables theories’, that might reduce quantum theory to a stochastic theory with classical flavour. The main achievement of Bell’s work was to put constraints on the HVMs that are possible, in particular proving the impossibility of local models.

### 3.1.2. Measurement in macro and microsystems

Essential in any branch of natural science are the experiments, consisting of just observations (as in astronomy) or preparations followed by measurements (as in laboratory work). Thus it is worth to study the differences between the experiments dealing with macroscopic and microscopic systems.

The most popular difference is of course the fact the some quantities may possess values only within a discrete set in microphysics. Hence the name quantum introduced by Planck. However more important for the hidden variables problem is the empirical fact that macroscopic systems are perturbed but slightly by measurements, while in the microscopic domain the state of the system is usually dramatically changed. One consequence is that the uncertainties in the results of the measurements in the macroscopic domain are usually small as compared with the result obtained. Thus it is assumed that the perturbation may be reduced indefinitely and therefore ignored in the formulation of classical theories, dealing with macroscopic systems. The logical consequence is to attach the observable quantities to the physical systems themselves, whence the belief that measurements just reveal intrinsic properties of the system under study. Thus classical theories dispense with a theory of measurement. In fact the process of measurement may be seen as just an interaction between the measured system and the measuring setup, and therefore a process to be studied within the theory itself.

In sharp contrast the measurement may give rise to a large perturbation in quantum physics. A consequence of this fact is that, while the simultaneous measurement of two observables on a macroscopic system perturb but slightly each other, the perturbation may be very relevant in joint measurements of microscopic systems. Therefore in classical theories
we may assume that all observables of a system are compatible, in the sense that they may be measured simultaneously. (In macroscopic systems there are examples of observables that cannot be measured simultaneously, but they are scarce and irrelevant for the formulation of classical theories). In sharp contrast for microscopic systems there are incompatible observables that cannot be measured with the same experimental setup. The quantum mechanical formalism incorporates this fact in a fundamental form, namely two observables are compatible if and only if the associated operators commute.

3.1.3. Incompatible observables. The existence of incompatible measurements in microscopic systems leads logically to the possibility that the results of joint measurements of two (or several) observables depend on the context. For instance, the measurement of the position of a particle requires a different experimental context than the measurement of momentum. Therefore the joint measurement of position and spin requires a context different from the one for a joint measurement of momentum and spin. Thus we might expect that the spin measurement (in the same state, that is after the same preparation procedure) gives different results in the two cases. This is not so in our example (in fact the result for the spin measurement would the same in both cases) but different results are actually obtained in more complex examples, involving measurements of more than two observables. The property is a consequence of the quantum formalism and it is formally stated by the Kochen–Specker theorem (see below).

3.1.4. Are there objective properties? As said above in classical physics it is assumed that measurements just reveal properties of the physical systems, existing with independence of observations. That is, in any actual (pure) state of a classical system all observable quantities have a well defined value. A mixed state (or statistical ensemble of systems) is associated to a joint probability distribution of the observables. This is not true in quantum physics, where it is not always possible to associate a joint probability distribution to all observables in a specific quantum state (i.e. to a given preparation procedure of the system). This fact has led some people to claim that, in contrast with classical physics, quantum systems have no properties whenever they are not measured. That is the observable properties ‘are created’ or ‘emerge’ as a result of measurements. This claim seemed bizarre to many people including Einstein, who criticized it with the rhetorical sentence: is the moon there when nobody looks?

In order to clarify the subject it is necessary to distinguish between the objective properties or ‘elements of reality’ [22], that exist independently of any observation, and the ‘observable quantities’ which appear only as a consequence of the observation or measurement. This is true in both classical and quantum physics. The point may be seen with an example. In playing dice, every die has spots on the six faces with the numbers 1–6. These are elements of reality, always present in the faces of the die. But if we play dice the value of our observables is only one number per die, namely the one printed in the upper face when the die becomes at rest on the table. The value of the observable, but not the objective property, is ‘created’ by the action of playing dice, i.e. the experiment. In summary we should distinguish between objective properties and observables both in classical and quantum physics. However the connection between both is rather obvious in classical physics but less clear in quantum physics.

I conclude that the differences between classical and quantum theories do not provide sufficient support for the claim that objective properties (or elements of reality) do not exist in quantum systems or that realistic interpretations of quantum physics are not possible, that is interpretations in terms of ‘pictures of the reality’ [22]. These pictures might be achieved by means of ontic or HVMs of quantum systems, to be studied in the next subsection.
3.2. Ontic models in classical and quantum physics

Following the advice of the initial paragraph of the EPR paper we may assume that a physical system is at any time in some 'ontic state’, that is ‘a real physical state not necessarily completely described by any theory (e.g. quantum), but objective and independent of the observer’ [24]. The observable properties of the system derive from the ontic state at the time of observation [25].

The standard classical assumption that measurements just reveal existing properties may be formalized stating that the observed result, \( a \), depends on the ontic state, \( \lambda \), of the system and the measuring setup, \( A \), appropriate for a given observable quantity. That is the observed result, \( a \), will be a function

\[
a = a(\lambda, A),
\]

(18)

At this moment I stress that this relation offers a ‘picture of the objective reality’ as demanded by EPR, in the sense that it provides a causal connection between system plus measuring setup and the result obtained. In contrast the assumption that the results of the measurement ‘emerge’ from the act of measurement would not provide any clear picture. This is true specially if it is assumed that the randomness of the result is ‘essential’ in the sense that it is not caused by the incomplete control of all the parameters in the measurement. Equation (18) explains the randomness of the result from the incomplete control of either the preparation procedure (leading to a statistical distribution of ontic states, \( \lambda \)) or the measuring setup, or both.

The assumption that the results of all observations on a system derive from functions like \( a(\lambda, A) \), equation (18), would allow getting the joint probability of any set of observables, \( \{A, B, ... C\} \). Without loss of generality we may consider that the set consists of observables with values \( \{0, 1\} \), because any observable may be defined in terms of yes–no questions. In this case any function like \( a(\lambda, A) \) takes on the value 1 for some ontic states \( \lambda \) and 0 for other states. Thus the joint probability distribution reduces to the knowledge of all expectations of products of observables, that is

\[
\langle AB ... C \rangle = \int f(\lambda) a(\lambda, A)b(\lambda, B)...c(\lambda, C)d\lambda.
\]

(19)

Here we have assumed that \( \lambda \) is a numerical parameter (or a set of parameters) attached to the ontic state so that the integral in \( \lambda \) makes sense. The function \( f(\lambda) \) gives the probability distribution of the ontic states in cases where we use a statistical ensemble of states (a ‘mixed state’) due to incomplete information. For a pure state the function \( f(\lambda) \) is zero except for some value of \( \lambda \) (the function \( f(\lambda) \) will have the form of a Dirac’s delta) which reduces the right side of equation (19) to a simple product (no integration is required).

The conclusion is that fixing the functions equation (18) for all possible observables of the system would determine their joint probability distribution. The reciprocal is also true. In fact, it is enough to identify the parameter \( \lambda \) with the set of values of all observables.

The above construction, equation (19), may be called ‘ontic model’ [24] and extended, with some modifications, to quantum physics. For some people the appeal to ontic states is a philosophical (methaphysical) assumption that should not enter physics and prefer to treat \( \lambda \) as a parameter in a model for an experiment and name it ‘hidden variable’. Thus the construction may be called HVM, which is the common name in the context of quantum physics.

Let us analyze whether HVM are possible in quantum mechanics. We consider a quantum system (pure or mixed), a state given by the density operator \( \hat{\rho} \), and the set of projection operators \( \{\hat{A}, \hat{B}, ...\} \). The projectors, fulfilling \( \hat{A}^2 = \hat{A} \), are the operators associated to dichotomic observables with values \( \{0, 1\} \). We may try to construct a HVM (or ontic...
model), as in the classical case, via a set of functions \( \{a(\lambda, A), b(\lambda, B), \ldots\} \) such that an equation similar to equation (19) holds true. At a difference with the classical case, quantum mechanics does not predict expectations for all products of observables (or more correctly, the expectations depend on the ordering of the operators), but only for products of ‘compatible’ observables, that is those whose associated operators commute with each other. For the sake of clarity I will start with a simple example.

Let us consider a state \( \hat{\rho} \) of a quantum system and a set of three observables \( \{A, B, C\} \) each with values \( \{0, 1\} \), represented in quantum mechanics by the projection operators \( \{\hat{A}, \hat{B}, \hat{C}\} \). We assume that \( \hat{A} \) commutes with both \( \hat{B} \) and \( \hat{C} \), but \( \hat{B} \) does not commute with \( \hat{C} \). (A physical instance would consists of the observables position, momentum and spin of a particle, although here we consider for simplicity that the possible values of every observable are 0 and 1.) In this case quantum mechanics predicts the following expectations, to be obtained via equations similar to equation (8)

\[
\langle A \rangle, \langle B \rangle, \langle C \rangle, \langle AB \rangle, \langle AC \rangle,
\]

but it does not predict the expectations (involving incompatible observables, which cannot be measured simultaneously)

\[
\langle BC \rangle, \langle ABC \rangle.
\]

We shall assume that commuting (noncommuting) operators correspond to compatible (incompatible) observables. Therefore the expectations \( \langle A \rangle, \langle B \rangle, \langle AB \rangle \) may be measured simultaneously in a context, say 1, and the observables \( \langle A \rangle, \langle C \rangle, \langle AC \rangle \) may be measured in another context, say 2, but it is not possible to measure the five expectations equation (20) in the same context (otherwise the observables \( A, B, C \) would be compatible). It is trivial to get functions \( \{a_1(\lambda, A), b_1(\lambda, B)\} \) giving the expectations \( \langle A \rangle, \langle B \rangle, \langle AB \rangle \) via equation (19). This would provide a HVM (or ontic model) for the context 1. Similarly we may obtain another HVM for the context 2 via the functions \( \{a_2(\lambda, A), c_2(\lambda, C)\} \). The two HVM together may be called a single HVM, that is noncontextual (contextual) if \( a_1(\lambda, A) = a_2(\lambda, A) \) (if \( a_1(\lambda, A) \neq a_2(\lambda, A) \)). The example shows that noncontextual are a particular kind of HVM. Studying whether they are possible for all pairs \( \{\hat{\rho}, P\} \) of a state, \( \hat{\rho} \), and a set of observables, \( P \), of a physical system is the aim of the next subsection.

It may be realized that a noncontextual HVM corresponds to a complete RVR of the pair \( \{\hat{\rho}, P\} \), as defined in the mathematical section 2. Indeed the knowledge of all expectations like equation (19) determines a joint probability distribution of all observables. I stress that the commutativity of all operators with each other is a sufficient condition for the existence of a noncontextual HVM, but it is not necessary.

### 3.3. Contextual hidden variables. Kochen–Specker theorem

In a Hilbert space of dimension 3 or more, noncontextual hidden variables theories are not possible in general.

The result is called Kochen–Specker theorem [26] for the authors who proved it in 1967, although it had been proved independently in 1966 by Bell [27, 28]. (It may be shown that for spaces of dimension 2 noncontextual HVM are always possible.) The original Kochen–Specker proof will not be given here, but once we know that noncontextual HVM correspond to complete RVRs, a simple proof consists of finding an example violating a Bell inequality. Such examples are given by equations (14) or (16). In fact in that case it is not possible a complete RVR, and therefore a noncontextual HVM. The proof of the Kochen–Specker theorem via Bell inequalities, here presented, has the advantage of suggesting empirical tests.
Indeed experiments have been performed showing the violation of noncontextual HVM. In particular empirical Bell tests refute noncontextual HVM (and even the particular class of noncontextual that are also local HVM if the experiments block the locality loophole). What makes necessary the use of contextual HVM is the fact that the operators associated to observables in quantum theory do not always commute with each other. But I shall point out that sometimes a noncontextual HVM is possible for systems where not all such operators commute. Indeed the possibility of noncontextual HVM is not a property of the system but it depends also on the quantum state (here we consider all possible states, including those not pure represented by density operators). For instance spin components of a spin-1/2 particle in different directions do not commute, but the state consisting of a mixture of spin up and spin down (along an arbitrary direction) with equal weights does admit a noncontextual HVM where the spin state is represented by a spherically symmetrical distribution of vectors. Indeed the density operator representing the state is a multiple of the identity. Also for some pure states noncontextual HVM are possible even if not all operators involved commute, provided that we do not consider all possible observables but only a subset. For instance in the ground state of a harmonic oscillator the Wigner function is positive whence it might be interpreted as a joint probability distribution of positions and momenta, thus defining a noncontextual HVM for these observables. However in this state of the oscillator no noncontextual HVM is possible if we include the Hamiltonian in addition to positions and momenta [29].

Noncontextual ontic models are always possible in classical physics, whence the impossibility of such models is a signal of quantum behaviour. Thus the Bell inequalities may be used, and have been used, in order to discriminate quantum versus classical phenomena. If a Bell inequality is violated the phenomenon cannot be interpreted via noncontextual HVM and it should be considered specifically quantum.

3.4. Bell’s theorem

Both in classical mechanics and in ordinary life the correlations between distant systems are assumed to derive from a common past. For instance the similarity between twins (possibly living in different cities) is a correlation between distant bodies. It is an obvious consequence of the common origin, which might be formalized stating

$$\langle AB \rangle = \int f(\lambda) a(\lambda, A)b(\lambda, B)d\lambda,$$  \hspace{1cm} (21)

where A and B label some feature of the twins, e.g. colour of the eyes, and f(\lambda) represents the (common) genetic code. But correlations cannot be interpreted in terms of equation (21) or (19) whenever a Bell inequality is violated. Hence a fundamental consequence of Bell’s work is to show that in nature there might be correlations between distant bodies not deriving from a common past; this would be the case if the correlations violated a Bell inequality. The result is a consequence of the laws of (standard) probabilities and it is therefore independent of the existence of quantum mechanics.

Quantum mechanics enters because it predicts the existence of such correlations, a result known as ‘Bell’s theorem’. Bell introduced HVMs, called local, that are partially contextual. He stressed that contextual HVM are not too strange provided contextuality means influence of the context (e.g. the measuring equipment) near the system under study. However it would be rather strange if the influence is at a distance, as in the experiments of the EPR type [22]. In those experiments two particles are produced in a source, each particle travels in a different direction and when they are far appart two different agents, Alice and Bob, measure their respective properties. In this case it is unplausible to assume that there may be an influence of the context of Alice (Bob) on the result of the measurement made by Bob (Alice). But those
influences should exist if a Bell inequality is violated by the four expectations 
\( \langle A_1 B_1 \rangle, \langle A_1 B_2 \rangle, \langle A_2 B_1 \rangle, \langle A_2 B_2 \rangle \), where \( A_j (B_k) \) are observables measured by Alice (Bob).

It may be realized that the relevance of Bell’s theorem is greater than just to refute a class of hidden variables theories of quantum mechanics. It proves that local realistic models of natural phenomena are not compatible with quantum mechanics. Indeed Bell himself reinterpreted the correlation between two measurements, one by Alice the other one by Bob within relativity theory [2]. To do that he considered the set of variables \( \lambda \) to be the union of two sets, \( \lambda_a \) and \( \lambda_b \), consisting each of all events in the past light cone of the measurement performed by Alice and Bob, respectively. Therefore the violation of a Bell inequality with space-like separated measurements, in the sense of relativity theory, would imply that the correlation does not derive from the intersection of the said past light cones.

Thus Bell’s theorem seems to prove the incompatibility of quantum mechanics with relativity theory. The contradiction looks dramatic, but most authors think that there is no real contradiction because quantum mechanics does not allow sending superluminal signals from Alice to Bob (or from Bob to Alice). Actually for many authors what is proved by Bell’s theorem is the nonexistence of hidden variables or maybe the unavoidability of an instrumentalistic interpretation of quantum mechanics, excluding realistic interpretations. For this reason it is frequent to state Bell’s theorem as proving the incompatibility between local realism and quantum mechanics.

In view of Bell’s theorem many experiments have been performed during the latter forty years in order to test empirically local realism against quantum mechanics. The experiments confirmed quantum mechanics but, due to the existence of loopholes, they were unable to refute local realism. Only recently has it been possible to perform loophole-free experiments [3–5].

### 3.5. Entanglement, the characteristic trait of quantum mechanics

One of the main difficulties for a realistic understanding of the quantum formalism derives from the phenomenon of entanglement. Known since the early days of quantum mechanics it was crucial in the EPR argument mentioned above [22]. Its relevance was emphasized by Schrödinger in a celebrated paper [30]. He stated that entanglement ‘is not one but the characteristic trait of quantum mechanics’. The problem is that getting a clear physical picture of entanglement is difficult. In fact, its standard definition is not made in terms of physical concepts, but requires abstract mathematics: ‘Several physical systems are entangled if their joint wavefunction (or statevector) cannot be written as the product of wavefunctions, one for each system’. As a consequence the common wisdom is that entanglement cannot be pictured in terms of physical models. Actually a physical interpretation has been proposed that considers entanglement as a correlation between the fluctuations of the vacuum fields acting on different systems (maybe separated at a long distance) [25]. However this interpretation is not popular.

One of the difficulties put by the phenomenon of entanglement may be seen with the following argument. Let us consider an entangled (singlet) state of two spin-1/2 particles. It may be represented (ignoring normalization) in the form

\[
\Psi = (\uparrow \downarrow - \downarrow \uparrow),
\]

which may be expressed in words saying that either electron \#1 has spin up and electron \#2 spin down or electron \#1 has spin down and electron \#2 spin up. The strange fact is that the state equation (22) is pure, which is interpreted as a complete knowledge of the whole system, and nevertheless we have no information about the spin of every electron. We only know that
there are 50%–50% probabilities for the two possibilities, up and dawn, and in fact these would be the probabilities of the results if measurements were performed. That is, according to the standard quantum view we have complete information about the whole two-spin system but no information about every part. (It is like a student claiming to have complete knowledge of a whole subject matter but no knowledge at all about every lesson). This is paradoxical because in ordinary life complete information just means information about every part. This statement is true not only in classical physics but in all sciences where standard (Kolmogorov) probability theory is used, e.g. in economics or biology. The fact may be put in quantitative form in terms of ‘entropy’, a measure of our ignorance (our lack of information). In fact in a system consisting of several parts the classical (Shannon) entropy fulfils the inequalities

\[ S_j \leq S \leq \sum_j S_j, \]  

where \( S_j \) is the entropy of subsystem \( j \) and \( S \) the entropy of the whole system. The latter inequality is called subadditivity and it is not too relevant here. The former means that our lack of information about a system cannot be smaller than about one of the parts. This is violated in quantum theory, where the standard measure of ignorance is von Neumann entropy \([31]\), defined by

\[ S = -\text{Tr}(\rho \log \rho), \quad S_j = -\text{Tr}(\tilde{\rho}_j \log \tilde{\rho}_j). \]  

In fact for the state equation \((22)\) the von Neumann entropy fulfils

\[ S = 0, \quad S_1 = S_2 = \log 2. \]

It is possible to show that the violation of the information inequality, former equation \((23)\), is closely related to the violation of a Bell inequality \([33]\). Indeed a necessary condition for both is that the quantum state is entangled.

A naive solution to the information problem would be assuming that the quantum state-equation \((22)\) represents a statistical mixture (of the two possibilities \(|\uparrow\rangle\downarrow\rangle\) and \(|\downarrow\rangle\uparrow\rangle\)). However this assumption poses well known difficulties because in quantum physics a linear combination like equation \((22)\) has empirically testable differences with a mixture. Of course in HVMs (possibly contextual) the von Neumann entropy equation \((24)\) is no longer the appropriate measure of information.

Entanglement is quite common in quantum physics. For instance in many-electron systems (atoms, molecules and solid bodies) the Pauli principle gives rise to entanglement. The fact could be illustrated with the statevector of a two-electron atom. If the total spin is zero (ignoring nuclear spin) the joint statevector might be written

\[ \Psi(1, 2) = \psi(r_1, r_2)(|\uparrow\downarrow - \downarrow\uparrow), \]  

where \( \psi(r_1, r_2) \) usually does not have the form of a product of one-electron wavefunctions. In this case equation \((25)\) represents an entangled state involving position and spin. However entanglement is most relevant when the systems are separated, as stressed in the EPR paper \([22]\).

We see that entanglement is a correlation, but a strange correlation different from the classical one by the fact that it may violate a Bell inequality. Indeed the close connection of entanglement with the Bell inequalities derives from the fact that entanglement is, in addition to necessary, a sufficient condition for the existence of a Bell inequality violated by (ideal) quantum predictions \([32]\).

An interesting question related to entanglement appears if quantum mechanics is universally valid, which is the common wisdom today. (Thus classical theories are seen as approximations to quantum theories, the passage from quantum to classical being an active
subject of research in recent years). If quantum mechanics is universally valid and we accept the unrestricted validity of the superposition principle then all statevectors correspond to physical states of the system. This assumption leads to bizarre consequences in the macroscopic domain as was emphasized by Schrödinger [30] with his celebrated cat example. It consists of a cat plus a radioactive atom such that the cat is killed when the atom decays. If every vector of the Hilbert space corresponds to a state of the system, then there exists amongst others a state represented by the linear combination

$$\frac{1}{\sqrt{2}} (|\text{livecat}, 1\rangle + |\text{deadcat}, 0\rangle),$$

where the first term represents the cat alive with the atom excited and the second term the cat dead with the atom in the ground state.

It is obvious that the description of the state of a macroscopic system cannot be made in practice associating to it a statevector and that a density operator is a more appropriate description. The usual solution to the difficulty is to take into account that a cat, as any macroscopic system, cannot be isolated from the environment. Therefore the state of the cat should not be represented by a pure state like equation (22), but by a reduced density operator resulting from taking the partial trace over the degrees of freedom of the environment. The reduced density operator might be interpreted as incomplete information, with some probability $P$ that the cat is alive and the probability $1 - P$ that she is dead. This approach to macroscopic quantum systems is named decoherence induced by the environment, and it leads from a pure state of a large system to a reduced density operator for any subsystem. The density operator is named an improper mixture, meaning that it is different from the mixture corresponding to an isolated system prepared in a state without complete control of all relevant parameters (complete control would lead to a pure state). However decoherence theory just translates the problem from the cat to the Universe. Indeed it is difficult to establish a rigorous spatial limit to the environment and if no limit is put we should consider the whole Universe. Thus the decoherence approach is closely connected to the many worlds interpretation of quantum mechanics [34]. In this case we should assume that, after the evolution of the cat interacting with the environment, the Universe would consist of a linear combination of statevectors everyone with a copy of the atom, the cat and everything else including ‘me’ (the observer). For many people a linear combination of two terms, one of them with ‘me’ seeing a dead cat and another one with ‘me’ seeing an alive cat is even more bizarre than a linear combination of living and dead cat. In spite of this the ‘wavefunction of the Universe’ has been advocated by respected physicists.

The Schrödinger cat example is also relevant for the discussion of measurement in quantum theory. In fact a measurement amounts to an interaction of a system (usually small) with a measuring device (usually macroscopic). If the state of system plus device is represented by a wavefunction, then after the interaction the composite system will be entangled. Indeed in the Schrödinger example the cat may be taken as a measuring device for the state of the atom. The measurement problem has been crucial for the interpretation of quantum mechanics, but a more detailed discussion is beyond the scope of this paper (see e.g. [25]).

References

[1] Bell J S 1964 On the Einstein–Podolsky–Rosen paradox Physics 1 195–200
[2] Bell J S 1987 Speakable and Unspeakable in Quantum Mechanics (Cambridge: Cambridge University Press) (this book contains reprints of most Bell’s papers about the foundations of quantum mechanics).
[3] Hensen B et al 2015 Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres Nature 526 682
[4] Shalm L K et al 2015 A strong loophole-free test of local realism Phys. Rev. Lett. 115 250402
[5] Giustina M et al 2015 A significant loophole-free test of Bell’s theorem with entangled photons Phys. Rev. Lett. 115 250401
[6] Brunner N, Cavalcanti D, Pironio S, Scarani V and Wehner S 2014 Bell nonlocality Rev. Mod. Phys. 86 419478
[7] Hooker C A (ed) 1975 The Logico-Algebraic Approach to Quantum Mechanics. Volume I: Historical Evolution (Dordrecht: Reidel)
Hooker C A (ed) 1979 The Logico-Algebraic Approach to Quantum Mechanics. Volume II: Contemporary Consolidation (Dordrecht: Reidel)
[8] Santos E 1986 The Bell inequalities as tests of classical logic Phys. Lett. A 115 363–5
[9] Pitovsky I 1989 Quantum Probability-Quantum Logic (Berlin: Springer)
[10] Beltrametti E and Maczynski M 1992 Problem of classical and nonclassical probabilities Int. J. Theor. Phys. 31 1849–56
[11] Dvurecenskij A and Langer H 1995 Bell-type inequalities in orthomodular lattices, I, II Int. J. Theor. Phys. 34 995–1024
[12] Pulmannova S 2002 Hidden variables and Bell inequalities on quantum logics Found. Phys. 32 193–216
[13] Pykacz J, Valaskova J and Nanasiova O 2015 Bell-type inequalities for bivariate maps on orthomodular lattices Found. Phys. 45 900–13
[14] Auletta G 2000 Foundations and Interpretation of Quantum Mechanics (Singapore: World Scientific)
[15] Laloe F 2012 Do we Really Understand Quantum Mechanics? Strange Correlations, Paradoxes, and Theorems (Cambridge: Cambridge University Press)
[16] Redhead M 1990 Incompleteness, Nonlocality and Reality (Oxford: Clarendon)
[17] d’Espagnat B 2003 Veiled Reality An Analysis of Present-Day Quantum Mechanical Concepts (Boulder, CO: Westview)
[18] d’Espagnat B 2006 On Physics and Philosophy (Princeton, NJ: Princeton University Press)
[19] Englert B G 2013 On quantum theory Eur. Phys. J. D 67 238
[20] Clauser F J and Horne M A 1974 Experimental consequences of objective local theories Phys. Rev. D 10 526–35
[21] Clauser F J, Horne M A, Shimony A and Holt R A 1969 Proposed experiment to test local hidden-variables theories Phys. Rev. Lett. 23 880–3
[22] Einstein A, Podolsky B and Rosen N 1935 Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. 47 777–80
[23] Bohr N 1935 Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. 48 696–670
[24] Harrigan N and Spekkens R W 2010 Einstein, incompleteness, and the epistemic view of quantum states Found. Phys. 40 125–57
[25] Santos E 2015 Towards a realistic interpretation of quantum mechanics providing a model of the physical world Found. Sci. 20 357–86
[26] Kochen S and Specker E 1967 The problem of hidden variables in quantum mechanics J. Math. Mech. 17 59–87
[27] Bell J S 1966 On the problem of hidden variables in quantum mechanics Rev. Mod. Phys. 38 447–52
[28] Mermin D N 1993 Hidden variables and the two theorems of John Bell Rev. Mod. Phys. 65 803–15
[29] Cetto A M, de la Peña L and Santos E 1985 A Bell inequality involving position, momentum and energy Phys. Lett. A 113 304
[30] Schrödinger E 1935 The present situation of quantum mechanics Naturwissenschaften 23 807–12
[31] von Neumann J 1932 Mathematische Grundlagen der Quantenmechanik (Berlin: Springer) von Neumann J 1955 (Princeton, NJ: Princeton University Press) (Engl. transl.)
[32] Gisin N 1991 Bell inequality holds for all non-product states Phys. Lett. A 154 201
[33] Santos E 2004 Entropy inequalities and bell inequalities for two-qubit systems Phys. Rev. A 69 022305
[34] de Witt B S and Graham N 1973 The Many- Worlds Interpretation of Quantum Mechanics (Princeton, NJ: Princeton University Press)