Chapter 5
Realism, Non-Contextuality, Local Causality, Entanglement

We have accumulated enough theoretical material to tackle some aspects of an important and intriguing issue regarding the theoretical interpretation of the quantum realm.

5.1 Hidden Variables and no-go Results

There exist approaches to quantum phenomenology, called *hidden-variable formulations* (see, e.g.,[BeCa81, Ghi07, Lan17, BeZe17, SEP], and [Red98] for the viewpoint of QFT), that compete with the standard interpretation of the formalism also known as *Copenhagen interpretation*, which is the one adopted in this book.

The most important exemplar of these alternative formulations is certainly the well-known *Bohmian mechanics* [DüTe09], a quite articulate and healthy theory. Also known as *pilot-wave theory* or *de Broglie–Bohm theory*, Bohmian mechanics posits that a quantum particle has a definite position at every time (in this sense it is a partially classic system and the position is the hidden variable) and moves according to an equation of motion subsuming a "quantum" interaction due to a wavefunction that evolves under the usual Schrödinger equation. Randomness arises from the fact that we do not know which trajectory the particle actually follows among the plethora permitted by the evolution law. Bohmian mechanics is named after David Bohm, who was the first physicist to frame (in 1952) into a definite form this alternate description, which had already been proposed in similar yet vague forms by other scientists like de Broglie, thus enabling it to make correct predictions. A thorough examination would deserve more than an entire chapter, so we shall not discuss it here (see also [Tum17] for a recent review).

Another classical subject concerns the celebrated *Bell theorem* apropos the *BCHSH inequality* and the role of *locality* (or *local causality*) in QM, in relationship to the phenomenology of *entangled states*. The reader may profitably consult

[BeZe17] for a recent review on Bell's achievements and the developments of his ideas on locality and entanglement in quantum theory—with regard to other topics discussed in the rest of this chapter—also including recent experimental achievements.

Although we will introduce two versions of Bell's analysis on the interplay between *entanglement*, *realism*, and *locality* in two sections of this chapter, we are also interested in discussing a different theoretical milestone about hidden-variable theories, known as the *Kochen–Specker theorem*, and the related notions of *realism* and *non-contextuality*. The last section tackles the interaction between entanglement and non-contextuality by addressing the *BCHSH inequality* from a different point of view.

5.1.1   Realistic Hidden-Variable Theories

The pivotal idea at the heart of hidden-variable formulations is that a quantum system is actually *partially* classic (quantum phenomenology and the constant $\hbar$ must however enter the theory, eventually) and the observed randomness of measurement outcomes is due to an *incomplete* knowledge of the system. There are in particular *hidden variables*, cumulatively denoted by $\lambda \in \Lambda$ usually, whose knowledge would completely fix a classical-like state of the system. For this school of thought it is implicit that *all observables always have definite values* when λ is given, even if we do not know them. *Measurements are thus simple observations of values which already exist.* This hypothesis goes under the name of *realism* after the celebrated analysis by Einstein et al. [EPR35] (though this notion of realism specifically refers to a theoretical context only, and should not be taken literally as a general philosophical assumption!). As we said above, due to reasons specified in concrete models, when we observe the quantum behaviour of our physical system, the knowledge of hidden variables is limited in a way similar to what happens in statistical mechanics. As a matter of fact, we only have access to a probability distribution of λ over Λ, which we shall denote by μ. The quantum fluctuations of the outcomes of a measurement are explained as *statistical fluctuations* related to μ. In this view, quantum randomness is merely *epistemic* rather than *ontic*, as in the Copenhagen interpretation.

5.1.2   The Bell and Kochen–Specker no-go Theorems

Let us get started with a non-existence theorem in the standard formulation of QM.

Under the hypotheses of Gleason's theorem, quantum-state operators and quantum probability measures correspond one-to-one, so the notion of *expectation value* and *standard deviation* of an observable can be ascribed to quantum probability measures $\rho \in \mathscr{M}(\mathsf{H})$. In particular $\langle A \rangle_\rho$ and ΔA_ρ can be defined when A is

bounded simply by replacing ρ with the corresponding state T and using the already known definitions (4.33) and (4.34).

Definition 5.1 If H is a Hilbert space, a quantum probability measure $\rho \in \mathscr{M}(\mathsf{H})$ is called **dispersion-free** if $\Delta A_\rho = 0$ for every observable $A \in \mathfrak{B}(\mathsf{H})$. ∎

Theorem 4.49 is the important consequence of Gleason's theorem discovered by Bell [Bel66] in 1966 (already known to von Neumann in 1932, however). Now we may rephrase it as a non-existence result for dispersion-free quantum probability measures.

Theorem 5.2 (Bell's Theorem (Alternative Statement)) *Let H be a Hilbert space, either of finite dimension* $\dim(\mathsf{H}) > 2$ *or infinite-dimensional and separable. There exist no dispersion-free quantum probability measures in* $\mathscr{M}(\mathsf{H})$.

Proof Suppose that such a $\rho \in \mathscr{M}(\mathsf{H})$ exists and let $T \in \mathscr{S}(\mathsf{H})$ be the associated quantum-state operator according to Gleason's theorem. Assuming $A = P \in \mathscr{L}(\mathsf{H})$, it follows $0 = (\Delta P_T)^2 = tr(TPP) - tr(TP)^2 = tr(TP) - tr(TP)^2$. As a consequence, either $tr(TP) = 0$ or $tr(TP) = 1$ for every $P \in \mathscr{L}(\mathsf{H})$. This is impossible by Theorem 4.49. □

Remark 5.3 The only technical difference with Theorem 4.49 is that now general bounded observables are considered, and not only elementary propositions. Notice that Theorem 5.2 easily implies Theorem 4.49 when we look at elementary observables. But is also uses Theorem 4.49 in its proof, so the two versions are indeed equivalent. ∎

If we specialise to the finite-dimensional case, we can recast the theorem in a form that has several implications for the hidden-variable theory. Improving on an earlier non-existence result due to von Neumann (1932), the famous 1967 *Kochen–Specker theorem* [KoSp67] is actually an elementary corollary of Gleason's theorem, as Bell realized, even though the original proof was direct and completely different (see, e.g., [Lan17, SEP]). We state and prove the theorem below, and then discuss the relevant theoretical consequences.

Notation 5.4 For a given Hilbert space H, $\mathfrak{B}(\mathsf{H})_{sa}$ indicates the real linear space of selfadjoint elements of $\mathfrak{B}(\mathsf{H})$. ∎

Theorem 5.5 (Kochen–Specker Theorem) *Let H be a finite-dimensional Hilbert space with* $\dim(\mathsf{H}) > 2$. *For any non-zero map* $v : \mathfrak{B}(\mathsf{H})_{sa} \to \mathbb{R}$, *the requirements*

(i) $v(A + B) = v(A) + v(B)$ *if* $A, B \in \mathfrak{B}(\mathsf{H})_{sa}$ *commute,*
(ii) $v(AB) = v(A)v(B)$ *if* $A, B \in \mathfrak{B}(\mathsf{H})_{sa}$ *commute,*

are incompatible.

Proof Every orthogonal projector $P \in \mathscr{L}(\mathsf{H})$ belongs to $\mathfrak{B}(\mathsf{H})_{sa}$. If a map v exists as in the hypotheses, then $v(P) = v(PP) = v(P)^2$ due to (ii), hence $v(P) \in \{0, 1\}$. In particular $v(I) = 1$, otherwise $v(A) = v(IA) = v(I)v(A) = 0$, which

is not permitted ($v \not\equiv 0$). Observing that $P_i P_j = 0$ implies $P_i P_j = P_j P_i$ for $P_i, P_j \in \mathscr{L}(\mathsf{H})$, it is easy to check that the map $\rho : \mathscr{L}(\mathsf{H}) \ni P \mapsto v(P)$ defines a quantum probability measure by (i) and $v(I) = 1$. Note that (i) implies the additivity of this map on $\mathscr{L}(\mathsf{H})$. In turn, additivity implies σ-additivity because H is finite-dimensional and hence only finite sequences of non-vanishing orthogonal projectors onto pairwise orthogonal subspaces exist, and also $v(0) = 0$ from $v(I) = 1$ and (i). Such ρ is not allowed by Theorem 4.49, since $\rho(\mathscr{L}(\mathsf{H})) \subset \{0, 1\}$ and $\dim \mathsf{H} > 2$. Consequently v cannot exist. $\qquad\Box$

Remark 5.6 If H is infinite-dimensional but is separable, the thesis of Theorem 5.5 is still valid if we add the requirement that (iii) v is continuous in the *strong operator topology*. In fact, according to the above proof of Theorem 5.5, the only extra fact to be proved is that $\rho : \mathscr{L}(\mathsf{H}) \ni P \mapsto v(P)$ is σ-additive. If P is the strong limit of $\sum_{k=1}^{N} P_k$ as $N \to +\infty$, where $P_k P_h = 0$ if $k \neq h$, the additivity of v together with its strong continuity force the σ-additivity of ρ. $\qquad\blacksquare$

We will use quite often in the rest of the chapter a technical lemma related to the hypotheses of the Kochen–Specker theorem.

Lemma 5.7 *Let H be a Hilbert space of any dimension, and $A \in \mathfrak{B}(\mathsf{H})_{sa}$. Take a non-zero real-valued map v defined on the unital Abelian algebra of real polynomials of A. If v fulfils the Kochen–Specker requirements (i) and (ii), then it also satisfies $v(I) = 1$ and $v(aA) = av(A)$ for $a \in \mathbb{R}$.*

Proof The first relation was shown during the proof of Theorem 5.5 (without using $\dim \mathsf{H} < +\infty$). To prove the other one, recall a known analysis result whereby the only *non-zero* additive and multiplicative map $f : \mathbb{R} \to \mathbb{R}$ ($f(a+b) = f(a)+f(b)$ and $f(ab) = f(a)f(b)$ for every $a, b \in \mathbb{R}$) is the identity $f(a) = a$. The function $f(a) := v(aI)$ satisfies the conditions above (in particular $f(1) = v(I) = 1 \neq 0$). Hence, $v(aA) = v(aI)v(A) = av(A)$ for $a \in \mathbb{R}$ and $A \in \mathfrak{B}(\mathsf{H})_{sa}$. $\qquad\Box$

Let us start discussing the physical repercussions of the Kochen–Specker no-go result. Theorem 5.5 imposes strong limitations on any *theory of hidden variables* which assumes the *realism* hypothesis, when taking the quantum phenomenology into account.

As already said, within these approaches it is supposed that a quantum system S is actually partially classic and the observed randomness of measurement outcomes is due to an incomplete knowledge of the system making quantum randomness merely epistemic. There exist *hidden variables* $\lambda \in \Lambda$ that completely fix a classical-like state of the system and the values of every observable, that are *always defined* (*realism* hypothesis). If we knew λ, we would know also the precise value $v_\lambda(A) \in \sigma(A)$ every observable A has. Here the quantum observables A are seen as classical quantities that attain real values, the same permitted by the quantum theory, depending on the value of the hidden variable.

However, it is by no means evident how the assignment $A \mapsto v_\lambda(A) \in \sigma(A)$ should encompass *functional relations* between observables when these relations

exist at quantum level. For instance, if $C = A + B$, we cannot in general assume that $v_\lambda(C) = v_\lambda(B) + v_\lambda(B)$, because it is not obvious how to interpret classically $C = A + B$ when the selfadjoint operators A and B do not commute, in other terms when these observables, in the quantum interpretation, cannot be measured simultaneously. (In this case also the relationship between the spectra of A, B, C is generally complicated and unexpected: think of $H = X^2 + P^2$ on $L^2(\mathbb{R}, dx)$.) Yet, there remains to explain how to interpret "A and B cannot be measured simultaneously" in a realistic hidden-variable theory, where we assume from the very beginning that every observable is always defined. In some sense, the values assumed to exist simultaneously for A and B in the hidden-variable theory cannot be measured (do they fluctuate wildly?).

The spirit of the Kochen–Specker theorem is just to *avoid* these difficult and subtle questions and concentrate on what we can reasonably assume. The eventual no-go result is independent of such nuanced details. Indeed, in the special case where *all the involved observables are pairwise compatible*, we expect that they can be treated as classical quantities measured on the system and thus, at least in this case, some functional relations may be preserved by the assignment v_λ. Observe in particular that, if H has finite dimension,

$$\sigma(A + B) \subset \{v + \mu \mid v \in \sigma(A) \,, \ \mu \in \sigma(B)\} \quad \text{when } A, B \in \mathfrak{B}(\mathsf{H})_{sa} \text{ commute,}$$

so maps $v_\lambda : \mathfrak{B}(\mathsf{H})_{sa} \to \mathbb{R}$ satisfying (i) $v_\lambda(C) = v_\lambda(B) + v_\lambda(B)$ are in principle conceivable. Condition (ii) can be similarly fulfilled, on the whole, since

$$\sigma(AB) \subset \{v\mu \mid v \in \sigma(A) \,, \ \mu \in \sigma(B)\} \quad \text{when } A, B \in \mathfrak{B}(\mathsf{H})_{sa} \text{ commute.}$$

The hypotheses of Theorem 5.5 concern the preservation of some very mild functional relations by the assignment of classical-like values $A \mapsto v_\lambda(A)$ fixed by the hidden variable λ when dealing with compatible observables. Even with such a minimal requirement, there can be no such map $\mathfrak{B}(\mathsf{H})_{sa} \ni A \mapsto v_\lambda(A) \in \sigma(A) \subset \mathbb{R}$. This is the powerfulness of the Kochen–Specker result.

The premises of the analogue 1932 no-go theorem by von Neumann can be phrased, in our setup, by making requirement (i) hold *also for incompatible observables* A, B—where v more generally represents an expectation value over a distribution of possible λ (including the assignment of a precise value, as before)—and weakening (ii) to $v(aA) = av(A)$, $a \in \mathbb{R}$. In 1966 Bell [Bel66] found a simple example showing that these stronger conditions cannot be fulfilled *regardless of the rest of von Neumann's argument*, thus proving the inadequacy of von Neumann's hypotheses. All that gave rise to an animated discussion to which the Kochen–Specker theorem put an end in 1967 [KoSp67] (see [Lan17] for a critical and historical discussion on the subject).

5.1.3   An Alternative Version of the Kochen–Specker Theorem

We present here an alternative version of the Kochen–Specker theorem which deals with the elementary observables, instead of insisting on functional identities of general observables. This is a formulation essentially analogous to Theorems 4.49 and 5.2 in the finite-dimensional case. Mild probabilistic requirements are assumed on a possible "probability distribution" p defined on a subset $\mathcal{P}$ (not necessarily the whole $\mathscr{L}(\mathsf{H})$) of elementary observables and only attaining sharp values 0 or 1. Such a distribution p cannot exist if $\mathcal{P}$ is sufficiently large, i.e. large enough to contain pairs of incompatible elementary observables. Assuming the standard interpretation of the quantum formalism regarding the notion of observable and its decomposition in elementary observables, this reformulation of the Kochen–Specker result is however equivalent to statement 5.5, as we prove below.

Theorem 5.8 (Kochen–Specker Theorem (Alternative Version))   *Let* H *be a Hilbert space with* $2 < \dim(\mathsf{H}) < +\infty$. *There exists a set of elementary observables* $\mathcal{P} \subset \mathscr{L}(\mathsf{H})$ *for which there is no map* $p : \mathcal{P} \to \{0, 1\}$ *satisfying the following requirements:*

(i') $p(P)p(P') = 0$ *if* $P, P' \in \mathcal{P}$ *define compatible and mutually exclusive elementary observables (i.e.* $PP' = 0$*),*

(ii') $\sum_{j \in J} p(P_j) = 1$ *for every subset* $\{P_j\}_{j \in J} \subset \mathcal{P}$ *made of compatible, pairwise exclusive elementary observables such that* $\vee_{j \in J} P_j = I$.

Proof Let us prove that Theorem 5.8 is a consequence of Theorem 5.5. Since the latter is true, this concludes the proof. Assume that Theorem 5.8 is false. Fix $\mathcal{P} := \mathscr{L}(\mathsf{H})$. There must exist a map $p : \mathcal{P} \to \{0, 1\}$ satisfying (i') and (ii'). Define the map $v : \mathfrak{B}(\mathsf{H})_{sa} \to \mathbb{R}$ such that $v(A) := \sum_{a \in \sigma(A)} a p(P_a^{(A)})$, where $A \in \mathfrak{B}(\mathsf{H})_{sa}$ and $P^{(A)}$ is the PVM of A. Notice that the map does not vanish because $v(I) = p(I) = 1$ by (ii'), with $\{P_j\}_{j \in J} := \{I\}$, and furthermore only one element of $\{p(P_a^{(A)})\}_{a \in \sigma(A)}$ does not vanish because the projectors $P_a^{(A)}$ are pairwise compatible and mutually exclusive and (i'), (ii') are assumed. Observe that, with this definition of v, $v(f(A)) = f(v(A))$ is satisfied for every $f : \mathbb{R} \to \mathbb{R}$ in view of the finite-dimensional version of the functional calculus, the uniqueness of the PVM of a selfadjoint operator, and the fact every $p(P_a^{(A)})$ vanishes but one. If $A, B \in \mathfrak{B}(\mathsf{H})_{sa}$ commute, using their spectral decompositions and the fact that $\dim(\mathsf{H}) < +\infty$, it is easy to prove that there exists $C \in \mathfrak{B}(\mathsf{H})_{sa}$ such that $A = f_A(C)$ and $B = f_B(C)$ for suitable functions $f_A, f_B : \mathbb{R} \to \mathbb{R}$. Indeed, the real number c is a discrete parameter which faithfully labels the finitely many $(a, b) \in \sigma(A) \times \sigma(B)$ and $P_c^{(C)} := P_{a_c}^{(A)} P_{b_c}^{(B)}$, $f_A(c) := a_c$, $f_B(c) := b_c$. The map v satisfies (i),(ii) of Theorem 5.5. In fact, $v(A + B) = v(f_A(C) + f_B(C)) = v((f_A + f_B)(C)) = (f_A + f_B)(v(C)) = f_A(v(C)) + f_B(v(C)) = v(A) + v(B)$ and a similar argument is valid for (ii). Hence Theorem 5.5 is false and this is not possible.   □

Proposition 5.9 *The statement of Theorem 5.8 is equivalent to the statement of Theorem 5.5.*

Proof It is sufficient to prove that Theorem 5.8 implies Theorem 5.5 since the converse is part of the proof of Theorem 5.8. Assume that Theorem 5.5 is false and let $v : \mathfrak{B}(\mathsf{H})_{sa} \to \mathbb{R}$ be a non-vanishing map which satisfies (i) and (ii). Since $\mathscr{L}(\mathsf{H}) \subset \mathfrak{B}(\mathsf{H})_{sa}$, we have in particular that $v(P) = v(PP) = v(P)v(P)$, so that (a) $v(P) \in \{0, 1\}$ for $P \in \mathscr{L}(\mathsf{H})$ and also $v(I) = 1$, otherwise $v(A) = v(AI) = v(A)0 = 0$ for every $A \in \mathfrak{B}(\mathsf{H})_{sa}$ which is not permitted. Iterating (ii), noticing that J must be finite ($\leq \dim \mathsf{H}$), we find (b) $\sum_{j \in J} v(P_j) = v(I) = 1$ for any set $\{P_j\}_{j \in J} \subset \mathscr{L}(\mathsf{H})$ such that $\sum_j P_j = I$ and $P_j P_h = 0$ when $j \neq h$ (notice that $P_j P_h = P_h P_j$ in this case). It is now easy to prove that the map $p := v|_{\mathcal{P}}$ satisfies (i') and (ii') of Theorem 5.8, for every $\mathcal{P} \subset \mathscr{L}(\mathsf{H})$ such that (i') and (ii') are eligible, invalidating Theorem 5.8. In fact, if $P, P' \in \mathcal{P} \subset \mathscr{L}(\mathsf{H})$ satisfies $PP' = 0$, we can augment the sequence to $P, P', Q_1, \ldots, Q_n$, where the operators project onto pairwise-orthogonal subspaces and their sum is I. This implies, from (b), that $v(P) + v(P') + \sum_k v(Q_k) = 1$. Since $v(P), v(P'), v(Q_k) \in \{0, 1\}$ by (a), then $p(P)p(P') = v(P)v(P') = 0$ and (i') is satisfied. Similarly, if $\{P_j\}_{j \in J} \subset \mathcal{P}$, with $P_j P_h = 0$ when $j \neq h$, satisfy $\sum_j P_j = I$, then $\sum_j p(P_j) = \sum_j v(P_j) = 1$ from (b), proving (ii'). □

Remark 5.10 For every dimension $\dim \mathsf{H} \geq 3$, there is numerical evidence that the set $\mathcal{P}$ violating (i') and (ii') is a *proper, finite* subset of $\mathscr{L}(\mathsf{H})$. As a matter of fact, the original proof in [KoSp67] for $\dim(\mathsf{H}) = 3$ establishes that there exists a subset $\mathcal{P} \subset \mathfrak{B}(\mathsf{H})_{sa}$ of cardinality 117, whose elements project onto one-dimensional subspaces, satisfying the thesis of Theorem 5.8. See [Cab06] for a discussion about the minimal cardinality of $\mathcal{P}$, and [AHANBSC13] for an interesting discussion of the experimental tests on version 5.8 of the Kochen–Specker theorem. ∎

Remark 5.11 In the rest of this chapter, the theorem quoted as 'Kochen–Specker theorem' will refer to Theorem 5.5, unless otherwise declared. ∎

5.2 Realistic (Non-)Contextual Theories

The simplest way out of the no-go result by Kochen and Specker, if one insists on a hidden-variable formulation, is to just reject the *realism* assumption and accept that *not all observables are simultaneously defined, even if we fix the hidden state λ.*

Another possibility is to assume that all observables are always and simultaneously defined, and is contingent on the idea of *contextuality*. It must be said that the same proposal was addressed by Bell in 1966 in his second celebrated paper [Bel66] in a more general context and with reference to the consequences of Gleason's theorem for the theories of hidden variables.

5.2.1 An Impervious Way Out: The Notion of Contextuality

First of all, observe that $\mathfrak{B}(\mathsf{H})_{sa}$ contains a profusion of *real* unital Abelian algebras S of mutually compatible observables (whose unit and structure are inherited from the complex algebra $\mathfrak{B}(\mathsf{H})$). From a practical point of view S represents observables we may measure simultaneously. Among the different choices for S many will be inequivalent. The observation playing a crucial role in the following discussion is that a generic $A \in \mathfrak{B}(\mathsf{H})_{sa}$ will belong to *different* algebras S, since compatibility is not a transitive relation.[1] Notice furthermore that the Kochen–Specker constraints (i) and (ii) concern only compatible observables, so they may be imposed on the elements of a given real unital Abelian algebra. To fulfil them without running into the negative result of Theorem 5.5, we could try the following: drop the main hypothesis of the Kochen–Specker theorem, thus foregoing the *unique* assignment of values v_λ on $\mathfrak{B}(\mathsf{H})_{sa}$, and allow instead for distinct values $v_\lambda(A|S)$ of the observable A, for every real unital Abelian algebra S containing A.

Remark 5.12

(a) $S \ni A$ can be taken to be the space of real polynomials $p(A)$ of A (where $A^0 := I$). This choice of S means in practice that we are measuring A alone. In this case, measuring only A automatically permits us to know also the values of the remaining observables in S: the values of the polynomials $p(A)$ satisfy $v_\lambda(p(A)) = p(v_\lambda(A))$ by virtue of (i), (ii) in Kochen–Specker's theorem and the relations of Lemma 5.7.

(b) $S \ni A$ may be defined by means of several substantially distinct observables $A_1, \ldots, A_n$ that we measure together with A. In this case, S coincides with the family of real polynomials $p(A, A_1, \ldots, A_n)$. The values of $p(A, A_1, \ldots, A_n)$ are known from the values of the generators $A_1, \ldots, A_n$, again by (i) and (ii) and Lemma 5.7.

(c) In any case, to know the values of all the observables of a generic unital Abelian algebra S it suffices to measure a linear basis $A_1, \ldots, A_m$ of S. Since the elements of S are linear functions of these, once more by (i), (ii) in the Kochen–Specker theorem and Lemma 5.7 we have $v_\lambda(\sum_{j=1}^{m} c_j A_j) = \sum_{j=1}^{m} v_\lambda(A_j)$. Such a basis always exists, see Remark 5.14 below. ∎

Let us now prove that it is possible to prescribe the values of any fixed A depending on the chosen $S \ni A$ satisfying (i) and (ii) in Kochen–Specker's theorem. The paradoxical aspect is that we are about to use the mathematical structure of quantum theory to corroborate the idea that a certain competitor theory is not mathematically contradictory!

---

[1] These sets of observables S represent the most classical structures one may extract form the whole set of observables of a quantum system. The fact that these structures are distinct and physically incompatible is one manifestation of *Bohr's complementarity principle*.

Proposition 5.13 *Assume* $\dim \mathsf{H} < +\infty$ *and let us denote by* $\mathfrak{C}$ *the family of real unital Abelian algebras* $S \subset \mathfrak{B}(\mathsf{H})_{sa}$. *For every given* $S \in \mathfrak{C}$, *there exists a non-zero map*

$$S \ni A \mapsto v(A|S) \in \sigma(A)$$

satisfying (i) and (ii) of the Kochen–Specker theorem and also

$$v(I|S) = 1 \quad and \quad v(aA|S) = av(A|S) \quad for \ A \in S \ and \ a \in \mathbb{R}.$$

Proof Since the selfadjoint operators in S commute with one another and $\dim \mathsf{H} = n < +\infty$, it is easy to prove that there exists a collection $\{P_k\}_{k=1,\ldots,m} \subset \mathscr{L}(\mathsf{H})$, $m \leq n$, of non-zero orthogonal projectors, with $\sum_{k=1}^{m} P_k = I$ and $P_r P_h = 0$ if $r \neq h$, such for every $A \in S$,

$$A = \sum_{k=1}^{m} a_k^{(A)} P_k \quad \text{for some } a_1^{(A)} \leq a_2^{(A)} \leq \cdots \leq a_m^{(A)} \in \mathbb{R}. \tag{5.1}$$

Notice that it may happen that $a_k^{(A)} = a_{k+1}^{(A)}$. By construction, $\{a_k^{(A)} \mid k = 1, \ldots, m\} = \sigma(A)$. Furthermore, every orthogonal projector $p_x = \langle x | \cdot \rangle x$, for $x \in P_k(\mathsf{H})$ of unit norm, satisfies $p_x A = A p_x$ for every $A \in S$. If $x \in \mathsf{H}$ is as above, define

$$S \ni A \mapsto v(A|S) := \langle x | Ax \rangle .$$

By construction $\langle x | Ax \rangle = a_k^{(A)} \in \sigma(A)$ for some $k = 1, 2, \ldots, m$, because x is a unit-norm eigenvector of A with eigenvalue $a_k^{(A)}$. Since this is valid for every $A \in S$, properties (i) and (ii) of Theorem 5.5 are immediate. Finally, $v(aA|S) = av(A|S)$ is due to linearity of the inner product, and $v(I|S) = 1$ because $\langle x|x \rangle = 1$. □

Remark 5.14 The m orthogonal projectors P_k appearing in the proof above are linearly independent because $P_k P_h = \delta_{hk} P_k$. Therefore (5.1) guarantees that $\{A_k\}_{k=1,\ldots,m}$ with $A_k := P_k$ is a linear basis of observables of the real unital Abelian algebra S. ∎

In this abstract context, a hidden variable can be defined as the choice of $\lambda = \{x_S\}_{S \in \mathfrak{C}}$, where $x_S \in \mathsf{H}$ is a common eigenvector of all the observables $A \in S$ picked out as prescribed above. Hence for every λ and every S, the maps

$$S \ni A \mapsto v_\lambda(A|S) := \langle x_S | Ax_S \rangle \in \sigma(A)$$

possess the desired properties. The price to pay when adopting this new framework—*circumventing the Kochen–Specker no-go result*—is that the value $v_\lambda(A|S) \in \sigma(A)$ of an observable $A \in S$ is determined not only by the hidden variable λ, but also by (finitely many) mutually compatible other observables that

we want to measure together with A (and that generate the chosen Abelian algebra S). This peculiar property a hidden-variable theory satisfies is called *contextuality*. Together with the *realism* assumption, the Kochen–Specker theorem only admits *realistic contextual* hidden-variable theories and denies *realistic non-contextual* ones.

Remark 5.15

(a) The existence of a finite linear basis of $S \in \mathfrak{C}$ is guaranteed if the Hilbert space is finite-dimensional and every element of $\mathfrak{B}(\mathsf{H})_{sa}$ represents an observable, whereas it is not warranted automatically if we relax these hypotheses.

(b) It has been argued that the standard formulation of QM is non-contextual, though this adjective is more often used do distinguish between theories of hidden variables alternative to the standard formulation. This means nothing but, when we fix the quantum state $T \in \mathscr{S}(\mathsf{H})$ of the system so that an observable A attains a definite value in that state ($\Delta A_T = 0$), this value does *not* depend on other possible observables we can measure simultaneously with A. The problem, so to speak, lies with the realism postulate: necessarily there exist other observables, different from A, that do not admit precise values for the quantum state T, as a consequence of Theorem 5.2.

(c) It is important to warn the reader that the notion of (non-)contextuality has acquired a wealth of different meanings originating in the debate on hidden variables. The rather cumbersome version discussed in this section is strictly pertaining to hidden variable theories in the framework of the Kochen–Specker theorem. The contextuality of Bohmian mechanics and the version dealing with Bell's inequality and entanglement have slightly different meanings. In all cases contextuality means that the value of one observable depends on the other observables (and their values) measured simultaneously; the specificities of this dependence may vary according to the notion of (non-)contextuality one adopts. ∎

5.2.2 The Peres–Mermin Magic Square

The Kochen–Specker theorem, in the form of Theorem 5.5, assumes that the set of observables considered is the whole $\mathfrak{B}(\mathsf{H})_{sa}$. However, in the spirit of the reformulation of Theorem 5.8, the no-go result can be obtained also by restricting the family of observables to a smaller set made of orthogonal projectors. After [KoSp67] many explicit proofs of that kind were produced. There are alternative, but theoretically equivalent formulations of the Kochen–Specker no-go result where the attention is placed on a minimal number of observables (not necessarily orthogonal projectors) violating some statement concerning the possibility to assign values to them in accordance with realism and non-contextuality. A popular and direct argument for $\dim(\mathsf{H}) = 4$ is provided by the well-known *Peres–Mermin magic square* [Per90, Mer90]. It refers to a system of two particles of spin $1/2$, and focuses

just on the spin part of the Hilbert space, $H = \mathbb{C}^2 \otimes \mathbb{C}^2$. One considers the 9 observables assembled in a square

$$
\begin{array}{|c|c|c|}
\hline
A_{11} & A_{12} & A_{13} \\
\hline
A_{21} & A_{22} & A_{23} \\
\hline
A_{31} & A_{32} & A_{33} \\
\hline
\end{array}
\quad = \quad
\begin{array}{|c|c|c|}
\hline
\sigma_x \otimes I & I \otimes \sigma_x & \sigma_x \otimes \sigma_x \\
\hline
I \otimes \sigma_y & \sigma_y \otimes I & \sigma_y \otimes \sigma_y \\
\hline
\sigma_x \otimes \sigma_y & \sigma_y \otimes \sigma_x & \sigma_z \otimes \sigma_z \\
\hline
\end{array}
\tag{5.2}
$$

The standard Hermitian *Pauli matrices* σ_k (see (1.12)) have eigenvalues ± 1 and satisfy the equations

$$
\sigma_x \sigma_y = i\sigma_z, \quad [\sigma_x, \sigma_y] = 2i\sigma_z \quad \text{for all cyclic permutations of } x, y, z.
$$

It is easy to prove that the three operators on each row or column are linearly independent and pairwise commuting.[2] Furthermore, the row and column of any given element contain a pair of incompatible elements (if we choose $\sigma_x \otimes I$ for example, $I \otimes \sigma_x$ and $I \otimes \sigma_y$ are incompatible).

For this special case, we will prove a Kochen–Specker-type theorem on $\mathfrak{B}(\mathbb{C}^2 \otimes \mathbb{C}^2)_{sa}$ with the further hypothesis that $v(A) \in \sigma(A)$.

Proposition 5.16 *Let* $H = \mathbb{C}^2 \otimes \mathbb{C}^2$ *and* $A_{ij} \in \mathfrak{B}(H)_{sa}$ *be defined as in (5.2). There exists no assignment of real values* $A_{ij} \mapsto v(A_{ij}) \in \sigma(A_{ij})$, *for* $i, j = 1, 2, 3$, *satisfying (ii) of the Kochen–Specker theorem and* $v(\pm I) = \pm 1$.

Proof The product of the values in all rows $\prod_{i=1}^{3} \prod_{j=1}^{3} v(A_{ij})$ equals the product of the values in all columns $\prod_{j=1}^{3} \prod_{i=1}^{3} v(A_{ij})$, so their product is 1. On the other hand, requirement (ii) implies that $\prod_{i=1}^{3} \prod_{j=1}^{3} v(A_{ij}) = \prod_{i=1}^{3} v(\prod_{j=1}^{3} A_{ij})$ and $\prod_{j=1}^{3} \prod_{i=1}^{3} v(A_{ij}) = \prod_{j=1}^{3} v(\prod_{i=1}^{3} A_{ij})$ since row elements are pairwise compatible, and column elements too. Therefore $\prod_{j=1}^{3} A_{ij} = I$ for $i = 1, 2, 3$ and $\prod_{i=1}^{3} A_{ij} = I$ for $j = 1, 2$, but $\prod_{i=1}^{3} A_{i3} = -I$. In summary, using $v(-I) = -1$, we find

$$
1 = \left[\prod_{j=1}^{3} \prod_{i=1}^{3} v(A_{ij}) \right] \prod_{i=1}^{3} \prod_{j=1}^{3} v(A_{ij}) = \left[\prod_{j=1}^{3} v\left(\prod_{i=1}^{3} A_{ij} \right) \right] \prod_{1=1}^{3} v\left(\prod_{j=1}^{3} A_{ij} \right)
$$

$$
= v(I)^3 v(I)^2 v(-I) = -1,
$$

which is impossible. $\qquad \square$

Remark 5.17 Proposition 5.16 automatically implies the thesis of the Kochen–Specker theorem *on the whole* $\mathfrak{B}(\mathbb{C}^2 \otimes \mathbb{C}^2)_{sa}$ (assuming also that $v(A_{ij}) \in$

---

[2]For instance, if $a\sigma_x \otimes I + bI \otimes \sigma_x + c\sigma_x \otimes \sigma_x = 0$, multiplying by $\sigma_a \otimes I$ or $I \otimes \sigma_a$ and computing the *partial trace* gives $a = b = c = 0$ easily, because $tr(\sigma_a) = 0$, $tr(\sigma_a \sigma_b) = 2\delta_{ab}$.

$\sigma(A_{ij})$), just because the restriction to the observables A_{ij} of the map v posited by the Kochen–Specker's theorem satisfies Proposition 5.16. However, here we are considering a smaller set of observables $A_{ij} \in \mathfrak{B}(\mathsf{H})_{sa}$ and we cannot say a priori that no assignment $v_\lambda(A_{ij}) \in \{\pm 1\}$ satisfies (some of the) requirements (i) and (ii) of Kochen–Specker and also Lemma 5.7. This is the relevance of the above proposition. ∎

5.2.3   A State-Independent Test on Realistic Non-Contextuality

The Peres–Mermin square can be used as experimental test for the no-go assertion of the Kochen–Specker theorem restricted to the only observables of a quantum physical system described on $\mathsf{H} = \mathbb{C}^2 \otimes \mathbb{C}^2$, interpreting the observables as classical quantities satisfying the *realism* and *non-contextuality* assumptions in a hidden-variable theory.

Consider a concrete physical system with Hilbert space $\mathbb{C}^2 \otimes \mathbb{C}^2$, and suppose we are able to give a definite interpretation to all observables A_{ij} in the Peres–Mermin square. If we measure the observables $A \in \mathfrak{B}(\mathbb{C}^2 \otimes \mathbb{C}^2)_{sa}$ repeatedly when the quantum state of the system $T \in \mathscr{S}(\mathsf{H})$ is fixed, the values will in general fluctuate. If we adopt a realistic non-contextual hidden-variable description, we are committed to assume that the fluctuation of the values $v_\lambda(A)$ is caused by a fluctuation of the state $\lambda \in \Lambda$, which is known only statistically and is described by a probability measure μ on a σ-algebra Σ of subsets of Λ (Σ obviously contains the singletons $\{\lambda\}$ as measurable sets). Quantum expectation values $tr(TA)$ must be interpreted as classical standard expectation values

$$\mathbb{E}_\mu(A) = \int_\Lambda v_\lambda(A) d\mu(\lambda) \,.$$

Suppose that the map $v_\lambda : \mathfrak{B}(\mathbb{C}^2 \otimes \mathbb{C}^2)_{sa} \to \mathbb{R}$ satisfies the very mild conditions of the Kochen–Specker theorem. There exists a quantity allowing, in principle, to choose between non-contextual hidden-variable models and a quantum description on the grounds of the experimental data. (Actually we already know that Proposition 5.16 rules out these assignments, but we will ignore this fact since we are interested in constructing an elementary experimental example.)

Consider the observable

$$\chi := A_{11}A_{12}A_{13} + A_{21}A_{22}A_{23} + A_{31}A_{32}A_{33} + A_{11}A_{21}A_{31}$$
$$+ A_{12}A_{22}A_{32} - A_{13}A_{23}A_{33}. \tag{5.3}$$

This is a selfadjoint operator because the selfadjoint operators in the products pairwise commute.

Remark 5.18 Notice that every observable A_{ij} appears in two different sets of pairwise compatible observables, yet these sets are not compatible to each other. E.g., A_{11}, A_{12}, A_{13} and A_{11}, A_{21}, A_{31} contain A_{11}, but $[A_{12}, A_{21}] \neq 0$. ∎

Now consider the experimental expectation value $\langle \chi \rangle$ obtained by collecting many measurement outcomes. There are two main possibilities:

1. fluctuations have a quantum nature, so that $\langle \chi \rangle = tr(T\chi)$,
2. fluctuations have a hidden-variable nature, hence $\langle \chi \rangle = \mathbb{E}_\mu(\chi)$.

In case (1), since $\chi = 2I \otimes I + 2I \otimes I + 2I \otimes I$, we should obtain

$$\langle \chi \rangle = 6 \,,$$

independently of the quantum state $T \in \mathscr{S}(\mathsf{H})$. In case (2), if we also assume the two Kochen–Specker hypotheses restricted to our observables—notice that the summands in (5.3) pairwise commute (each equals $\pm I \otimes I$!) so we may assume both (i) and (ii) in Kochen–Specker theorem—we have that

$$v_\lambda \left(\prod_{i=1}^3 A_{ij} \right) = \prod_{i=1}^3 v_\lambda \left(A_{ij} \right) \quad \text{and} \quad v_\lambda \left(\prod_{j=1}^3 A_{ij} \right) = \prod_{j=1}^3 v_\lambda \left(A_{ij} \right) \,.$$

Hence using Lemma 5.7 on the polynomials of $A_{13}A_{23}A_{33}$,

$$v_\lambda(\chi) := v_\lambda(A_{11})v_\lambda(A_{12})v_\lambda(A_{13}) + v_\lambda(A_{21})v_\lambda(A_{22})v_\lambda(A_{23}) + v_\lambda(A_{31})v_\lambda(A_{32})v_\lambda(A_{33})$$

$$+ v_\lambda(A_{11})v_\lambda(A_{21})v_\lambda(A_{31}) + v_\lambda(A_{12})v_\lambda(A_{22})v_\lambda(A_{32}) - v_\lambda(A_{13})v_\lambda(A_{23})v_\lambda(A_{33}) \,.$$

Remark 5.19 It is very important to stress that we have explicitly made use of *non-contextuality* since each observable A_{ij} appears simultaneously in two sets that contain incompatible observables. Nonetheless, we have given A_{ij} a *unique* value $v_\lambda(A_{ij})$ independently of the set to which it belongs. ∎

Each value $v_\lambda \left(A_{ij} \right) \in \{-1, +1\}$ is completely determined by λ, in some unknown way. It is however possible to prove that, in all cases, $-4 \leq v_\lambda(\chi) \leq 4$, so that the integration with respect to the probability measure μ gives

$$-4 \leq \langle \chi \rangle \leq 4 \,.$$

This is consequence of the following more general proposition.

Proposition 5.20 *Let $M(3, \mathbb{R})$ denote the algebra of real 3×3 matrices and define the map $f : M(3, \mathbb{R}) \ni X \to f(X) \in \mathbb{R}$ by*

$$f(X) := X_{11}X_{12}X_{13} + X_{21}X_{22}X_{23} + X_{31}X_{32}X_{33} + X_{11}X_{21}X_{31}$$
$$+ X_{12}X_{22}X_{32} - X_{13}X_{23}X_{33}. \tag{5.4}$$

Then $|f(X)| \leq 4$ if $X \in [-1, 1]^9$, where we have identified $M(3, \mathbb{R})$ with $\mathbb{R}^9$.

Proof The map f is continuous on $[-1, 1]^9$ and $\Delta f = 0$ on $(-1, 1)^9$. As a consequence of the maximum principle, $f \upharpoonright_{[-1,1]^9}$ attains its extremal values on the boundary of $[-1, 1]^9$. The boundary is the union of the 18 sets $Q_{ij}^{\pm} := \{X \in [-1, 1]^9 \mid X_{ij} = \pm 1\}$. It is evident that the restriction of f to $Q_{ij}^{\pm}$ is continuous and harmonic in the interior of $Q_{ij}^{\pm} \subset \mathbb{R}^8$. The argument can be iterated, and eventually the extreme values of f belong in the discrete set $D = \{X \in [-1, 1]^9 \mid X_{ab} = \pm 1 \text{ for } a, b = 1, 2, 3\}$. Therefore it is sufficient to prove that $|f(X)| \leq 4$ if $X_{ij} \in \{-1, 1\}$. First of all, if $X_{ij} = 1$ then $f(X) = 4$. Let us prove that a larger value is impossible to achieve when $X_{ij} \in \{-1, 1\}$. From the expression of f it immediately follows the only possible value greater than 4 which f could attain if all $X_{ij} \in \{-1, 1\}$ is 6. This value would be reached iff the first 5 summands in (5.4) had value 1 and the last one ($X_{13}X_{23}X_{33}$) were -1. In turn this would mean that: (1) in each of first 5 addends an even number of factors X_{ij} (or none) take value -1; (2) in the last term an odd number take value -1. In summary, f attains value > 4 iff an *odd* number of factors X_{ij} in (5.4) take the value -1. This is impossible because every X_{ij} occurs twice with the same value. We conclude that $f(X) \leq 4$ in $[-1, 1]^9$. Since $f(-X) = -f(X)$ and $[-1, 1]^9$ is invariant under $X \mapsto -X$, we also have $-4 \leq f(X)$ in $[-1, 1]^9$. $\square$

To recap:

1. quantum mechanics implies $\langle \chi \rangle = 6$, independently of the quantum state;
2. realistic non-contextual hidden-variable models (assuming (i) and (ii) of the KS theorem) imply $-4 \leq \langle \chi \rangle \leq 4$, independently of the hidden-variable distribution μ.

It is evident that quantum theory is incompatible with realistic non-contextual hidden-variable models, and $\langle \chi \rangle$ could be exploited to test the difference experimentally.

Real experiments have been performed to test the Kochen–Specker theorem on concrete physical systems (photons [MWZ00, HLZPG03], neutrons [HLBBR06, BKSSCRH09] and trapped ions [KZGKGCBR09]) using observables similar to χ and possibly dealing with suitably prepared quantum states. State-independent tests have been studied in [AHANBSC13].

5.3 Entanglement and the BCHSH Inequality

According to Sect. 4.4.8, if a quantum system is made of two subsystems, the overall Hilbert space has the form $H_1 \otimes H_2$, where H_1 and H_2 are the Hilbert spaces of the two subsystems. $\mathscr{S}(H_1 \otimes H_2)$ contains the so-called (pure) **entangled states**: by definition these are represented by unit vectors that are *not* factorized as $\psi_1 \otimes \psi_2$, but rather *linear combinations* of such vectors

$$\Psi = \sum_{k=1}^{n} c_k \psi_{1k} \otimes \psi_{2k} \, ,$$

where at least two c_k do not vanish. As first observed by Einstein, Podolski and Rosen in a celebrated 1935 paper [EPR35], this sort of state gives rise to very peculiar phenomena—often mentioned as the *EPR paradox*—as soon as one assumes the postulate of collapse of the state after a measurement (see Sect. 4.4.7) with post-measurement state (4.23). Suppose the whole state is represented by the entangled vector

$$\Psi = \frac{1}{\sqrt{2}} \left(\psi_a \otimes \phi + \psi_{a'} \otimes \phi' \right) \, ,$$

where $\psi_a, \psi_{a'} \in H_1$ and $\phi, \phi' \in H_2$ are of unit norm. We also assume that $A_1 \psi_a = a \psi_a$ and $A_1 \psi_{a'} = a' \psi_{a'}$ for a certain observable $A_1 \in \mathfrak{B}(H_1)_{sa}$ belonging to part S_1 of the total system and such that $a, a' \in \sigma_p(A_1)$. Due to the collapse of the state, when performing a measurement of A_1 on S_1 we actually act on the whole state, hence *also* on the part describing S_2. As a matter of fact,

 (i) if the outcome of the measurement of $A_1 \otimes I$ is a, then the state of the full
      system after the measurement will be described by $\psi_a \otimes \phi$;
(ii) if the outcome of the measurement of $A_1 \otimes I$ is a' then the state of the full
      system after the measurement will be described by $\psi_{a'} \otimes \phi'$.

Therefore as we act on S_1 by measuring A_1, we "instantaneously" produce a change of S_2 which, in principle, can be observed by performing measurements on it. All of this happens even if the measuring apparatus of S_2 is very far from the instrument measuring S_1. It is further possible to realize a more subtle version of the experiment where we can measure different observables on each side of the experiment, and the (possibly random) choice of these observables and the associated measurement are made in such a short lapse of time that any non-superluminal exchange of information between the two sides is prevented (see [BeZe17] for up-to-date theoretical and experimental discussions). This seems to stand in flat contradiction to the *locality* postulate of Relativity (whereby a maximal speed exists, the speed of light, for propagating physical information) in connection with the *realism assumption* that the values of the observables pre-exist the measurements and can be changed only through sub-luminal interactions.

5.3.1 BCHSH Inequality from Realism and Locality

We shall give an outline of Bell's 1964 analysis [Bel64] of an improved version of the EPR phenomenon proposed by Aharonov and Bohm, for a physical system consisting of a pair of spin 1/2 particles, so that

$$\mathsf{H} = \mathsf{H}_{\text{orbital}} \otimes \mathsf{H}_{1,\text{spin}} \otimes \mathsf{H}_{2,\text{spin}}$$

where $\mathsf{H}_{\text{orbital}} = L^2(\mathbb{R}^3, dx_1) \otimes L^2(\mathbb{R}^3, dx_2) \simeq L^2(\mathbb{R}^3 \times \mathbb{R}^3, dx_1 \otimes dx_2)$ and $\mathsf{H}_{i,\text{spin}} \simeq \mathbb{C}^2$ for $i = 1, 2$, and the entanglement takes place in the space of spins,

$$\Psi = \phi_1 \otimes \phi_2 \otimes \frac{1}{\sqrt{2}} \left(\psi_1 \otimes \psi_2 + \psi_1' \otimes \psi_2' \right) \quad \text{with} \quad \phi_i \in L^2(\mathbb{R}^3, dx_i) \quad \psi_i, \psi_i' \in \mathsf{H}_{i,\text{spin}}.$$

Once created into sharply separated wavepackets, the particles ϕ_1, ϕ_2 move along opposite directions towards the detectors where the spin observables are eventually measured.[3]

Bell's analysis considered the possibility of explaining the phenomenology of entanglement in terms of a hidden-variable theory and, most importantly, he proposed an experiment capable of checking if local realism is satisfied.

As in the previous sections, it is supposed that there exists a hidden variable $\lambda \in \Lambda$ which completely fixes the state of the couple of particles *when they are spacelike separated.* As before, we do not have direct access to λ but we do know its probability distribution μ over Λ, and this statistical description should be in agreement with (actually it should explain it!) the stochastic behaviour of measurement outcomes of QM. To be precise, λ generally indicates a *set* of hidden variables, and the state of S_1 only depends on a subset of these parameters while the state of S_2 depends on another subset. In a complete theory, one could also assume that hidden variables have a deterministic dynamical evolution. If so, our λ would represent the initial values of that evolution.

We are in particular interested in the value $A(\mathbf{a}|\lambda) \in \{\pm 1\}$ of the spin along the direction $\mathbf{a} \in \mathbb{S}^2$ (the unit sphere in $\mathbb{R}^3$) detected on particle S_1, and in the value $B(\mathbf{b}|\lambda) \in \{\pm 1\}$ of the spin along the direction $\mathbf{b} \in \mathbb{S}^2$ detected on particle S_2. (Actually the true spin values amount to $\hbar A(\mathbf{a}|\lambda)/2$ and $\hbar B(\mathbf{b}|\lambda)/2$ but we shall henceforth ignore the factor $\hbar/2$.)

Remark 5.21 As opposed to previous sections, we are *not* directly assuming that the spin is a *quantum observable*, i.e. a selfadjoint operator on a Hilbert space. It

---

[3]A more complete model would include the state's skew-symmetry (the electrons may be swapped), but we shall disregard details such as this one. When dealing with photons the spin must be replaced by the *polarization*, which is still described on $\mathbb{C}^2$, and the positions x_i by the momenta k_i; in this case the state must be symmetric when swapping the photons.

is just a quantity, taking values in $\{\pm 1\}$, that we can measure on both sides of the system depending on the choice of direction.   ∎

Let us make explicit two assumptions involved in Bell's picture.

1. *Realism.* The values of A and B exist at every time and for every choice of the directions $\mathbf{a}, \mathbf{b} \in \mathbb{S}^2$, independently of their explicit observation.
2. *Locality.* When measurements are performed on S_1 and S_2 by devices placed in causally separated regions of spacetime, the choice of $\mathbf{a} \in \mathbb{S}^2$ cannot have any influence on the outcome $B(\mathbf{b}|\lambda)$, and the choice of $\mathbf{b} \in \mathbb{S}^2$ cannot have any influence on the outcome $A(\mathbf{a}|\lambda)$; moreover, these outcomes are (pre-)determined by the hidden variable λ. (This is the reason why we write $A(\mathbf{a}|\lambda)$ but not, say, $A(\mathbf{a}|\lambda, \mathbf{b})$.)

Let us consider the quantity, obtained by measurements,

$$\chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda) := A(\mathbf{a}|\lambda)B(\mathbf{b}|\lambda) + A(\mathbf{a}'|\lambda)B(\mathbf{b}|\lambda) + A(\mathbf{a}'|\lambda)B(\mathbf{b}'|\lambda) - A(\mathbf{a}|\lambda)B(\mathbf{b}'|\lambda)$$

which depends on four choices of directions $\mathbf{a}, \mathbf{a}'$ for S_1 and $\mathbf{b}, \mathbf{b}'$ for S_2. Since

$$\chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda) = A(\mathbf{a}|\lambda)\left[B(\mathbf{b}|\lambda) - B(\mathbf{b}'|\lambda)\right] + A(\mathbf{a}'|\lambda)\left[B(\mathbf{b}|\lambda) + B(\mathbf{b}'|\lambda)\right] ,$$

and $B(\mathbf{b}|\lambda), B(\mathbf{b}'|\lambda) \in \{\pm 1\}$, only one summand survives. As $A(\mathbf{a}|\lambda), A(\mathbf{a}'|\lambda) \in \{\pm 1\}$, we conclude that

$$- 2 \leq \chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda) \leq 2 . \tag{5.5}$$

If we take the expectation value of $\chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda)$ when λ varies in Λ according with its probability distribution μ,

$$\mathbb{E}_\mu(\chi) := \int_\Lambda \chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda)d\mu(\lambda) ,$$

we find $-2 \leq \mathbb{E}_\mu(\chi) \leq 2$ since the measure is positive and the total integral is 1. Defining

$$E_\mu(\mathbf{a}, \mathbf{b}) := \int_\Lambda A(\mathbf{a}|\lambda)B(\mathbf{b}|\lambda)d\mu(\lambda) \quad \mathbf{a}, \mathbf{b} \in \mathbb{S}^2 ,$$

we obtain the famous **BCHSH inequality**, after J. Bell, J. Clauser, M. Horne, A. Shimony, and R. Holt[4]:

$$- 2 \leq E_\mu(\mathbf{a}, \mathbf{b}) + E_\mu(\mathbf{a}', \mathbf{b}) + E_\mu(\mathbf{a}', \mathbf{b}') - E_\mu(\mathbf{a}, \mathbf{b}') \leq 2 \quad \text{for every } \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}' \in \mathbb{S}^2.$$
$$\tag{5.6}$$

---

[4]The original paper of Bell [Bel64] presented a slightly less general inequality.

*The BCHSH inequality—regarding correlations of measurements of spin compo-
nents of pair of particles—must be satisfied by every realistic local theory.*

What is the quantum prevision instead? First of all, the *spin observable along*
$\mathbf{a} \in \mathbb{S}^2$ must be defined as the selfadjoint operator in $\mathfrak{B}(\mathbb{C}^2)_{sa}$

$$\mathbf{a} \cdot \sigma := \sum_{k=x,y,z} a_k \sigma_k \ . \tag{5.7}$$

In this context, we have to interpret $E_\mu(\mathbf{a}, \mathbf{b})$ as an expectation value with respect to
a quantum state $T \in \mathscr{S}(\mathbb{C}^2 \otimes \mathbb{C}^2)$ (neglecting the state's orbital part, which plays
no role at present):

$$E_T(\mathbf{a}, \mathbf{b}) = tr\left[T(\mathbf{a} \cdot \sigma \otimes \mathbf{b} \cdot \sigma)\right] \ . \tag{5.8}$$

We restrict the choice of state to *entangled* pure states $T_\pm = \langle \Psi_\pm | \cdot \rangle \Psi_\pm$ of a
particular type, called **Bell states**,

$$\Psi_+ := \frac{1}{\sqrt{2}}\left(\psi_+ \otimes \psi_+ + \psi_- \otimes \psi_-\right) \ , \quad \Psi_- := \frac{1}{\sqrt{2}}\left(\psi_+ \otimes \psi_- - \psi_- \otimes \psi_+\right) \ , \tag{5.9}$$

where $\psi_\pm \in \mathbb{C}^2$ are ± 1-eigenvectors of σ_z: $\sigma_z \psi_\pm = \pm \psi_\pm$. If $\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z \in \mathbb{S}^2$ are the
unit vectors along three orthogonal axes of the physical rest space of the laboratory,
we choose

$$\mathbf{a} = \mathbf{e}_x \ , \quad \mathbf{a}' = \mathbf{e}_z \ , \quad \mathbf{b} = \frac{\mathbf{e}_x + \mathbf{e}_z}{\sqrt{2}} \ , \quad \mathbf{b}' = \frac{\mathbf{e}_z - \mathbf{e}_x}{\sqrt{2}} \tag{5.10}$$

An elementary but lengthy computation based on (1.12) yields

$$E_{T_\pm}(\mathbf{a}, \mathbf{b}) + E_{T_\pm}(\mathbf{a}', \mathbf{b}) + E_{T_\pm}(\mathbf{a}', \mathbf{b}') - E_{T_\pm}(\mathbf{a}, \mathbf{b}') = \pm 2\sqrt{2} \ . \tag{5.11}$$

Since $2\sqrt{2} > 2$, we conclude that *the result predicted by Quantum Theory, with said
choices of directions and entangled states, is incompatible with realism and locality.*

The strong empirical evidence is that *local realism* is rejected by experimental
data accumulated, over the years, in several very delicate experiments performed to
test the BCHSH inequality on couples of particles in entangles states. See [GaCh08]
for a review on the various experiments and [Han15] for a recent important
experimental achievement on the subject. The non-locality of QM—with the above
specific meaning due to Bell [BeZe17]—is nowadays widely accepted as a real and
fundamental feature of Nature [Ghi07, SEP, Lan17].

Remark 5.22

(a) We stress, without entering in details, that the quantum violation of locality together with the stochastic nature of measurement outcomes do not permit superluminal propagation of physical information [Bell75, Ghi07].
(b) Incidentally, $2\sqrt{2}$ is the maximum value attainable for a quantum state $T \in \mathscr{S}(\mathsf{H})$ violating the BCHSH inequality [Tsi80], and is known as *Tsirelson's bound*. ∎

5.3.2 BCHSH Inequality and Factorized States

Let us examine what happens to the BCHSH inequality if $T = \langle \Psi | \ \rangle \Psi$ is not entangled, i.e., if

$$\Psi := \psi_1 \otimes \psi_2 \tag{5.12}$$

is a product of unit vectors ψ_i. We need at technical proposition.

Proposition 5.23 *Let $f : \mathbb{R}^4 \to \mathbb{R}$ be the map $f(x_1, x_2, x_3, x_4) = x_1 x_3 + x_2 x_3 + x_2 x_4 - x_1 x_4$. Then $|f(x_1, x_2, x_3, x_4)| \le 2$ if $(x_1, x_2, x_3, x_4) \in [-1, 1]^4$.*

Proof The map f is continuous on $[-1, 1]^4$ and satisfies $\Delta f = 0$ on the interior of $[-1, 1]^4$, so the maximum principle implies it is extremized on the boundary. The latter is the union of the eight sets $Q_i^\pm := \{(x_1, x_2, x_3, x_4) \in [-1, 1]^4 \mid x_i = \pm 1\}$. It is evident that the restriction of f to each $Q_i^\pm$ is still continuous and harmonic on the interior of $Q_i^\pm \subset \mathbb{R}^3$. Iterating the argument we eventually find that the extreme values of f are achieved on $D := \{(x_1, x_2, x_3, x_4) \in [-1, 1]^4 \mid x_i = \pm 1, i = 1, 2, 3, 4\}$. Since $f(x_1, x_2, x_3, x_4) = x_1(x_3 - x_4) + x_2(x_3 + x_4)$, when $x_3, x_4 = \pm 1$ only one of the summands is non-zero. Further imposing $x_1, x_2 = \pm 1$ tells $f(x_1, x_2, x_3, x_4) = \pm 2$ for every $(x_1, x_2, x_3, x_4) \in D$. Since $\max\{|f(z_1, z_2, z_3, z_4)| \mid (z_1, z_2, z_3, z_4) \in [-1, 1]^4\} = |f(x_1, x_2, x_3, x_4)|$ for some $(x_1, x_2, x_3, x_4) \in D$, the claim is proved. □

Given Ψ as in (5.12) and $T = \langle \Psi | \ \rangle \Psi$, a trivial computation proves that

$$E_T(\mathbf{a}, \mathbf{b}) + E_T(\mathbf{a}', \mathbf{b}) + E_T(\mathbf{a}', \mathbf{b}') - E_T(\mathbf{a}, \mathbf{b}')$$
$$= \langle \psi_1 | \mathbf{a} \cdot \sigma \psi_1 \rangle \langle \psi_2 | \mathbf{b} \cdot \sigma \psi_2 \rangle + \langle \psi_1 | \mathbf{a}' \cdot \sigma \psi_1 \rangle \langle \psi_2 | \mathbf{b} \cdot \sigma \psi_2 \rangle$$
$$+ \langle \psi_1 | \mathbf{a}' \cdot \sigma \psi_1 \rangle \langle \psi_2 | \mathbf{b}' \cdot \sigma \psi_2 \rangle - \langle \psi_1 | \mathbf{a} \cdot \sigma \psi_1 \rangle \langle \psi_2 | \mathbf{b}' \cdot \sigma \psi_2 \rangle . \tag{5.13}$$

But $||\mathbf{a} \cdot \sigma|| = \sup\{|\nu| \mid \nu \in \sigma(\mathbf{a} \cdot \sigma)\} = 1$, so $|\langle \psi_1 | \mathbf{a} \cdot \sigma \psi_1 \rangle| \le ||\mathbf{a} \cdot \sigma|| \, ||\psi_1||^2 = 1$, and then $\langle \psi_1 | \mathbf{a} \cdot \sigma \psi_1 \rangle$, $\langle \psi_1 | \mathbf{a}' \cdot \sigma \psi_1 \rangle$, $\langle \psi_2 | \mathbf{b} \cdot \sigma \psi_2 \rangle$, $\langle \psi_2 | \mathbf{b}' \cdot \sigma \psi_2 \rangle \in [-1, 1]$. In

summary, in view of Proposition 5.23, the absolute value of the right-hand side of (5.13) is bounded by 2. Therefore

$$-2 \le E_T(\mathbf{a}, \mathbf{b}) + E_T(\mathbf{a}', \mathbf{b}) + E_T(\mathbf{a}', \mathbf{b}') - E_T(\mathbf{a}, \mathbf{b}') \le 2 \quad \text{for every } \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}' \in \mathbb{S}^2.$$

Hence, factorized pure states satisfy the BCHSH inequality. An incoherent superposition of factorized pure states gives rise to the same result by construction. The lesson this story teaches us is:

Factorized pure states, and incoherent superpositions of them, do not violate the BCHSH inequality.

In a sense, they are more classical than entangled states.

Remark 5.24

(a) The natural question arising from our discovery is whether or not there exist pure *entangled* states satisfying the BCHSH inequality. As a matter of fact they do exist, and there also exist pure entangled states which violate the BCHSH inequality without reaching the maximum value $2\sqrt{2}$ [GaCh08, BeZe17].

(b) As a byproduct, the violation of the BCHSH inequality can be used to *detect entanglement*, paying attention that it only gives sufficient but not necessary conditions. ∎

5.3.3 BCHSH Inequality from Relativistic Local Causality and Realism

In order to derive the BCHSH inequality, Bell presented [Bell75] the very general approach[5] we set out to introduce now (see also [Jar84] and [Shi90]).

We remind the reader that in a *time-oriented* spacetime M, such as Minkowski's spacetime, the **causal past** $J^-(O)$ (resp. **causal future** $J^+(O)$) of $O \subset M$ is the set of points $p \in M$ which admit a curve from p to O whose tangent vector is either timelike or lightlike, and future-directed (resp. past-directed). Since these curves represent causal interactions (at the macroscopic level at least), O cannot be influenced by anything that happens outside $J^-(O)$. Two subsets $O, O' \subset M$ are **causally separated** if $J^\pm(O) \cap O' = \varnothing$ (which is equivalent to $J^\pm(O') \cap O = \varnothing$): no causal relation can exist between them.

In Bell's view, a general relativistic physical system is described in terms of physical quantities, named **beables** by Bell in opposition to *observables*. These objects are supposed to always exist independently of our measurements, they ought to have objective properties and satisfy locality, *local causality* to be precise, in the

---

[5]The author is grateful to S.Mazzucchi for many clarifications and discussions on subtleties related to the content of this section.

sense we shall discuss below. Every physical description ought to be based on them. This is the strongest form of *realism* and *locality*.

In a typically stochastic description of a physical system S, a beable is a *random variable* $X : \Omega \rightarrow R_X$ defined on a probability measure space $(\Omega, \Sigma(\Omega), \mathbb{P})$ common to all beables, where R_X is any measurable space characteristic of X, typically a subset of some $\mathbb{R}^n$. The overall stochastic state of the system is represented by the probability measure $\mathbb{P}$ over Ω.

Remark 5.25

(a) Included in this are deterministic descriptions where (some) beables have definite values, simply by assuming that $\mathbb{P}$ is such that the physically relevant random variable attains the chosen value with probability 1.
(b) It is clear that this description is completely classic, as it relies on Kolmogorov's notion of probability and not on the quantum notion used in Gleason's theorem. ∎

Beables are also *localized in spacetime regions* (Fig. 5.1) where they satisfy *causal locality* requirements, as we proceed to explain. We are interested in systems made of two parts S_1 and S_2, whose beables are localized in two *causally separated* regions O_1, O_2 of spacetime. In the following $P := J^-(O_1) \cap J^-(O_2)$ denotes the **common causal past** of the regions. As in the specific case of the EPR

Fig. 5.1 Causally separated regions O_1 and O_2
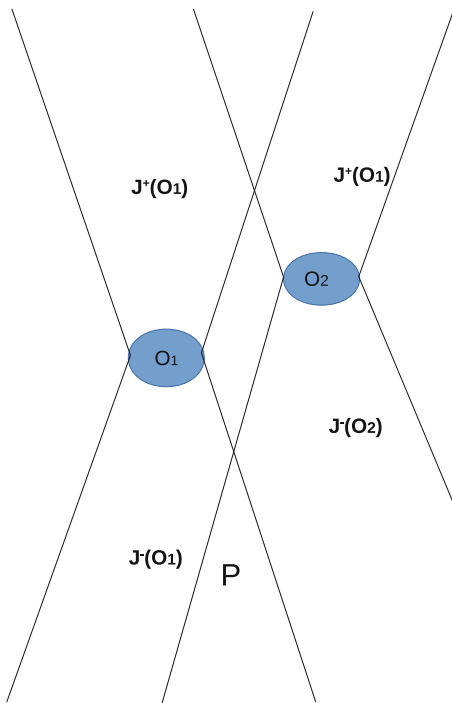
phenomenology, where S consists of two entangled particles S_1 and S_2 localized in causally separated regions O_1 and O_2, we assume that beables are of three types:

(a) s_1 is a random variable localized at O_1 and taking values in V_1, and s_2 is a random variables localized at O_2 and taking values in V_2. We also assume that V_1 and V_2 are discrete subsets of $[-1, 1]$;

(b) $\mathbf{n}_1$ is a random variable localized at $J^-(O_1) \setminus P$ taking values in N_1, and $\mathbf{n}_2$ is a random variable localized at $J^-(O_2) \setminus P$ taking values in N_2;

(c) λ is a random variable localized in the common causal past P taking values in some measurable space Λ.

The physical interpretation (not the only one) goes as follows:

1. s_1 is the (normalized) value of the component of the spin of S_1 along the direction $\mathbf{n}_1$, s_2 is the (normalized) value of the spin of S_2 along the direction $\mathbf{n}_2$. The value of s_1 cannot have any influence on the value of s_2, for O_1 and O_2 are causally separated.

2. The random variables $\mathbf{n}_1$ and $\mathbf{n}_2$ represent the choice we made of the components of the spin we intend to measure on S_1 in O_1 and on S_2 in O_2. The possible directions of the spins are taken in subsets N_1, N_2 of $\mathbb{S}^2$.

    These choices are made in the causal past of O_1 and O_2 respectively. We also assumed that the choice of $\mathbf{n}_1$ cannot have any influence on what happens in O_2 and *vice versa*, since both beables are localized outside P.

    (The $\mathbf{n}_i$ appear here as stochastic variables—in real measurements of EPR correlations the components of the spin to be measured are actually chosen randomly—but non-random choices are subsumed by assuming that the probability of a certain choice is 1, see Remark 5.25.)

3. The role of the beable λ as a *hidden variable* is less precise than in the previous section: it lives in the common causal past P and represents a potential *common cause* responsible for possible *correlations* of the beables localized at O_1 and O_2, since no direct causal relations are permitted between them as O_1 and O_2 are causally separated. The measure μ introduced in the previous sections, which betrays our ignorance about the precise value of λ, can be defined here as $\mu(L) := \mathbb{P}(\lambda^{-1}(L))$, where $L \subset \Lambda$ is any measurable set.

Remark 5.26 Let us emphasize that we are not assuming that the particles have spin 1/2, and the following reasoning would go through, with trivial adjustments, even if s_2 and s_2 were continuous on $[-1, 1]$. The rest of the argument is actually valid provided (a), (b), (c) are true regardless of the particle-spin interpretation when assuming the statistical interpretation of local causality (5.15)–(5.17) below.   ∎

By assuming (a), (b) and (c) the discussion goes on in terms of *conditional probabilities*. We want to prove an inequality about the *expectation value*

$$E(\lambda_0, \mathbf{a}, \mathbf{b}) := \mathbb{E}(s_1 s_2 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b})$$

of the product $s_1 \cdot s_2$ under the conditions $\lambda = \lambda_0$, $\mathbf{n}_1 = \mathbf{a}$, $\mathbf{n}_2 = \mathbf{b}$, where

$$\mathbb{E}(s_1 s_2 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b})$$

$$:= \sum_{\alpha \in V_1, \, \beta \in V_2} \alpha\beta \, \mathbb{P}(s_1 = \alpha, s_2 = \beta | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b}) \,. \tag{5.14}$$

We start from the observation that, in a *locally causal* theory as the one presented above, the following relations declaring statistical independence of the two subsystems must be true:

$$\mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b}, s_2 = \beta) = \mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}) \,, \tag{5.15}$$

$$\mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b}) = \mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}) \,, \tag{5.16}$$

$$\mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, s_2 = \beta) = \mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}) \,. \tag{5.17}$$

This is because the values of $\mathbf{n}_2$ and s_2 cannot have any influence on what happens in O_1, see (1) and (2) above. The same holds if we swap the beables of S_1 and S_2. Let us therefore consider the joint conditional probability

$$\mathbb{P}(s_1 = \alpha, s_2 = \beta | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b})$$

$$= \mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b}, s_2 = \beta)\mathbb{P}(s_2 = \beta | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b}) \,.$$

Using (5.15)–(5.17) and the analogous formulas with subsystems interchanged, we finally have

$$\mathbb{P}(s_1 = \alpha, s_2 = \beta | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}, \mathbf{n}_2 = \mathbf{b})$$

$$= \mathbb{P}(s_1 = \alpha | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a})\mathbb{P}(s_2 = \beta | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}) \,.$$

Inserting the result in (5.14) gives

$$E(\lambda_0, \mathbf{a}, \mathbf{b}) = \mathbb{E}(s_1 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a})\mathbb{E}(s_2 | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}) \,. \tag{5.18}$$

Since the values of s_1 and s_2 are bounded by 1 in absolute value, we also have

$$-1 \le \mathbb{E}(s_1 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}) \le 1 \quad \text{and} \quad -1 \le \mathbb{E}(s_2 | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}) \le 1 \,. \tag{5.19}$$

As a consequence, using Proposition 5.23, we conclude that no matter how we fix $\mathbf{a}, \mathbf{a}' \in N_1$ and $\mathbf{b}, \mathbf{b}' \in N_2$, the absolute value of

$$\mathbb{E}(s_1 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a})\mathbb{E}(s_2 | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}) + \mathbb{E}(s_1 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a})\mathbb{E}(s_2 | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}')$$

$$+\mathbb{E}(s_1 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}')\mathbb{E}(s_2 | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}) - \mathbb{E}(s_1 | \lambda = \lambda_0, \mathbf{n}_1 = \mathbf{a}')\mathbb{E}(s_2 | \lambda = \lambda_0, \mathbf{n}_2 = \mathbf{b}')$$

is bounded by 2. In other words, from (5.18),

$$-2 \leq E(\lambda_0, \mathbf{a}, \mathbf{b}) + E(\lambda_0, \mathbf{a}, \mathbf{b}') + E(\lambda_0, \mathbf{a}', \mathbf{b}) - E(\lambda_0, \mathbf{a}', \mathbf{b}') \leq 2 . \quad (5.20)$$

We can get rid of $\lambda_0 \in \Lambda$ by taking the expectation value with respect to the probability measure μ over Λ introduced in (3) above:

$$E(\mathbf{a}, \mathbf{b}) := \int_\Lambda E(\lambda, \mathbf{a}, \mathbf{b}) d\mu(\lambda) .$$

Using this definition in (5.20), the linearity of the integral and the fact that the total integral is 1, we eventually obtain the BCHSH inequality:

$$-2 \leq E(\mathbf{a}, \mathbf{b}) + E(\mathbf{a}, \mathbf{b}') + E(\mathbf{a}', \mathbf{b}) - E(\mathbf{a}', \mathbf{b}') \leq 2$$

under the hypotheses (a), (b), (c) and the natural interpretation of local causality (5.15)–(5.17).

5.3.4 BCHSH Inequality from Realism and Non-Contextuality

We do not wish to insist again on the interplay between entanglement, realism and locality, so we switch to the relationship between entanglement, realism, and *non-contextuality* instead.

Let us consider again a *quantum* system S made of two *independent* parts S_1 and S_2 which are not necessarily spatially separated. A physical example of such a system is a spin-1/2 massive particle, or a photon, where the polarization's two degrees of freedom are exploited in place of the two degrees of freedom of the spin. In principle, according to Sect. 4.4.8, the Hilbert space of this system is the Hilbert tensor product $L^2(\mathbb{R}^3, d^3k) \otimes \mathbb{C}^2$ (momentum picture). However, we can restrict the possibilities in the momentum space $L^2(\mathbb{R}^3, d^3k)$ to a 2-dimensional subspace. In practice, through a suitable experimental filter only the span of two states labelled by two momenta $k_1, k_2 \in \mathbb{R}^3$ is accessible to the system. These two pure states are defined by a pair of unit-norm vectors ψ_{k_1} and ψ_{k_1}. In terms of L^2 functions, these vectors are wavefunctions typically living in $\mathscr{S}(\mathbb{R}^3)$, whose support in momentum space is strictly concentrated around k_1 and k_2 respectively. Since $k_1 \neq k_2$, it is reasonable to assume $\langle \psi_{k_1} | \psi_{k_1} \rangle = 0$. In this way the span of the vectors is isomorphic to $\mathbb{C}^2$, the effective Hilbert space of the system is

$$\mathsf{H} = \mathbb{C}^2_{\text{momentum}} \otimes \mathbb{C}^2_{\text{polarization/spin}} ,$$

and observables corresponding to real linear combinations of $\sigma_1, \sigma_2, \sigma_3$ can be introduced also on the first factor. From the experimental point of view all these observables correspond to devices like *beam-splitters*, *mirrors*, *polarization*

analyzers and so on. A typical apparatus dealing with photons whose momentum states are confined to the $\mathbb{C}^2$ space is the *Mach–Zehnder interferometer* [GaCh08].

In contrast to Bell's analysis, we know a priori that the observables of S_1 are compatible with the observables of S_2, and *this fact has nothing to do with locality*.

We want to show that, in this context, the BCHSH inequality can be used to distinguish between the hidden-variable descriptions assuming realism and non-contextuality and the ones that *do not*. The difference with the similar discussion of Sect. 5.2.3 is that here we will obtain distinct results depending on the states used. In particular, entangled states will play a crucial role even if locality does not enter the game.

Referring again to notation (5.7), we define spin-like observables for each side of the system (whose meaning is not that of spin components in general):

$$A(\mathbf{a}) := \mathbf{a} \cdot \sigma \in \mathfrak{B}(\mathsf{H}_1)_{sa} \quad \text{and} \quad B(\mathbf{b}) := \mathbf{b} \cdot \sigma \in \mathfrak{B}(\mathsf{H}_2)_{sa}$$

so that $\sigma(A(\mathbf{a})) = \sigma(B(\mathbf{b})) = \{\pm 1\}$ in particular.

Let us now suppose that a quantum state $T \in \mathscr{S}(\mathsf{H})$ is given. If we believe in a realistic non-contextual hidden-variable theory, exactly as in Sect. 5.2.3, we must first assume that this state corresponds to a probability measure μ over the space Λ of hidden variables $\lambda \in \Lambda$. Realism and non-contextuality act as follows.

1. *Realism* prescribes that all observables $A(\mathbf{a})$, $B(\mathbf{b})$, for every $\mathbf{a}, \mathbf{b} \in \mathbb{S}^2$, attain a definite value $v_\lambda(A(\mathbf{a})) \in \{\pm 1\}$ and $v_\lambda(B(\mathbf{b})) \in \{\pm 1\}$, for $\lambda \in \Lambda$.
2. *Non-contextuality* demands that the value $v_\lambda(A(\mathbf{a}))$ does not depend on the choice of observables $B(\mathbf{b})$ and $B(\mathbf{b}')$ which can be measured simultaneously with $A(\mathbf{a})$, when $\mathbf{b}' \neq \mathbf{b}$ are such that $B(\mathbf{b})$ and $B(\mathbf{b}')$ are not compatible.

In the previous discussion, when we were considering a pair of *entangled particles*, this independence was due to locality; here, instead, locality cannot be imposed any longer.

As in Bell's analysis of entangled particles, it is convenient to introduce the quantity

$$\chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda) = v_\lambda(A(\mathbf{a}))v_\lambda(B(\mathbf{b})) + v_\lambda(A(\mathbf{a}'))v_\lambda(B(\mathbf{b})) + v_\lambda(A(\mathbf{a}'))v_\lambda(B(\mathbf{b}'))$$

$$-v_\lambda(A(\mathbf{a}))v_\lambda(B(\mathbf{b}')) . \tag{5.21}$$

If we take the expectation value of $\chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda)$ when λ varies in Λ according with its probability distribution μ,

$$\mathbb{E}_\mu(\chi) := \int_\Lambda \chi(\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'|\lambda)d\mu(\lambda) ,$$

with the same reasoning as in the previous section we find again $-2 \leq \mathbb{E}_\mu(\chi) \leq 2$. Defining

$$E_\mu(\mathbf{a}, \mathbf{b}) := \int_\Lambda v_\lambda(A(\mathbf{a})) v_\lambda(B(\mathbf{b})) d\mu(\lambda) \quad \mathbf{a}, \mathbf{b} \in \mathbb{S}^2 \, ,$$

produces the *BCHSH inequality*

$$-2 \leq E_\mu(\mathbf{a}, \mathbf{b}) + E_\mu(\mathbf{a}, \mathbf{b}') + E_\mu(\mathbf{a}', \mathbf{b}') - E_\mu(\mathbf{a}, \mathbf{b}') \leq 2 \quad \text{for every } \mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}' \in \mathbb{S}^2. \quad (5.22)$$

This inequality regarding correlations of measurements of the spin-like components of a bipartite system must be satisfied by every realistic non-contextual theory.

Passing to the quantum side, we can proceed exactly as in the previous section: restrict to *entangled* pure *Bell states* (5.9), take $T_\pm = \langle \Psi_\pm | \cdot \rangle \Psi_\pm$ and fix axes $\mathbf{a}, \mathbf{a}', \mathbf{b}, \mathbf{b}'$ as in (5.10). Then we find (5.11) again:

$$E_{T_\pm}(\mathbf{a}, \mathbf{b}) + E_{T_\pm}(\mathbf{a}, \mathbf{b}') + E_{T_\pm}(\mathbf{a}', \mathbf{b}') - E_{T_\pm}(\mathbf{a}, \mathbf{b}') = \pm 2\sqrt{2} \, .$$

Remark 5.27 The type of entanglement we are considering here is called **intra-particle entanglement**, as it is built with a unique particle entangling the orbital degrees of freedom described on $\mathbb{C}_{\text{orbital}}$ and the spin/polarization freedom degrees described on $\mathbb{C}^2_{\text{polarization/spin}}$. ∎

Since $2\sqrt{2} > 2$, we conclude that *the result predicted by Quantum Theory, with the given choices of observables and Bell's intraparticle entangled states, is incompatible with non-contextual realism.*