The physical origin of quantum nonlocality and contextuality

Adán Cabello
Departamento de Física Aplicada II, Universidad de Sevilla, E-41012 Sevilla, Spain
(Dated: November 20, 2018)

What is the physical principle that singles out the quantum correlations for Bell and contextuality scenarios? Here we show that, if we restrict our attention to correlations that, as is the case for all correlations in classical and quantum physics, can be produced by measurements that: (i) yield the same outcome when repeated, (ii) only disturb measurements that are not jointly measurable, and (iii) all their coarse-grainings have realizations that satisfy (i) and (ii), then the question has a surprising answer. The set of quantum correlations is singled out by the following principle: There is no law governing the outcomes of the measurements; for any scenario made of these measurements, every not inconsistent behavior does take place. “Inconsistent” behaviors are those that violate a condition that holds for measurements satisfying (i)–(iii), namely, that the sum of the probabilities of any set of pairwise exclusive events is bounded by one. Two events are exclusive if they correspond to different outcomes of the same measurement. To prove the result, we begin by characterizing the sets of not inconsistent probability assignments for each “graph of exclusivity,” without referring to any particular scenario, but treating all Bell and contextuality scenarios at once. The restrictions of each scenario are introduced at the end of the proof and then we obtain the set of behaviors that satisfies the above principle for each scenario. Each of these sets is equal to the corresponding set in quantum theory.

I. BACKGROUND

A. The problem of the physical origin of quantum correlations

According to quantum theory (QT), the outcome of some experiments is unpredictable. Nature only allows us to calculate the probability of each possible outcome. For that, we use a set of rules that are collectively known as the Born rule [1]. A long-standing open problem is identifying the physical principle that explains why the behaviors for Bell and contextuality scenarios are the ones allowed by the Born rule (see, e.g., [2–13]). In other words, what is the physical reason responsible for all forms of Bell nonlocality [14] and contextuality [15–17] that, according to QT, are possible in nature. This is the problem we address in this paper.

B. Bell and contextuality scenarios

Definition 1 (Measurement scenario). A measurement scenario is characterized by a set of measurements, \( \{M^{(i)}\}_{i=1}^n \), their respective outcomes, \( \{x_j^{(i)}\}_{j=1}^{k_i} \), and the subsets of \( \{M^{(i)}\}_{i=1}^n \) that are jointly measurable (as defined in Appendix A).

In this paper we will consider two types of measurement scenarios.

Bell scenarios [18, 19]. They involve two or more experimenters. In each round, each experimenter freely chooses and performs one among several possible measurements. The region of spacetime defined from one experimenter’s choice of measurement to the recording of its outcome is spacelike separated from the analogous regions of the other experimenters. Therefore, measurements performed by different experimenters are not only jointly measurable but mutually nondisturbing (as defined in Appendix A). Recall that, in general, joint measurability does not imply nondisturbance [20–22].

Contextuality scenarios [23–27]. They generalize Bell scenarios to situations where jointly measurable measurements are not spacelike separated, allowing, e.g., that measurements can be performed sequentially on the same indivisible system. However, in order to guarantee that joint measurements are mutually nondisturbing, as occurs in Bell scenarios (and for other reasons explained in Appendix B), in these scenarios, measurements are assumed to be “ideal” in the following sense.

Definition 2 (Ideal (aka sharp) measurement [28, 29]). A measurement is ideal if

(i) it gives the same outcome when performed consecutive times,

(ii) only disturbs measurements that are not jointly measurable with it,

(iii) all its coarse-grainings have realizations that satisfy (i) and (ii).

(Coarse-graining is defined in Appendix A). In QT, ideal measurements are those represented by projection-valued measures (PVMs), i.e., self-adjoint projections on a Hilbert space, and are called “projective measurements.” General measurements in QT are represented by positive operator-valued measures (POVMs), i.e., non-negative self-adjoint operators on
a Hilbert space. In QT, general measurements admit implementations as ideal measurements since, by Neumark’s (aka Naimark’s) theorem [30–32], every POVM can be implemented as a PVM. In Appendix B we collect the reasons why the definition of contextuality scenarios is restricted to ideal measurements and why any reasonable physical theory must yield probability distributions for scenarios with ideal measurements.

Definition 3 (Context). For any Bell and contextuality scenario (as defined in this paper) a context is any subset of measurements that only contains jointly measurable (and therefore mutually nondisturbing) measurements.

Definition 4 (Behavior [33]). For a given measurement scenario, a behavior is a set of probability distributions, one for each context.

In the literature, a behavior is also called a “probability model” [34].

C. The problem in one example

To explain the problem of the physical origin of quantum correlations, let us focus on the contextuality scenario with the smallest number of measurements in which the predictions of QT differ from those of classical physics. This scenario admits a realization as a Bell scenario and is called the Bell-CHSH scenario (after Refs. [18, 19]) or (2, 2) Bell scenario, as it involves two parties, traditionally called Alice and Bob, each of them having to choose in each round between two measurements, Alice between \( M^A_1 \) and \( M^A_2 \) and Bob between \( M^B_1 \) and \( M^B_2 \), and each of these measurements only has two possible outcomes: 0 and 1.

A behavior in the Bell-CHSH scenario is characterized by the set \( \{ P(M^A_1, M^B_1 | \psi \rangle | \psi \rangle \} \), with \( x, y \in \{ 1, 2 \} \) and \( a, b \in \{ 0, 1 \} \), of conditional probabilities of obtaining outcomes \( a \) and \( b \) when measurements \( M^A_x \) and \( M^B_y \) are performed, respectively, on a system in state \( \psi \). We will refer to \( \{ P(M^A_1, M^B_1 | \psi \rangle | \psi \rangle \} \) as an “event” (see also Appendix A).

For the Bell-CHSH scenario, the set \( \{ P(M^A_1, M^B_1 | \psi \rangle | \psi \rangle \} \) that determines a behavior has 16 conditional probabilities, but not all of them are independent. On one hand, there are four normalization constraints, since, for each pair of jointly measurable quadruples \((M^A_x, M^B_y)\) (i.e., for each context),

\[
\sum_{a,b} P(M^A_x, M^B_y | \psi \rangle | \psi \rangle) = 1. \tag{1}
\]

On the other hand, the assumption of spacelike separation (or the assumption of mutual nondisturbance) imposes the following eight no-signaling constraints (called no-disturbance constraints when there is timelike separation): for all pairs \((M^A_x, a)\),

\[
\sum_b P(M^A_x, M^B_y | \psi \rangle | \psi \rangle) = \sum_b P(M^A_x, M^B_y | \psi \rangle | \psi \rangle). \tag{2}
\]

and, for all pairs \((M^B_y, b)\),

\[
\sum_a P(M^A_x, M^B_y | \psi \rangle | \psi \rangle) = \sum_a P(M^A_x, M^B_y | \psi \rangle | \psi \rangle). \tag{3}
\]

Only eight of these 12 normalization and no-signaling constraints are independent. For this reason, the whole set \( \{ P(M^A_x, M^B_y | \psi \rangle | \psi \rangle) \} \) is determined by eight suitably chosen conditional probabilities. For example, \( P(M^A_x, M^B_y | \psi \rangle | \psi \rangle) : a \oplus b = (x-1)(y-1), \) where \( \oplus \) denotes sum modulo two.

The question is what is the physical principle that singles out the quantum set of correlations for this scenario. A behavior \( \{ P(M^A_x, M^B_y | \psi \rangle | \psi \rangle) \} \) belongs to the quantum set if and only if it satisfies all the following three conditions:

(i) The initial state of the system \( \psi \) can be represented by a vector with unit norm \( |\psi\rangle \) in a Hilbert space \( \mathcal{H} \) over the complex numbers.

(ii) The state of the system after measurement \( M^{A_x} \) with outcome \( a \) and measurement \( M^{B_y} \) with outcome \( b \) can be represented by a vector with unit norm that can be expressed as

\[
|\psi'\rangle = E^{A_x}_a E^{B_y}_b |\psi\rangle, \tag{4}
\]

where \( E^{A_x}_a \) and \( E^{B_y}_b \) are two projection operators acting on \( \mathcal{H} \), \( E^{A_x}_a \) corresponding to measurement \( M^{A_x} \) with outcome \( a \), and \( E^{B_y}_b \) corresponding to measurement \( M^{B_y} \) with outcome \( b \). These projection operators satisfy

\[
E^{A_x}_a E^{A_x}_a = \delta_{a',a} E^{A_x}_a, \tag{5a}
\]

\[
E^{B_y}_b E^{B_y}_b = \delta_{b',b} E^{B_y}_b. \tag{5b}
\]

That is, projection operators corresponding to different outcomes of the same measurement are orthogonal. In addition, they satisfy

\[
E^{A_x}_a + E^{A_x}_b = 1, \tag{6a}
\]

\[
E^{B_y}_a + E^{B_y}_b = 1, \tag{6b}
\]

where \( 1 \) denotes the identity operator in \( \mathcal{H} \). They also satisfy that, for any \( x, y \),

\[
[E^{A_x}_a, E^{B_y}_b] = 0 \quad \forall a, b. \tag{7}
\]

That is, \( E^{A_x}_a E^{B_y}_b = E^{B_y}_b E^{A_x}_a \).

(iii) The probability of obtaining outcomes \( a \) and \( b \) when measuring \( M^{A_x} \) and \( M^{B_y} \) on state \( \psi \) is equal to \( |\langle \psi' | \psi \rangle|^2 \), where \( \langle \psi' | \psi \rangle \) denotes the inner product of \( |\psi\rangle \) and \( |\psi'\rangle \).

Arguably, conditions (i)–(iii) hardly qualify as a physical principle. The problem is identifying the physical principle that enforces them.
D. The problem in general

For a given Bell or contextuality scenario, a behavior is characterized by the sets of conditional probabilities of obtaining each of the possible sets of outcomes for each of the contexts when the system is in a given state. Probability distributions satisfy normalization constraints (i.e., the sum of the probabilities of all possible outcomes of any context must be equal to one) and nondisturbance constraints between measurements that are jointly measurable and mutually nondisturbing (i.e., the probability of any particular measurement outcome must be independent of which other jointly measurable measurements are performed together).

According to QT, a behavior for a Bell or contextuality scenario is physically realizable if and only if it satisfies the following three conditions:

(I) The initial state of the system $\psi$ can be represented by a vector with unit norm $|\psi\rangle$ in a Hilbert space $\mathcal{H}$ over the complex numbers.

(II) The state of the system after performing a set of jointly measurable and mutually nondisturbing measurements $\{\mathcal{M}^{(i)}\}$ with respective outcomes $\{x_i\}$ can be represented by a vector with unit norm that can be expressed as

$$|\psi'\rangle = \prod_i E^{(i)}_{x_i} |\psi\rangle,$$

(8)

where $\{E^{(i)}_{x_i}\}$ satisfies the following conditions:

(Iia) $\{E^{(i)}_{x_i}\}$ is a set of projection operators acting on $\mathcal{H}$. $E^{(i)}_{x_i}$ corresponds to measurement $\mathcal{M}^{(i)}$ with outcome $x_i$.

(Iib) Projection operators corresponding to different outcomes of the same measurement are orthogonal. That is,

$$E^{(i)}_{x_j} E^{(i)}_{x_k} = \delta_{j,k} E^{(i)}_{x_k}.$$

(9)

(Iic) For any measurement, the sum of the projection operators of all its outcomes is the identity operator in $\mathcal{H}$. That is,

$$\sum_k E^{(i)}_{x_k} = \mathbb{I}.$$

(10)

(Id) Projection operators corresponding to the outcomes of measurements that are jointly measurable and mutually nondisturbing, commute. That is, if $\mathcal{M}^{(i)}$ and $\mathcal{M}^{(k)}$ belong to the same context,

$$[E^{(i)}_{x_j}, E^{(k)}_{x_m}] = 0 \quad \forall j, m.$$

(11)

(III) The probability of obtaining the set of outcomes $\{x_i\}$ when respectively measuring the set of measurements $\{\mathcal{M}^{(i)}\}$ on state $\psi$ is equal to $|\langle \psi' | \psi \rangle|^2$, where $\langle \psi' | \psi \rangle$ denotes the inner product of $|\psi\rangle$ and $|\psi'\rangle$. The conditions that characterize the quantum set of correlations are common to Bell and contextuality scenarios because every quantum correlation in a Bell scenario can be achieved with ideal measurements. This is due to Neumark’s theorem and to the fact that, due to the spacelike separation between jointly measurable measurements characteristic of Bell scenarios, any local POVM admits a local dilation to a PVM that is the same for every context in which the POVM appears.

We will refer collectively to conditions (I)–(III) as the Born rule of QT. Our aim here is singling out a physical principle that enforces these conditions.

E. The try-and-fail approach

Historically, the problem of the physical origin of quantum correlations has been addressed following a try-and-fail approach consisting of several steps (see, e.g., [2–13]). Step 1: A nonquantum behavior $\tilde{\rho}$ is noticed to satisfy the laws of probability and some compelling principles. Then, the question is: what do we learn from the fact that $\tilde{\rho}$ is not allowed by QT? Step 2: After some time, someone notices that $\tilde{\rho}$ violates principle $X$. Step 3: One tries to find a good reason why nature should satisfy $X$ and tries to prove that $X$ singles out the quantum correlations. So far, we have failed in step 3.

For example, in Ref. [35], it is shown that the following behavior for the Bell-CHSH scenario:

$$\tilde{\rho}_{\text{CHSH}} = \{ P(M^{A_1}_0 M^{B_1}_0 | \psi \rangle \langle \psi |), P(M^{A_1}_1 M^{B_1}_0 | \psi \rangle \langle \psi |),$$

$$P(M^{A_1}_0 M^{B_1}_2 | \psi \rangle \langle \psi |), P(M^{A_1}_1 M^{B_1}_2 | \psi \rangle \langle \psi |),$$

$$P(M^{A_0}_0 M^{B_2}_0 | \psi \rangle \langle \psi |), P(M^{A_0}_1 M^{B_2}_0 | \psi \rangle \langle \psi |),$$

$$P(M^{A_0}_0 M^{B_2}_1 | \psi \rangle \langle \psi |), P(M^{A_0}_1 M^{B_2}_1 | \psi \rangle \langle \psi |) \}$$

(12)

satisfies the following principles: no-signalling [6], nontrivial communication complexity [7, 8], no advantage for nonlocal computation [9], information causality [10], macroscopic locality [11], and local orthogonality [13]. In Ref. [36] it is shown that $\tilde{\rho}_{\text{CHSH}}$ fails to satisfy Specker’s principle (i.e., that, if in a set of measurements every pair is jointly measurable, then all the measurements are jointly measurable) [37] and in Ref. [38] it is shown that $\tilde{\rho}_{\text{CHSH}}$ fails to satisfy the no-restriction hypothesis (i.e., that the set of measurements is the dual of the set of states). However, neither Specker’s principle nor the no-restriction hypothesis qualify as physical principles. Moreover, there is no proof that any of them singles out conditions (I)–(III) for every Bell and contextuality scenario.

F. The information-theoretical reconstructions

A different approach to the problem of the physical origin of quantum correlations is that of reconstructing the formalism of QT, and thus conditions (I)–(III), from information-theoretic axioms within the framework of general probabilistic theories (e.g., [39–46]). One problem of all reconstructions
to date is the “inclusion of doubtful principles and the exclusion of infinite quantum theories [such as quantum field theories]” [47]. Other problem is that any of the sets of proposed axioms hardly qualify as a set of physical principles. There is always some axiom whose presence is difficult to justify on physical grounds. If one thinks that “the essential core of quantum mechanics” is “[s]omething that’s written in physical, nonoperational, noninformation-theoretic terms,” then all information-theoretic reconstructions to date “miss the mark for an ultimate understanding of quantum theory” [48].

Perhaps, this second problem is related to the fact that these operational reconstructions have either intentionally or effectively eluded the interpretational problems of QT. For example, Müller and Masanes emphasize that they “do not address (…) [the] [i]nterpretation of quantum mechanics” [49]. Similarly, D’Ariano, Chiribella, and Perinotti point out that “opening the Pandora’s jar of interpretations of quantum theory (…) is not in the purposes of the[ir] book” [50]. Also Hardy confesses that “[w]hen [he] went into this approach, [he] hoped it would help to resolve these interpretational problems. But [he] would say it hasn’t” [51]. This is not strange, since, as pointed out by Ball, “precisely because the reconstructionist program is inherently ‘operational,’ meaning that it focuses on the ‘user experience’ – probabilities about what we measure – it may never speak about the ‘underlying reality’ that creates those probabilities” [52]. Perhaps the neutrality with respect to interpretations and a purely operational character are drawbacks [53]. Perhaps a better option would be reconstructing QT from a specific picture of the underlying reality.

G. The law-without-law hypothesis and the totalitarian principle

In the search for a specific picture of the underlying reality, it is useful to recall two ideas proposed in the past.

In the context of the origin of the universe and the laws in physics, Wheeler argued as follows [54–56]. Suppose that there is an ultimate principle \( P \) from which everything that is knowable about the world follows. If \( P \) were itself a law of physics, then the problem of the origin of \( P \) would be insoluble and hence \( P \) would not be all-explanatory within physics. Therefore, \( P \) cannot be a law of physics. Particles, fields, spacetime, and the initial conditions of the universe emerge from the fact that, at bottom, there is no law. “In contrast to the view that the universe is a machine governed by some magic equation, (…) the world is a self-synthesizing system” [56], “[e]verything is built on the unpredictable outcomes of billions upon billions of elementary quantum phenomena” [54]. What does govern elementary quantum phenomena? Nothing, elementary quantum phenomena are “lawless events” [57]. Physics arises out of this chaos by the action of a regulating principle, “the discovery and proper formulation of which is the number one task of the coming third era of physics” [54].

In the context of the interactions between baryons, antibaryons, and mesons, Gell-Mann made the observation that any process which is not forbidden by a conservation law is not only allowed but must be included in the sum over all paths which contribute to the outcome of the interaction. Gell-Mann name it the “principle of compulsory strong interactions” [58] and commented that “is related to the state of affairs that is said to prevail in a perfect totalitarian state. Anything that is not compulsory is forbidden” [58]. After that, the principle was called “Gell-Mann’s totalitarian principle” [59, 60] and reformulated as “anything not forbidden is compulsory.” However, Trigg [61] and Weinberg [62] have pointed out that the one who deserves the credit for the totalitarian principle is T. H. White because, as Trigg remarks, “[i]n The Sword in the Stone, part 1 of The Once and Future King, Wart, the boy who will later be King Arthur, is being educated by Merlin by being transformed into various animals. One of his experiences is as an ant, and he finds that the ant hill is run on the totalitarian principle” [61]. Specifically, in Chapter XIII of The Sword in the Stone [63].

EVE R YTH ING NOT FORBIDDEN IS COMPULSORY

is the slogan carved over the entrance to each tunnel in the ant fortress.

Recall that our aim is to identify a picture of the underlying reality that explains quantum correlations. For that, we will take literally neither the law-without-law hypothesis nor the totalitarian principle. Instead, we will combine them somehow.

II. MAIN RESULT

Our starting point is the framework of general probabilistic theories, which is the framework common to the approaches commented in Subsec. I F. The relevant concepts of this framework are summarized in Appendix A. Within this framework, let us make the following assumption:

Axiom 1 (No-law/totalitarian principle). There is no law governing the outcomes of some measurements; for any scenario made of these measurements, every not inconsistent behavior does take place.

The definition of “not inconsistent” behavior is presented in Subsec. III B. At first sight, it may seem that nothing can be deduced from Axiom 1. However, we will see that it is, in fact, a powerful axiom. We call Axiom 1 the “no-law/totalitarian principle” because it is the totalitarian principle applied to behaviors for the scenarios made of some measurements and motivated by the law-without-law hypothesis.

The right perspective to understand Axiom 1 is the following: Suppose that there no low governing the outcomes of measurements. What would be the signature of that? Our answer is that, in that case, every behavior that is not inconsistent (as defined in Subsec. III B) should take place.

But a behavior is a set of probability distributions, one for each context. And every probability distribution is a signature of a repeatable experimental procedure consisting of a preparation followed by some measurements. Therefore, the statement that every not inconsistent behavior should take place
implies that it should take place following a set of repeatable experimental procedures of this type.

Classical probability theory does not satisfy Axiom 1. This can be proven by noticing that QT contains classical probability theory as a particular case and, at the same time, allows for nonclassical not inconsistent (as defined in Subsec. III B) behaviors.

The existence of apparently not inconsistent behaviors that are not allowed by QT (see, e.g., [2–6, 35]) would suggest that QT neither satisfies Axiom 1. In fact, we typically think of QT as an island in the space of general probabilistic theories [64], but neither the only island nor the largest one.

Our aim is to prove the following result:

**Theorem 1.** For each Bell and contextuality scenario (as defined in this paper), the set of behaviors allowed by QT is equal to the set of behaviors allowed by Axiom 1.

### III. FUNDAMENTAL CONCEPTS

#### A. The set of probability assignments for a graph of exclusivity

The concept of set of probability assignments for a graph of exclusivity is the key tool for proving Theorem 1. To introduce it, we first need the following definition.

**Definition 5** (Graph of exclusivity). An $n$-vertex graph of exclusivity $G$ represents the relations of mutual exclusivity between $n$ events. Two events are mutually exclusive if they are produced by the same measurement and each of them corresponds to a different outcome of the measurement. The interpretation of the graph is the following: a pair of adjacent vertices represent two mutually exclusive events and a pair of nonadjacent vertices represents two events that are not mutually exclusive.

Let us illustrate the concepts of mutually exclusive events and graph of exclusivity with two examples.

**Example 1**

In Fig. 1 (a), we consider five events produced in the Bell-CHSH scenario. $(M_0^{A1} M_0^{B1} | \psi \rangle)$ denotes the event in which measurements $M_0^{A1}$ and $M_0^{B1}$ are performed on state $\psi$ and outcome 0 is obtained for both measurements. Since in the Bell-CHSH scenario measurements are spacelike separated, $(M_0^{A1} M_0^{B1} | \psi \rangle)$ and $(M_0^{B1} M_0^{A1} | \psi \rangle)$ are equivalent events. The graph of exclusivity of these five events is a pentagon. Adjacent vertices represent mutually exclusive events. That is, there is a measurement that produces both events and each event corresponds to a different outcome. For example, $(M_0^{A1} M_0^{B1} | \psi \rangle)$ and $(M_1^{A1} M_1^{B1} | \psi \rangle)$ are mutually exclusive because the following measurement $M'$ produces each of them: $M_0^{A1}$ is measured on state $\psi$; if the outcome is 0, then $M_0^{B1}$ is measured; if the outcome of $M_0^{A1}$ is 1, then $M_1^{B1}$ is measured instead. Each of $(M_0^{A1} M_0^{B1} | \psi \rangle)$ and $(M_1^{A1} M_1^{B1} | \psi \rangle)$ corresponds to a different outcome of $M'$.

**Example 2**

In Fig. 1 (b), we consider five events whose graph of exclusivity is a also a pentagon. However, this time, the five events are produced in the Klyachko-Can-Binicio˘glu-Shumovsky (KCBS) contextuality scenario [23]. The KCBS scenario is the simplest contextuality scenario in which a qutrit (the quantum system of smallest dimension that can produce contextuality as defined in this paper) produces contextuality. The KCBS scenario has five ideal measurements $M_j^{C_i}$, with $j = 1, \ldots, 5$, with possible outcomes 0 and 1, and such that $M_1^{C_j}$ and $M_{j+1}^{C_j}$ are jointly measurable, with the sum taken modulo five. Since measurements are ideal and $M_1^{C_j}$ and $M_{j+1}^{C_j}$ are jointly measurable, $(M_1^{C_j} M_0^{C_{j+1}} | \varphi \rangle)$ and $(M_0^{C_{j+1}} M_1^{C_j} | \varphi \rangle)$ are equivalent events. $(M_1^{C_j} M_0^{C_{j+1}} | \varphi \rangle)$ and $(M_1^{C_j+1} M_0^{C_{j+2}} | \varphi \rangle)$ are mutually exclusive events be-
cause the following measurement \( M'' \) produces each of them: \( M_{C+1}^{(1)} \) is measured on state \( \varphi \); if the outcome is 0, then \( M_{C}^{(2)} \) is measured; if the outcome of \( M_{C+1}^{(2)} \) is 1, then \( M_{C+2}^{(2)} \) is measured instead.

**Definition 6** (Probability assignment for a graph of exclusivity). A probability assignment for a graph of exclusivity \( G \) is a map

\[
p : V(G) \to [0, 1]
\]

for every \( i \in V(G) \) and such that \( p_i + p_j \leq 1 \) whenever \((i, j) \in E(G)\), where \( V(G) \) and \( E(G) \) are the sets of vertices and edges of \( G \), respectively. A compact way to represent a probability assignment for \( G \) is through a vector of probabilities \( \vec{p} = \{p_i\}_{i=1}^{V(G)} \).

**Definition 7** (The set of probability assignments for a graph of exclusivity [65, 66]). For a given physical theory or set of principles, the set \( S(G) \) of probability assignments for the vertices of the graph of exclusivity \( G \) is the set of all probability assignments for \( G \) allowed by the theory or the set of principles (regardless of which scenario produces each assignment).

It is important to explain how the concept of set of probability assignments for a graph of exclusivity, \( S(G) \), differs from the concept of set of behaviors for a measurement scenario, \( S(S) \). The latter is generated by fixing one scenario, \( S \), and considering all possible initial states and sets of measurements satisfying the conditions characterizing \( S \). In contrast, \( S(G) \) is generated by considering all scenarios, and within each of them, considering all possible initial states and sets of measurements satisfying the conditions characterizing each scenario. Both \( S(G) \) and \( S(S) \) are signatures of a given physical theory. In Appendix D, we have the explicit characterization of \( S(G) \), for any \( G \), in classical probability theory, QT, and a more general theory.

The advantage of studying the sets of probability assignment for each graph of exclusivity (rather than the sets for each scenario) is that we can identify restrictions common to all scenarios, without specifying any particular scenario, but considering all scenarios at once.

**B. Not inconsistent behaviors**

Here we define what is a “not inconsistent” behaviors for a scenario. We use “not inconsistent” rather than “consistent” on purpose. It is easier to agree on a sufficient condition for a behavior to be “inconsistent” than to agree on conditions for a behavior to be “consistent.” We need some definitions before defining a “not inconsistent” behavior for a scenario.

**Definition 8** (Composite experiment). Two independent experiments, one producing \( n \) events with probability assignment \( \vec{p} \) and the other producing \( m \) events with probability assignment \( \vec{q} \), can be seen as a single composite experiment producing \( n \times m \) events with probability assignment given by the tensor product \( \vec{p} \otimes \vec{q} \).

Similarly, one can define a composite experiment composed of three or more experiments.

**Definition 9** (The exclusivity principle). The sum of the probabilities of any set of pairwise exclusive events is bounded by one.

The exclusivity “principle” [12, 65–71] (aka global exclusivity [12, 67] and consistent exclusivity [69, 71]) is related [71] to the concept of orthocoherence in quantum logic (see Appendix F 3 for details), to the principle of local orthogonality [13], and can be derived from different axioms (see Appendix F 3). Here we do not assume the exclusivity principle. Instead, we will prove that it holds for ideal measurements. However, at this point, it is more convenient to introduce it as a definition.

**Definition 10** (Self-inconsistent assignment for a graph of exclusivity). A probability assignment \( \vec{p} \) for a graph of exclusivity \( G \) is self-inconsistent if the exclusivity principle is violated when is applied to a composite experiment composed of a finite number of experiments each of them producing events with graph of exclusivity \( G \) and probability assignment \( \vec{p} \). Otherwise, the assignment is self-consistent.

**Definition 11** (Assignment inconsistent with a self-consistent assignment). A probability assignment \( \vec{q} \) for a graph of exclusivity \( G \) is inconsistent with a self-consistent assignment \( \vec{p} \) for \( G \) if the exclusivity principle is violated when is applied to a composite experiment composed of one experiment producing events with graph of exclusivity \( G \) and probability assignment \( \vec{q} \) and another experiment producing events with graph of exclusivity \( G \) and probability assignment \( \vec{p} \).

**Definition 12** (Not inconsistent assignment for a graph of exclusivity). A probability assignment \( \vec{p} \) for a graph of exclusivity \( G \) is not inconsistent if \( \vec{p} \) is neither self-inconsistent nor inconsistent with any self-consistent probability assignment for \( G \).

**Definition 13** (Not inconsistent behavior for a scenario). A behavior \( \vec{p} \) for the events of a measurement scenario is not inconsistent if \( \vec{p} \) is a not inconsistent assignment for the graph of exclusivity of the basic events and satisfies the restrictions that characterize the scenario.

**IV. STRUCTURE OF THE PROOF OF THEOREM 1**

The proof of Theorem 1 has six steps. Step 1 is presented in Sec. V. There we show that ideal measurements can be constructed in a world in which Axiom 1 holds. Then, the question is whether, for ideal measurements, Axiom 1 defines a unique set of correlations for every Bell and contextuality scenario. To answer this question, first, in step 2, in Sec. VI, we prove that ideal measurements satisfy the exclusivity principle. Then, in steps 3–5, we prove the following result:

**Theorem 2.** For measurements satisfying the exclusivity principle, for each graph of exclusivity, Axiom 1 singles out a unique set of not inconsistent probability assignments and this set is the same allowed by QT.
For that, in step 3, we first identify some self-inconsistent probability assignments for graphs of exclusivity. This is the content of in Sec. VII. Then, in Sec. VIII, we show that some probability assignments for a graph of exclusivity can be proven to be self-consistent.

In step 4, in Sec. IX, we show that, for any self-complementary graph $G_{SC}$ of exclusivity (see Appendix C for the definition of self-complementary graph), we can identify probability assignments for $G_{SC}$ that are inconsistent with self-consistent probability assignments for $G_{SC}$. This turns out to be enough to characterize the set of not inconsistent assignments for any self-complementary graph of exclusivity.

Step 5 is presented in Sec. X. There, we characterize the set of inconsistent assignments for any graph of exclusivity. At this point, we find out that the set of not inconsistent assignments is already identical to the corresponding set in QT. This proves Theorem 2.

Step 6 of the proof of Theorem 1 is presented in Sec. XI. For any measurement scenario $S$, if we denote by $G_S$ the graph of exclusivity associated to the basic events in $S$, then the set of not inconsistent behaviors for $S$, denoted $S(S)$, is the subset of $S(G_S)$ that satisfies the restrictions that characterize $S$. We show that, for any $S$, Theorem 2 and these restrictions make $S(S)$ equal to the quantum set of correlations for $S$.

The conclusions are presented in Sec. XII. We also include some appendices where we collect definitions from the framework of general probabilistic theories (Appendix A), reasons for the importance of ideal measurements (Appendix B), basic notions of graph theory (Appendix C), and results of the graph-theoretic approach to correlations of Refs. [65, 66] used in the proof of Theorem 1 (Appendix D). The appendices also include a proof that ideal measurements satisfy the exclusivity principle (Appendix E) and some remarks on the connection between Theorem 1 and the program of reconstructing QT, the interpretations of QT, and quantum logic (Appendix F).

V. STEP 1 OF THE PROOF: IDEAL MEASUREMENTS IN A WORLD WHERE AXIOM 1 HOLDS

At first sight, it seems to be difficult to do science in a world where Axiom 1 holds. Therefore, the first step of the proof is proving the following.

Lemma 1. Ideal measurements can be constructed in a world where Axiom 1 holds.

Proof. In a world satisfying Axiom 1, “every behavior that is not inconsistent does take place.” This implies that in such a world there are preparations and measurements producing any behavior allowed by classical probability theory for any scenario where all the measurements are jointly measurable. This follows from the fact that any of these behaviors satisfies the definition of “not inconsistent” (Definition 13). Therefore, even if all measurements in such a world were destructive, an experimenter can use suitably chosen preparations and destructive measurements to construct ideal measurements. For example, suppose that an experimenter wants to construct all ideal measurements in a scenario where there are only two jointly measurable three-outcome measurements: one with possible outcomes $x_i$ with $i = 1, 2, 3$, whose ideal implementation is denoted by $M_S$, and the other with possible outcomes $y_j$, with $j = 1, 2, 3$, whose ideal implementation is denoted by $M_C$. For constructing $M_S$ and $M_C$, the experimenter has to identify a nine-outcome destructive measurement $M$ and nine preparations $\{\psi_{(i,j)}\}$ such that, when $M$ is performed on $\psi_{(i,j)}$, the probability of obtaining outcome $(x_i, y_j)$ is one. Then, $M_S$ can be effectively constructed as follows: Measure $M$; if the outcome is $(x_i, y_j)$, then the outcome of $M_S$ is defined to be $x_i$. In addition, after performing the destructive measurement of $M$ with outcome $(x_i, y_j)$, prepare a physical system in state $\psi_{(i,j)}$. The whole process implements $M_S$. Similarly, the experimenter can construct $M_C$. To construct an ideal coarse-graining of $M_S = x_1$, the experimenter can implement $M_S$ and then relabel the outcomes as $x_2$ and $x_3$, as $\pi_1$. Similarly, the experimenter can construct all other ideal measurements in such a toy universe. \(\square\)

VI. STEP 2: IDEAL MEASUREMENTS SATISFY THE EXCLUSIVITY PRINCIPLE

Ideal measurements satisfy the exclusivity principle in the following sense:

Lemma 2 ([28]). For any set $S$ of events produced by ideal measurements and such that every two events in $S$ are mutually exclusive, the sum of the probabilities of all the events in $S$ is bounded by one.

We include the proof in Appendix E.

VII. STEP 3A: IDENTIFYING SELF-INCONSISTENT ASSIGNMENTS

The exclusivity principle, when applied to experiments composed of a number of experiments sharing the same graph of exclusivity $G$ and the same probability assignment $\tilde{p}$, allows us to discover that apparently plausible probability assignments are actually self-inconsistent according to Definition 10.

Let us consider two independent experiments with the same probability assignment (or two mental copies of the same probability assignment). Experiment $A$, producing events $\{\langle M_A \psi | \psi \rangle \}$, with $x = 1, \ldots, n$, whose graph of exclusivity is the one in Fig. 2 (a), and experiment $B$, producing events $\{\langle N_B \psi' | \psi' \rangle \}$, with $y = 1, \ldots, n$, whose graph of exclusivity is the one in Fig. 2 (b). We are assuming that the probability assignments are equal, i.e., $P(M_A \psi | \psi) = P(N_B \psi' | \psi')$. From the perspective of an observer witnessing both experiments (or mental copies), the two experiments constitute a single composite experiment producing events that we denote as $\langle (M_A \psi, N_B \psi' | \psi, \psi') \rangle$, and whose graph of exclusivity is the one in Fig. 2 (c). This graph is obtained by making the OR product (defined in Appendix C) of the graphs in Figs. 2 (a) and (b). By definition of independent experiments
ASSOCIATIONS

FIG. 2. (a) Graph of exclusivity of five events produced in an experiment A. (b) Graph of exclusivity of five events produced in an experiment B. (c) Graph of exclusivity of the 25 events produced in the composite experiment defined by considering experiments A and B as a single experiment.

(or by construction, if we refer to the mental copies),

$$P(M_x|\psi, \psi') = P(M_x|\psi)P(N_y|\psi').$$

To identify self-inconsistent probability assignments, the perspective of Fig. 2 (c) is more powerful than the perspective of Fig. 2 (a).

A. Example 1

Consider the following probability assignment for the events in each of Figs. 2 (a) and (b):

$$\tilde{p}_{C_5} = \{p_i\}_{i=1}^5,$$

$$= \{x, x, x, x, x\},$$

assuming that $i$ and $i + 1$, with the sum modulo 5, are adjacent vertices. The assignment $\tilde{p}_{C_5}$, with $x = \frac{1}{2}$ does not violate the exclusivity principle from the perspective of Fig. 2 (a). However, from the perspective of Fig. 2 (c), every $\tilde{p}_{C_5}$, with $x > \frac{1}{\sqrt{5}}$ is self-inconsistent, as it violates the exclusivity principle. This follows from the fact that, if we consider, e.g., the clique $\{(1,4), (2,2), (3,5), (4,3), (5,1)\}$ of the graph in Fig. 2 (c) and Eq. (14), then the exclusivity principle implies $5x^2 < 1$ [12].

The power of the exclusivity principle for revealing self-inconsistent assignments increases as we consider experiments composed of an increasing number of independent experiments (or mental copies of a probability assignment).

B. Example 2

Consider the following assignment for the vertices of the pentagon of exclusivity:

$$\tilde{q}_{C_5} = \{q_i\}_{i=1}^5 = \{\frac{1}{3}, \frac{2}{3}, y, \frac{2}{3}, \frac{1}{3}\},$$

assuming that $i$ and $i + 1$, with the sum modulo 5, are adjacent vertices. From the perspective of Fig. 2 (a), for any $y > \frac{1}{3} \approx 0.333$, $\tilde{q}_{C_5}$ is self-inconsistent, as it violates the exclusivity principle when we consider, e.g., the clique $\{(3,4)\}$ of the graph in Fig. 2 (a). Then, the exclusivity principle enforces that $y + \frac{1}{3} \leq 1$. However, by considering the perspective of Fig. 2 (c), we observe that, in fact, for any $y > \frac{2}{5} \approx 0.222$, $\tilde{q}_{C_5}$ is self-inconsistent, as it violates the exclusivity principle when we consider Eq. (14) and, e.g., the clique $\{(1,4), (2,2), (3,5), (4,3), (5,1)\}$ of the graph in Fig. 2 (c). Then, the exclusivity principle implies that $y + \frac{2}{5} \leq 1$. No other clique of Fig. 2 (c) enforces stronger restrictions on $y$. The benefit of a wider perspective does not stop here. By considering three copies of $\tilde{q}_{C_5}$, we observe that, for any $y > \frac{1}{3\sqrt{5}} \approx 0.192$, $\tilde{q}_{C_5}$ is self-inconsistent, as it violates the exclusivity principle when we consider Eq. (14) and, e.g., the clique $\{(1,1), (1,2,4), (2,3,3), (2,4,1), (2,5,4), (3,2,3), (3,3,2), (4,1,2), (4,2,5), (5,4,2)\}$ of the graph obtained by making the OR product of three copies of the graph in Fig. 2 (a). Then, the exclusivity principle implies that $2y^2 + \frac{2}{25} \leq 1$. No other clique of the OR product of three pentagons enforces stronger restrictions on $y$. Notice that the clique we have used is neither of the form of a clique of $C_5 \ast C_5$ OR product with a clique of $C_5$ nor of form a clique of $C_5$ OR product with a clique of $C_5 \ast C_5$.

Presumably, the value of $y$ in $\tilde{q}_{C_5}$ allowed by the exclusivity principle becomes more restricted as we consider more copies. The problem is that computing these restrictions is increasingly hard. In fact, even computing the cardinality of the clique of maximal size of $G^{\ast n}$, for $n \geq 4$ constitutes a field of research in itself (see, e.g., [72]).

VIII. STEP 3B: IDENTIFYING SELF-CONSISTENT ASSIGNMENTS

However, there are assignments that are self-consistent (as defined in Definition 10), as it is possible to prove that, no matter how many copies of them we take, they will never violate the exclusivity principle. This leads us to the following lemma whose proof allows us to introduce tools that will be useful later.
Lemma 3. For each graph of exclusivity $G$, the sets of classical and quantum assignments for $G$ are self-consistent.

Proof. Consider a assignment $\vec{p}$ for the graph of exclusivity $G$. Consider $n$ independent experiments, each of them having $\vec{p}$. If one observer witnesses all these experiments and considers them a single composite experiment, it will observe the probability assignment $\vec{p}_n = \bigotimes^n \vec{p}$, denoting the tensor product of $n$ copies of $\vec{p}$. As shown in Ref. [65] (see Appendix D), $\vec{p}$ being quantum is equivalent to say that $\vec{p} \in \text{TH}(G)$ (see Appendix D for the definition and history of TH($G$)). Similarly, $\vec{p}$ being classical means $\vec{p} \in \text{STAB}(G)$ (see Appendix D).

Since, for every $G$, $\text{STAB}(G) \subseteq \text{TH}(G)$, then it is enough to prove that the set of quantum assignments for $G$ is self-consistent. Another result in Ref. [65] is that the set of assignments for $G$ satisfying the exclusivity principle applied only to $G$ is $\text{QSTAB}(G)$ (see Appendix D) and that, for any $G$, $\text{TH}(G) \subseteq \text{QSTAB}(G)$.

If $\vec{p} \in \text{TH}(G)$, then $\vec{p}_n \in \text{TH}(G^{*n})$ [73], where $G^{*n}$ denotes the OR product of $n$ copies of $G$. Therefore, for all $n$, $\vec{p}_n \in \text{QSTAB}(G^{*n})$. \hfill \square

For example, for $\vec{p}_{C_5} = \{x, x, x, x, x\}$, any $x \leq \frac{1}{\sqrt{5}}$ is in $\text{TH}(C_5)$ and therefore is self-consistent [12]. The case $x = \frac{1}{\sqrt{5}}$ has been the target of several experiments [74–76]. On the other hand, $\vec{q}_{C_5} = \{1, \frac{2}{3}, y, \frac{2}{3}, \frac{1}{3}\}$ is in $\text{TH}(C_5)$ if and only if $y \leq \frac{1}{3}$. The case $y = \frac{1}{3}$ has been experimentally addressed in Ref. [77].

IX. STEP 4: CHARACTERIZING THE SET OF NOT INCONSISTENT ASSIGNMENTS FOR ANY SELF-COMPLEMENTARY GRAPH OF EXCLUSIVITY

Consider again $\vec{q}_{C_5} = \{1, \frac{2}{3}, y, \frac{2}{3}, \frac{1}{3}\}$. For $y = \frac{1}{3} + \epsilon$, with $\epsilon > 0$ and suitably small, we cannot classify $\vec{q}_{C_5}$ as self-consistent or self-inconsistent. This is due to the practical impossibility of identifying all the cliques of the graphs resulting from the OR product of many copies of a given graph, which is related to a famous open problem in graph theory and zero-error information theory (see Appendix D). More generally, there are many probability assignments for many graphs of exclusivity which we cannot classify as self-consistent or self-inconsistent. We do not even know whether it may exist a third class of assignments that can never be proved to be self-consistent or self-inconsistent. Therefore, we have to consider a new approach.

In physics, we can fix the measurements and change the preparations. This does not change the relations of exclusivity of the events, but changes their probability assignments. This operation opens a new possibility for identifying inconsistent assignments. Suppose a particular probability assignment $\vec{q}$ for $G$ that is self-consistent. Now consider another probability assignment $\vec{q}'$, for the same $G$, that we cannot classify as self-consistent or self-inconsistent. Now recall the definition of probability assignment inconsistent with a self-consistent assignment (Definition 11). Then we can check whether $\vec{q}'$ is not inconsistent with every self-consistent assignment for $G$. This approach leads to the following fundamental discovery.

Lemma 4. For every graph of exclusivity $G$ such that $G = \overline{G}$ (i.e., $G$ is self-complementary as defined in Appendix C), every probability assignment is either self-consistent or inconsistent with a self-consistent assignment.

Proof. Let $\vec{x}$ be a probability assignment that is not in $\text{TH}(G)$ (e.g., one that we cannot classify as self-inconsistent). Sec. 3 in Ref. [78] shows that $\text{TH}(G)$ can also be characterized as

$$\text{TH}(G) = \{ \vec{p} \in \mathbb{R}^n ; p_i \geq 0, \vartheta(G, \vec{p}) \leq 1 \},$$

where, using Lemma 4.3 in Ref. [78],

$$\vartheta(G, \vec{p}) = \max \left\{ \sum_i p_i q_i ; \vec{q} \in \text{TH}(\overline{G}) \right\}. \quad (18)$$

This implies that, if $G = \overline{G}$, then $\text{TH}(G)$ contains all the assignments $\vec{p}$ that are not inconsistent with any of the assignments in $\text{TH}(G)$. And, according to Lemma 3, all the assignments in $\text{TH}(G)$ are self-consistent. Therefore, if $\vec{x}$ is not in $\text{TH}(G)$, there is at least one $\vec{q} \in \text{TH}(G)$ such that, when we consider the assignment $\vec{x} \otimes \vec{q}$ of the composite experiment, then

$$\sum_i x_i y_i > 1. \quad (19)$$

That is, $\vec{x}$ and $\vec{y}$ violate the exclusivity principle. Hence, $\vec{x}$ is inconsistent with a self-consistent assignment. \hfill \square

Therefore, from the proof of Lemma 4, taking into account the definition of not inconsistent assignment (Definition 12), we can state the following:

Lemma 5. For measurements satisfying the exclusivity principle, for each graph of exclusivity $G$ such that $G = \overline{G}$, Axiom 1 singles out a unique set of not inconsistent probability assignments for $G$; this set is $\text{TH}(G)$.

This is quite amazing since, as mentioned earlier and shown in Appendix D, for any $G$, $\text{TH}(G)$ is exactly the quantum set of probability assignments for the graph of exclusivity $G$. Therefore, at least for a particular type of graphs of exclusivity (the self-complementary graphs), we have been able to single out the quantum set by only keeping the not inconsistent probability assignments.

X. STEP 5: CHARACTERIZING THE SET OF NOT INCONSISTENT ASSIGNMENTS FOR ANY GRAPH OF EXCLUSIVITY

The problem is that Lemma 5 only applies to a very specific type of graphs (those satisfying $G = \overline{G}$). However, now we have all the tools needed to prove our first main result, namely, Theorem 2.

Theorem 2. For measurements satisfying the exclusivity principle, for each graph of exclusivity, Axiom 1 singles out a unique set of not inconsistent probability assignments and this set is the same allowed by QT.
a producing two mutually exclusive events: measurable ideal measurements, A and C, each of them producing two mutually exclusive events: A producing events \( a_0 \) or \( a_1 \), \( B \) producing \( b_0 \) or \( b_1 \), and \( C \) producing \( c_0 \) or \( c_1 \).

For example, \( A, B, \) and \( C \) can be tossing three different coins. Suppose that the experimenter defines the following \( 4n \) events: \( \{(a_0, e_k), (a_1, b_0, x_k), (b_1, c_0, y_k), (c_1, z_k)\}_{k=1}^n \), where, e.g., \((a_0, e_1)\) is the event in which measurement \( A \) gives \( a_0 \) and experiment \( E \) gives \( e_1 \). We denote by \( H(G) \) the graph of exclusivity of these \( 4n \) events. Fig. 3 (c) shows \( H(G) \) for \( G = C_7 \). The graph \( H(G) \) in Fig. 3 (c) is the generalized composition \( \mathcal{G}[G_1, G_2, G_3, G_4] \), where \( G_1, G_2, G_3, G_4 \) are the graphs in Fig. 3 (b) and \( \mathcal{G} \) is the graph in Fig. 3 (d). See Appendix C for the definition of generalized composition of graphs.

There are two properties of \( H(G) = \mathcal{G}[G, \overline{G}, \overline{G}, \overline{G}] \) that are crucial for the proof. The first property is that \( \mathcal{G} \) is a perfect graph (as defined in Appendix C). For perfect graphs, the set of probability assignments satisfying the exclusivity principle is equal to the set of probability assignments in classical probability theory (see Appendix D). Therefore, the fact that \( \mathcal{G} \) is a perfect graph implies that, without loss of generality, any possible probability assignment \( \vec{h} \) for \( H(G) \) can be implemented by suitably choosing a probability assignment \( \vec{p} \) for \( \{e_k\}_{k=1}^n \), a probability assignment \( \vec{x} \) for \( \{x_k\}_{k=1}^n \), a probability assignment \( \vec{y} \) for \( \{y_k\}_{k=1}^n \), a probability assignment \( \vec{z} \) for \( \{z_k\}_{k=1}^n \), a probability assignment \( \vec{a} \) for \( \{a_0, a_1\} \), a probability assignment \( \vec{b} \) for \( \{b_0, b_1\} \), and a probability assignment \( \vec{c} \) for \( \{c_0, c_1\} \). In other words, the set of not inconsistent probability assignments for \( H(G) \) can be written as

\[
S[H(G)] = \text{convex hull}\{\vec{h} = (\vec{p}, \vec{x}, \vec{y}, \vec{z})
\in\{(S(G), 0^{1|V(G)}], 0^{1|V(G)]}, S(G)\}, (S(G), 0^{1|V(G)]}, S(\overline{G}), 0^{1|V(G)]}), (0^{1|V(G)]}, S(\overline{G}), 0^{1|V(G)]})\},
\]

(20)

where, e.g., \( S(G) \) is the set of not inconsistent probability assignments for \( G \). Notice that the convex hull plays the role of \( \vec{a}, \vec{b}, \) and \( \vec{c} \), and the vectors inside the curly brackets are enforced by the mutual exclusivity relations depicted in Fig. 4 (a).

The second crucial property of \( H(G) \) is that, for any \( G \), \( H(G) = \overline{H(G)} \) (i.e., \( H(G) \) is self-complementary). This can be seen in Fig. 4. Therefore, Lemma 4 characterizes the set of not inconsistent probability assignments for \( H(G) \) defined by Axiom 1. That is,

\[
S[H(G)] = TH[H(G)].
\]

(21)
In addition,

$$\text{TH}[H(G)] = \text{convex hull}(\vec{h} = (\vec{p}, \vec{x}, \vec{g}, \vec{z})) \in \{(\text{TH}(G), 0^{\vert V(G)\vert}, 0^{\vert V(G)\vert}, \text{TH}(G)),$$

$$\text{(TH}(G), 0^{\vert V(G)\vert}, \text{TH}(G), 0^{\vert V(G)\vert}),$$

$$\{0^{\vert V(G)\vert}, \text{TH}(G), 0^{\vert V(G)\vert}, \text{TH}(G))\}\}.$$  \hspace{1cm} (22)

Therefore, Eqs. (20), (21), and (22) imply

$$S(G) = \text{TH}(G).$$ \hspace{1cm} (23)

And this holds for any $G$. \hspace{1cm} \Box

Before this proof, the quantum set, i.e., TH(G), was one among infinitely many possible sets satisfying the exclusivity principle. In particular, there was the possibility that, for a nonself-complementary graph $G$, a theory produced $S(G)$ containing $\vec{p} \notin \text{TH}(G)$ (of course, $\vec{p}$ such that it cannot be proven self-inconsistent) and $S(G) = \text{TH}(G) \setminus \{\vec{q}\}$, that is, $S(G)$ minus some probability assignments $\{\vec{q}\}$ [68]. After our proof, $\vec{p} \notin \text{TH}(G)$ is not possible anymore because, by Eq. (20), would produce $\vec{h} \notin \text{TH}[H(G)]$, i.e., an inconsistent assignment for $H(G)$. Therefore, the proof shows that the set of probability assignments defined by Axiom 1 is unique and equal to the set in QT.

XI. STEP 6: CHARACTERIZING THE SET OF NOT INCONSISTENT BEHAVIORS FOR ANY BELL AND CONTEXTUALITY SCENARIO

So far, the strategy for proving Theorem 1 has consisted in studying all Bell and contextuality scenarios at once and characterizing the set of not inconsistent probability assignments for each graph of exclusivity. Intentionally, we have paid no attention to which measurement scenario originates the assignments. Instead, we have considered them all at the same time. Thanks to this perspective, after step 5, we have discovered that, for any graph of exclusivity $G$, the only not inconsistent probability assignments are those that belong to TH($G$), i.e., those that satisfy the following conditions:

(I') The initial state of the system can be associated to a unit vector $|\psi\rangle$ in a finite-dimensional vector space $\mathcal{V}$ with an inner product.

(II') The state after a measurement $\mathcal{M}^{(i)}$ with outcome $x_i$ can be associated to a unit vector $|\mathcal{M}^{(i)}_{x_i}|\psi\rangle$ in $\mathcal{V}$. Post-measurement states corresponding to mutually exclusive events can be associated to mutually orthogonal vectors.

(III') The probability of the event $\langle \mathcal{M}^{(i)}_{x_i}|\psi|\psi\rangle$ is equal to $|\langle \mathcal{M}^{(i)}_{x_i}|\psi\rangle|^2$, where $\langle \mathcal{M}^{(i)}_{x_i}|\psi\rangle$ denotes the inner product of $|\psi\rangle$ with $|\mathcal{M}^{(i)}_{x_i}|\psi\rangle$.

Conditions (I')–(III') mirror conditions (I)–(III) in Subsec. 1D, with the following nuances:

In (I'), vector spaces with an inner product are not required to be Hilbert spaces over the complex numbers as in (I). However, it is known that for Bell and contextuality scenarios quantum correlations also allow for representations in real vector spaces [79–82] (see also Appendix F 1). Therefore, this can be taken to be a virtue rather than a weakness.

Collectively, (I'), (II'), and (III') imply that, associated to each event $\langle \mathcal{M}^{(i)}_{x_i}|\psi\rangle$, there is a map $E^{(i)}_{x_i}$ that transforms the unit vector $|\psi\rangle$ into the unit vector $|\mathcal{M}^{(i)}_{x_i}|\psi\rangle$ and satisfies the following conditions:

(II'a) $E^{(i)}_{x_i}E^{(j)}_{x_j}|\psi\rangle = E^{(j)}_{x_j}E^{(i)}_{x_i}|\psi\rangle$, since measurement $\mathcal{M}^{(i)}$ must not disturb itself. Therefore, for any outcome $x_j$, the map $E^{(i)}_{x_j}$ must be idempotent.

(II'b) For any measurement $\mathcal{M}^{(i)}$ with outcomes $\{x_j\}$, $\{E^{(i)}_{x_j}|\psi\rangle\}_j$ is a set of orthonormal vectors.

(II'c) For any measurement $\mathcal{M}^{(i)}$ with outcomes $\{x_j\}$, $\sum_j |\langle E^{(i)}_{x_j}|\psi\rangle|^2 = 1$.

(II'd) If $\mathcal{M}^{(i)}$ and $\mathcal{M}^{(k)}$ are jointly measurable and mutually nondisturbing, then $E^{(i)}_{x_j}E^{(k)}_{x_m}|\psi\rangle = E^{(k)}_{x_m}E^{(i)}_{x_j}|\psi\rangle$, $\forall j, m$.

Conditions (II'a)–(II'd) are thus equivalent to conditions (IIa)–(IIId) in Subsec. 1D.

A. Restrictions that characterize a scenario

Our aim now is to characterize the set $S(S)$ of not inconsistent behaviors for each scenario $S$. According to Definition 13, on the one hand, the elements of $S(S)$ must be “not inconsistent for the graph of exclusivity of the basic events” of $S$. Therefore, if we consider the graph of exclusivity $G_S$ of the basic events in $S$, the behaviors in $S(S)$ must satisfy (I')–(III'). That is, $S(S)$ must be a subset of $S(G_S)$. In other words, $S(S) \subseteq \text{TH}(G_S)$. As an example of $G_S$, Fig. 5 (a) shows the graph of exclusivity of the basic 16 events of the Bell-CHSH scenario. On the other hand, the behaviors in $S(S)$ must “satisfy the restrictions that characterize the scenario.”

For each Bell or contextuality scenario as defined in this paper, there are three types of restrictions characteristic of each scenario:

(A) The normalization constraints. That is, the sum of the probabilities of all possible outcomes of any set of measurements that are jointly measurable must be equal to one.

(B) The nondisturbance constraints between jointly measurable measurements. That is, the probability of any particular measurement outcome must be independent of which other jointly measurable measurements are performed.

(C) The exclusivity constraints.
(C1) Each event must be produced by a specific combination of outcomes of a specific subset of jointly measurable measurements of the set of measurements of that scenario.

(C2) The relations of mutual exclusivity between any two events must be explained by a specific measurement constructed from the measurement outcomes that produce each of the two events.

For example, for the Bell-CHSH scenario, restrictions of type (A) are the four normalization constraints given by Eq. (1), and restrictions of type (B) are the eight no-signaling constraints given by Eqs. (2) and (3). Conditions (I’)—(III’) plus restrictions of type (A) enforce the extra constraint given by Eq. (6) on the projection operators. Similarly, conditions (I’)—(III’) plus restrictions of type (B) enforce extra constraints on the projection operators.

To explain what the restrictions of type (C) add to conditions (I’)—(III’) for the Bell-CHSH scenario, it is useful to present two examples.

Example 1

Consider event \( (M_{0}^{B2}M_{0}^{A1}|\psi\rangle) \). According to (I’)—(III’) and (C1), \( P(M_{0}^{B2}M_{0}^{A1}|\psi\rangle) \) must be expressed as \( |\langle E_{0}^{B2}E_{0}^{A1}|\psi\rangle|^{2} \), where \( |\psi\rangle \) and \( |E_{0}^{B2}E_{0}^{A1}\rangle = E_{0}^{B2}E_{0}^{A1}|\psi\rangle \) are unit vectors and \( E_{0}^{B2}E_{0}^{A1} = E_{0}^{A1}E_{0}^{B2} \). Now consider event \( (M_{1}^{A1}M_{1}^{B1}|\psi\rangle) \). Similarly, according to (I’)—(III’) and (C1), \( P(M_{1}^{A1}M_{1}^{B1}|\psi\rangle) \) must be expressed as \( |\langle E_{1}^{A1}E_{1}^{B1}|\psi\rangle|^{2} \), where \( |\psi\rangle \) and \( |E_{1}^{A1}E_{1}^{B1}\rangle = E_{1}^{A1}E_{1}^{B1}|\psi\rangle \) are unit vectors and \( E_{1}^{A1}E_{1}^{B1} = E_{1}^{B1}E_{1}^{A1} \). Notice that the measurements that characterize each of the events are not the same. But, since \( (M_{0}^{B2}M_{0}^{A1}|\psi\rangle) \) and \( (M_{1}^{A1}M_{1}^{B1}|\psi\rangle) \) are mutually exclusive, there must be a measurement \( M' \) that produces both events so that each of the events is associated to a different outcome of \( M' \). However, \( M' \) is not any measurement but, due to (C2), a specific measurement made from the measurement outcomes that produce these two events. Namely, \( M' \) is: measure \( M^{A1} \), if the outcome is 0, then measure \( M^{B2} \); if the outcome is 1, then measure \( M^{B1} \). Measurement \( M' \) has four outcomes and produces four events: two of them are the ones we are interested in and the other two are \( (M_{1}^{B2}M_{0}^{A1}|\psi\rangle) \) and \( (M_{0}^{A1}M_{1}^{B1}|\psi\rangle) \). Therefore, according to (I’)—(III’), (C1), and (C2), \( \{ E_{0}^{B2}E_{0}^{A1}|\psi\rangle, E_{1}^{A1}E_{1}^{B1}|\psi\rangle, E_{0}^{B2}E_{0}^{A1}|\psi\rangle, E_{1}^{A1}E_{1}^{B1}|\psi\rangle \} \) must be a set of orthonormal vectors.

Example 2

Consider the events \( (M_{0}^{A1}M_{0}^{B1}|\psi\rangle) \) and \( (M_{0}^{A1}M_{1}^{B1}|\psi\rangle) \). They are mutually exclusive. However, in this case, the measurement that produces both events is the same. Therefore, the measurement \( M' \) that explains why they are mutually exclusive is that measurement. That is, \( M'' \) is: measure \( M^{A1} \) and \( M^{B1} \). If we take into account (I’)—(III’), (C1), and (C2), we obtain the corresponding restrictions on \( P(M_{1}^{A1}M_{1}^{B1}|\psi\rangle) \) and \( P(M_{0}^{A1}M_{1}^{B1}|\psi\rangle) \).

B. Graphical representation of the restrictions of type (C)

The reason why restrictions of type (C) strongly constrain the behaviors is because, for a given scenario, there is a limited number of measurements, a limited number of measurement outcomes (and thus of projection operators), and a limited number of joint measurabilities. With these ingredients, and only with them, one must produce every event and construct every measurement that explains every relation of mutual exclusivity. For example, in the Bell-CHSH scenario, there are only eight measurement outcomes, two for each of the four local measurements. Each event is produced by two specific measurement outcomes. Each measurement that explains why two events are mutually exclusive must be made of a specific combination of them (see the examples of \( M' \) and \( M'' \) above).

To understand the importance of the restrictions of type (C), it is useful to represent them graphically by adding colors to the edges and vertices of the graph of exclusivity of the basic events of the scenario as follows. Each measurement is represented by a color. Therefore, there are as many colors as measurements the scenario has. For example, in the Bell-CHSH scenario, there are four colors. Each vertex has as many colors as measurements produce the corresponding event. Therefore, in the Bell-CHSH scenario, each vertex has two colors, since each event is produced by the outcomes of two jointly measurable measurements. To distinguish the outcomes of each measurement we use different shapes. In the Bell-CHSH scenario, since each local measurement only has two outcomes, we represent outcome 0 by half circumference with white inside, and the outcome 1 by half circumference with the color of the measurement inside. Finally, each edge is colored by the colors of the local measurements common to the connected events. This way, the measurement that explains why two events are mutually exclusive is characterized by the color(s) of the edge and the colors and shapes of the connected vertices. For example, the graph of exclusivity of the basic 16 events of the Bell-CHSH scenario with the restrictions of type (C) is shown in Fig. 5 (b).

While in Fig. 5 (a) there are 56 edges, each of them representing a relation of mutual exclusivity explained by one measurement and each of the 56 measurements could be different, Fig. 5 (b) forces us to explain each of the 56 relations of mutual exclusivity by a specific combination of the eight measurement outcomes available in the Bell-CHSH scenario.

To appreciate the role of the restrictions of type (C), it is worth pointing out that the behavior in Eq. (12) satisfies conditions (I’)—(III’) and all the restrictions of type (A) and (B) for the Bell-CHSH scenario. However, it fails to satisfy all the restrictions of type (C) for the Bell-CHSH scenario (see Ref. [83] for a proof). This means that the vector of probabilities in Eq. (12) is allowed by QT, yet not in the Bell-CHSH scenario, but in a different contextuality scenario (see
Bell [18, 19] and Kochen-Specker [2, 84–86] theorems tell us how we cannot explain the world described by QT. However, they do not tell us why. They do not answer the question of what is the physical principle that singles out which behaviors are allowed by QT. Nevertheless, when we combine both theorems, the notion of contextuality for ideal measurements naturally emerges as the simplest way to generalize Bell scenarios in order to include correlations between events that are not spacelike separated [23–27]. This notion enables us to formulate the problem of the physical origin of quantum correlations in a sufficiently comprehensive manner. For each Bell and contextuality scenario, QT asserts that only a specific set of behaviors is physically realizable. Why these sets and not larger sets? Here we have answered this question. We have shown that there is an ontological message in the specific way QT is nonlocal and contextual, as each of these sets is identical to the one produced in a world where no law governs the outcomes of the measurements and, as a consequence, any not inconsistent behavior takes place.

On the one hand, from this result we learn that, for ideal measurements (or, more generally, for measurements satisfying the exclusivity principle), QT is not merely an island in the space of general probabilistic theories, but the largest possible island. This island has smaller subsets like classical probability theory, but there is no consistent physical probabilistic theory beyond this island. At the light of this result, “the message of the quantum” [87] and “the characteristic trait of quantum mechanics” [88] is, in fact, that, for any scenario, all not inconsistent behaviors take place. This is what determines which forms of incompatibility, entanglement, randomness, nonlocality, and contextuality are possible in QT.

This result makes us realize that we do know how to produce any not inconsistent behavior for any Bell and contextuality scenario [89–91] with a wide variety of physical systems (see Appendix F2). So far, we have never failed to produce any of these behaviors, even when we aimed to the most extreme forms of nonlocality [92] and contextuality [76]. Therefore, these experimental results can be taken as evidence supporting that the lack of laws governing the outcomes is the physical reason that explains what we observe. This explanation seems to us simpler and more satisfactory than any other proposed explanation (see Appendix F2).

Alternatively, one may ignore Axiom 1 and step 1 of the proof and keep the “technical” result: In any world that admits measurements satisfying the exclusivity principle and that can be combined both temporally and in parallel, the constraints arising from the laws of probability, in conjunction with these possibilities of combination, are sufficient to constrain the behaviors for each measurement scenario to those of QT. No further constraints are needed – there is “no additional law”...
rather than what has already been specified. However, since the already specified “laws” constitute a framework that is common to any physical probabilistic theory that has classical probability theory as a particular case, it is difficult to not interpret this technical result as a crucial insight on how our world is.

ACKNOWLEDGMENTS

The author thanks S. Abramsky, B. Amaral, M. Araújo, E. Cavalcanti, G. Chiribella, C. A. Fuchs, P. Grangier, M. J. W. Hall, G. Jaeger, M. Kleinmann, M. P. Müller, M. Navascués, J. R. Portillo, M. F. Pusey, A. Ribeiro de Carvalho, R. Shack, K. Svozil, M. Terra Cunha, H. M. Wiseman, Z.-P. Xu, M. Żukowski, and W. H. Zurek for help with the preparation. G. Cañas and A. J. López-Tarrida for help with Figs. 1, 2, Z.-P. Xu for help with Fig. 1, S. López-Rosa for help with Fig. 2, S. López-Rosa for help with Fig. 5, A. J. López-Tarrida for help with Fig. 6, and C. Budroni, S. López-Rosa, and A. J. López-Tarrida for extensive comments on the successive versions of the manuscript. This work was supported by the Foundational Questions Institute (FQXi) Large Grant “The Observer Observed: A Bayesian Route to the Reconstruction of Quantum Theory” (FQXi-RFP-1608), the MINECO project “Advanced Quantum Information” (FIS2014-60843-P) with FEDER funds, and the Knut and Alice Wallenberg Foundation project “Photonic Quantum Information.” The author is also supported by the MINECO-MCINN project “Quantum Tools for Information, Computation and Research” (FIS2017-89609-P) with FEDER funds.

Appendix A: Basic concepts in general probabilistic theories

Here we recall some definitions taken from the framework of general probabilistic theories (see, e.g., Refs. [39–41, 50]) that are used in this paper.

Definition 14 (Probabilistic theory). A probabilistic theory is one that specifies the probabilities of each possible outcome of each possible measurement given each possible preparation.

Definition 15 (Preparation). A preparation is a sequence of unambiguous and reproducible experimental procedures on a physical system.

Definition 16 (State). The state of a system describes a preparation of that system. The state provides a probability distribution for the outcomes of each possible subsequent measurement on the system. The state can be represented by a vector of probabilities. Transformations of a system produce transformations of this vector. A unified account of transformations and measurements can be given by allowing probabilistic transformations.

Definition 17 (Measurement (aka non-destructive or non-demolition measurement)). A measurement $\mathcal{M}$ is an interaction, between a system and a measurement device, that transforms the state of the system into a new state and produces an outcome. A measurement is thus characterized by a set $\{\mathcal{M}_x\}_{x \in X}$ of transformations, where $X$ is the set of outcomes of $\mathcal{M}$.

$\mathcal{M}_x$ denotes the state of the system after performing measurement $\mathcal{M}$ and obtaining outcome $x$ if the state of the system before the measurement was $\psi$.

Definition 18 (Event). An event $(\mathcal{M}_x \psi|\psi)$ is the process of transformation from the state $\psi$ into the state $\mathcal{M}_x \psi$ due to the measurement $\mathcal{M}$ with outcome $x$.

$P(\mathcal{M}_x \psi|\psi)$ denotes the probability of $(\mathcal{M}_x \psi|\psi)$.

Two transformations $\mathcal{M}_x$ and $\mathcal{N}_y \psi$ are equivalent, denoted $\mathcal{M}_x \sim \mathcal{N}_y$, if $P(\mathcal{M}_x \psi|\psi) = P(\mathcal{N}_y \psi|\psi)$ for all $\psi$.

Two events $(\mathcal{M}_x \psi|\psi)$ and $(\mathcal{N}_y \psi|\psi)$ are equivalent, denoted $(\mathcal{M}_x \psi|\psi) \sim (\mathcal{N}_y \psi|\psi)$, if $\mathcal{M}_x \sim \mathcal{N}_y$.

Definition 19 (Destructive measurement). A measurement is destructive if it forbids any further measurement on the system.

Hereafter, unless specified, we will assume that measurements are non-destructive.

Definition 20 (Disturbance). A measurement $\mathcal{M}$ disturbs a (possibly destructive) measurement $\mathcal{N}$ if, from the outcome statistics of $\mathcal{N}$, one can detect whether $\mathcal{M}$ was performed.

That is, two measurements $\{\mathcal{M}_x\}_{x \in A}$ and $\{\mathcal{M}_y\}_{y \in B}$ are mutually nondisturbing if, for any state $\psi$, for any outcome $x \in A$, $P(\mathcal{M}_x \psi|\psi) = \sum_{y \in B} P(\mathcal{M}_x \mathcal{M}_y \psi|\psi) = \sum_{b \in B} P(\mathcal{M}_x \mathcal{M}_b \psi|\psi)$, and, for any outcome $y \in B$, $P(\mathcal{M}_y \psi|\psi) = \sum_{a \in A} P(\mathcal{M}_a \mathcal{M}_y \psi|\psi) = \sum_{a \in A} P(\mathcal{M}_a \mathcal{M}_y \psi|\psi)$.

When measurements are timelike separated, $(\mathcal{M}_a \mathcal{M}_b \psi|\psi)$ denotes the event in which $\mathcal{M}_a$ is performed after $\mathcal{M}_b$.

Definition 21 (Joint measurability (aka Compatibility)). Two (possibly destructive) measurements, $\{\mathcal{M}_x\}_{x \in X}$ and $\{\mathcal{N}_y\}_{y \in Y}$, are jointly measurable (or compatible) if there is a (possibly destructive) measurement $\{\mathcal{L}_{x,y}\}_{x \in X, y \in Y}$ such that, for all states $\psi$, for all states $x \in X$, for all states $y \in Y$, the probability $P(\mathcal{L}_{x,y} \psi|\psi) = \sum_{x \in X} P(\mathcal{L}_{x,y} \psi|\psi)$ and, for all $\psi$ and all $y \in Y$, $P(\mathcal{N}_y \psi|\psi) = \sum_{x \in X} P(\mathcal{L}_{x,y} \psi|\psi)$.

Nondisturbance implies joint measurability, but the reverse implication does not hold in general [20–22]. Since ideal measurements only disturb measurements that are not jointly measurable, then for ideal measurements joint measurability implies mutual nondisturbance.

Definition 22 (Coarse-graining). A (possibly destructive) measurement $\{\mathcal{C}_y\}_{y \in Y}$ is a coarse-graining of a (possibly destructive) measurement $\{\mathcal{F}_x\}_{x \in X}$ if for all $y \in Y$ exists $X_y \subseteq X$ such that, for all $\psi$, $P(\mathcal{C}_y \psi|\psi) = \sum_{x \in X_y} P(\mathcal{F}_x \psi|\psi)$ and $X_y \cap X_{y'} = \emptyset$ if $y \neq y'$.

Hence, a measurement and its coarse-grainings are jointly measurable. Also, two measurements are jointly measurable if both of them are coarse-grainings of the same measurement.
Definition 23 (Mutual exclusivity of events). Two events $(\mathcal{M}_z \psi | \psi)$ and $(\mathcal{N}_y \psi | \psi)$ are mutually exclusive, denoted $(\mathcal{M}_z \psi | \psi) \perp (\mathcal{N}_y \psi | \psi)$, if there is a measurement $\{\mathcal{L}_z\}_{z \in \mathbb{Z}}$ with two different outcomes $z$ and $z'$ such that $(\mathcal{M}_z \psi | \psi) \sim (\mathcal{L}_z \psi | \psi)$ and $(\mathcal{N}_y \psi | \psi) \sim (\mathcal{L}_{z'} \psi | \psi)$.

If $(\mathcal{M}_z \psi | \psi) \sim (\mathcal{N}_y \psi | \psi)$ and $(\mathcal{M}_z \psi | \psi) \perp (\mathcal{L}_z \psi | \psi)$, then $(\mathcal{N}_y \psi | \psi) \perp (\mathcal{L}_z \psi | \psi)$.

Appendix B: Contextuality and ideal measurements

1. Why the definition of contextuality scenarios is restricted to ideal measurements

The restriction to ideal measurements in the definition of contextuality scenarios is motivated by the following reasons:

(1a) It assures that the “contextuality” of a behavior (indicated by the fact that it does not belong to the set of noncontextual behaviors for the contextuality scenario considered) can be taken as a signature of nonclassicality. As pointed out in Ref. [93], the assumption that the outcome of a measurement depends deterministically on the ontic state (which is the assumption satisfied by the extreme points of the set of noncontextual behaviors) is reasonable if and only if the measurement is ideal. In particular, it is not a physically plausible assumption when applied to a noisy measurement, since, in this case, the outcome may have an indeterministic dependence on the ontic state of the measured system [93]. It is worth mentioning that there are other notions of contextuality [16, 69, 94] that apply to more general measurements.

(1b) It provides a quantitative measure of nonclassicality for correlations between not spacelike separated events, as the classical simulation of this contextuality has quantifiable memory [95] and thermodynamical overcosts [96, 97].

(1c) It leads to the notion of contextuality that has been proven to be necessary for quantum speedup in universal quantum computation via magic state distillation [98].

(2a) It allows us to extend the concept of Bell scenario to situations without spacelike separation, as it allows us to define contexts in which, as naturally occurs in Bell scenarios, measurements in the same context do not disturb each other (as, for ideal measurements—but not for general measurements—joint measurability implies mutual nondisturbance [21]).

(2b) It assures that the classical and quantum sets of behaviors for contextuality scenarios are identical regardless of whether there is timelike or spacelike separation. The set of classical behaviors is a polytope called “local polytope” in the case of Bell scenarios [99–102] and “noncontextual polytope” in the case of contextuality scenarios [26, 27]. The set of tight Bell and noncontextuality inequalities define the facets of the local and noncontextual polytopes, respectively. The corresponding set of quantum behaviors is a convex set that, in general, is not a polytope [5, 33, 102–106]. All Bell and contextuality scenarios (as defined in this paper) for which the quantum set of behaviors differs from the classical set of behaviors can be identified using a method described in Ref. [107].

(3) It has historical roots. The notion of contextuality was born connected to the proofs that QT cannot be explained in terms of hidden variables proposed by Kochen and Specker [2, 85] and Bell [84] (which later derived in the emphasis on Bell scenarios [18]). The definition of contextuality scenarios adopted in this paper combines the assumption of outcome noncontextuality for ideal measurements made in the proofs in Refs. [2, 84, 85] with the concept of set of correlations for a Bell scenario [33, 99–106]. However, our definition of contextuality scenario does not assume that the measurement outcomes satisfy functional relations inherited from QT as the Kochen-Specker theorem does. This extra assumption leads to the so-called “Kochen-Specker inequalities” [108, 109], which, as discussed in Ref. [110], are different from the “non-contextuality inequalities” used here (and in Refs. [23–27]), which bound the set of noncontextual behaviors.

2. Why any physical theory must yield probability distributions for ideal measurements

Every measurement in classical physics and in QT can be implemented with ideal measurements [28, 30–32]. Consequently, all classical and quantum correlations can be produced with ideal measurements. These observations raise three questions.

(Q1) Can physical theories produce correlations that cannot be produced by ideal measurements? [36]. We do not know. There are examples of behaviors that cannot be produced by ideal measurements [36]. However, we do not know whether there is a consistent physical theory that produces them. In any case, we can set aside question (Q1) since we are addressing the problem of singling out the physical principle that explains correlations that can be produced by ideal measurements.

(Q2) Should any physical theory yield probability distributions for ideal measurements? Every probabilistic theory that includes classical probability theory or QT as a particular case must yield probability distributions for the outcomes of ideal measurements.

(Q3) Why physical theories have ideal measurements? In a nutshell, because ideal measurements are useful. Physics looks for invariants that allow for predicting the outcomes of future experiments. Even if a measurement $\mathcal{M}$ giving outcome $x$ is destructive (i.e., it forbids any further measurement on the system), Physics is still interested in this measurement as far as it can make sense of what $\mathcal{M} = x$ means. That is, as far as there is a preparation such that outcome $x$ is obtained with certainty after measurement $\mathcal{M}$ is performed. Then, if, after a destructive measurement $\mathcal{M}$ giving outcome $x$, one prepares a state associated to $\mathcal{M} = x$, the whole process is equivalent to an experimental procedure that “gives the same outcome when performed consecutive times.” This explains why condition (i) in Definition 2 is useful in Physics. The interest of condition (ii) comes from the fact that a desirable property for a measurement $\mathcal{M}$ is that performing $\mathcal{M}$ limits as little as possible the possibility to get additional information by performing additional measurements. For that,
Definition 24 (Graph). A graph $G$ is an ordered pair comprising a set $V(G)$ of vertices and a set $E(G)$ of edges, the latter being two-element subsets of $V(G)$.

Definition 25 (Adjacent vertices). Two vertices $i$ and $j$ of $G$ are adjacent if $(i,j) \in E(G)$ (i.e., when there is an edge between them).

Definition 26 (Clique). A clique of a graph $G$ is a set of vertices every pair of which are adjacent. The set of cliques of $G$ is denoted $C(G)$.

Definition 27 (OR product of graphs). The OR product (aka disjunctive or co-normal product) of the graphs $G$ and $G'$, denoted $G \ast G'$, is the graph with $V(G \ast G') = V(G) \times V(G')$ and $((i,i'),(j,j')) \in E(G \ast G')$ if and only if $(i,j) \in E(G)$ or $(i',j') \in E(G')$. The OR product of $n$ copies of $G$ is denoted $G^n$.

Definition 28 (Complement of a graph). The complement of $G$, denoted $\overline{G}$, is the graph with the same vertices as $G$ and such that two distinct vertices of $\overline{G}$ are adjacent if and only if they are not adjacent in $G$.

Definition 29 (Self-complementary graph). A graph $G$ is self-complementary if $G$ and $\overline{G}$ are isomorphic.

Definition 30 (Generalized composition [112]). If $G$ is a graph with $n$ vertices, then the graph $G[G_1, \ldots, G_n]$ is constructed by taking the disjoint graphs $G_1, \ldots, G_n$ and joining every vertex of $G_i$ with every vertex of $G_j$ whenever $v_i$ and $v_j$ are adjacent vertices in $G_i$.

Definition 31 (Induced subgraph). An induced subgraph of $G$ is a graph with vertex set $S \subseteq V(G)$ and edge set comprising all the edges of $G$ with both ends in $S$.

Definition 32 (Cycle graph). The cycle graph with $n$ vertices $C_n$ is a graph with $n$ vertices connected in a closed chain.

E.g., $C_4$ is the square and $C_5$ is the pentagon.

Definition 33 (Perfect graph [113]). A graph $G$ is perfect if and only if it does not contain a cycle graph $C_n$ with $n$ odd and $\geq 5$, or its complement, as an induced subgraph.

Appendix D: Graph theory and quantum correlations

Here we outline the story of the connection between graph theory and quantum correlations.

Motivated by the problem of determining the Shannon capacities of graphs, which measures the optimal zero-error capacity of an associated memoryless classical communication channel [114] (for a survey of zero-error information theory see, e.g., Ref. [115]), Berge introduced the notion of a perfect graph. A graph is perfect if, in all its induced subgraphs, the size of a largest clique is equal to the chromatic number (which is the smallest value of $k$ for which there exists a partition of the vertices into $k$ sets such that, in each set, no two vertices are adjacent) [116]. Perfect graphs are known to satisfy that their Shannon capacity equals their independence number (which is the weighted independence number defined later, in the case that all the weights are one). Berge conjectured that a graph is perfect if and only if it does not contain, as induced subgraph, an odd cycle $C_n$ of length five or more, or the complement of such a graph. This was a long-standing conjecture, known as the strong perfect graph conjecture, which was proved by Chudnovsky, Robertson, Seymour, and Thomas [113] (for an overview of the proof, see Ref. [117]). Now the conjecture has been renamed the strong perfect graph theorem. It follows from this result that the odd cycles of length five or more and their complements are the minimal graphs for which the determination of the Shannon capacity is nontrivial.

Lovász proved that the Shannon capacities of these graphs are upper bounded by the so-called Lovász theta function and used it to prove the Shannon capacity of the pentagon [118]. However, the Shannon capacities of odd cycles and the complements of odd cycles on seven or more vertices remain unknown.

In Ref. [65] (see also Ref. [66]) we show that there is a fundamental connection between quantum correlations, a graph invariant introduced by Lovász in Ref. [118], and a convex set introduced by Grötschel, Lovász, and Schrijver in Ref. [78] (see also Ref. [111]). Firstly, we show that, for any given graph of exclusivity $G$, the set of probability assignments allowed by classical probability theory is the stable set polytope [111] (aka vertex packing polytope [78]) of $G$, which is defined (see, e.g., [73, 111]) as follows:

$$\text{STAB}(G) = \text{convex hull} \{ p \in \{0,1\}^{V(G)} : p_i p_j = 0 \text{ if } (i,j) \in E(G) \}. \quad (D1)$$

To our knowledge, the first systematic study of $\text{STAB}(G)$ was undertaken by Padberg in Ref. [119]. $\text{STAB}(G)$ is related to
the study of the weighted independence number of a graph, denoted $\alpha(G, \vec{w})$, since $\alpha(G, \vec{w})$ is the maximum of the linear function $\vec{w}^T \vec{p}$ for $\vec{p} \in \text{STAB}(G)$, where $\vec{w}^T$ is the transpose of the vector $\vec{w}$ of the weights of the vertices.

The most curious result in Ref. [65] is that, for any graph of exclusivity $G$, the set of probability assignments allowed by QT is equal to $\text{TH}(G)$, the theta body of $G$, a convex set introduced by Grötschel, Lovász, and Schrijver in Ref. [78] with no reference to QT or physics. Grötschel, Lovász, and Schrijver were interested in $\text{TH}(G)$ because it is a convex body generated by “a polynomially computable upper bound the weighted Lovász number for the weighted independence number of a graph.” That is, an easily computable bound to a quantity that is NP-hard to compute [78] (see also Refs. [73, 111]). Using the characterization of $\text{TH}(G)$ given by Theorem 3.5 in Ref. [78] and adopting the notation introduced by Dirac for quantum mechanics [120], the theta body of $G$ is defined:

\[
\text{TH}(G) = \{\vec{p} \in [0, 1]^{V(G)} : p_i = |\langle M_x^{(i)}(\psi) | \psi \rangle|^2, \ |\langle \psi | \psi \rangle| = 1,
\langle M_x^{(i)}(\psi) | M_x^{(j)}(\psi) \rangle = 0, \forall (i, j) \in E(G)\}\).
\]

(\text{D2})

In other words, the elements of $\text{TH}(G)$ are those that satisfy all the following properties:

(I’) The initial state of the system can be associated to a unit vector $|\psi\rangle$ in a finite-dimensional vector space $\mathcal{V}$ with an inner product.

(II’) The state after a measurement $M^{(i)}$ with outcome $x_i$ can be associated to a unit vector $|M_x^{(i)}(\psi)\rangle$ in $\mathcal{V}$. Post-measurement states corresponding to mutually exclusive events can be associated to mutually orthogonal vectors.

(III’) The probability of the event $\langle M_x^{(i)}(\psi) | \psi \rangle$ is equal to $|\langle M_x^{(i)}(\psi) | \psi \rangle|^2$, where $\langle M_x^{(i)}(\psi) | \psi \rangle$ denotes the inner product of $|\psi\rangle$ and $|M_x^{(i)}(\psi)\rangle$.

Ref. [65] left as a challenge identifying the principle that singles out $\text{TH}(G)$. Arguably, Theorem 2 solves this problem.

In Ref. [65] we also identify a set which is strictly larger than the quantum set and we motivate it as the set of assignments that satisfy the exclusivity principle applied to one copy of the assignment. Surprisingly, for any graph of exclusivity $G$, that set of probability assignments was exactly another well-known set in graph theory, first defined by Shannon in Ref. [114], the clique-constrained stable set polytope [111] (aka fractional stable set polytope [111] or fractional vertex packing polytope [78]) of $G$, defined (see, e.g., [111]) as follows:

\[
\text{QSTAB}(G) = \{\vec{p} \in [0, 1]^{V(G)} : \sum_{i \in c} p_i \leq 1, \forall c \in C(G)\}\).
\]

(\text{D3})

Once we have stated this one-to-one correspondence between physical sets of correlations and sets in graph theory, in Ref. [65] we used well-known results in graph theory to prove that, for perfect graphs of exclusivity, the sets of classical, quantum, and more general assignments satisfying the exclusivity principle are equal. This follows from the fact that, for any perfect graph $G$, $\text{STAB}(G) = \text{TH}(G) = \text{QSTAB}(G)$ [111], while this is not the case only for imperfect graphs, where $\text{STAB}(G) \subseteq \text{TH}(G) \subseteq \text{QSTAB}(G)$ [111].

It follows from this result that the cyclic graph $C_5$ (the pentagon) is of particular interest, as it is the smallest graph that is not perfect. Therefore, a challenge was to identify the principle that explains the quantum set of probability assignments for the pentagon. Fortunately, the pentagon is not only imperfect but also self-complementary, which allowed us to identify the exclusivity principle as the principle that explains the quantum correlations for the pentagon of exclusivity [12, 67, 68].

Appendix E: Proof that ideal measurements satisfy the exclusivity principle

\textbf{Lemma 2 ([28])}. For any set $S$ of events produced by ideal measurements and such that every two events in $S$ are mutually exclusive, the sum of the probabilities of all the events in $S$ is bounded by one.

\textbf{Proof}. The set of events can be written as $S = \{(M_x^{(i)}(\psi))_{i=1}^n\}$, where $\{M_x^{(i)}\}_{i=1}^n$ is a set of ideal measurements. By condition (iii) in the definition of ideal measurement, for each event $\langle M_x^{(i)}(\psi) | \psi \rangle \in S$, there is an ideal [i.e., satisfying conditions (i)–(iii)] realization

\[
\overline{M}_x^{(i)} = \{\overline{M}_x^{(i)} := M_x^{(i)}, \overline{M}_y^{(i)} \}
\]

of the two-outcome coarse-graining of $M^{(i)}$ in which all the outcomes different than $x_i$ are coarse-grained into outcome $\overline{x_i}$. The corresponding transformation $\overline{M}_x^{(i)}$ is defined by the
fact that \( \overline{\mathcal{M}}^{(i)} \) is an ideal coarse-graining of \( \mathcal{M}^{(i)} \).

Since every two events \( (\mathcal{M}^{(i)}_{x_1} \psi|\psi) \) and \( (\mathcal{M}^{(j)}_{x_2} \psi|\psi) \) in \( S \) are mutually exclusive, there is a measurement \( \mathcal{M}^{(ij)} \) that produces both events so that each of the events is associated to a different outcome of \( \mathcal{M}^{(ij)} \). Consider the following three-outcome coarse-graining of \( \mathcal{M}^{(ij)} \):

\[
\mathcal{M}^{(ij)} = \{ \mathcal{M}^{(ij)}_{x_1} := \mathcal{M}^{(i)}_{x_1}, \mathcal{M}^{(ij)}_{x_2} := \mathcal{M}^{(j)}_{x_2}, \mathcal{M}^{(ij)}_{x_3} \},
\]

where the explicit form of the transformation \( \mathcal{M}^{(ij)}_{x_3} \) is not needed for the proof.

From the Definition of \( \overline{\mathcal{M}}^{(ij)} \) in Eq. (E2), \( \forall \psi \),

\[
P(\mathcal{M}^{(ij)}_{x_1} | \psi) = P(\mathcal{M}^{(j)}_{x_2} | \psi).
\]

Therefore,

\[
P(\mathcal{M}^{(ij)}_{x_1} | \mathcal{M}^{(ij)}_{x_2} | \psi) = P(\mathcal{M}^{(j)}_{x_2} | \mathcal{M}^{(j)}_{x_2} | \psi).
\]

Since \( \mathcal{M}^{(i)} \) is ideal and a coarse-graining of \( \mathcal{M}^{(ij)} \), \( \forall \psi \),

\[
P(\mathcal{M}^{(ij)}_{x_1} | \mathcal{M}^{(ij)}_{x_2} | \psi) = P(\mathcal{M}^{(i)}_{x_1} | \mathcal{M}^{(i)}_{x_1} | \psi),
\]

Therefore, from Eqs. (E3)–(E5), \( \forall \psi \),

\[
P(\mathcal{M}^{(ij)}_{x_1} | \mathcal{M}^{(ij)}_{x_2} | \psi) = P(\mathcal{M}^{(ij)}_{x_1} | \psi).
\]

Now we can define the following \((n + 1)\)-outcome measurement \( \mathcal{M} = \{ \mathcal{M}_o \}_{i=1}^{n+1} \):

\[
\mathcal{M}_{o_1} := \overline{\mathcal{M}}^{(1)}_{x_1},
\]

\[
\mathcal{M}_{o_2} := \overline{\mathcal{M}}^{(2)}_{x_2},
\]

\[
\mathcal{M}_{o_3} := \overline{\mathcal{M}}^{(i-1)}_{x_1} \overline{\mathcal{M}}^{(1)}_{x_1}, \text{ for } i = 3, \ldots, n.
\]

\[
\mathcal{M}_{o_{n+1}} := \overline{\mathcal{M}}^{(n-1)}_{x_1} \overline{\mathcal{M}}^{(1)}_{x_1}.
\]

Fig. 6 illustrates how \( \mathcal{M} \) is constructed. Then, \( \forall \psi \),

\[
P(\mathcal{M}_{o_i} | \psi) = P(\mathcal{M}^{(1)}_{x_1} | \mathcal{M}^{(1)}_{x_1} | \psi) = P(\mathcal{M}^{(1)}_{x_1} | \mathcal{M}^{(1)}_{x_1} | \psi),
\]

where \( \psi_j = \mathcal{M}^{(ij)}_{x_j} \cdot \overline{\mathcal{M}}^{(1)}_{x_1} \psi \).

Taking into account that, by Eq. (E1), \( \overline{\mathcal{M}}^{(i)}_{x_i} = \mathcal{M}^{(i)}_{x_i} \), then

\[
P(\mathcal{M}_{o_i} | \psi) = P(\mathcal{M}^{(1)}_{x_1} | \mathcal{M}^{(1)}_{x_1} | \psi) = P(\mathcal{M}^{(1)}_{x_1} | \mathcal{M}^{(1)}_{x_1} | \psi).
\]

Taking Eq. (E6) into account, \( \forall \psi \),

\[
P(\mathcal{M}_{o_i} | \psi) = P(\mathcal{M}^{(1)}_{x_1} | \psi) = P(\mathcal{M}^{(1)}_{x_1} | \psi).
\]

Hence, taking Eq. (E12) into account,

\[
\sum_{i=1}^{n} P(\mathcal{M}_{o_i} | \psi) \leq 1.
\]

Notice that in the proof we have made a limited use of conditions (i)–(iii), which suggests that Lemma 2 can be extended to a more general class of measurements.
Appendix F: Connections

Here we outline some connections between the result presented in this paper and other lines of research.

1. Reconstruction of QT using only physical principles

We have presented a physical explanation for quantum correlations. But we have not yet reconstructed the whole formalism of QT. In particular, because quantum correlations admit representations both in real and complex vector spaces [79–82], while QT is formulated in a complex vector space. The next step in our program must be adding physical details to the picture of the reality we have proposed until we reconstruct the whole QT. For example, while in the operational reconstructions of QT the complex-vector-space representation is enforced by imposing the axiom of “local tomography” [39], a purely physical reason replacing this operational axiom could be that the laws of nature stay the same for all observers that are moving with respect to one another within an inertial frame. This enforces any theory to be Lorentz-invariant. For that, the theory needs to be local and free of holistic states inaccessible to the experimenter. In contrast to that, real-vector-space representations of quantum correlations seem to require a holistic inaccessible ontological space that permeates every point of the universe [82]. This will be the subject of a future research. In any case, the intuition that the solution to the real-vs-complex problem may come from a different physical requirement than the solution to the problem of the physical origin of quantum correlations justifies presenting the solution to the latter independently.

2. Interpretations of QT

How is it possible to answer the question “what the hell are we talking about [when we use quantum mechanics]?” [121]?, i.e., how should we interpret QT, without first answering the question “where does it come from?” [122]?. Trying to be consequent with this point of view, in this paper we have shown that, in a world in which some experiments follow the no-law总论制原则 given by Axiom 1, the sets of correlations for Bell and contextuality scenarios are identical to the ones predicted by QT.

Inevitably, our starting point, a world that has experiments producing outcomes for which the world has no laws, already contrasts with the pictures of the underlying reality proposed by some interpretations of QT. In addition, our explanation of the Born rule is in striking contrast to the explanations proposed within these interpretations. Here we outline some of these differences.

In the Bohmian interpretation of QT [123, 124], measurement outcomes are determined by the quantum potential, a field that permeates the whole universe. If the Bohmian interpretation is determinist, why then is there a particular probability rule at all? Because, although the particle is at a certain position, this position is governed by variables that are hidden to the experimenters, so experimenters are restricted to calculating the probability density to observe that the particle is at some position. But then why outcomes happen with relative frequencies given by the Born rule? Because of the initial state of the quantum potential and the initial positions of all particles [124]. That is, the Bohmian interpretation blames the initial state of the world for the Born rule.

In the Everettian interpretation of QT [125], a measurement is a unitary transformation of a universal wave function that gives rise to a multiverse. Everettians and proponents of the decoherent (or consistent) histories interpretation of QT [126] have made several attempts to justify the Born rule from simpler assumptions (see, e.g., [127–135]). However, each of these attempts in turn has attracted critical attention and purported refutation (see, e.g., [136–140]).

In fact, there is a risk of circularity in any derivation of the Born rule that assumes that measurements are related to self-adjoint operators in a vector space \( \mathcal{V} \) in the sense that measurement outcomes correspond to their eigenvalues and each orthonormal basis corresponds to the mutually exclusive results of a measurement. Because then, for every \( \mathcal{V} \) of dimension \( d \) greater than two, Gleason’s theorem [141] shows that the only possible states are vectors in \( \mathcal{V} \) and the only possible choice to assign a probability \( P(v_i) \) to every vector \( v_i \) in \( \mathcal{V} \) for any given vector \( v \) in \( \mathcal{V} \) such that, for every orthonormal basis \( \{v_i\}_{i=1}^d \) of \( \mathcal{V} \), the sum of the probabilities satisfies \( \sum_{i=1}^d P(v_i) = 1 \) and \( P(v_i) \) only depends on \( v_i \) (and not on the orthonormal basis of \( \mathcal{V} \) considered) is \( P(v_i) = |\langle v_i | v \rangle|^2 \).

Therefore, the key question is what is the physical reason that leads us to assume that measurements are represented self-adjoint operators in a vector space. In other words, explaining the Born rule is not only explaining \( P(v_i) = |\langle v_i | v \rangle|^2 \), but explaining conditions (I)–(III) in Subsec. 1D.

QBism [142] is a different case as it does not assume any particular picture of the underlying reality. For QBism, QT is a personal tool for each agent. For example, adopting a QBist perspective, Mermin writes: “When I take an action on the world, quantum mechanics tells me the likelihood of the kind of experience the world will induce back in me, in response to my action. ‘Likelihood’ means my personal expectation, based on how the world has responded to me in the past” [121]. Still, QBism does not explain what is the physical reason why the world responded that way in the past and keeps responding the same way to any agent. According to QBism, the Born rule is not a law of nature in the usual sense, but “an empirical addition to the laws of Bayesian probability” [142] that a wise agent should follow in addition to the Bayesian coherence conditions. But QBism does not clarify where does this empirical addition come from.

In contrast, here we have identified a physical reason behind the Born rule. The Born rule itself is a “law without law.” An effective law that reflects that, at bottom, there is no law governing the outcomes of some experiments. This fits within the QBist view that “even if [QT] (.) is not a theory of the world itself, it is conditioned by the particular character of this world” [142]. However, our result is quite explicit on why and how the world conditions QT. In particular, it recognizes Born’s rule as an emerging law of nature, which portrays QT
as something more than a personal tool: QT is a consequence of a property of nature.

Therefore, our result answers Jaynes’ challenge of unscrambling to what extent QT is about properties of nature and to what extent is about how we organize experiences [143] as follows: QT is about a fundamental property of nature, the lack of laws governing the outcomes of some experiments, and how agents have to organize experiences because of this property. A task of Physics in a world where Axiom I holds is to classify all reproducible experimental procedures according to the probability distributions they generate. The insight our result provides is that the classification will not be complete until every not inconsistent behavior is associated to a reproducible experimental procedure (or until we identify why some not inconsistent behaviors are not realizable [144]). This is very much a task physicists have been doing in our world. For example, Ref. [90] can be seen as comprehensive catalog of experimental procedures generating every not inconsistent behavior by using a source of single photons (or “electrons, neutrons, atoms, or any other type of radiation”), a mirror and a set of suitable beam splitters and phase shifters. Ref. [91] gives a similar comprehensive catalog for a single trapped ion using adiabatic passage techniques. Ref. [89] provides a similar catalog for a source of single particles of a suitable spin passing through adequate magnetic fields. Moreover, our result provides a simple explanation of why the outcomes produced by photons, electrons, neutrons, and ions submitted to different interactions follow the same “law” (the Born rule): the world is equally “economic in laws” for any of them.

3. The exclusivity principle in quantum logic

As pointed out in Ref. [71], the exclusivity principle is related to the notion of orthocoherence that appeared in quantum logic (“an orthoalgebra is orthocoherent if and only if finite pairwise summable subsets are jointly summable”[145]). The origin of orthocoherence can be traced back to Mackey’s axiom V in Ref. [146]. Interestingly, in the literature of quantum logic, since the end of the 1970’s, orthocoherence is presented as “suspicious ( . . . ) as a fundamental principle” [147]. Three related reasons are offered for that:

(a) “There has never been given any entirely compelling reason for regarding orthocoherence as an essential feature of all reasonable physical models” [145].

(b) “[I]t is quite easy to manufacture simple and plausible toy examples that are not orthocoherent” [147] (see, e.g., [2–6]).

(c) “[Orthocoherence] is not stable under formation of the tensor product” [147] (see [148, 149]).

In contrast, in this paper, we have seen that

(a’) A compelling reason why the exclusivity principle is an essential feature of any reasonable physical theory is that (as proven in Appendix E) the exclusivity principle holds for events produced by ideal measurements. Ideal measurements must exist in any theory that contains classical probability theory or QT as particular cases. Moreover, so far, in any physical theory, any measurement can be implemented using an ideal measurement. There are also other approaches that lead to the exclusivity principle. For example, in Ref. [46], it is shown that two postulates are sufficient to guarantee the exclusivity principle. The first postulate says that every state can be represented as a probabilistic mixture of perfectly distinguishable pure states. The second postulate says that every set of perfectly distinguishable pure states of a given dimension can be reversibly transformed to any other such set of the same dimension. In Ref. [71], it is shown that irreducible third-order interference (a generalization of the idea that no probabilistic interference remains unaccounted for once we have taken into account interference between pairs of slits in an n-slit experiment) also implies the exclusivity principle. Finally, in Ref. [150], it is shown that every Bayesian framework must include a set of ideal experiments and that the latter must satisfy the exclusivity principle.

(b’) The toy examples in Refs. [2–6] are not entirely plausible. Common to all of them is that they can be implemented with measurements satisfying conditions (i) and (ii) in Definition 2, and also satisfying (i) and (ii) for some of the coarse-grainings of the measurements involved. This makes the examples apparently plausible. However, mysteriously, conditions (i) and (ii) do not hold for all coarse-grainings of the measurements (as proven in Appendix E). This makes these toy examples physically implausible.

(c’) Observation (c) is not an obstacle but an opportunity. The fact that a probability assignment satisfies the exclusivity principle when the principle is applied to n copies but fails to satisfy it when is applied to n + 1 copies indicates that such a probability assignment is inconsistent in a physical theory.

[1] M. Born, Zur Quantenmechanik der Stoßvorgänge, Z. Physik 37, 863 (1926) [English version: On the quantum mechanics of collisions, in Quantum Theory and Measurement, edited by J. A. Wheeler and W. H. Zurek (Princeton University Press, Princeton, NJ, 1983), p. 52.]
[2] E. P. Specker, Die Logik nicht gleichzeitig entscheidbarer Aussagen, Dialectica 14, 239 (1960) [English version: The logic of non-simultaneously decidable propositions,
Cabello, Memory cost of quantum contextuality, New J. Phys. 13, 113011 (2011).
[96] A. Cabello, M. Gu, O. Gühne, J.-Á. Larsson, and K. Wiesner, Thermodynamical cost of some interpretations of quantum theory, Phys. Rev. A 94, 052127 (2016).
[97] A. Cabello, M. Gu, O. Gühne, and Z.-P. Xu, Optimal Classical Simulation of State-Independent Quantum Contextuality, Phys. Rev. Lett. 120, 130401 (2018).
[98] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the ‘magic’ for quantum computation, Nature (London) 510, 351 (2014).
[99] P. Suppes and M. Zanotti, When are probabilistic explanations possible?, Synthese 48, 191 (1981).
[100] M. Froissart, Constructive generalization of Bell’s inequalities, Nuovo Cimento Soc. Ital. Fis. B 64, 241 (1981).
[101] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, Phys. Rev. Lett. 48, 291 (1982).
[102] I. Pitowsky, Quantum Probability - Quantum Logic (Springer-Verlag, Berlin, 1989).
[103] M. Navascués, S. Pironio, and A. Acín, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, New J. Phys. 10, 073013 (2008).
[104] B. S. Cirel’son [Tsirelson], Quantum generalizations of Bell’s inequality, Lett. Math. Phys. 4, 93 (1980).
[105] B. S. Cirel’son [Tsirelson], Zap. Nauchn. Semin LOMI 142, 174 (1985) [English version: Quantum analogues of the Bell inequalities. The case of two spatially separated domains, J. Sov. Math. 36, 557 (1987).]
[106] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, Phys. Rev. Lett. 48, 291 (1982).
[107] Z.-P. Xu and A. Cabello, Necessary and sufficient condition for contextuality from incompatibility, arXiv:1805.02032.
[108] C. Simon, Č. Brukner, and A. Zeilinger, Hidden-Variable Theorems for Real Experiments, Phys. Rev. Lett. 86, 4427 (2001).
[109] J.-Á. Larsson, A Kochen-Specker inequality, EPL 58, 799 (2002).
[110] I. Bengtsson, K. Blanchfield, and A. Cabello, A Kochen-Specker inequality from a SIC, Phys. Lett. A 376, 374 (2012).
[111] M. Grötschel, L. Lovász, and A. Schrijver, Geometric Algorithms and Combinatorial Optimization (Springer, Berlin, 1988).
[112] A. Schwenk, Computing the characteristic polynomial of a graph, in Graphs Combinatorics, Lecture Notes in Mathematics, vol. 406, edited by R. Bary and F. Harary (Springer-Verlag, Berlin, 1974), p. 153.
[113] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas, The strong perfect graph theorem, Ann. Math. 164, 51 (2006).
[114] C. E. Shannon, The zero-error capacity of a noisy channel, IRE Trans. Inf. Theory 2, 8 (1956).
[115] J. Körner and A. Orlitsky, Zero-error information theory, IEEE Trans. Inf. Theory 44, 2207 (1998).
[116] C. Berge, Färbung von Graphen, deren sämtliche bzw. deren ungerade Kreise starr sind (Zusammenfassung), Wiss. Zeitschrift der Martin-Luther Universität Halle-Wittenberg 10, 114 (1961).
[117] G. Cornuéjols, The strong perfect graph conjecture, Proceedings of the International Congress of Mathematicians (2002, Beijing). Vol. III: Invited Lectures (Higher Education Press, Beijing, 2002), p. 547.
[118] L. Lovász, On the Shannon capacity of a graph, IEEE Trans. Inf. Theory 25, 1 (1979).
[119] M. W. Padberg, On the facial structure of set packing polyhedra, Mathematical Programming 5, 199 (1973).
[120] P. A. M. Dirac, A new notation for quantum mechanics, Mathematical Proceedings of the Cambridge Philosophical Society 35, 416 (1939).
[121] N. D. Mermin, Mysl, mysyl, svět, arXiv:1710.05229.
[122] J. A. Wheeler in an interview conducted by K. Ford in December 1996, available at the Web of Stories. The quote is from a fragment of the interview available at this link.
[123] D. Bohm, A suggested interpretation of the quantum theory in terms of “hidden” variables. I, Phys. Rev. 85, 166 (1952); A suggested interpretation of the quantum theory in terms of “hidden” variables. II, Phys. Rev. 85, 180 (1952).
[124] D. Bohm, Proof that probability density approaches $|\psi|^2$ in causal interpretation of quantum theory, Phys. Rev. 89, 458 (1953).
[125] H. Everett, III, “Relative state” formulation of quantum mechanics, Rev. Mod. Phys. 29, 454 (1957).
[126] M. Gell-Mann and J. B. Hartle, Quantum mechanics in the light of quantum cosmology, in Entropy, and the Physics of Information, SFI Studies in the Sciences of Complexity, vol. VIII, edited by W. H. Zurek (Addison-Wesley, Reading, MA, 1990) p. 425.
[127] D. Finkelstein, The logic of quantum physics, Trans. New York Acad. Sciences 25, 621 (1965).
[128] J. B. Hartle, Quantum mechanics of individual systems, Am. J. Phys. 36, 704 (1968).
[129] E. Farhi, J. Goldstone, and S. Gutmann, How probability arises in quantum mechanics, Ann. Phys. 192, 368 (1989).
[130] D. Deutsch, Quantum theory of probability and decisions, Proc. Roy. Soc. London A455, 3129 (1999).
[131] D. Wallace, Everettian rationality: defending Deutsch’s approach to probability in the Everett interpretation, Stud. Hist. Phil. Sci. B 34, 415 (2003).
[132] D. Wallace, Quantum probability from subjective likelihood: Improving on Deutsch’s proof of the probability rule, Stud. Hist. Phil. Sci. B 38, 311 (2007).
[133] W. H. Zurek, Environment-Assisted Invariance, Entanglement, and Probabilities in Quantum Physics, Phys. Rev. Lett. 90, 120404 (2003).
[134] W. H. Zurek, Probabilities from entanglement, Born’s rule $p_k = |\psi_k|^2$ from envariance, Phys. Rev. A 71, 052105 (2005).
[135] L. Vaidman, Probability in the many-worlds interpretation of quantum mechanics, in Probability in Physics, edited by Y. Ben-Menahem and M. Hemmo (Springer, Berlin 2012), p. 299.
[136] E. J. Squires, On an alleged “proof” of the quantum probability law, Phys. Lett. A 145, 67 (1990).
[137] A. Cassinello and J. L. Sánchez-Gómez, On the probabilistic postulate of quantum mechanics, Found. Phys. 26, 1357 (1996).
[138] H. Barnum, C. M. Caves, J. Finkelstein, C. A. Fuchs, and R. Schack, Quantum probability from decision theory?, Proc. Roy. Soc. London A456, 1175 (2000).
[139] C. Caves and R. Schack, Properties of the frequency operator do not imply the quantum probability postulate, Ann. Phys. (N.Y.) 315, 123 (2005).
[140] M. Schlosshauer and A. Fine, On Zurek’s derivation of the Born rule, Found. Phys. 35, 197 (2005).
[141] A. M. Gleason, Measures on the closed subspaces of a Hilbert space, J. Math. Mech. 6, 885 (1957).
[142] C. A. Fuchs and R. Schack, Quantum-Bayesian coherence, Rev. Mod. Phys. 85, 1693 (2013).
[143] E. T. Jaynes, Probability in quantum theory, in Complexity, Entropy, and the Physics of Information, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA, 1990), p. 381.
[144] G. C. Wick, A. S. Wightman, and E. P. Wigner, The intrinsic
parity of elementary particles, Phys. Rev. 88, 101 (1952).

[145] A. Wilce, Quantum logic and probability theory, in The Stanford Encyclopedia of Philosophy, edited by E. N. Zalta (2017).

[146] G. W. Mackey, The Mathematical Foundations of Quantum Theory (W. A. Benjamin, New York, 1963).

[147] A. Wilce, Test spaces and orthoalgebras, in Current Research in Operational Quantum Logic edited by B. Coecke, D. Moore, and A. Wilce (Springer Netherlands, 2000), p. 81.

[148] C. H. Randall and D. J. Foulis, Tensor products of quantum logics do not exist, Notices of the American Math. Soc. 26, 557 (1979).

[149] D. J. Foulis and C. H. Randall, Empirical logic and tensor products, in Interpretations and Foundations of Quantum Mechanics, edited by H. Neumann (Wissenschaftsverlag, Mannheim, Germany, 1981), p. 9.

[150] G. Chiribella, A. Cabello, M. Kleinmann, and M. P. Müller, General Bayesian theories and the emergence of the exclusivity principle (in preparation).