

Classical theories with entanglement

Giacomo Mauro D'Ariano ^{*}, Marco Erba [†] and Paolo Perinotti [‡]

Università degli Studi di Pavia, Dipartimento di Fisica, QUIT Group, and INFN Gruppo IV, Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy



(Received 16 September 2019; revised manuscript received 10 February 2020; accepted 16 March 2020; published 20 April 2020)

We investigate operational probabilistic theories where the pure states of every system are the vertices of a simplex. A special case of such theories is that of *classical theories*, i.e., simplicial theories whose pure states are jointly perfectly discriminable. The usual classical theory satisfies also local discriminability. However, simplicial theories—including the classical ones—can violate local discriminability, thus admitting entangled states. First, we prove sufficient conditions for the presence of entangled states in arbitrary probabilistic theories. Then we prove that simplicial theories are necessarily causal, and this represents a no-go theorem for conceiving noncausal classical theories. We then provide necessary and sufficient conditions for simplicial theories to exhibit entanglement and classify their system-composition rules. We conclude by proving that, in simplicial theories, an operational formulation of the superposition principle cannot be satisfied, and that—under the hypothesis of n -local discriminability—no mixed state admits a purification. Our results hold also in the general case where the sets of states fail to be convex.

DOI: [10.1103/PhysRevA.101.042118](https://doi.org/10.1103/PhysRevA.101.042118)

I. INTRODUCTION

Entanglement is the quantum feature marking the starkest departure of quantum theory from classical theory (in the following, these will be referred to as QT and CT, respectively). The phenomenon is commonly popularized as so-called *quantum nonlocality*, although the two concepts are not coincident. Indeed, it is known that the mere existence of entangled states is not sufficient for nonlocality [1,2]. On the other hand, CT does not allow for any kind of entanglement or nonlocality.

States in CT have a very simple geometrical structure; namely, the set of states for every system is a *simplex*. In the present work, we argue that the absence of entanglement in CT is due not only to the simplicial structure of the set of states of the single systems, but also to the composition rule of the systems, which satisfies local discriminability [3]. Indeed, as we will show, *simplicial theories* can exhibit entanglement, including general *classical theories*, namely, those simplicial theories where the pure states are jointly perfectly discriminable [4–6]. Besides classical theories, the definition of simplicial theory encompasses even more general cases—e.g., noisy versions of classical theories, where pure states cannot be reliably distinguished due to a limited set of measurements [7]. The characteristic trait of a simplicial theory is that every state has a unique convex decomposition into pure states. The results of the present work hold for all simplicial theories, encompassing even cases where the sets of states are not convex. In contrast to the common understanding of the quantum-classical divide, as a consequence of our results we show that a classical theory generally admits entangled states.

The paper is organized as follows: In Sec. II we provide a review of the framework of operational probabilistic theories (OPTs). In Sec. III we point out the relevance of the notion of parallel composition in a probabilistic theory, focusing on the consequences regarding the presence of entanglement in arbitrary OPTs. In particular, we prove sufficient conditions for the presence of entangled states in a probabilistic theory. In Sec. IV, we specialize to simplicial theories, discussing various features of this class of OPTs. First, as a consequence of the geometric structure of the sets of states, we show that a simplicial theory is necessarily causal, highlighting that causality is intrinsic to classical theories—and, in particular, to CT. Then, we provide necessary and sufficient conditions for the presence of entangled states in a simplicial theory, along with a classification of simplicial theories in terms of their composition rules for systems. We conclude by showing that, despite the possibility of exhibiting entanglement, no simplicial theory can admit of superposition (in an operational sense), or purifications of (any) mixed states (under the hypothesis of n -local discriminability). Finally, in Sec. V we discuss some information-theoretic features shared by simplicial theories with entanglement, making a comparison with the existing literature on probabilistic theories, and we draw our conclusions.

II. REVIEW OF FRAMEWORK OF OPERATIONAL PROBABILISTIC THEORIES

The primitive notions of an operational theory are those of *systems*, *tests*, and *events*. A system S represents the physical entity which is probed in the laboratory, such as a radiation field, a molecule, or an elementary particle. A test \mathcal{E} is characterized by a collection of events $\{\mathcal{E}_x\}_{x \in X}$ and represents the single occurrence of a physical process, such as the use of a physical device or a measuring apparatus. The *outcome space* X associated with the test \mathcal{E} collects all the possible outcomes

^{*}dariano@unipv.it

[†]marco.erba@unipv.it

[‡]paolo.perinotti@unipv.it

that can occur within the test. Each test \mathcal{E} is characterized by an input system A and an output system B . For instance, think of electron-proton scattering: both the input and the output systems are an electron-proton pair, the test contains only one event corresponding to the two-particle interaction, and finally the outcome space is a singleton. On the other hand, a fundamental feature for a physical theory is the power to make predictions. Accordingly, the operational structure needs to be endowed with a suitable set of rules for computing the probability distributions associated with each experiment.

The above structure can be formalized by defining the notion of an operational probabilistic theory (OPT).¹ In the present section, we recall some relevant concepts in the scope of OPTs. Given an OPT Θ , $\mathbf{Sys}(\Theta)$ will denote the set of the systems of the theory. We will denote by roman letters $A, B, \dots \in \mathbf{Sys}(\Theta)$ the systems, and by $\mathcal{E} = \{\mathcal{E}_x\}_{x \in X}$ the tests, where each \mathcal{E}_x for $x \in X$ is a possible event that can occur in the test. Given a test \mathcal{E} , each possible event \mathcal{E}_x is a map from an input system A to an output system B and corresponds to the outcome x . Whenever the output of a test $\mathcal{E}_1 = \{\mathcal{E}_x\}_{x \in X}$ coincides with the input of another test $\mathcal{E}_2 = \{\mathcal{E}_y\}_{y \in Y}$, the *sequential composition* $\mathcal{E}_2 \circ \mathcal{E}_1 := \mathcal{E}_2 \mathcal{E}_1 \equiv \{\mathcal{E}_y \mathcal{E}_x\}_{(y,x) \in Y \times X}$ of the two tests can be defined, being an allowed test for the theory. Thus, physical systems define the connection rules for tests. In an OPT, two important classes of tests are those of *preparations* and *observations*, i.e., tests with no input or no output system, respectively. Accordingly, preparations ρ and observations a , where A is, respectively, the output or the input system, are conveniently denoted by using the Dirac-like notation $|\rho\rangle_A$ and $\langle a|_A$. Given some arbitrary preparation event $|\rho_i\rangle_A$, event \mathcal{E}_x from system A to system B , and observation event $\langle a_k|_B$, the purpose of an OPT is to compute joint probabilities of the form

$$p(i, x, k | \rho, \mathcal{E}, a) := \langle a_k|_B \mathcal{E}_x |\rho_i\rangle_A. \quad (1)$$

Two events from a system X to a system Y are equivalent if all their joint probabilities of the form (1), given the same set of other events appearing in Eq. (1), are equal. We call equivalent classes of preparation and observation events *states* and *effects*, respectively. Besides, equivalent classes of arbitrary events from A to B will be called *transformations*. Notice that states and effects will be considered special cases of transformations. This can be done simply by introducing the notion of *trivial system* I , defined as their, respectively, input and output system. For every $A, B \in \mathbf{Sys}(\Theta)$, $\mathbf{St}(A)$, $\mathbf{Transf}(A \rightarrow B)$, and $\mathbf{Eff}(B)$ will denote, respectively, the sets of states of A , of transformations from A to B , and of effects of B . Clearly, every event $\mathcal{E}_x \in \mathbf{Transf}(A \rightarrow B)$ is a map from $\mathbf{St}(A)$ to $\mathbf{St}(B)$, or, dually, from $\mathbf{Eff}(B)$ to $\mathbf{Eff}(A)$. According to the above definitions, effects are separating for states, i.e., given two states $|\rho_1\rangle_A \neq |\rho_2\rangle_A$, there exists $\langle a|_A \in \mathbf{Eff}(A)$ such that $\langle a|\rho_1\rangle_A \neq \langle a|\rho_2\rangle_A$. Similarly, an effect is the equivalence class of those observation events that give the same probabilities for every state, and thus states are separating for effects.

Within an operational perspective, an agent is allowed to perform a test—say with outcome space X —disregarding the

single outcomes within a subset $Y \subseteq X$, and then merging events in Y into a single event: this possibility is captured by the notion of *coarse-graining*. According to probability theory, the probability of the coarse-grained event Y amounts to the sum of the probabilities of all the outcomes in the subset Y . Then, for each test $\{\mathcal{E}_x\}_{x \in X} \subseteq \mathbf{Transf}(A \rightarrow B)$ and every subset $Y \subseteq X$, the coarse-grained event is symbolically given by $\sum_{y \in Y} \mathcal{E}_y$, where sequential composition distributes over sums. The converse procedure of a coarse-graining is called *refinement*. An event with trivial refinement, i.e., which cannot be further refined modulo a rescaling by a probability, is called *atomic*. Being framework probabilistic, one may also want to consider convex combinations of states, transformations, and effects, corresponding to a randomization, i.e., a statistical mixture of events.

In general, states, transformations, and effects can be thought of as embedded in convex spaces. However, in general the set of states of a system might not be convex. Then, one can conveniently remind two important notions from conic and convex analysis, i.e., those of atomic and extremal points. Let $\mathbf{St}(A)$ be the set of states of a system A . The null state $|\varepsilon\rangle_A$, which occurs with null probability in every context, is always included in $\mathbf{St}(A)$. Given $|x_1\rangle, |x_2\rangle \in \mathbf{St}(A)$ and $p \in (0, 1)$, a state $|x\rangle \in \mathbf{St}(A)$ is called *atomic* if the condition $|x\rangle = |x_1\rangle + |x_2\rangle$ implies $|x_1\rangle \propto |x_2\rangle$, while it is called *extremal* if the condition $|x\rangle = p|x_1\rangle + (1-p)|x_2\rangle$ implies $|x_1\rangle = |x_2\rangle$. In the case where $\mathbf{St}(A)$ is a convex cone for every system A , the theory is called *convex*. CT and QT are both convex theories. Let now $\mathbf{ConvH}(A)$ denote the convex hull of $\mathbf{St}(A)$. It might happen that a state is not atomic in $\mathbf{St}(A)$, whereas it is in $\mathbf{ConvH}(A)$. We will denote by $\mathbf{ExtSt}(A)$ the set of extremal points of $\mathbf{ConvH}(A)$. Notice that $\mathbf{ExtSt}(A)$ contains the null state $|\varepsilon\rangle_A$. The event associated with a singleton test—namely a test where the outcome space is a singleton, i.e., having a single outcome—is called *deterministic*. The interpretation of a deterministic event is that the physical process considered happens with certainty, i.e., with probability 1. For instance, a state is deterministic if and only if it gives probability 1 on every deterministic effect, or, in other words, if and only if it is normalized. The deterministic extremal states are historically called *pure states*. Finally, the *mixed states* are all states that are not proportional to some extremal state.

It is often convenient to consider the real span of sets of states $\mathbf{St}(A)$, which is a linear space denoted by $\mathbf{St}_{\mathbb{R}}(A) := \mathbf{Span}_{\mathbb{R}} \mathbf{St}(A)$. The null state $|\varepsilon\rangle_A$ is represented in $\mathbf{St}_{\mathbb{R}}(A)$ by the null vector 0. Every system A is then associated with a quantity $D_A := \dim \mathbf{St}_{\mathbb{R}}(A)$, which is called *the dimension of the system* A . Since effects are separating for states and viceversa, one has that $\dim \mathbf{St}_{\mathbb{R}}(A) = \dim \mathbf{Eff}_{\mathbb{R}}(A)$. As usual throughout the literature, we consider finite-dimensional OPTs, namely, theories where $D_A < +\infty$ for all systems $A \in \mathbf{Sys}(\Theta)$. The latter assumption means that we are considering systems whose states can be completely probed via the statistics of a finite number of experiments. In CT, for instance, D_A is the number of perfectly distinguishable states of a system A , while in QT one has $D_A = d_A^2$, where d_A is the dimension of the Hilbert space associated with the system A .

We conclude the present review by considering a fundamental structure of operational theories: *parallel composition*.

¹For a thorough presentation of the OPT framework, we refer the interested reader to Refs. [3,8–10].

Indeed, the last piece of information one needs to characterize an OPT is a recipe to form compounds out of systems and events available to local experimenters. We denote parallel composition by the symbol \boxtimes , writing, for two arbitrary tests $\mathcal{E}_1 = \{\mathcal{E}_x\}_{x \in X}$ and $\mathcal{E}_2 = \{\mathcal{E}_y\}_{y \in Y}$, $\mathcal{E}_1 \boxtimes \mathcal{E}_2 \equiv \{\mathcal{E}_x \boxtimes \mathcal{E}_y\}_{(x,y) \in X \times Y}$. The latter, from a probabilistic point of view, represent uncorrelated tests. The main property of parallel composition is the following:

$$(\mathcal{A} \boxtimes \mathcal{B}) \circ (\mathcal{C} \boxtimes \mathcal{D}) = (\mathcal{A} \circ \mathcal{C}) \boxtimes (\mathcal{B} \circ \mathcal{D}). \quad (2)$$

Property (2) states that the two operations of sequential and parallel composition commute. Also parallel composition distributes over sums. In the case of systems, states, and effects, we use the following notation: $AB := A \boxtimes B$, $|\rho\rangle_A |\sigma\rangle_B := |\rho\rangle_A \boxtimes |\sigma\rangle_B$, and $(a|_A (b|_B := (a|_A \boxtimes (b|_B$. One has then to specify, for all transformations $\mathcal{E}_x \in \text{Transf}(A \rightarrow B)$ and $\mathcal{D}_y \in \text{Transf}(C \rightarrow D)$, how the composite event $\mathcal{E}_x \boxtimes \mathcal{D}_y$ embeds into the total space of transformations $\text{Transf}(AC \rightarrow BD)$. In QT, for instance, this operation is given by the standard tensor product \otimes . Notice that both the sequential and the parallel composition of two deterministic transformations are deterministic. However, in principle, the property of atomicity may be not preserved under both kinds of composition.

III. PARALLEL COMPOSITION AND PRESENCE OF ENTANGLED STATES IN ARBITRARY OPERATIONAL PROBABILISTIC THEORIES

The existence of parallel composition entails a prescription to assign a dimension to a composite system AB as a function of the dimensions D_A and D_B of the local systems A, B . $\text{St}(AB)$ contains at least the parallel composition of the states of A and B , which can be composed independently. By virtue of property (2), for every OPT Θ one has the following inequality (see e.g., Ref. [11]):

$$D_{AB} \geq D_A D_B \quad \forall A, B \in \text{Sys}(\Theta). \quad (3)$$

This leads us to introduce an “excess dimension” of the composite system AB as follows:

$$\Delta_{AB}^{(2)} := D_{AB} - D_A D_B \quad \forall A, B \in \text{Sys}(\Theta).$$

From Eq. (3), one can see that $\Delta_{AB}^{(2)}$ is in general a non-negative integer for all systems A, B . In both CT and QT one has $\Delta_{AB}^{(2)} = 0$. We then see how this relates to the degree of holism required in the task of state discrimination.

Property 1: n-local discriminability [11]. Let $n \leq m$. The effects obtained as a conic combination of the parallel compositions of effects a_1, a_2, \dots, a_l , where a_j is k_j partite with $k_j \leq n$ for all values of j , are separating for m -partite states.

In the simplest case where $n = 1$, the property is called *local discriminability*. In both CT and QT, the parallel compositions of local effects are separating for multipartite states; namely, both theories satisfy local discriminability. The next result, whose proof can be found in Ref. [3], characterizes the composition rules for the dimensions of local systems in theories satisfying local discriminability.

Proposition 1. Let Θ be an OPT, then Θ satisfies local discriminability if and only if the following rule holds:

$$D_{AB} = D_A D_B \quad \forall A, B \in \text{Sys}(\Theta). \quad (4)$$

As a consequence of the previous result, if a theory Θ does not satisfy local discriminability, then there exist two systems $A, B \in \text{Sys}(\Theta)$ such that $\Delta_{AB}^{(2)} > 0$. Another important feature shared by both CT and QT is the following:

*Property 2: atomicity of state composition.*² The parallel composition of two atomic states is atomic.

In this work, we relate the properties of local discriminability and atomicity of state composition to the presence of entangled states. We now recall the definition of separable and entangled states. Given two systems A, B , the *separable states* of the bipartite system AB are those of the form

$$|\sigma\rangle_{AB} = \sum_{i \in I} |\alpha_i\rangle_A |\beta_i\rangle_B, \quad (5)$$

with $\text{St}(A) \ni |\alpha_i\rangle_A \neq |\varepsilon\rangle_A$, $\text{St}(B) \ni |\beta_i\rangle_B \neq |\varepsilon\rangle_B$ for every $i \in I$. This set of states contains all states that can be prepared using only local operations and classical communication (LOCC)—but is generally larger than the set of LOCC states. In CT and QT the converse is also true. By negation, *entangled states* are those states that are *nonseparable*. The two following results hold for arbitrary OPTs:

Proposition 2. Let Θ be an OPT. If Θ does not satisfy local discriminability, then it admits entangled states.

Proof. By hypothesis, from Eqs. (3) and (4), we have that there exist $A, B \in \text{Sys}(\Theta)$ such that $D_{AB} > D_A D_B$. Since product states generate a subspace of $\text{St}_{\mathbb{R}}(AB)$ of dimension $D_A D_B$, containing all separable states, $\text{St}(AB)$ must contain at least one state that is not separable. Then Θ admits entangled states. ■

Proposition 3. Let Θ be an OPT. If Θ does not satisfy atomicity of state composition, then it admits entangled states.

Proof. By hypothesis, there exist two systems A, B and two atomic states $|\rho\rangle_A, |\tau\rangle_B$ whose parallel composition $|\sigma\rangle_{AB} = |\rho\rangle_A |\tau\rangle_B$ is not atomic. By contradiction, suppose that the states of the composite system AB are all separable, namely of the form (5). By hypothesis, $|\sigma\rangle_{AB}$ is not atomic and then admits a decomposition of the form

$$|\sigma\rangle_{AB} = \sum_{i=1}^p |\alpha_i\rangle_A |\beta_i\rangle_B, \quad (6)$$

where there are two indices $1 \leq l < k \leq p$ such that $|\alpha_l\rangle_A |\beta_l\rangle_B \not\propto |\alpha_k\rangle_A |\beta_k\rangle_B$. This implies that $|\alpha_l\rangle_A \not\propto |\alpha_k\rangle_A$ or $|\beta_l\rangle_B \not\propto |\beta_k\rangle_B$ —say the former without loss of generality. Now, since the decomposition in Eq. (6) can be taken such that $|\beta_i\rangle_B \neq |\varepsilon\rangle_B$ for every $1 \leq i \leq p$ without loss of generality, then for every $1 \leq i \leq p$ there must exist a deterministic effect $(\tilde{e}_i|_B$ such that $(\tilde{e}_i|\beta_i\rangle_B > 0$. Let us now choose one of the deterministic effects $(\tilde{e}_j|_B$. Let also $|\tilde{\tau}\rangle_B$ be such that

²One may argue that the property of atomicity of state composition should be renamed *purity of state composition*. However, as opposed to the notion of *atomicity* (from conic analysis) and *extremality* (from convex analysis), the notion of *purity* is a historical one introduced only for extremal deterministic states. In causal theories (see Property 4 in Sec. IV), where every state is proportional to a deterministic one, there is no need for a distinction between atomicity and purity. With a slight abuse of terminology, in several situations, purity is used as a synonym of atomicity.

$\{|\tau\rangle_B, |\tilde{\tau}\rangle_B\}$ is a test. Then

$$\begin{aligned} & (\tilde{e}_j|_B|\sigma)_{AB} + (\tilde{e}_j|\tilde{\tau})_B|\rho\rangle_A \\ &= (\tilde{e}_j|\tau)_B|\rho\rangle_A + (\tilde{e}_j|\tilde{\tau})_B|\rho\rangle_A = |\rho\rangle_A \\ &= \sum_{i=1}^p (\tilde{e}_j|\beta_i)_B|\alpha_i\rangle_A + (\tilde{e}_j|\tilde{\tau})_B|\rho\rangle_A. \end{aligned}$$

For a fixed choice of j , the term $(\tilde{e}_j|\beta_j)_B|\alpha_j\rangle_A$ is non-null, and by the atomicity of $|\rho\rangle_A$, one must then have $|\alpha_j\rangle_A \propto |\rho\rangle_A$. Since this is true for every $1 \leq j \leq p$, we come to a contradiction with the hypothesis that $|\alpha_l\rangle_A \not\propto |\alpha_k\rangle_A$. This implies that there exists a state of AB which is not separable, namely Θ admits entangled states. ■

A counterexample to the converse of both Propositions 2 and 3 is simply given by QT, which contains entangled states and also satisfies both local discriminability and atomicity of state composition.

IV. ENTANGLEMENT IN SIMPLICIAL THEORIES

In the present section, we characterize simplicial theories from the point of view of the admissibility of entangled states, finding that the converse of Proposition 2 holds in general for this family of theories, whereas the converse of Proposition 3 holds in the case of simplicial theories satisfying n -local discriminability for some positive integer n .

Definition 1: simplicial theory. A simplicial theory Θ is a finite-dimensional OPT where the extremal states of every system $A \in \mathbf{Sys}(\Theta)$ are the vertices of a D_A simplex.

Notice that a D_A simplex is the convex hull of $D_A + 1$ vertices, which in the present context are the elements of $\mathbf{ExtSt}(A)$ [including also the null state $|\varepsilon\rangle_A$]. The property of simpliciality for an OPT Θ implies that, for every system $A \in \mathbf{Sys}(\Theta)$: (i) A has exactly D_A non-null extremal states, and (ii) every state of A has a unique decomposition as a convex combination of non-null extremal states. We make extensive use of the two previous properties in the remainder of the paper.

Property 3: joint perfect discriminability. Let Θ be an OPT and $A \in \mathbf{Sys}(\Theta)$. A set of states $\{|\rho_i\rangle_A\}_{i=1}^n$ is jointly perfectly discriminable if there exists an observation test $\{(a_i|_A)\}_{i=1}^n$ such that

$$(a_i|\rho_j)_A = \delta_{ij} \quad \forall i, j \in \{0, 1, \dots, n\}.$$

As has been recalled, *classical theories* are those simplicial theories where the pure states satisfy Property 3. On the other hand, CT is a simplicial theory satisfying convexity, local discriminability, and joint perfect discriminability of pure states for every system. Furthermore, the following property, i.e., *causality*, is usually also assumed for CT.

Property 4: causality. An OPT is causal if every system admits a unique deterministic effect.

The above definition of causality is in fact equivalently formulated in a more intuitive way as follows: A theory is causal if the probabilities of the outcomes of preparations are independent of the choice of the observations connected at their output. The latter statement, indeed, can be recognized as the most popular formulation of the notion of causality in a physical theory. However, its content is equivalent to

Property 4 (for a proof, see Ref. [3]). We now prove that the property of causality is in fact intrinsic to that of simpliciality, thus establishing that all simplicial theories—CT included—are indeed inherently causal.

Theorem 1. Simplicial theories are causal.

Proof. Let Θ be a simplicial theory, $A \in \mathbf{Sys}(\Theta)$, and $|i\rangle_A$ be an extremal and non-null state. Suppose that $|i\rangle_A$ is not deterministic: then it can be completed to a deterministic one; namely, there exists a non-null state $|\sigma\rangle_A$ such that $|\rho\rangle_A = |i\rangle_A + |\sigma\rangle_A$ is deterministic. Now, let $(b_j|_A$ for $j = 1, \dots, D_A$ be the linear functionals such that $(b_j|j')_A = \delta_{jj'}$ for every non-null extremal state $|j'\rangle_A$: this implies, by simpliciality, that $0 \leq (b_j|\tau)_A \leq 1$ for all $|\tau\rangle_A \in \mathbf{St}(A)$ and all j . One then has

$$(b_i|\rho)_A = (b_i|i)_A + (b_i|\sigma)_A = 1 + (b_i|\sigma)_A \leq 1;$$

namely, $(b_i|\sigma)_A = 0$ and $(b_i|\rho)_A = 1$. Then, posing $\mathbf{ExtSt}(A) := \{|k\rangle_A\}_{k=0}^{D_A}$, with $|0\rangle_A := |\varepsilon\rangle_A$, one has the following:

$$|\sigma\rangle_A = \sum_{k=0}^{D_A} p_k |k\rangle_A, \quad p_k \geq 0 \quad \forall k, \quad p_i = 0, \quad \sum_{k=0}^{D_A} p_k \leq 1.$$

Let us now pose

$$(e|_A := \sum_{j=1}^{D_A} (b_j|_A.$$

Clearly $0 \leq (e|\tau)_A \leq 1$ for all $|\tau\rangle_A \in \mathbf{St}(A)$, and $(e|i)_A = 1$, implying $(e|\sigma)_A = 0$, i.e., $p_k = 0 \quad \forall k \neq 0$. This shows that $|i\rangle_A$ is deterministic. Being independent of i , the above argument proves that all non-null extremal states are deterministic. Extending then the above argument to arbitrary convex combinations of non-null extremal states of the form $\sum_{k=0}^{D_A} p_k |k\rangle_A$, it is easy to see that a state is deterministic if and only if $\sum_{k=1}^{D_A} p_k = 1$. In other words, the set of deterministic states coincides with the convex hull of extremal non-null states. Now, the effect $(e|_A$ amounts to unity on all deterministic states, hence it is deterministic. Since the non-null extremal states of a simplicial theory are complete and linearly independent, there exists a unique effect $(\tilde{e}|_A$ such that $(\tilde{e}|j)_A = 1$ for all extremal non-null states $|j\rangle_A$. Thus the deterministic effect $(e|_A$ is unique. ■

Now we show that, for any simplicial theory, the converse of Proposition 2 is also true.

Theorem 2. Let Θ be a simplicial theory. Then Θ admits entangled states if and only if it does not satisfy local discriminability.

Proof. (\Leftarrow) The implication holds true by Proposition 2. (\Rightarrow) By hypothesis, the theory Θ is simplicial. Let us denote the non-null extremal states of a system X by $|l\rangle_X$, for $l = 1, \dots, D_X$. As a straightforward consequence of Theorem 1, such states are also deterministic. Then, for all $A, B \in \mathbf{Sys}(\Theta)$ and all non-null $|i\rangle_A \in \mathbf{ExtSt}(A)$, $|j\rangle_B \in \mathbf{ExtSt}(B)$, there exists a nonempty set $I_{ij} \subseteq \{1, \dots, \}$ such that

$$|i\rangle_A |j\rangle_B = \sum_{k \in I_{ij}} p_k^{ij} |k\rangle_{AB}, \quad (7)$$

with $p_k^{ij} > 0$ for $k \in I_{ij}$ and $\sum_{k \in I_{ij}} p_k^{ij} = 1$. Let $(e|_A \in \text{Eff}(A))$ be the (unique) deterministic effect of A . Since for all $k \in \{1, \dots, D_{AB}\}$, $(e|_A|k)_{AB} \in \text{St}(B)$ is deterministic, it is also nonvanishing for all k . Moreover, let us suppose that there exist two different pairs of indices $(\tilde{i}, \tilde{j}), (\tilde{i}', \tilde{j}')$ such that $\tilde{J} := I_{\tilde{i}\tilde{j}} \cap I_{\tilde{i}'\tilde{j}'} \neq \emptyset$. Without loss of generality, we can assume $\tilde{j} \neq \tilde{j}'$. Choose $\tilde{k} \in \tilde{J}$: being the convex decomposition into non-null extremal states unique by simpliciality, from Eq. (7) we conclude that $(e|_A|\tilde{k})_{AB} = |\tilde{j}\rangle_B = |\tilde{j}'\rangle_B$, which is absurd. Thus $I_{ij} \cap I_{i'j'} = \emptyset$ for every different pair of indices. Now, let us suppose that Θ satisfies local discriminability. Then, by Proposition 1, Eq. (4) holds, and $I_{ij} \subseteq \{1, 2, \dots, D_A D_B\} \forall i, j$. By conditions $I_{ij} \neq \emptyset$ and $I_{ij} \cap I_{i'j'} = \emptyset$, every set I_{ij} must be a singleton. This implies that, for every pair i, j , there exists $|k\rangle_{AB}$ such that $|k\rangle_{AB} = |i\rangle_A |j\rangle_B$; namely, the parallel composition of non-null extremal states is a non-null extremal state. Accordingly, Eq. (4) implies that the states of the form $|i\rangle_A |j\rangle_B$ exhaust the set of non-null extremal states of the composite system AB . As a consequence, Θ does not admit entangled states. ■

Next, we prove that, for a simplicial theory satisfying n -local discriminability for some positive integer n , the converse of Proposition 3 is also true. To do this, we need the following result, which is proved in Appendix A:

Proposition 4. Let Θ be a simplicial OPT satisfying n -local discriminability for some positive integer n . For every pair of systems $A, B \in \text{Sys}(\Theta)$, every non-null extremal state $|\lambda\rangle_{AB} \in \text{ExtSt}(AB)$ convexly refines the parallel composition of some pair of pure states $|\rho\rangle_A |\sigma\rangle_B \in \text{St}(AB)$.

We are now in position to prove the following theorem:

Theorem 3. Let Θ be a simplicial theory satisfying n -local discriminability for some positive integer n . Then Θ admits entangled states if and only if it does not satisfy atomicity of state composition.

Proof. (\Leftarrow) The implication holds true by Proposition 3. (\Rightarrow) According to Proposition 4, under the hypothesis of the theorem every non-null extremal state of a bipartite system refines some parallel composition of extremal states. Since by hypothesis the theory admits entangled states, by Theorem 2 and Proposition 1 there exists a pair of systems $A, B \in \text{Sys}(\Theta)$ such that $D_{AB} > D_A D_B$. Then, there are more non-null extremal states of the composite system than parallel compositions of (non-null) extremal states. Thus, there must be a parallel composition of extremal states $|i\rangle_A |j\rangle_B$ that is refined by more than one non-null extremal states, i.e., $|i\rangle_A |j\rangle_B = \sum_{k \in I_{ij}} p_k^{ij} |k\rangle_{AB}$, namely atomicity of state composition does not hold. ■

The following lemma provides a characterization of the parallel composition of states in simplicial theories satisfying n -local discriminability for some positive integer n :

Lemma 1. Let Θ be a simplicial theory satisfying n -local discriminability for some positive integer n . Then, for all systems $A, B \in \text{Sys}(\Theta)$ and non-null extremal states $|k\rangle_{AB} \in \text{ExtSt}(AB)$, there exists a unique product of non-null extremal states $|i_k j_k\rangle_{AB} = |i_k\rangle_A |j_k\rangle_B$ such that $|k\rangle_{AB}$ convexly refines $|i_k j_k\rangle_{AB}$.

Proof. Existence is provided by Proposition 4, and uniqueness has been proved in the proof of Theorem 2. ■

Notice that the map $k \mapsto i_k j_k$ of the above lemma is not injective in general. In particular, the map is not injective as long as the theory does not satisfy atomicity of state composition. As a straightforward consequence of Lemma 1, we are now in position to provide a general classification of the composite-state spaces in simplicial theories satisfying n -local discriminability for some integer n .

Theorem 4: classification of composite-state spaces in simplicial theories. Let Θ be a simplicial theory satisfying n -local discriminability for some integer n . For every pair of systems $A, B \in \text{Sys}(\Theta)$, the state-space $\text{St}(AB)$ is classified by the following: (i) a choice of the dimension of the composite system D_{AB} ; (ii) an unambiguous labeling for the pure states of AB as $|(ij)_k\rangle_{AB}$ —for $i \in \{1, \dots, D_A\}, j \in \{1, \dots, D_B\}$, and k in finite sets I_{ij} —and a choice of probability distributions p_k^{ij} , such that

$$|i\rangle_A |j\rangle_B = \sum_{k \in I_{ij}} p_k^{ij} |(ij)_k\rangle_{AB}, \tag{8}$$

with $p_k^{ij} > 0$ and $\sum_{k \in I_{ij}} p_k^{ij} = 1$ for all non-null extremal states $|i\rangle_A \in \text{ExtSt}(A), |j\rangle_B \in \text{ExtSt}(B)$.

Notice that not all the choices of the coefficients p_k^{ij} in Eq. (8) necessarily lead to a consistent theory. However, there are cases of classical theories beyond CT, proving that a consistent choice of p_k^{ij} is possible [12].

We conclude our investigation on simplicial theories by proving that these theories, whereas they may admit of entangled states, cannot admit of superposition (in an operational sense), or purification (under the hypothesis of n -local discriminability for some positive integer n). We formulate the superposition principle in three possible ways.

Property 5: superposition principle. Let $\mathcal{D} = \{|\rho_i\rangle_A\}_{i=1}^d$, with $d \geq 2$, be any maximal set of jointly perfectly discriminable pure states of a system A . Let $\mathcal{A} = \{(a_i)\}_{i=1}^d$ denote an observation such that $(a_i|\rho_j) = \delta_{ij}$ for all $i, j \in \{1, \dots, d\}$, $\mathbf{p} = \{p_i\}_{i=1}^d$ a probability distribution, and $|\sigma\rangle_A$ a pure state of A .

(i) *Ultraweak:* For every choice of \mathbf{p} , there exist \mathcal{A} and $|\sigma\rangle_A$ such that the following holds:

$$p_i = (a_i|\sigma)_A \forall i = 1, \dots, d. \tag{9}$$

(ii) *Weak:* For every choice of \mathbf{p} and \mathcal{A} , there exists $|\sigma\rangle_A$ such that Eq. (9) holds.

(iii) *Strong:* For every choice of \mathbf{p} , there exists $|\sigma\rangle_A$ such that, for every choice of \mathcal{A} , Eq. (9) holds.

Property 6: purification principle. Let $|\rho\rangle_A$ be a deterministic state. Then there exists a system B , a pure state $|\Sigma_\rho\rangle_{AB}$, and a deterministic effect $(e_\rho|_B)$, such that

$$|\rho\rangle_A = (e_\rho|_B|\Sigma_\rho)_{AB}.$$

Both the ultraweak superposition and the purification principles are satisfied by QT—but, clearly, not by CT. Interestingly, a simplicial or classical theory where the states of every system are exhausted by the pure ones—e.g., this is the case of deterministic classical theory (DCT) [13]—trivially satisfies the purification principle, since all states are pure.

Theorem 5: no superposition for simplicial theories. Let Θ be an OPT. If Θ is simplicial, then there is no system in Θ

satisfying the weak formulation of the superposition principle. If Θ is classical, then there is no system in Θ satisfying any formulation of the superposition principle.

Proof. Let Θ be a simplicial theory, and A a system of Θ . First, we notice that, if there does not exist a set $\mathcal{D} = \{|i\rangle_A\}_{i=1}^d$ of jointly perfectly discriminable pure states with $d \geq 2$, Property 5 is empty. Accordingly, we assume that \mathcal{D} exists, that it is maximal, and, as a necessary condition, that $D_A \geq 2$. Let us pose $I := \{1, \dots, d\}$ and $K := \{d+1, \dots, D_A\}$. Let $\{(f_i|_A)\}_{i \in I \cup K}$ be the set of linear functionals such that $(f_i|_j)_A = \delta_{ij}$ for all non-null extremal states $|j\rangle_A \in \text{ExtSt}(A)$ and all $i \in I \cup K$. Then, every observation $\mathcal{A} = \{(a_i)\}_{i \in I}$ such that $(a_i|i') = \delta_{ii'}$ for all $i, i' \in I$ has the following form:

$$(a_i|_A = (f_i|_A + \sum_{k \in K} q_k^i (f_k|_A, \quad (10)$$

with $q_k^i \geq 0$ and $\sum_{i \in I} q_k^i = 1$ for all $i \in I, k \in K$, since it must be $\sum_{i \in I} (a_i|_A = (e|_A \equiv \sum_{i \in I \cup K} (f_i|_A$. If the theory Θ is not classical, then $|I| \equiv d < D_A$ and $|K| \geq 1$. Being the number of pure states finite, for every choice of $\mathcal{A} = \{(a_i)\}_{i \in I}$ of the form (10), Eq. (9) might be satisfied only for a finite number of choices of probability distributions $\mathbf{p} = \{p_i\}_{i \in I}$, namely, those with $p_i = \delta_{i i_0}$ and $i_0 \in I$ or those with $p_i = q_k^i$ for $k \in K$. Therefore, no simplicial theory satisfies the weak formulation (ii) of Property 5. If the theory Θ is classical, then $|I| \equiv d = D_A$ and $K = \emptyset$ by definition. This means that an observation \mathcal{A} of the form (10) is unique—in particular, $\mathcal{A} = \{(f_i|_A)\}_{i \in I}$. Accordingly, Eq. (9) can be satisfied for a unique (modulo permutations of the indices) choice of probability distribution $\mathbf{p} = \{p_i\}_{i=1}^d$, namely, $p_i = \delta_{i i_0}$ for some $i_0 \in I$ and all $i \in I \setminus \{i_0\}$. Therefore no classical theory satisfies the ultraweak formulation (i) of Property 5. ■

Theorem 6: no purification for simplicial theories with n -local discriminability. Let Θ be a simplicial theory satisfying n -local discriminability for some n . Then, there is no mixed state in Θ having a purification.

Proof. By Theorem 4, every pure state of a composite system convexly refines the parallel composition $|i\rangle_A |j\rangle_B$ of two extremal states. In particular, denoting the (unique) deterministic effect of B by $(e|_B$, by direct inspection of Eq. (8) one concludes that, for any pure state $|(ij)_k\rangle_{AB}$, it must be $(e|_B |(ij)_k\rangle_{AB} = |i\rangle_A$, since the convex decomposition into non-null extremal states is unique by simpliciality. Accordingly, any marginal state of a pure state is pure, implying that no mixed state can be purified in Θ . ■

V. DISCUSSION AND CONCLUSIONS

In Propositions 2 and 3 we showed that, in arbitrary probabilistic theories, the presence of entangled states is a consequence of the failure of local discriminability or, independently, of atomicity of state composition. Then we specialized to simplicial theories, proving the converse of Propositions 2 and 3 for this class of theories. In particular, Theorem 2 asserts that a simplicial theory admits entangled states if and only if it does not satisfy local discriminability. Theorems 3 and 4 are proved under the hypothesis of n -local discriminability for some integer n . On the one hand, a simplicial theory with n -local discriminability contains entangled states if and only if it does not satisfy atomicity of state composition. This implies

that there exists at least one pair of systems A, B such that at least a product of pure states of A and B is not pure. On the other hand, Theorem 4 provides a simple classification of the parallel composition rule for such theories, as a consequence of the failure of atomicity of composition. Equation (8), in turn, implies that the marginal state of a pure entangled state in a simplicial theory with n -local discriminability is always pure.

As far as classical theories are concerned, on the one hand, the above results entail that the usual notion of classicality does not prevent the presence of entanglement in a physical theory. Indeed, CT is the only classical theory without entanglement.³ On the other hand, Theorem 1 asserts that the notion of classicality given by the simplicial structure of states constrains a theory to be necessarily causal. Interestingly, in the above sense, this means that the notion of causality is inherent to that of classicality, representing a no-go theorem for conceiving *noncausal* classical theories.

Finally, Theorems 5 and 6 highlight two relevant properties that every simplicial theory shares with CT: superposition is not admitted, and no mixed state has a purification if n -local discriminability holds for some n . Interestingly, at this stage it cannot be excluded that there might exist nonclassical simplicial theories satisfying the ultraweak operational formulation of the superposition principle (see Property 5). Besides, we notice that Theorem 6 has a different content from the no-go theorem proven in Ref. [14], although the latter holds for convex discrete theories—which strictly include the simplicial ones. Indeed, the no-go theorem of Ref. [14] states that, in a convex discrete theory, for every system *only a finite number of (mixed) states may possibly have a purification*. However, our Theorem 6 generalizes the latter in the case of simplicial theories with n -local discriminability, stating that *not a single (mixed) state admits a purification*.

The general results of the present work do not rely on any additional structure—such as, convexity, the no-restriction hypothesis [3,15], or other properties—beyond the simplicial one. If a theory is simplicial, there is no complementarity—and it is thus impossible to violate Bell's inequalities.⁴ Moreover, in the presence of local discriminability, the converse is also true [16]. While the nonviolation of any probability bound means that *every single correlation* of the theory can be described by a local realistic hidden-variable model, this in principle does not imply that there exists a *coherent and complete ontological model* [17] describing *the theory as a whole* in a local realistic fashion. Then, an interesting open question is whether there exist a convenient ontological model for such theories.

All simplicial theories with entanglement contain states with non-null discord (see Ref. [18]), despite being simplicial. Furthermore, as shown in Ref. [13], simplicial theories are

³More precisely, CT is the only classical theory that is (i) without entanglement, (ii) convex, and (iii) such that every transformation which is compatible with simpliciality, convexity, and local discriminability is allowed.

⁴Indeed, if there is no complementarity, there exists a joint probability distribution for the outcomes of any pair of measurements, which then excludes a violation of any probability bound.

the only ones where full information can be extracted by measurements that do not disturb the measured system—and this in particular holds despite the presence of entanglement. Moreover, simplicial theories with entanglement feature *hy-persignaling* [19], since for such theories one has $D_{AB} > D_A D_B$ for at least a pair of systems A, B. At this stage, it is not possible to compare simplicial theories exhibiting entanglement with results on *broadcasting* [5] or *teleportation* [20] in arbitrary probabilistic theories, since local discriminability was therein assumed.

It is interesting to notice that all classical theories have no *dimension mismatch* [21], while in the case of simplicial theories the problem is open. Interestingly, simplicial theories satisfy both information causality (IC) [22] and the information content principle (ICP) [23]. This is a consequence of the fact that, in a simplicial theory, the mixing entropy satisfies the sufficient properties guaranteeing that IC and ICP hold in a general simplicial theory [23–25]. In particular, IC and ICP cannot single out CT among arbitrary classical theories.

More generally, are there device-independent principles [26,27] ruling out simplicial theories with entanglement, complete with the set of transformations, would allow one to study the simplicial scenario as far as, for example, communication complexity [28] is concerned. One could also examine the relation between simplicial theories with entanglement and the so-called *epistemically restricted* (or “*epirestricted*”) theories—such as Spekkens toy theory [29]—or consider them in the light of the literature on the notion of classicality (e.g., see Refs. [17,30,31]). We conclude by pointing out that our results highlight the relevance of the notion of compositionality [32] in the scope of probabilistic theories. Indeed, our results strongly rely on the general properties of the parallel composition of systems in a theory. It would be interesting to investigate the applicability of the methods exploited to broader contexts, e.g., extending them to theories where the sets of states are quantum, or even to post-quantum theories.

ACKNOWLEDGMENTS

This publication was made possible through the support of a grant from the John Templeton Foundation, ID # 60609 “Quantum Causal Structures.” The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

APPENDIX: PROOF OF PROPOSITION 4

We here recall that, in the main text, we proved that (i) any simplicial theory is causal (see Theorem 1), and (ii) the non-null extremal states of a simplicial theory are deterministic (see proof of Theorem 1). We make use of the above results in the present Appendix. We also introduce the definition of separable state for an arbitrary number of systems, which will be of crucial relevance in the remainder. Let $S = S_1 S_2 \cdots S_n \in \text{Sys}(\Theta)$, and $|\rho\rangle \in \text{St}(S)$. We say that $|\rho\rangle$ is *separable* if there exist (finitely many) disjoint nontrivial bipartitions $S_0^a := \{j_1^a, j_2^a, \dots, j_k^a\}$, $S_1^a := \{j_1^a, j_2^a, \dots, j_{N-k}^a\}$ of $\{1, 2, \dots, n\}$ such

that

$$|\rho\rangle = \sum_{a \in A} p_a |\sigma\rangle_{S_0^a} |\tau\rangle_{S_1^a},$$

with $p_a > 0$ for all $a \in A$.

Lemma 2. Let Θ be a simplicial OPT satisfying n -local discriminability for some positive integer n . For all $(n + 1)$ -partite system $S = S_1 S_2 \cdots S_{n+1} \in \text{Sys}(\Theta)$, every state $|\rho\rangle \in \text{St}(S_1 S_2 \cdots S_{n+1})$ admits a convex decomposition into states each of which convexly refines some separable state of $S_1 S_2 \cdots S_{n+1}$.

Proof. Take the subset $\mathcal{E} \subseteq \text{ExtSt}(S_1 \cdots S_{n+1})$ of all non-null extremal states of $\text{St}(S_1 \cdots S_{n+1})$ which convexly refine some separable state. Since Θ satisfies n -local discriminability, this is a spanning set for the space $\text{St}_{\mathbb{R}}(S_1 \cdots S_{n+1})$. Moreover, since Θ is simplicial, the elements of \mathcal{E} are linearly independent. As a consequence, the dimension of $\text{St}_{\mathbb{R}}(S_1 \cdots S_{n+1})$ amounts to the cardinality of \mathcal{E} , which is then, by simpliciality, a complete set of states convexly generating every state $|\rho\rangle \in \text{St}(S_1 \cdots S_{n+1})$. Equivalently, $\text{ExtSt}(S_1 \cdots S_{n+1}) = \{\mathcal{E}\} \cup \mathcal{E}$. ■

Lemma 3. Let Θ be a simplicial OPT. Let $|\pi\rangle \in \text{St}(S_1 \cdots S_n)$ with $n \geq 2$, so that $|\pi\rangle = |\pi_I\rangle |\pi_J\rangle$, for some states $|\pi_I\rangle \in S_I$, $|\pi_J\rangle \in S_J$ with $I \cup J = \{1, \dots, n\}$, $I, J \neq \emptyset$, $I \cap J = \emptyset$, and $S_K = S_{k_1} S_{k_2} \cdots S_{k_l}$ for every l -tuple $K \subseteq \{1, \dots, n\}$.

Let $|\phi\rangle \in \text{ExtSt}(S_1 \cdots S_n)$ be a non-null extremal state that convexly refines $|\pi\rangle$. Finally, given $i \in I$ and $j \in J$, let $(e_{K_{ij}} |$ denote the deterministic effect on $S_{K_{ij}}$ with $K_{ij} = \{1, \dots, n\} \setminus \{i, j\}$. Then $(e_{K_{ij}} | \pi\rangle)$ is a product state, and $(e_{K_{ij}} | \phi\rangle) \in \text{St}(S_i S_j)$ is a physical state that convexly refines $(e_{K_{ij}} | \pi\rangle)$.

Proof. The case $n = 2$ is trivially true, and we then assume $n \geq 3$ in the following. By hypothesis, we can pose $|\pi\rangle = p|\phi\rangle + (1 - p)|\sigma\rangle$, where $|\sigma\rangle$ is a deterministic state and $p \in (0, 1]$. By construction we have that

$$(e_{K_{ij}} | \pi\rangle) = p(e_{K_{ij}} | \phi\rangle) + (1 - p)(e_{K_{ij}} | \sigma\rangle) \in \text{St}(S_i S_j), \quad (\text{A1})$$

and clearly both $(e_{K_{ij}} | \phi\rangle)$ and $(e_{K_{ij}} | \sigma\rangle)$ are deterministic states of $S_i S_j$. Moreover, by causality, $(e_{K_{ij}} | = (e_{I \setminus \{i\}} | (e_{J \setminus \{j\}} |$, then $(e_{K_{ij}} | \pi\rangle)$ is a product state of $S_i S_j$. Since the convex decomposition into non-null extremal states is unique by simpliciality, from Eq. (A1) we conclude that $(e_{K_{ij}} | \phi\rangle)$ —that is non-null, although it may possibly be nonextremal—convexly refines $(e_{K_{ij}} | \pi\rangle)$. ■

Proposition 4. Let Θ be a simplicial OPT satisfying n -local discriminability for some positive integer n . For every pair of systems $A, B \in \text{Sys}(\Theta)$, every non-null extremal state $|\lambda\rangle_{AB} \in \text{ExtSt}(AB)$ convexly refines the parallel composition of some pair of non-null extremal states $|\rho\rangle_A |\sigma\rangle_B \in \text{St}(AB)$.

Proof. By contradiction, let us suppose that there exists a pair of systems $A, B \in \text{Sys}(\Theta)$ and an extremal state $|\lambda\rangle_{AB} \in \text{ExtSt}(AB)$ that does not convexly refine any state of the form $|\rho\rangle_A |\sigma\rangle_B \in \text{St}(AB)$ with $|\rho\rangle_A$ and $|\sigma\rangle_B$ non-null and extremal. Let us denote

$$|\lambda\rangle_{AB} := \left(\begin{array}{c} \vdots \\ \vdots \end{array} \right). \quad (\text{A2})$$

The theory Θ satisfies n -local discriminability. The case $n = 1$ is trivial, since all states are separable (see Theorem 2

in the main text) and $|\lambda\rangle_{AB}$ must be vanishing. In the following we assume $n \geq 2$.

Let us now define the following $(n + 1)$ -partite state:

$$|\Psi\rangle := \left(\begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right) \in \text{St}((A_1 \cdots A_n) B_1 \cdots B_n), \quad (\text{A3})$$

where the systems A_m and $B_{m'}$ are copies of, respectively, A and B for every $m, m' \in \{1, \dots, n\}$. By Lemma 2, the state $|\Psi\rangle$ must be in the convex hull of some states refining the separable states $|\Lambda\rangle$ of $(A_1 \cdots A_n) B_1 \cdots B_n$. The latter must be of one of the following two types:

$$|\Lambda\rangle = \left(\begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right), \quad \left(\begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right), \quad (\text{A4})$$

where in the first case $(A_1 \cdots A_n)$ is factorized from $B_1 \cdots B_n$, while in the second case there must exist a state of some proper subsystem of $B_1 \cdots B_n$ that is factorized. Then there

exist coefficients $\alpha_i \in [0, 1]$ such that

$$|\Psi\rangle = \sum_i \alpha_i |\phi_i\rangle, \quad (\text{A5})$$

where the $|\phi_i\rangle$ are non-null extremal states in the convex refinement of some separable state $|\Lambda\rangle$ of one of the two types in Eq. (A4).

By construction, in both cases we can always find at least a subsystem $S = A_j B_j$ of $A_1 \cdots A_n B_1 \cdots B_n$ (now considered as a $2n$ -partite system) such that the marginal state $(e_{\bar{S}}|\Lambda)$ (where \bar{S} is the complementary subsystem of S in $A_1 \cdots A_n B_1 \cdots B_n$) is a product state of $A_j B_j$. On the other hand the marginal state of $|\Psi\rangle$ on S is $(e_{\bar{S}}|\Psi) = |\lambda\rangle_{A_j B_j}$. For each term on the right-hand side in Eq. (A5), one can then find the above-mentioned subsystem S and apply $(e_{\bar{S}}|$ to both sides. This gives an equation of the form

$$|\lambda\rangle_{A_j B_j} = \alpha_i |\chi_i\rangle_{A_j B_j} + |\omega_i\rangle_{A_j B_j}, \quad (\text{A6})$$

where $|\chi_i\rangle_{A_j B_j}$ is in the convex refinement of some product state by Lemma 3, and $|\omega_i\rangle_{A_j B_j}$ is a physical state. Since $|\lambda\rangle_{A_j B_j}$ is an extremal point of a simplex, its convex decomposition must be trivial, and then either $\alpha_i = 0$ or

$$|\lambda\rangle_{A_j B_j} = |\chi_i\rangle_{A_j B_j}. \quad (\text{A7})$$

Finally, either $\alpha_i = 0$ holds for every i , and then by direct inspection of the definitions (A2) and (A3) $|\lambda\rangle_{AB}$ is vanishing, or identity (A7) holds for some i , which is a contradiction. \blacksquare

[1] R. F. Werner, Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model, *Phys. Rev. A* **40**, 4277 (1989).

[2] J. Barrett, Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a Bell inequality, *Phys. Rev. A* **65**, 042302 (2002).

[3] G. M. D'Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, Cambridge, 2017).

[4] J. Barrett, Information processing in generalized probabilistic theories, *Phys. Rev. A* **75**, 032304 (2007).

[5] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Generalized No-Broadcasting Theorem, *Phys. Rev. Lett.* **99**, 240501 (2007).

[6] C. Pfister and S. Wehner, An information-theoretic principle implies that any discrete physical theory is classical, *Nat. Commun.* **4**, 1851 (2013).

[7] P. Janotta and H. Hinrichsen, Generalized probability theories: What determines the structure of quantum theory? *J. Phys. A: Math. Theor.* **47**, 323001 (2014).

[8] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, *Phys. Rev. A* **81**, 062348 (2010).

[9] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Informational derivation of quantum theory, *Phys. Rev. A* **84**, 012311 (2011).

[10] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum from principles, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016), pp. 171–221.

[11] L. Hardy and W. K. Wootters, Limited holism and real-vector-space quantum theory, *Found. Phys.* **42**, 454 (2012).

[12] G. M. D'Ariano, M. Erba, and P. Perinotti, Classicality without local tomography: Decoupling complementarity and entanglement (unpublished).

[13] G. M. D'Ariano, P. Perinotti, and A. Tosini, Information and disturbance in operational probabilistic theories, [arXiv:1907.07043](https://arxiv.org/abs/1907.07043).

[14] M. Winczewski, T. Das, K. Horodecki, P. Horodecki, Ł. Pankowski, M. Piani, and R. Ramanathan, No purification in all discrete theories and the power of the complete extension, [arXiv:1810.02222](https://arxiv.org/abs/1810.02222).

[15] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, *Phys. Rev. A* **87**, 052131 (2013).

[16] M. Plávala, All measurements in a probabilistic theory are compatible if and only if the state space is a simplex, *Phys. Rev. A* **94**, 042108 (2016).

[17] D. Schmid, J. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, The characterization of noncontextuality in the framework of generalized probabilistic theories, [arXiv:1911.10386](https://arxiv.org/abs/1911.10386).

[18] P. Perinotti, Discord and Nonclassicality in Probabilistic Theories, *Phys. Rev. Lett.* **108**, 120502 (2012).

[19] M. Dall'Arno, S. Brandsen, A. Tosini, F. Buscemi, and V. Vedral, No-Hypersignaling Principle, *Phys. Rev. Lett.* **119**, 020401 (2017).

- [20] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Teleportation in General Probabilistic Theories, *Mathematical Foundations of Information Flow: Proceedings of the Clifford Lectures 2008, Proceedings of Symposia in Applied Mathematics* (American Mathematical Society, 2012), Vol. 71, pp. 25-47.
- [21] N. Brunner, M. Kaplan, A. Leverrier, and P. Skrzypczyk, Dimension of physical systems, information processing, and thermodynamics, *New J. Phys.* **16**, 123050 (2014).
- [22] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski, Information causality as a physical principle, *Nature (London)* **461**, 1101 (2009).
- [23] L. Czekaj, M. Horodecki, P. Horodecki, and R. Horodecki, Information content of systems as a physical principle, *Phys. Rev. A* **95**, 022119 (2017).
- [24] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, Entropy and information causality in general probabilistic theories, *New J. Phys.* **12**, 033024 (2010).
- [25] S. W. Al-Safi and A. J. Short, Information causality from an entropic and a probabilistic perspective, *Phys. Rev. A* **84**, 042323 (2011).
- [26] V. Scarani, The device-independent outlook on quantum physics, *Acta Phys. Slovaca* **62**, 347 (2012).
- [27] A. Grinbaum, How device-independent approaches change the meaning of physical theory, *Stud. Hist. Phil. Mod. Phys.* **58**, 22 (2017).
- [28] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, Limit on Nonlocality in Any World in Which Communication Complexity is Not Trivial, *Phys. Rev. Lett.* **96**, 250401 (2006).
- [29] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, *Phys. Rev. A* **75**, 032110 (2007).
- [30] M. Navascués, Macroscopic locality, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016), pp. 439–463.
- [31] J. G. Richens, J. H. Selby, and S. W. Al-Safi, Entanglement is Necessary for Emergent Classicality in All Physical Theories, *Phys. Rev. Lett.* **119**, 080503 (2017).
- [32] B. Coecke, Quantum picturalism, *Contemp. Phys.* **51**, 59 (2010).