

**Classicality without local discriminability: Decoupling entanglement and complementarity**Giacomo Mauro D'Ariano <sup>\*</sup>, Marco Erba <sup>†</sup>, and Paolo Perinotti <sup>‡</sup>*Università degli Studi di Pavia, Dipartimento di Fisica, QUIT Group, and INFN Gruppo IV, Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy*

(Received 8 August 2020; accepted 16 October 2020; published 11 November 2020)

An operational probabilistic theory where all systems are classical, and all pure states of composite systems are entangled, is constructed. The theory is endowed with a rule for composing an arbitrary number of systems, and with a nontrivial set of transformations. Hence, we demonstrate that the presence of entanglement is independent of the existence of incompatible measurements. We then study a variety of phenomena occurring in the theory, some of them contradicting both classical and quantum theories, including cloning, entanglement swapping, dense coding, additivity of classical capacities, nonmonogamous entanglement, hypersignaling. We also prove the existence, in the theory, of a universal processor. The theory is causal and satisfies the no-restriction hypothesis. At the same time, it violates a number of information-theoretic principles enjoyed by quantum theory, most notably, local discriminability, purity of parallel composition of states, and purification. Moreover, we introduce an exhaustive procedure to construct generic operational probabilistic theories, and a sufficient set of conditions to verify their consistency. In addition, we prove a characterization theorem for the parallel composition rules of arbitrary theories, and specialize it to the case of bilocal-tomographic theories. We conclude pointing out some open problems. In particular, on the basis of the fact that every separable state of the theory is a statistical mixture of entangled states, we formulate a no-go conjecture for the existence of a local-realistic ontological model.

DOI: [10.1103/PhysRevA.102.052216](https://doi.org/10.1103/PhysRevA.102.052216)**I. INTRODUCTION**

In Ref. [1], a paper dating back to 1935, Schrödinger provides a seminal description of the phenomenon of entanglement: “When two systems, of which we know the states by their respective representatives, enter into temporary physical interaction due to known forces between them, and when after a time of mutual influence the systems separate again, then they can no longer be described in the same way as before, viz., by endowing each of them with a representative of its own. I would not call that *one* but rather *the* characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought.” Indeed, in classical theory (CT), the state of affairs of every system admits of a suitable description as a statistical mixture of products of pure states. This means that in CT any bipartite state can be prepared by two experimenters just by using local operations and shared randomness. In other words, every state in CT is *separable*. On the contrary, quantum theory (QT) allows for states which are not separable, namely, which are *entangled*. Actually, in the broad landscape of probabilistic theories [2], entanglement is far from being *the* characteristic trait of QT. Most notably, the so-called *PR boxes* [3,4] provided the first example of a probabilistic model beyond QT and featuring entangled states, initiating a fruitful research field in the scope of foundations of physics. On top of this, it has been even

argued that entanglement is *an inevitable feature of any theory superseding classical theory* [5].

While entanglement can be considered a ubiquitous feature in the scope of probabilistic theories, one of the points of this work is to question to which extent the presence of entanglement in a physical theory *enforces its entire departure from classical lines of thought*. One widespread notion of *classicality* for a physical theory [2,5–7] is based on the set of states of the theory. A theory is deemed classical if the pure states of states for every system are (i) the vertices of a simplex, and (ii) jointly perfectly discriminable. We present a complete operational probabilistic theory, which we call bilocal classical theory (BCT), that, in spite of being classical in the above sense, features entangled states. Classical theories with entanglement have been characterized in Ref. [8]. Moreover, it is interesting to notice that BCT also complies with the notion of classicality proposed in Ref. [9], admitting of a noncontextual ontological model.

BCT represents the proof of concept that entanglement is compatible with the absence of *complementarity*, i.e., with the existence of incompatible measurements. The theory is causal and satisfies the no-restriction hypothesis. However, BCT violates the principles of local discriminability, purity of parallel composition of states, and purification. The theory also features nonmonogamous entanglement and hypersignaling [10]. Furthermore, we show that in BCT it is possible to perform entanglement swapping and dense coding, and we prove a theorem of universal programmability for the theory.

The paper is organized as follows. In Sec. II we provide a conceptual preview of the main results of the paper in a nontechnical way. In Sec. III we review the framework of

<sup>\*</sup>dariano@unipv.it<sup>†</sup>marco.erba@unipv.it<sup>‡</sup>paolo.perinotti@unipv.it

operational probabilistic theories (OPTs), which captures the building blocks necessary to construct a theory of physical processes. This can be done by resorting to *category theory* [11]. Indeed, the categorical framework provided a powerful toolbox for deriving quantum theory [12], reformulating it [13], and also for modeling generic physical theories [14–16], or more general kinds of theories [17,18]. In Sec. IV we make a survey of the operational probabilistic structures that are relevant to this work. We prove two first results, providing in particular a characterization for the composition rule of the system sizes in bilocal-tomographic theories. We then move on to considering the construction of a probabilistic theory. When a mathematical structure is constructed, one needs to provide a procedure not only to build it, but also to make sure that it is consistent. Accordingly, in Sec. V we set the problem of consistency for the construction of an operational probabilistic theory, proposing a building procedure, and identifying a sufficient set of conditions to check well posedness and coherence. Thus, in Sec. VI we present bilocal classical theory, also verifying its consistency, while in Sec. VII we analyze a variety of information-theoretic properties of the theory. In Sec. VIII we discuss our findings in the light of the existing literature. Finally, in Sec. IX we draw our conclusions, pointing out some open problems of foundational and interpretative relevance.

## II. CONCEPTUAL PREVIEW

In this paper we exhibit a complete probabilistic theory that embodies features studied in Ref. [8], where classical theories without local discriminability were investigated. Such theories, in particular, feature entanglement, and allow to study its logical dependence on other properties that are commonly associated with it. The first step that we take is to review the framework of operational probabilistic theories (OPTs). OPTs allow for the study of statistics of events, as in generalized probabilistic theories (GPTs), but they also enable us to describe the transformations and their composition in parallel and in sequence. Unlike most of the literature on GPTs, with remarkable exceptions [9], OPTs formalize in a thorough manner the compositional structure, and can be thus thought as a completion of the framework of GPTs, emphasizing the role of transformations. Introducing new structure on top of the statistical content of a GPT, an OPT has to satisfy further consistency constraints, that regard not only its coherence as a probability theory [19], but, most importantly, the interplay of compositional structures<sup>1</sup> among themselves, and with the probabilistic one.

Classical theories are OPTs that are *locally classical*, i.e., the state space of every system is classical, but their composition rule is not the usual one. Constructing such a theory presents remarkable difficulties, in particular, abiding by consistency constraints. For this reason the paper presents an exhaustive discussion of such constraints, that are then checked after introducing in detail an example of classical theory with entanglement.

We now provide a brief survey of the main results of the paper. Theorem 1 provides necessary conditions that the parallel composition rules for arbitrary theories must satisfy, and Theorem 2 provides a characterization of those rules in the case of bilocal-tomographic theories [21]. Bilocal discriminability is particularly relevant in this work since the theory here presented is bilocal tomographic. In Sec. V we formulate a procedure to check the consistency of a given theory, that turns out to be crucial in establishing the validity of the theory constructed in Sec. VI. The latter, that we call bilocal classical theory (BCT), represents the main result of the paper. It constitutes a proof of concept that incompatibility of measurements and entanglement are independent properties. In subsequent sections we analyze other relevant features of BCT, some in common with quantum theory, e.g., dense coding, entanglement swapping, some in common with classical theory, e.g., perfect clonability, full information without disturbance, no Bell nonlocality, universal programming, insecure cryptography, and finally some which supersede both theories, such as the violation of atomicity of parallel composition and of entanglement monogamy, hypersignaling.

## III. OPERATIONAL PROBABILISTIC THEORIES: A REVIEW

The primitive notions of an operational probabilistic theory (OPT) are those of *systems*, *tests*, *events*, and *probabilities*. Systems represent the physical entities which are probed in a laboratory (e.g., an electron, a molecule, a radiation field, etc. ...). Tests represent the physical processes, occurring between two systems, which experiments are made up of (such as the single use of a physical device). Accordingly, an *outcome space* is associated with each test, collecting all the possible outcomes of the test itself. On the other hand, an *event* is associated with each outcome, representing a possible occurrence in a physical process. Finally, the goal of a physical theory is to associate some *probability distributions* with each event.

In this section, we will provide a review of the main general properties of an OPT. In particular, we will present the operational and compositional properties of the primitives, their probabilistic structure, and the resulting linear structure.

### A. Operational structure: Compositional properties of a theory

Let  $\text{Sys}(\Theta)$ ,  $\text{Test}(\Theta)$ , and  $\text{Event}(\Theta)$  denote the classes of, respectively, the systems, the tests, and the events of a theory  $\Theta$ . Systems will be denoted using latin characters  $A, B, C, \dots \in \text{Sys}(\Theta)$ , while tests from a system  $A$  to a system  $B$  will be denoted by  $T_X^{A \rightarrow B} \in \text{Test}(A \rightarrow B)$ , where  $X$  is the corresponding outcome space. A test having a single outcome  $*$  will be called a *singleton test*, and the singleton set will be denoted by  $\star := \{*\}$ . Without loss of clarity, sometimes we will use the shorthand notation  $T \equiv T_X \equiv T_X^{A \rightarrow B}$ . Each test  $T_X^{A \rightarrow B}$  is a collection of events from  $A$  to  $B$ , namely,

$$T_X^{A \rightarrow B} = \{\mathcal{T}_x\}_{x \in X}; \quad \forall x \in X, \mathcal{T}_x : A \rightarrow B.$$

The class of events from a system  $A$  to a system  $B$  will be denoted by  $\text{Event}(A \rightarrow B)$ . Events  $\mathcal{T} \in \text{Event}(A \rightarrow B)$  are represented as wired boxes, where the source and target systems are the labels of the input and, respectively, the output

<sup>1</sup>The compositional structure of an OPT, in technical terms, corresponds to that of a *monoidal category* [20].

wires:

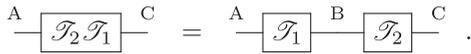


As it is clear from the diagram above, the input-output direction is conventionally represented as going from left to right.

One requires that every test with, say, output system B, can be sequentially composed with any other test having B as input. This operation represents the consecutive occurrence of two physical processes. That is, for all  $A, B, C \in \text{Sys}(\Theta)$  and all  $\mathcal{T}_1 \in \text{Event}(A \rightarrow B)$ ,  $\mathcal{T}_2 \in \text{Event}(B \rightarrow C)$ , there exists an associative map  $\circ$ , called *sequential composition*, and an event  $\mathcal{T}_2 \mathcal{T}_1 := \mathcal{T}_2 \circ \mathcal{T}_1 \in \text{Event}(A \rightarrow C)$ , such that the sequential composition of two tests  $T_X^{A \rightarrow B}, T_Y^{B \rightarrow C}$  is defined as

$$(T \circ T')_{X \times Y}^{A \rightarrow C} \equiv T_Y^{B \rightarrow C} T_X^{A \rightarrow B} := \{\mathcal{T}'_y \circ \mathcal{T}_x\}_{(x,y) \in X \times Y}.$$

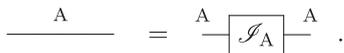
Sequential composition is pictorially represented by the horizontal juxtaposition of boxes from left to right, connecting the two consecutive input and output wires which carry the same label:



Moreover, for all  $S \in \text{Sys}(\Theta)$  there exists a (unique) singleton test, denoted by  $I_S^S = \{\mathcal{I}_S\}$  and called *the identity of S*, satisfying, for all systems A and B, the following property:

$$\mathcal{I}_B \mathcal{I}_A = \mathcal{I}_A = \mathcal{I}_A \mathcal{I}_B, \quad \forall T_X^{A \rightarrow B}, \forall \mathcal{I}_x \in T_X^{A \rightarrow B}. \quad (1)$$

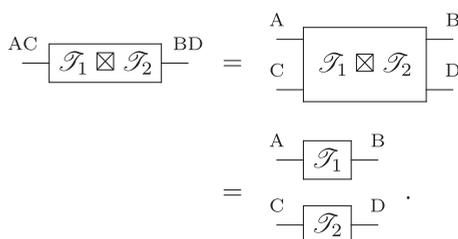
In the following, we will denote the identity family also by using the symbol  $\mathcal{I}$ . Identity processes are equivalent to doing nothing, and can be then conveniently represented just as an extended wire carrying the respective system's label:



One can compose two systems A and B to make the new composite system AB. Correspondingly, any two arbitrary events  $\mathcal{T}_1 \in \text{Event}(A \rightarrow B)$ ,  $\mathcal{T}_2 \in \text{Event}(C \rightarrow D)$  can be composed in parallel. This operation corresponds to an associative map  $\boxtimes$ , called *parallel composition*, that produces the composite event  $\mathcal{T}_1 \boxtimes \mathcal{T}_2 \in \text{Event}(AC \rightarrow BD)$ . The parallel composition of tests is then straightforwardly defined as

$$(T \boxtimes T')_{X \times Y}^{AC \rightarrow BD} \equiv T_X^{A \rightarrow B} \boxtimes T_Y^{C \rightarrow D} := \{\mathcal{T}_x \boxtimes \mathcal{T}'_y\}_{(x,y) \in X \times Y}.$$

This can be thought as a composite test which is completely described by two single tests performed in different laboratories. Parallel composition is pictorially represented by vertically juxtaposing transformations from top to bottom:

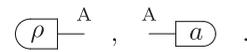


Moreover, the following properties are required for all  $A, B, C, D, E, F \in \text{Sys}(\Theta)$  and all  $\mathcal{T}_1 \in \text{Event}(A \rightarrow B)$ ,  $\mathcal{T}_2 \in \text{Event}(B \rightarrow C)$ ,  $\mathcal{T}_3 \in \text{Event}(D \rightarrow E)$ ,  $\mathcal{T}_4 \in \text{Event}(E \rightarrow F)$ :

$$\frac{A}{B} = \frac{AB}{}, \quad (2)$$

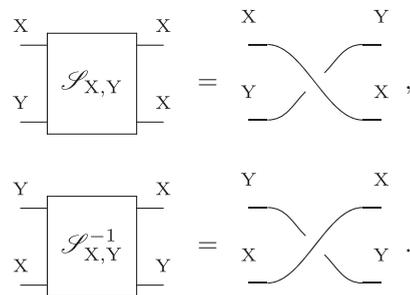
Equation (2) asserts that the parallel composition of the identities  $\mathcal{I}_A$  and  $\mathcal{I}_B$  is the identity  $\mathcal{I}_{AB}$  of the compound system AB. Property (3) states that the operations of sequential and parallel composition commute.

Furthermore, one requires the possibility to consider physical processes having no input or output. The corresponding tests are those where the experimenter disregards everything that, from the viewpoint of the input-output direction, has occurred before or, respectively, will occur after, the physical process considered. In order to capture this notion, there exists a (unique) system I, called the *trivial system*, satisfying  $IA = A = AI$  for all  $A \in \text{Sys}(\Theta)$ . Also, the parallel composition of any test with the identity of I amounts to doing nothing. Indeed, one can conveniently omit an explicit diagrammatic representation of both I and  $\mathcal{I}_I$ , leaving blank spaces. Accordingly, events  $\rho \in \text{Event}(I \rightarrow A)$ ,  $a \in \text{Event}(A \rightarrow I)$  will be represented, respectively, as



Such events are called, respectively, *preparations* and *observations*.

Finally, we introduce one last relevant feature. One can think to each system as being controlled by an agent. Accordingly, one also requires the possibility of exchanging systems between agents. This operation is captured by the notion of *braiding*, namely, a family  $\mathbf{S}$  of invertible singleton tests defined in the following way. For any two systems X, Y, the braiding  $\mathbf{S}$  contains two tests,  $\mathbf{S}^{XY \rightarrow YX} = \{\mathcal{S}_{X,Y}\}$  and its inverse  $(\mathbf{S}^{-1})^{YX \rightarrow XY} = \{\mathcal{S}_{X,Y}^{-1}\}$ , whose associated events will be denoted as follows:



The above graphical representation as a twisting of systems is due to the fact that the braiding is required to obey, for all  $A, B, C, D \in \text{Sys}(\Theta)$  and all  $\mathcal{T}_1 \in \text{Event}(A \rightarrow B)$ ,  $\mathcal{T}_2 \in$

Event( $C \rightarrow D$ ), to the following ‘‘sliding property’’:

$$(4)$$

In the following, we will denote a braiding also by using the symbol  $\mathcal{S}$ , identifying it with the family of events that defines  $\mathbf{S}$ . In the case where the members of the braiding satisfy  $\mathcal{S}_{A,B}^{-1} = \mathcal{S}_{B,A}$  for all  $A, B \in \mathbf{Sys}(\Theta)$ , the OPT is called *symmetric*. Notice that all the theories developed so far are symmetric. In analogy to the cases of CT and QT, symmetry has been assumed in the literature, for both physical and process theories.

**B. Probabilistic structure: Linear properties of a theory**

Tests  $P_X = \{p_x\}_{x \in X} \in \mathbf{Test}(I \rightarrow I)$  are probability distributions. For all  $p \in \mathbf{Event}(I \rightarrow I)$  and all  $\mathcal{T} \in \mathbf{Event}(A \rightarrow B)$ , one can define an operation, called *multiplication by a scalar*, as follows:

$$(5)$$

Notice that the equality can be shown to hold using the properties of parallel composition and of the trivial system. In the case of preparations  $\rho$  and observations  $a$ , one can also show that  $p \boxtimes \rho = \rho \circ p$  and  $p \boxtimes a = p \circ a$ . On the other hand, the following is required:

$$p \boxtimes q := pq, \quad \forall p, q \in \mathbf{Event}(I \rightarrow I), \quad (6)$$

where  $pq$  is the usual multiplication on real numbers. Moreover, an experimenter is allowed to perform a test  $T_X^{A \rightarrow B}$  disregarding the different outcomes within an arbitrary subset  $Y \subseteq X$ . This amounts to merging the events in  $Y$  into a single event. Such a possibility is captured by the notion of *coarse graining*. 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  $T_X^{A \rightarrow B}$  and every subset  $Y \subseteq X$ , there exists a coarse-grained event formally given by

$$\mathcal{T}_Y := \sum_{y \in Y} \mathcal{T}_y, \quad (7)$$

where sequential and parallel composition distribute over sums. The coarse-graining operation (7) clearly boils down to the usual sum over real when  $A = I = B$ .

Now, let  $\Theta$  be an OPT. For all systems  $A, B \in \mathbf{Sys}(\Theta)$ , we define the following equivalence relation for every  $\mathcal{T}_1, \mathcal{T}_2 \in \mathbf{Event}(A \rightarrow B)$ :  $\mathcal{T}_1 \sim \mathcal{T}_2$  if for all  $E \in \mathbf{Sys}(\Theta)$ ,  $\rho \in \mathbf{Event}(I \rightarrow AE)$ ,  $a \in \mathbf{Event}(BE \rightarrow I)$  one has

$$(8)$$

The members of the quotient class

$$\mathbf{Transf}(A \rightarrow B) := \mathbf{Event}(A \rightarrow B) / \sim$$

are called the *transformations from A to B*. Quotient classes of preparations  $\mathbf{St}(A) := \mathbf{Transf}(I \rightarrow A)$  are called *the states of A*, while those of observations  $\mathbf{Eff}(A) := \mathbf{Transf}(A \rightarrow I)$  *the effects of A*. Finally, equivalence classes of tests are called *instruments*, and are collected by the quotient class  $\mathbf{Instr}(\Theta)$ . We will often denote the states of A as  $|\rho\rangle_A$ , and the effects of A as  $\langle a|_A$ . In the case of the parallel composition of states or effects, we will safely omit the symbol  $\boxtimes$  without loss of clarity, simply writing  $|\rho_1\rangle_{A_1} |\rho_2\rangle_{A_2}$  or  $\langle a_1|_{A_1} \langle a_2|_{A_2}$ , while the sequential composition between a state and an effect will be given by the pairing  $\langle a|\rho\rangle_A := \langle a|_A \circ |\rho\rangle_A \in [0, 1]$ .

*Remark 1.* We observe that an arbitrary transformation  $\mathcal{T} \in \mathbf{Transf}(A \rightarrow B)$  is indeed defined by the entire class of transformations  $\{\mathcal{T} \boxtimes \mathcal{S}_E\}_{E \in \mathbf{Sys}(\Theta)}$ . This means that, given two transformations  $\mathcal{T}_1, \mathcal{T}_2 \in \mathbf{Transf}(A \rightarrow B)$ , the following holds:

$$\mathcal{T}_1 = \mathcal{T}_2 \iff \mathcal{T}_1 \boxtimes \mathcal{S}_E = \mathcal{T}_2 \boxtimes \mathcal{S}_E, \quad \forall E \in \mathbf{Sys}(\Theta). \quad (9)$$

*Remark 2.* Two tests associated with the same probability distribution in every possible experiment are said to be *operationally equivalent*. Importantly, in the construction and characterization of an OPT, one is interested in equivalence classes of tests, namely, in instruments. Thus, as it will be explicitly shown in Secs. V and VI, an OPT can be defined by specifying (i) the systems  $\mathbf{Sys}(\Theta)$ , (ii) a parallel composition rule  $\boxtimes$  for systems and states, and (iii) the instruments  $\mathbf{Instr}(\Theta)$  and their parallel composition  $\boxtimes$ .

Now, say that we have a preparation with output system A, an observation with input system B, and in-between a test from A to B: this scenario corresponds to a joint probability distribution of the events that may occur in the experiment. More generally, the goal of a physical theory is to associate a probability with each event of a composite test as the above-described one. In particular,

$$(10)$$

where  $\rho_x, \mathcal{T}_y, a_z$  are transformations of, respectively, the instruments  $\rho_X^{I \rightarrow A}, T_Y^{A \rightarrow B}, a_Z^{B \rightarrow I}$ . The above definition shows the parametric dependence of the joint probability distribution of outcomes on the whole closed circuit representing a test  $I \rightarrow I$ . Along with condition (6), Eq. (10) amounts to say that disconnected circuits represent statistically independent processes or experiments.

By virtue of Eq. (8), effects separate states and, *vice versa*, states separate effects. This means that, for every pair of states  $|\rho\rangle_A, |\sigma\rangle_A \in \mathbf{St}(A)$  such that  $|\rho\rangle_A \neq |\sigma\rangle_A$ , there exists an effect  $\langle a|_A \in \mathbf{Eff}(A)$  such that  $\langle a|\rho\rangle_A \neq \langle a|\sigma\rangle_A$  (and vice versa for every pair of different effects). The latter amounts to say that, given two different states (or effects), there exists an experiment producing different statistics for them. States (effects) can be seen as functionals from effects (states) to probabilities. Consequently, in the light of definitions (5), (6), and (7), one can take linear combinations of them. In particular, the pairing between states and effects defines a complete class of linearly independent vectors in  $\mathbf{St}(A)$ , spanning a real vector space  $\mathbf{St}_{\mathbb{R}}(A) := \mathbf{Span}_{\mathbb{R}} \mathbf{St}(A)$ . Similarly,  $\mathbf{Eff}(A)$  is a class of non-negative linear functionals on  $\mathbf{St}(A)$ , and spans the dual space  $\mathbf{Eff}_{\mathbb{R}}(A) := \mathbf{Span}_{\mathbb{R}} \mathbf{Eff}(A) = \mathbf{St}_{\mathbb{R}}(A)^\vee$ .

The members of  $\text{St}_{\mathbb{R}}(A)$  and  $\text{Eff}_{\mathbb{R}}(A)$  are called, respectively, *generalized states* and *generalized effects*. The dimension  $D_A := \dim \text{St}_{\mathbb{R}}(A)$  is called *the size (or dimension) of the system A*. In the following, every system A will be always thought as accompanied by its associated size  $D_A$ . A *type of system* is an infinite collection of systems with a given size  $D$ . In an OPT,  $D_A$  is the minimum number of probabilities that must be ascertained to determine the state of a system A. For instance, in QT, for a system A whose associated Hilbert space has dimension  $d_A$ , one has  $D_A = d_A^2$ . In this work we will restrict to the case of *finite-dimensional* OPTs, namely, theories  $\Theta$  where  $D_A < +\infty$  for all  $A \in \text{Sys}(\Theta)$ . In this case, for all systems A one has  $\dim \text{St}_{\mathbb{R}}(A) = \dim \text{Eff}_{\mathbb{R}}(A)$ .

Every transformation  $\mathcal{T}$  from A to B maps  $\text{St}(A)$  to  $\text{St}(B)$ . Accordingly, transformations from A to B span a real vector space  $\text{Transf}_{\mathbb{R}}(A \rightarrow B) := \text{Span}_{\mathbb{R}} \text{Transf}(A \rightarrow B)$ , whose members are called *generalized transformations*. Given  $A, B, C, D \in \text{Sys}(\Theta)$ , the transformations of the form  $\mathcal{T} \boxtimes \mathcal{T}' \in \text{Transf}(AC \rightarrow BD)$  are called *local transformations*. The transformations of the form  $\mathcal{T} \boxtimes \mathcal{I}_C \in \text{Transf}(AC \rightarrow BC)$  and  $\mathcal{I}_A \boxtimes \mathcal{T}' \in \text{Transf}(AC \rightarrow AD)$  are called, respectively, *local transformations from the system A to the system B* and *from the system C to the system D*. A transformation  $\mathcal{R} \in \text{Transf}(A \rightarrow B)$  is called *reversible* if there exists a transformation  $\mathcal{R}^{-1} \in \text{Transf}(B \rightarrow A)$  such that  $\mathcal{R}^{-1}\mathcal{R} = \mathcal{I}_A$  and  $\mathcal{R}\mathcal{R}^{-1} = \mathcal{I}_B$ . The class of reversible transformations of a theory  $\Theta$  will be denoted by  $\text{RevTransf}(\Theta)$ .

Now, for all  $p \in \text{St}(I)$  and  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$ , we will write the following:

$$\begin{array}{c} A \\ \hline p \mathcal{T} \\ \hline B \end{array} = p \begin{array}{c} A \\ \hline \mathcal{T} \\ \hline B \end{array}, \quad (11)$$

where  $p\mathcal{T}$  represents the scalar multiplication on the vector space  $\text{Transf}_{\mathbb{R}}(A \rightarrow B)$ . One has that  $\text{St}_{\mathbb{R}}(I) \cong \mathbb{R}$ , implying  $D_I = 1$ . Then, for all  $p \in \text{St}(I)$ ,  $A, B \in \text{Sys}(\Theta)$ , and  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$ , one has  $I \boxtimes A = I \otimes A = A = A \boxtimes I = A \otimes I$  and  $p \boxtimes \mathcal{T} = p \otimes \mathcal{T}$ , where  $\otimes$  is the standard tensor product. The identity of the trivial system is clearly given by the unit, i.e.,  $\mathcal{I}_I = 1$ . Finally, the braiding  $\mathcal{S}$  satisfies the following relation:

$$\mathcal{S}_{A,I} = \mathcal{I}_A = \mathcal{S}_{A,I}^{-1}, \quad \forall A \in \text{Sys}(\Theta). \quad (12)$$

The transformations from a system A to a system B are contained in a convex and bounded subset of  $\text{Transf}_{\mathbb{R}}(A \rightarrow B)$  [22]. Then, it is often convenient to consider the convex cone generated by the convex combinations of transformations in  $\text{Transf}(A \rightarrow B)$ . The latter will be denoted by  $\text{Transf}_+(A \rightarrow B)$ ; clearly,  $\text{St}_+(A)$  and  $\text{Eff}_+(A)$  will be used in the case of the convex cones generated by, respectively,  $\text{St}(A)$  and  $\text{Eff}(A)$ . A transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is called *atomic* if, given  $\mathcal{T}_1, \mathcal{T}_2 \in \text{Transf}(A \rightarrow B)$ , one has the following implication:

$$\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2 \Rightarrow \mathcal{T}_1 \propto \mathcal{T}_2. \quad (13)$$

A transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is called *extremal* if, given  $\mathcal{T}_1, \mathcal{T}_2 \in \text{Transf}(A \rightarrow B)$  and  $p \in (0, 1)$ , the condition  $\mathcal{T} = p\mathcal{T}_1 + (1 - p)\mathcal{T}_2$  implies  $\mathcal{T}_1 = \mathcal{T}_2$ . On the other hand, we call a *refinement* of  $\mathcal{T}$  a set  $\{\mathcal{T}_i\}_{i \in \mathcal{R}} \subseteq \text{Transf}(A \rightarrow B)$  such that  $\mathcal{T} = \sum_{i \in \mathcal{R}} \mathcal{T}_i$ , and similarly we define a *convex refinement* of  $\mathcal{T}$  a set  $\{\mathcal{T}_i\}_{i \in \mathcal{R}} \subseteq \text{Transf}(A \rightarrow B)$  such that

$\mathcal{T} = \sum_{i \in \mathcal{R}} p_i \mathcal{T}_i$  for some probability distribution  $\{p_i\}_{i \in \mathcal{R}}$ . Finally,  $\text{Ref}(\mathcal{T})$  and  $\text{ConvRef}(\mathcal{T})$  will denote the union of, respectively, all the refinements and all the convex refinements of  $\mathcal{T}$ .

We assume that  $0 \in \text{Transf}(I \rightarrow I)$ , so that events associated with a zero probability can be considered in a theory. Accordingly, the existence of parallel composition implies that the null generalized transformation  $\varepsilon_{A \rightarrow B} \in \text{Transf}_{\mathbb{R}}(A \rightarrow B)$ , such that  $(a|_{\text{BE}}(\varepsilon_{A \rightarrow B} \boxtimes \mathcal{I}_E)|\rho)_{\text{AE}} = 0$  for every  $E \in \text{Sys}(\Theta)$ ,  $|\rho\rangle_{\text{AE}} \in \text{St}(\text{AE})$ ,  $(a|_{\text{BE}} \in \text{Eff}(\text{BE}))$ , is included in the transformations  $\text{Transf}(A \rightarrow B)$  for all A, B. A null transformation always occurs with null marginal probability. A transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  such that there exists a singleton instrument  $\mathcal{T}_* = \{\mathcal{T}\} \in \text{Instr}(A \rightarrow B)$  is called *deterministic*. In continuity with the past literature, we will often call *channels* those deterministic transformations which are not states or effects. Oppositely to the case of null transformations, a deterministic physical process happens with certainty, i.e., with marginal 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. We will denote by  $\text{Transf}_1(A \rightarrow B)$  the class of deterministic transformations from system A to system B; clearly,  $\text{St}_1(A)$  and  $\text{Eff}_1(A)$  will be used in the case of, respectively, deterministic states and effects of system A. For any given instrument  $\mathcal{T}_X^{A \rightarrow B}$ , one has clearly that  $\sum_{x \in X} \mathcal{T}_x \in \text{Transf}_1(A \rightarrow B)$ , and, as a particular case,  $\text{Transf}_1(I \rightarrow I) = \{\mathcal{I}_I\} = \{1\}$ . In the light of the definition of null transformation, we will assume that  $\mathcal{T}_X^{A \rightarrow B} \in \text{Instr}(\Theta)$  if and only if  $\varepsilon_{A \rightarrow B} \cup \mathcal{T}_X^{A \rightarrow B} \in \text{Instr}(\Theta)$ . The reason why this is a convenient assumption is the following characterization of the deterministic transformations. If  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is a deterministic transformation,  $\mathcal{T} \boxtimes \mathcal{I}_E$  clearly maps  $\text{St}_1(\text{AE})$  to  $\text{St}_1(\text{BE})$  for all E. Conversely, let  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  be such that  $\mathcal{T} \boxtimes \mathcal{I}_E$  maps  $\text{St}_1(\text{AE})$  to  $\text{St}_1(\text{BE})$  for all E. Suppose now that there exists an instrument  $\mathcal{T}_{\mathcal{K} \cup \star}^{A \rightarrow B}$  of the form  $\mathcal{T}_{\mathcal{K} \cup \star}^{A \rightarrow B} = \{\mathcal{T}_k\}_{k \in \mathcal{K}} \cup \{\mathcal{T}\}$  with  $\mathcal{T}_k \in \text{Transf}(A \rightarrow B)$  for every  $k \in \mathcal{K}$ . Let  $\{|\rho_j\rangle_{\text{AE}}\}_{j \in \mathcal{J}}$  and  $\{(a|_{\text{BE}})_{l \in \mathcal{L}}\}$  be, respectively, a preparation instrument of AE and an observation instrument of BE. By coarse graining,  $|\rho\rangle_{\text{AE}} := \sum_{j \in \mathcal{J}} |\rho_j\rangle_{\text{AE}}$  and  $(\tilde{e}|_{\text{B}} := \sum_{l \in \mathcal{L}} (a|_{\text{BE}}$  are deterministic. Then, by hypothesis,  $(\mathcal{T} \boxtimes \mathcal{I}_E)|\rho\rangle_{\text{AE}} \in \text{St}_1(\text{BE})$  also holds. Accordingly,  $\text{Transf}_1(I \rightarrow I) \ni (\tilde{e}|_{\text{BE}}(\mathcal{T} \boxtimes \mathcal{I}_E)|\rho)_{\text{AE}} = 1$ . On the other hand, it must also be

$$(\tilde{e}|_{\text{BE}}(\mathcal{T} \boxtimes \mathcal{I}_E)|\rho)_{\text{AE}} + \sum_{k \in \mathcal{K}} (\tilde{e}|_{\text{BE}}(\mathcal{T}_k \boxtimes \mathcal{I}_E)|\rho)_{\text{AE}} = 1,$$

which in turn implies

$$\sum_{l \in \mathcal{L}, k \in \mathcal{K}, j \in \mathcal{J}} (a|_{\text{BE}}(\mathcal{T}_k \boxtimes \mathcal{I}_E)|\rho_j)_{\text{AE}} = 0,$$

with  $(a|_{\text{BE}}(\mathcal{T}_k \boxtimes \mathcal{I}_E)|\rho_j)_{\text{AE}} \geq 0$  for every  $l \in \mathcal{L}, k \in \mathcal{K}, j \in \mathcal{J}$  by definition. Accordingly, since the previous argument does not depend on the choice of instruments  $\{|\rho_j\rangle_{\text{AE}}\}_{j \in \mathcal{J}}$  and  $\{(a|_{\text{BE}})_{l \in \mathcal{L}}\}$ , we conclude that it must be  $\mathcal{T}_k = \varepsilon_{A \rightarrow B}$  for all  $k \in \mathcal{K}$ , namely,  $\mathcal{T} \in \text{Transf}_1(A \rightarrow B)$ . Thus, a transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is deterministic if and only if  $\mathcal{T} \boxtimes \mathcal{I}_E$  maps  $\text{St}_1(\text{AE})$  to  $\text{St}_1(\text{BE})$  for all systems E. By the same argument as above, one can show that, equivalently, a

transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is deterministic if and only if  $\mathcal{T} \boxtimes \mathcal{I}_E$  maps  $\text{Eff}_1(AE)$  to  $\text{Eff}_1(BE)$  for all systems  $E$ . The property of being deterministic is clearly preserved under both sequential and parallel composition.

As a first consequence, it is clear why in Sec. III A the identity process  $\mathcal{I}$  has been, by definition, associated with a family of singleton tests. In the remainder of this work, we will make extensive use of the above characterization, which we now use in order to prove the following useful result.

*Proposition 1.* Let  $\Theta$  be an OPT, and  $\mathcal{R} \in \text{RevTransf}(A \rightarrow A')$  be a reversible transformation from a system  $A \in \text{Sys}(\Theta)$  to a system  $A' \in \text{Sys}(\Theta)$ . Then,  $\mathcal{R} \in \text{Transf}_1(A \rightarrow A')$ , and preserves both atomicity and extremality via sequential composition. In particular,  $\text{RevTransf}(A \rightarrow A)$  is a group of permutations on  $\text{PurSt}(A)$ .

*Proof.* By hypothesis, there exists  $\mathcal{R}^{-1} \in \text{RevTransf}(A' \rightarrow A)$  such that  $\mathcal{R}^{-1} \circ \mathcal{R} = \mathcal{I}_A$  holds. By definition, there also exist  $\mathcal{D} \in \text{Transf}_1(A \rightarrow A')$  and  $\mathcal{D}' \in \text{Transf}_1(A' \rightarrow A)$  such that  $\mathbb{T} := \{\mathcal{R}, \mathcal{D} - \mathcal{R}\}$  and  $\mathbb{T}' := \{\mathcal{R}^{-1}, \mathcal{D}' - \mathcal{R}^{-1}\}$  are instruments of  $\Theta$ . Their sequential composition  $\mathbb{T}' \circ \mathbb{T} = \{\mathcal{I}_A, \mathcal{R}^{-1} \mathcal{D} - \mathcal{I}_A, \mathcal{D}' \mathcal{R} - \mathcal{I}_A, \mathcal{D}' \mathcal{D} - \mathcal{D}' \mathcal{R} - \mathcal{R}^{-1} \mathcal{D} + \mathcal{I}_A\}$  is also an instrument. Since  $\mathcal{I}_A$  is associated with a singleton instrument by definition (see Sec. III A),  $\mathcal{I}_A \in \text{Transf}_1(A \rightarrow A)$  holds. Accordingly, by direct inspection of  $\mathbb{T}' \circ \mathbb{T}$  and using the characterization of the deterministic transformations, one must have  $\mathcal{D} = \mathcal{R}$ , namely,  $\mathcal{R} \in \text{Transf}_1(A \rightarrow A')$ . We now just prove that if  $\mathcal{A} \in \text{Transf}(C \rightarrow A)$  is atomic, then  $\mathcal{R}\mathcal{A}$  is also atomic. The case of  $\mathcal{B} \in \text{Transf}(A' \rightarrow C)$  atomic implies  $\mathcal{B}\mathcal{R}$  atomic will be then a straightforward adaptation of the former case; on the other hand, preservation of extremality via sequential composition can be proven in an analogous manner, just by using the definition of extremal transformation instead of that of atomic one. Let then  $\mathcal{A} \in \text{Transf}(C \rightarrow A)$  be an atomic transformation. By contradiction, suppose that  $\mathcal{R}\mathcal{A}$  is not atomic. Then, by definition, there exist  $\mathcal{A}_1, \mathcal{A}_2 \in \text{Transf}(C \rightarrow A)$  with  $\mathcal{A}_1 \not\propto \mathcal{A}_2$  such that

$$\mathcal{R}\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 \implies \mathcal{A} = \mathcal{R}^{-1}\mathcal{A}_1 + \mathcal{R}^{-1}\mathcal{A}_2.$$

However, by atomicity of  $\mathcal{A}$ , the above equation implies  $\mathcal{R}^{-1}\mathcal{A}_1 \propto \mathcal{R}^{-1}\mathcal{A}_2$ , i.e.,  $\mathcal{A}_1 \propto \mathcal{A}_2$ , that is absurd. ■

By Proposition 1, it is clear why in Sec. III A also the braiding  $\mathcal{S}$  has been, by definition, associated with a family of singleton tests.

When  $\text{St}(A)$  coincides with its convex hull for every  $A \in \text{Sys}(\Theta)$ , the theory  $\Theta$  is called *convex*. We will denote by  $\text{ExtSt}_0(A)$  the set of extremal points of the convex hull of  $\text{St}(A)$ . Since  $|\varepsilon\rangle_A \in \text{ExtSt}_0(A)$  for every system  $A$ , we also define the set of non-null extremal points  $\text{ExtSt}(A) := \text{ExtSt}_0(A) \setminus \{|\varepsilon\rangle_A\}$ . The deterministic extremal states are those historically called *the pure states*. Deterministic states which are not extremal are those historically called *mixed states*. More generally, we will call *mixed* those states which are neither extremal nor atomic. This nomenclature is extended to arbitrary transformations: accordingly, *extremal channels* will be called *pure (transformations)*. We will denote by  $\text{PurSt}(A) \subseteq \text{ExtSt}(A)$  the set of pure states of a system  $A$ . In a convex theory, every mixture of pure states can be

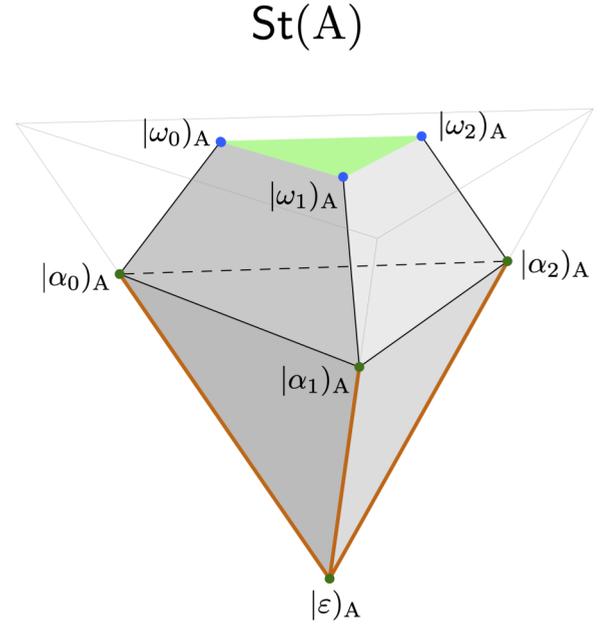


FIG. 1. Example of a set of states for a system  $A$  of size 3.  $\text{St}(A)$  is convex and contained in the complete state space of a classical trit, namely, a tetrahedron. The elements depicted in orange are the atomic ones, while those in dark green (the null state  $|\varepsilon\rangle_A$  and  $|\alpha_i\rangle_A$  for  $i \in \{0, 1, 2\}$ ) are both atomic and extremal. The elements in blue ( $|\omega_i\rangle_A$  for  $i \in \{0, 1, 2\}$ ) are the pure states of  $\text{St}(A)$ , which are extremal but not atomic. The elements in light green, i.e., those contained in the upper face, are the deterministic mixed states of  $A$ . Finally, all the remaining states (in gray) are mixed.

*deterministically prepared*. CT and QT are examples of convex theories.

*Remark 3.* One may argue that *atomic* should be used as a synonym of *pure* (see, e.g., Ref. [23]). However, in general we keep the two notions of atomicity and purity distinct. The only case where there is no need for a distinction between atomicity and purity, at least for states, is that of theories where every state is proportional to a deterministic one. Moreover, the theory presented in this work provides an explicit motivation to take the two notions distinct. Indeed, as it will be discussed in Sec. VII, in the case of transformations of the theory which are not states or effects, the property of purity is preserved by parallel composition, while the one of atomicity is not.

Figures 1 and 2 illustrate the main notions related to atomicity and convexity introduced in this section.

#### IV. OPERATIONAL PROBABILISTIC STRUCTURE: PRELIMINARY DEFINITIONS AND RESULTS

In this section, we will review some important properties in the scope of operational probabilistic theories, also proving a first relevant result related to the so-called *bilocal-tomographic theories*. Then, we will set and discuss the notions of no-restriction hypothesis and causality in the OPT framework. Finally, we will focus on the family of *simplicial theories*, which comprises the probabilistic theory introduced in this work.

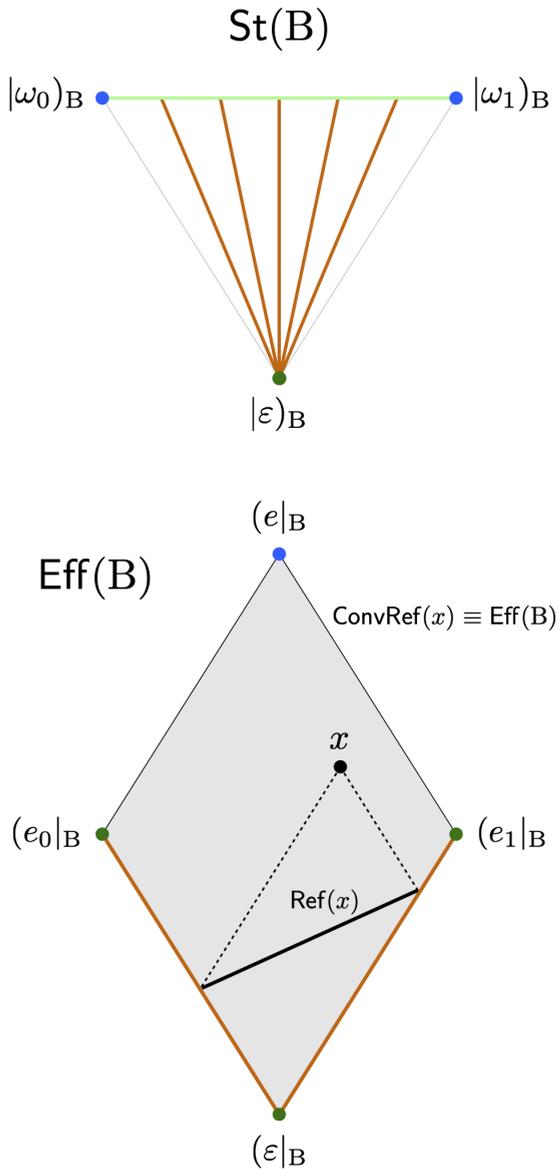


FIG. 2. Example of sets of states and effects of a system  $B$  of size 2.  $\text{St}(B)$  is contained in the complete state space of a classical bit, namely, a triangle, and is not convex. The colors are associated to the same properties as in Fig. 1. The atomic rays of  $\text{St}(B)$  provide the example of atomic elements which are not proportional to extremal ones. One has  $(e_i|\omega_j)_B = \delta_{ij}$  for every  $i, j \in \{0, 1\}$ .  $\text{Eff}(B)$  contains a unique deterministic effect, namely,  $(e|_B = (e_0|_B + (e_1|_B$ , which is pure and has a trivial convex refinement set. The refinement and convex refinement sets of a mixed (internal) point  $x \in \text{Eff}(B)$  are depicted.

### A. Parallel composition and properties of $n$ -local-tomographic theories

We first recall some definitions and properties which are relevant to this work. Let us start by defining the following map:

$$f : (A, B) \mapsto D_{AB} := f(A, B).$$

The above map  $f$  defines a rule for the size of composite systems. The following must always hold for all  $A, B, C \in \text{Sys}(\Theta)$ :

$$f(I, A) = D_{IA} = D_A = D_{AI} = f(A, I), \quad (14)$$

$$f(AB, C) = f(A, BC). \quad (15)$$

Furthermore, since a theory is endowed with a braiding, one has the additional constraint

$$f(A, B) = f(B, A). \quad (16)$$

We stress that when  $f(A, B) = g(D_A, D_B)$ , i.e., the dimension of the composite system depends only on the dimensions of the components, the map  $f$  boils down to an operation  $g$  on the dimensions of systems. In this case, the above conditions (14)–(16) take the form

$$g(D_A, 1) = g(1, D_A) = D_A,$$

$$g(g(D_A, D_B), D_C) = g(D_A, g(D_B, D_C)),$$

$$g(D_A, D_B) = g(D_B, D_A).$$

In general, one has the following inequality [8]:

$$D_{AB} \geq D_A D_B, \quad \forall A, B \in \text{Sys}(\Theta). \quad (17)$$

*Property 1 ( $n$ -local discriminability [8]).* Let  $n \leq m$ . The effects obtained as a conical 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 the  $m$ -partite states.

The above property was introduced in Ref. [21], where it was named  *$n$ -local tomography*. The case of  $n$ -local discriminability for  $n = 1$  is called *local discriminability*. A theory  $\Theta$  satisfies local discriminability if and only if

$$D_{AB} = D_A D_B, \quad \forall A, B \in \text{Sys}(\Theta) \quad (18)$$

(see Ref. [22] for a proof). In this case, for every composite system  $AB$ , the vector space  $\text{St}_{\mathbb{R}}(AB)$  is spanned by the *local (or product) states*, given by the parallel composition of single-system states. This is the case, e.g., of QT. Given two systems  $A$  and  $B$ , the *separable states* of the bipartite system  $AB$  are those of the form

$$|\sigma\rangle_{AB} = \sum_{i \in I} |\alpha_i\rangle_A |\beta_i\rangle_B, \quad (19)$$

where  $\text{St}(A) \ni |\alpha_i\rangle_A \neq |\varepsilon\rangle_A$ ,  $\text{St}(B) \ni |\beta_i\rangle_B \neq |\varepsilon\rangle_B$  for every  $i \in I$ . By negation, the *entangled states* are those that are *not separable*. Notice that the above definition is straightforwardly generalized to arbitrary transformations. If a theory does not satisfy local discriminability, then it necessarily has entangled states [8]. The converse is not true: QT is the example of a theory having entangled states and satisfying local discriminability.

Let now  $\Theta$  be an OPT. In the light of Eq. (17), we can define the non-negative integer quantity

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

The “excess”  $\Delta_{AB}^{(2)}$  is the dimension of any subspace  $\mathcal{B}_{AB}$  of  $\text{St}_{\mathbb{R}}(AB)$  which is linearly independent of the span of the separable states, denoted by  $\mathcal{S}_{AB}$ , and such that every element

of  $\text{St}_{\mathbb{R}}(AB)$  can be uniquely decomposed as the sum of an element of  $\mathcal{S}_{AB}$  and one of  $\mathcal{B}_{AB}$ . In formula

$$\text{St}_{\mathbb{R}}(AB) = \mathcal{S}_{AB} \oplus \mathcal{B}_{AB}.$$

From now on, we will choose one space  $\mathcal{B}_{AB}$ , where  $\Delta_{AB}^{(2)} = \dim \mathcal{B}_{AB}$ . For the sake of clarity, we will adopt the following convenient diagrammatic notation:

$$\begin{array}{c} \leftarrow \\ \leftarrow \end{array} := \mathcal{S}_{AB}, \quad \left( \begin{array}{c} \leftarrow \\ \leftarrow \end{array} \right) := \mathcal{B}_{AB}.$$

For any tripartite system ABC, we can introduce the subspaces

$$\mathcal{S}_{ABC}^{(1)} := \begin{array}{c} \leftarrow \\ \leftarrow \\ \leftarrow \end{array}, \quad \mathcal{S}_{ABC}^{(2)} := \left( \begin{array}{c} \leftarrow \\ \leftarrow \\ \leftarrow \end{array} \right), \quad \mathcal{S}_{ABC}^{(3)} := \left( \begin{array}{c} \leftarrow \\ \leftarrow \\ \leftarrow \end{array} \right), \quad \mathcal{S}_{ABC}^{(4)} := \left( \begin{array}{c} \leftarrow \\ \leftarrow \\ \leftarrow \end{array} \right). \quad (20)$$

Let us then define  $I := \{1, 2, 3, 4\}$ , and

$$\mathcal{S}_{ABC}^{(I)} := \text{Span}_{\mathbb{R}} \bigcup_{i \in I} \mathcal{S}_{ABC}^{(i)}.$$

Then,  $\text{St}_{\mathbb{R}}(ABC)$  is completely spanned by  $\mathcal{S}_{ABC}^{(I)}$  along with any subspace  $\mathcal{T}_{ABC}$  of  $\text{St}_{\mathbb{R}}(ABC)$  which is linearly independent of  $\mathcal{S}_{ABC}^{(I)}$ , and such that every element of  $\text{St}_{\mathbb{R}}(ABC)$  can be uniquely decomposed as the sum of an element of  $\mathcal{S}_{ABC}^{(I)}$  and one of  $\mathcal{T}_{ABC}$ . In formula

$$\text{St}_{\mathbb{R}}(ABC) = \mathcal{S}_{ABC}^{(I)} \oplus \mathcal{T}_{ABC}.$$

We will arbitrarily choose  $\mathcal{T}_{ABC}$  once for all, and shall denote it diagrammatically as follows:

$$\left( \begin{array}{c} \leftarrow \\ \leftarrow \\ \leftarrow \end{array} \right) := \mathcal{T}_{ABC}.$$

The dimension of  $\mathcal{T}_{ABC}$  will be denoted by  $\Delta_{ABC}^{(3)} := \dim \mathcal{T}_{ABC}$ . We are now in position to prove the following general result.

*Theorem 1 (Composition rules for system sizes in arbitrary theories).* Let  $\Theta$  be an OPT, and ABC any tripartite system in  $\Theta$ . Then, the following identity holds:

$$D_{ABC} = D_A D_B D_C + \Delta_{AB}^{(2)} D_C + \Delta_{BC}^{(2)} D_A + \Delta_{AC}^{(2)} D_B + \Delta_{ABC}^{(3)}. \quad (21)$$

*Proof.* From linear independence of  $\mathcal{S}_{AB}$  and  $\mathcal{B}_{AB}$ , combined with the fact that effects separate states, we conclude that for every  $|\sigma\rangle_{AB} \in \mathcal{S}_{AB}$  and for every  $|\beta\rangle_{AB} \in \mathcal{B}_{AB}$  there exist generalized effects  $(b|_{AB}, (b'|_{AB} \in \text{Eff}_{\mathbb{R}}(AB)$  such that  $(b|\sigma)_{AB} = 1$  and  $(b|\beta')_{AB} = 0$  for all  $|\beta'\rangle_{AB} \in \mathcal{B}_{AB}$ , and similarly  $(b'|\beta)_{AB} = 1$  and  $(b'|\sigma')_{AB} = 0$  for all  $|\sigma'\rangle_{AB} \in \mathcal{S}_{AB}$ . Using the above property, combined with the facts that every  $|\sigma\rangle_{AB} \in \mathcal{S}_{AB}$  is a linear combination of separable states, and that states separate effects, one has also that  $(b'|_{AB}|\rho\rangle_A = (\varepsilon|_B$  and  $(b'|_{AB}|\rho'\rangle_B = (\varepsilon|_A$  for all  $|\rho\rangle_A \in \text{St}_{\mathbb{R}}(A), |\rho'\rangle_B \in \text{St}_{\mathbb{R}}(B)$ . Moreover, by construction,  $(b'|_{AB}$  cannot be a

combination of separable effects. Diagrammatically,

$$\begin{array}{c} \leftarrow \\ \leftarrow \end{array} := (b'|_{AB}, \quad \left( \begin{array}{c} \leftarrow \\ \leftarrow \end{array} \right) = 0, \\ \Rightarrow \quad \left( \begin{array}{c} \leftarrow \\ \leftarrow \end{array} \right) = (\varepsilon|_B, \quad \left( \begin{array}{c} \leftarrow \\ \leftarrow \end{array} \right) = (\varepsilon|_A. \quad (22)$$

The four subspaces  $\mathcal{S}_{ABC}^{(i)}$  for  $i \in I = \{1, 2, 3, 4\}$  are spanned by the (tripartite) product states, and their dimensions are given by

$$\dim \mathcal{S}_{ABC}^{(1)} = D_A D_B D_C, \quad \dim \mathcal{S}_{ABC}^{(2)} = \Delta_{AB}^{(2)} D_C, \\ \dim \mathcal{S}_{ABC}^{(3)} = \Delta_{BC}^{(2)} D_A, \quad \dim \mathcal{S}_{ABC}^{(4)} = \Delta_{AC}^{(2)} D_B.$$

Now, it is easy to show that the five subspaces  $\mathcal{S}_{ABC}^{(i)}$  for  $i \in I$  and  $\mathcal{T}_{ABC}$  are separated by effects, and then in fact they are linearly independent. This can be done using the following argument. Let us pick a complete linearly independent set  $\{|\alpha_j^{(5)}\rangle_{ABC}\}_{j \in J_5}$  in  $\mathcal{T}_{ABC}$ , where  $J_5 := \{1, 2, \dots, \Delta_{ABC}^{(3)}\}$ . By linear independence of  $\mathcal{T}_{ABC}$  and  $\mathcal{S}_{ABC}^{(I)}$ , combined with the fact that effects separate states, one can construct a set of generalized effects  $\{(a_j^{(5)}|_{ABC})_{j \in J_5} \subseteq \text{Eff}_{\mathbb{R}}(ABC)$  such that  $(a_j^{(5)}|\alpha_k^{(5)})_{ABC} = \delta_{jk}$  while  $(a_j^{(5)}|\sigma)_{ABC} = 0$  for all  $|\sigma\rangle_{ABC} \in \mathcal{S}_{ABC}^{(I)}$ . Let us now pick a complete linearly independent set  $\{|\alpha_j^{(i)}\rangle_{ABC}\}_{j \in J_i}$  in every subspace  $\mathcal{S}_{ABC}^{(i)}$ , where  $J_i := \{1, 2, \dots, \dim \mathcal{S}_{ABC}^{(i)}\}$ . By construction, every  $|\rho\rangle_{ABC} \in \text{St}_{\mathbb{R}}(ABC)$  can be written as

$$|\rho\rangle_{ABC} = \sum_{i=1}^5 \sum_{j \in J_i} r_j^{(i)} |\alpha_j^{(i)}\rangle_{ABC}.$$

Let us now suppose that  $|\rho\rangle_{ABC} = |\varepsilon\rangle_{ABC}$ . Then, we have

$$0 = (a_j^{(5)}|\rho)_{ABC} = r_j^{(5)}, \quad \forall j \in J_5,$$

namely,

$$|\rho\rangle_{ABC} = \sum_{i=1}^4 \sum_{j \in J_i} r_j^{(i)} |\alpha_j^{(i)}\rangle_{ABC}.$$

Now, using formulas (20) and (22), for all  $i \in I$  one can construct a set of generalized effects  $(a_j^{(i)}|_{ABC} \in \text{Eff}_{\mathbb{R}}(ABC)$  such that  $(a_j^{(i)}|\alpha_k^{(i)})_{ABC} = \delta_{jk}$  and  $(a_j^{(i)}|\alpha_{k'}^{(l)})_{ABC} = 0$  for all  $i, l \in I$  with  $l \neq i$ , all  $j, k \in J_i$ , and all  $k' \in J_l$ . Thus, we have

$$0 = (a_j^{(i)}|\rho)_{ABC} = r_j^{(i)}, \quad \forall i \in I, \forall j \in J_i.$$

Accordingly,  $|\rho\rangle_{ABC} = |\varepsilon\rangle_{ABC}$  if and only if  $r_j^{(i)} = 0$  for all  $1 \leq i \leq 5$  and all  $j \in J_i$ . This shows that the five subspaces  $\mathcal{S}_{ABC}^{(i)}$  for  $i \in I$  and  $\mathcal{T}_{ABC}$  are not only complete, but also linearly independent. Then, in general Eq. (21) holds. ■

Notice that, for every positive integer  $n$ ,  $n$ -local discriminability implies  $(n+1)$ -local discriminability (see Property 1). For this reason, we will call the property of  $n$ -local discriminability, in the case where  $(n-1)$ -local discriminability does not hold, *strict  $n$ -local discriminability*. The case of  $n$ -local discriminability for  $n=2$  is called *bilocal discriminability*. Relevantly, strict bilocal discriminability is

a property satisfied by real quantum theory and fermionic quantum theory [21,24]. The next result provides the general form of the composition rules for the system dimensions in a theory satisfying bilocal discriminability.

*Theorem 2 (Composition rules for system sizes in bilocal-tomographic theories).* An OPT  $\Theta$  satisfies bilocal discriminability if and only if the following equality holds for any tripartite system  $ABC \in \text{Sys}(\Theta)$ :

$$D_{ABC} = D_A D_B D_C + \Delta_{AB}^{(2)} D_C + \Delta_{BC}^{(2)} D_A + \Delta_{AC}^{(2)} D_B \quad (23)$$

or, equivalently,  $\Delta_{ABC}^{(3)} = 0$ .

*Proof.* We recall Property 1. On the one hand, if Eq. (23) holds, for all  $A, B, C \in \text{Sys}(\Theta)$  the space  $\text{St}_{\mathbb{R}}(ABC)$  has the same dimension as  $\mathcal{S}_{ABC}^{(I)}$ . On the other hand, since  $\mathcal{S}_{ABC}^{(I)}$  is a subspace of  $\text{St}_{\mathbb{R}}(ABC)$ , being all its elements allowed by the composition rules of OPTs, it must be  $\text{St}_{\mathbb{R}}(ABC) \equiv \mathcal{S}_{ABC}^{(I)}$ . This implies that the collection of effects  $\bigcup_{i=1}^4 \{(a_j^i |)_{j \in J_i}\}$  introduced in the proof of Theorem 1 separates states in  $\text{St}_{\mathbb{R}}(ABC)$ , namely,  $\Theta$  satisfies Property 1 with  $n = 1$  or  $2$ , depending on whether  $\Delta_{XY}^{(2)} = 0$  or  $\Delta_{XY}^{(2)} \neq 0$ , respectively. On the other hand, if  $\Theta$  satisfies Property 1 with  $n = 1$  or  $2$ , then, by definition, the effects generated by the collection  $\bigcup_{i=1}^4 \{(a_j^i |)_{j \in J_i}\}$  separate states. By direct inspection of the collection  $\bigcup_{i=1}^4 \{(a_j^i |)_{j \in J_i}\}$ , and recalling that for all systems  $A$  one has  $D_A = \dim \text{Eff}_{\mathbb{R}}(A)$ , then Eq. (23) holds. We conclude that  $\Theta$  satisfies bilocal discriminability if and only if Eq. (23) holds or, equivalently [see Eq. (21)], if and only if  $\Delta_{ABC}^{(3)} = 0$ . ■

Equation (23) generalizes Eq. (18), holding for theories satisfying local discriminability. Indeed, when local discriminability holds in Eq. (23) one has  $\Delta_{XY}^{(2)} = 0$  for all  $X, Y \in \text{Sys}(\Theta)$ . Then, in the latter case, Eq. (23) represents just a restatement of Eq. (18) for the tripartite scenario. More importantly, when strict bilocal discriminability holds, implying  $\Delta_{XY}^{(2)} \neq 0$  for some  $X, Y \in \text{Sys}(\Theta)$ , the statement of Theorem 2 is nontrivial. Interestingly, Theorem 2 implies that the upper bound given in Refs. [21,24], holding for the dimension of tripartite systems in an OPT with bilocal discriminability, is in fact always saturated. In Ref. [24], a theory was defined to be *maximally bilocal tomographic* if it satisfies strict bilocal discriminability and Eq. (23). However, Theorem 2 states that if a theory satisfies strict bilocal discriminability, then it necessarily satisfies also Eq. (23), namely, *every theory satisfying strict bilocal discriminability is in fact maximally bilocal tomographic*.

### B. On the notions of no-restriction hypothesis and causality

The following two important properties, called *the no-restriction hypothesis* and *causality* [22], are extensively assumed throughout the literature on probabilistic theories.

*Property 2 (No-restriction hypothesis).* Let  $\Theta$  be an OPT,  $A, B \in \text{Sys}(\Theta)$ , and  $T_X \subset \text{Transf}_{\mathbb{R}}(A \rightarrow B)$  be a collection of generalized transformations. If, for all  $E \in \text{Sys}(\Theta)$ ,  $T_X \boxtimes |_{\star}^{E \rightarrow E}$  maps preparation instruments of  $AE$  to preparation instruments of  $BE$ , then  $T_X \in \text{Instr}(\Theta)$ , namely,  $T_X$  is an instrument of the theory  $\Theta$ .

For any given theory  $\Theta$ , one is always able to check whether  $\Theta$  satisfies Property 2, by definition. If a theory satisfies the no-restriction hypothesis, this means that it cannot be

extended, *in a consistent way*, by adding further transformations or effects.

A brief comment on the formulation of Property 2 is in order. The no-restriction hypothesis is usually formulated stating that “all the positive functionals on states are physical,” or something along these lines. First, we explicitly included also general transformations. Second, to make a sensible statement, in general one has to be sure that a collection of generalized transformations not only maps states to states, but also that this is the case when the collection is extended to a composite system. The previous observation is relevant in the absence of local discriminability (see Remark 1).

*Remark 4.* From a constructive perspective, namely, when a theory  $\Theta$  is constructed out of a number of postulates, one can use Property 2 as a postulate, and this is indeed very common in the literature on probabilistic theories. However, despite being a viable constructive postulate, in general the no-restriction hypothesis is not *selective*, in the sense that, by assigning the preparation instruments and imposing Property 2, one does not generally end up with a single theory, but with a *family of different theories*.<sup>2</sup> The only case where the no-restriction hypothesis can be safely taken as a selective postulate is when the choice of the preparation instruments satisfies local discriminability (which is in fact the case in vast majority of the literature on probabilistic theories). Indeed, in the latter case, the local states of every composite system  $AB$  generate the whole space  $\text{St}_{\mathbb{R}}(AB)$ , parallel composition being given by the standard tensor product, i.e.,  $\boxtimes \equiv \otimes$ . As a consequence, every transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow C)$  is solely defined by its action on  $\text{St}(A)$  (see Remark 1) and the braiding  $\mathcal{S}$  is *uniquely determined* by the mere geometry of states, so that  $\mathcal{T} \in \text{Transf}(A \rightarrow C)$  maps  $\text{St}(A)$  to  $\text{St}(C)$  if and only if  $\mathcal{T} \boxtimes \mathcal{I}_B$  maps  $\text{St}(AB)$  to  $\text{St}(CB)$  for all  $B \in \text{Sys}(\Theta)$ .

*Property 3 (Causality).* The probability distributions of preparation instruments do not depend on the choice of observation instrument at their output.

We remind that the above statement of causality is equivalent to the uniqueness of the deterministic effect for every system [22], namely, an OPT is causal if and only if every system has a unique deterministic effect. The unique deterministic effect is often called *the unit effect (or measure)* [25]. The causality condition in Property 3 is known as *no signaling from the future*. The consequences of this requirement have been extensively studied in Ref. [26]. Incidentally, it also implies another kind of no-signaling, namely, the *no signaling at a distance (without interaction)* [22], which is the standard assumption of *spatial no signaling* made in the spirit of nonlocal boxes [4]. Indeed, in a bipartite scenario for, say, system  $AB$ , the uniqueness of the deterministic effect always allows one to define a unique marginal state for both  $A$  and  $B$ .

<sup>2</sup>This is due to the fact that a class of generalized instruments may fail to be *uniquely determined* by the mere geometry of states. A significant example is given by the theory of superselected qubits and fermionic quantum theory [24], sharing the same class of preparation instruments and both satisfying the no-restriction hypothesis, despite being different theories.

*Property 4 (Conditional instruments).* For all systems  $A, B, C \in \text{Sys}(\Theta)$ , all instruments  $A_X \in \text{Instr}(A \rightarrow B)$ , and all choices of maps

$$\begin{aligned} X &\longrightarrow \text{Instr}(B \rightarrow C), \\ x &\longmapsto B_{Y(x)}^{(x)}, \end{aligned}$$

the following holds:

$$\bigcup_{x \in X} \{B_y^{(x)} \circ A_x\}_{y \in Y(x)} \in \text{Instr}(\Theta).$$

In other words, the generalized conditional instrument associated to every  $A_X \in \text{Instr}(A \rightarrow B)$  and every choice of labeled collection  $\{B_{Y(x)}^{(x)}\}_{x \in X} \subset \text{Instr}(B \rightarrow C)$  is an instrument of the theory.

Property 4 affects the geometry of instruments of a theory as follows. Suppose that an arbitrary instrument  $A_X \in \text{Instr}(A \rightarrow B)$  is performed. Property 4 guarantees that it is always possible to use any outcome  $x \in X$  of  $A_X$  in order to freely choose a second instrument  $B_{Y(x)}^{(x)} \in \text{Instr}(B \rightarrow C)$  to be performed *at the output* of the transformation  $A_x \in A_X$ . This has two major consequences [22,27]. In any theory satisfying Property 4, the existence of a probability which is not 0 or 1 implies that every convex combination of instruments is itself an instrument. As a consequence, the theory is convex. Moreover, Property 4 selects theories where signals can propagate *solely* in the input-output direction (see Property 3), namely, causality is implied. By the above reasons, in Ref. [27] it is argued that Property 4 should be considered as a (strong) notion of causality.

The no-restriction hypothesis and causality are often *simultaneously* assumed throughout the literature on probabilistic theories, being a defining part of the framework itself. However, we observe that these properties might be incompatible, even in the presence of local discriminability. Indeed, assuming the no-restriction hypothesis for a class of states may give rise, depending on their geometry, to a class of effects containing more than one deterministic effect, namely, to a theory which is not causal. The possibility of formulating probabilistic theories without the no-restriction hypothesis has been explored in the literature [25], whereas causality is therein implicitly assumed.

### C. Simplicial and classical theories

We conclude this section introducing some final definitions and results which are particularly relevant to the scope of this work.

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

Notice that a  $D_A$  simplex is the convex hull of  $D_A + 1$  affinely independent vertices, being in the present context the elements of  $\text{ExtSt}_0(A)$ , which includes the null state  $|\varepsilon\rangle_A$ .

*Property 5 (Joint perfect discriminability).* Let  $\Theta$  be an OPT and  $A \in \text{Sys}(\Theta)$ . A set of states  $\{|\rho_i\rangle_A\}_{i=1}^n$  is jointly perfectly discriminable if there exists an observation instrument  $\{(a_i|_A)\}_{i=1}^n$  such that

$$(a_i|\rho_{i'})_A = \delta_{ii'}, \quad \forall i, i' \in \{1, 2, \dots, n\}.$$

Notice that a necessary condition for a set of states  $\{|\rho_i\rangle_A\}_{i=1}^n$  to satisfy Property 5 is that  $\{|\rho_i\rangle_A\}_{i=1}^n \subseteq \text{St}_1(A)$ .

*Definition 2 (Classical theories).* A *classical theory* is a simplicial theory where the pure states of every system are jointly perfectly discriminable.

Simplicial theories are not necessarily convex. However, the set of states of a simplicial theory is relatively simple to treat since every state admits of a unique decomposition into non-null extremal states. Moreover, simplicial theories are inherently causal, and for such theories  $\text{ExtSt}(A) \equiv \text{PurSt}(A)$  for every system  $A$  (see Theorem 1 and its proof in Ref. [8]).

*Definition 3 (Classical theory).* *Classical theory (CT)* is the OPT  $\Theta$  satisfying the following properties: (i)  $\Theta$  is simplicial and convex, (ii) local discriminability holds, (iii) the preparation instruments of every system  $A \in \text{Sys}(\Theta)$  are all the collections of states of  $A$  that add up to a point in the convex hull of  $\text{PurSt}(A)$ , and (iv) the no-restriction hypothesis holds.

Notice that, by Remark 4, in the above case the no-restriction hypothesis singles out a unique theory. CT and QT, despite being very different theories, feature some relevant common properties, most notably causality and local discriminability. They also share convexity and the no-restriction hypothesis. Indeed, both CT and QT also satisfy a stronger property, known as (*strong*) *self-duality* [28]. This means that, for every system  $A$ , one has the equality  $\text{Eff}_+(A) = \text{St}_+(A)^\vee$  (using Riesz's representation).

We finally discuss another property which also holds in both CT and QT.

*Property 6 (Atomicity of parallel composition).* The parallel composition of two atomic transformations is atomic.

Analogously to the case of local discriminability, if a theory does not satisfy atomicity of *state composition*, then it necessarily has entangled states [8]. In the simplicial case, local discriminability and atomicity of *state composition* (provided that  $n$ -local discriminability is satisfied for some  $n$ ) are in fact equivalent [8]. Indeed, the simplicial theory presented in this paper satisfies strict bilocal discriminability, and not atomicity of parallel composition. In Appendix D we prove a structure theorem for the parallel composition rule of strictly bilocal simplicial theories satisfying a certain property on the reversible transformations (see Property 8 in Sec. VII B).

## V. CONSTRUCTING AN OPT

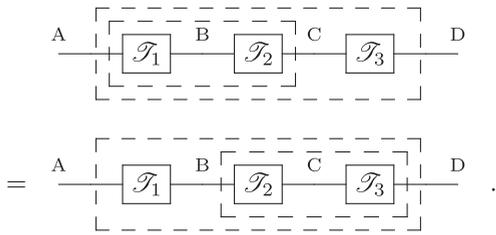
We begin the present section treating the consistency of an OPT. We will provide an account of the general consistency conditions which any OPT must comply with in order to be coherent and well posed. In the light of Remark 1, the careful assessment of coherence conditions, indeed, is particularly relevant in the absence of local discriminability, as it will be clear in the following. Then we shall present an exhaustive procedure to construct an OPT and check its consistency.

### A. Coherence and well posedness

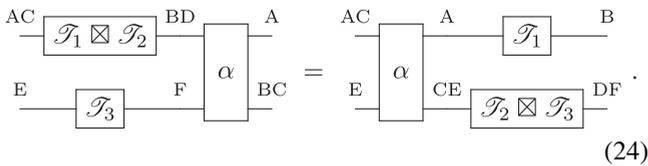
When a mathematical structure is constructed, one needs to make sure that it is consistent. Accordingly, this section will be devoted to present the consistency conditions that must be imposed on the construction of an OPT: as a category,

the OPT must abide by coherence conditions [11], while the further probabilistic structure must be well posed, and the two structures must be compatible.

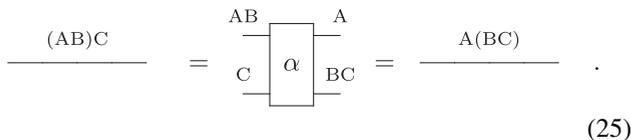
Both sequential and parallel composition of tests are required to be associative operations. The reason is that when an experimenter considers composed physical tests, the choice of a particular association is just a formal action, not corresponding to a physical operation. In an OPT, transformations are defined as maps between sets: two transformations are the same if they are represented by the same function. Accordingly, sequential composition coincides with *function composition*, which is always associative. Therefore, in the construction of an OPT, as long as transformations are defined as linear maps between vector spaces, the following always holds:



The previous argument does not apply to parallel composition. Accordingly, in general, when one chooses a rule for parallel composition, one has to assign an invertible map  $\alpha$  from  $(\text{Sys}(\Theta) \boxtimes \text{Sys}(\Theta)) \boxtimes \text{Sys}(\Theta)$  to  $\text{Sys}(\Theta) \boxtimes (\text{Sys}(\Theta) \boxtimes \text{Sys}(\Theta))$ , called *the associator*, such that

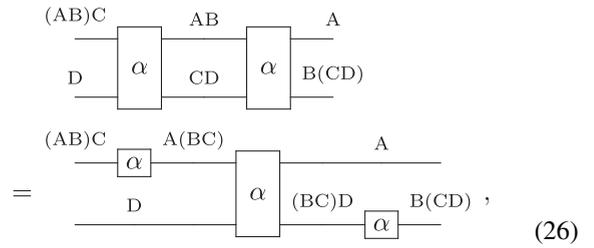


This map allows one to switch between different associations respecting the associativity condition. Namely, the associator is given by the identity transformation itself:



Now, from a constructive perspective, when an associative parallel composition rule is chosen, not only one has to assign an associator, but also one needs to check whether it is *coherent*, namely, self-consistent when extended to the composition of more than three objects. In order to verify this important requirement, it is sufficient to check the so-called *pentagon*

*identity*:<sup>3</sup>



for all  $A, B, C, D \in \text{Sys}(\Theta)$ , where one sequentially uses the appropriate assigned association rule. Equations (24) and (25), along with the pentagon identity (26), ensure that the operation of parallel composition is associative in a coherent way.

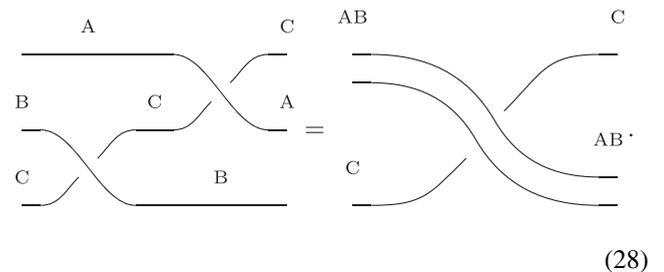
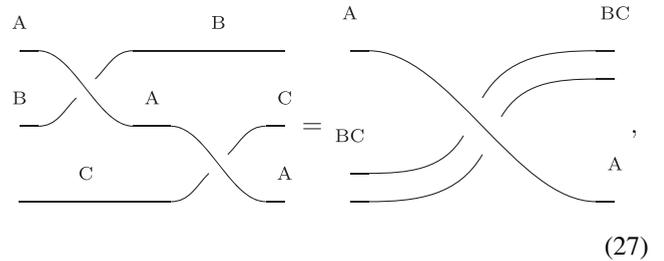
The second coherence requirement is related to the possibility of exchanging systems between agents. For instance, one demands that the single exchange

$$A(BC) \mapsto (BC)A$$

is equivalent, for operational consistency, to the two sequential exchanges

$$A(BC) \equiv (AB)C \mapsto (BA)C \equiv B(AC) \mapsto B(CA) \equiv (BC)A.$$

In other words, the braiding  $\mathcal{S}$  is required to compose as in the *braid group*. In order to verify this, it is sufficient to check the two so-called *hexagon identities*:<sup>4</sup>



for all  $A, B, C \in \text{Sys}(\Theta)$  (where we omitted an explicit graphical representation of the associator). However, in the case where the theory is *symmetric*, it is easy to verify that the two hexagon identities (27) and (28) are in fact equivalent [11].

<sup>3</sup>This nomenclature is due to the fact that the corresponding commutative diagram, generally holding for a nonstrict monoidal category, has five vertices (see, e.g., Ref. [11]).

<sup>4</sup>Analogously to the case of the pentagon identity, this terminology is related to the fact that the corresponding commutative diagrams have six vertices (see, e.g., Ref. [11]).

In a general category, one also needs to assign two suitably defined invertible maps  $\lambda : IA \mapsto A$  and  $\rho : AI \mapsto A$ , called, respectively, *the left* and *the right unitors* (see, e.g., Ref. [11]). Accordingly,  $\lambda$  and  $\rho$  must be not only well posed, but also abide by some coherence conditions involving the associator  $\alpha$  and the braiding  $\mathcal{S}$ . In fact, in the light of the linear structure derived in Sec. III B, in particular, by virtue of Eqs. (5), (11) and what follows: well posedness and coherence for the unitors are respected, and do not need to be checked.

We conclude with some final requirement of compatibility with the probabilistic structure. First, states must be separating for effects and vice versa. Instruments  $P_X \in \text{Instr}(I \rightarrow I)$  must be probability distributions, and the null transformation  $\varepsilon_{A \rightarrow B}$  must be included in  $\text{Transf}(A \rightarrow B)$  for every  $A, B$ . The coarse-graining operation must be well posed for every instrument. Finally, for compatibility with sequential composition, every instrument  $T_X^{A \rightarrow B} \boxtimes I_*^{E \rightarrow E}$  must map preparation instruments of  $AE$  to preparation instruments of  $BE$  for every system  $A, B, E$ .

We are now in position to present a possible procedure for the construction of an OPT  $\Theta$ , and for the check of its well posedness.

### B. Setting the postulates

In this section, we elaborate the procedure sketched in Remark 2. First, we can specify the class  $\text{Sys}(\Theta)$ , endowed with the binary composition operation  $(A, B) \mapsto AB \in \text{Sys}(\Theta)$  for each pair  $A, B \in \text{Sys}(\Theta)$ . A dimension  $D_A$  for the real vector space  $\text{St}_{\mathbb{R}}(A)$  and a class of states  $\text{St}(A) \subset \text{St}_{\mathbb{R}}(A)$  are associated with each system  $A \in \text{Sys}(\Theta)$ . Now, we can assign a composition rule for systems, choosing a map  $f : (A, B) \mapsto D_{AB}$  for each pair  $A, B \in \text{Sys}(\Theta)$ .

We can then proceed by choosing, for all  $A, B, E \in \text{Sys}(\Theta)$ , the action on  $\text{St}(AE)$  of the local transformations from  $A$  to  $B$ :

$$\frac{AE}{\text{---}} \boxed{\mathcal{T} \boxtimes \mathcal{S}_E} \frac{BE}{\text{---}} = \frac{A}{\text{---}} \boxed{\mathcal{T}} \frac{B}{\text{---}} \frac{E}{\text{---}}$$

These span a real vector space of linear functions from  $\text{St}_{\mathbb{R}}(AE)$  to  $\text{St}_{\mathbb{R}}(BE)$ . On the one hand, one should always include the identity family  $\mathcal{I}$ . On the other hand, the two elementary cases  $E = I$  and  $A = B = I$ , corresponding to scalar multiplication, have been already set in Sec. III B [see, in particular, Eq. (11)]. Then, we can proceed with the cases  $A = I$  and  $B \neq I$ , specifying the action of all the local transformations of the form  $|\sigma\rangle_B \boxtimes \mathcal{S}_E \in \text{Transf}(E \rightarrow BE)$  on all  $\rho \in \text{St}(E)$ , namely, a rule for composing states in parallel:

$$\frac{\rho}{\text{---}} \frac{\sigma}{\text{---}} \frac{E}{\text{---}} \frac{B}{\text{---}} \in \text{St}(BE).$$

This is done by choosing a decomposition of every product state into linearly independent vectors of the composite system. Since it must be  $\text{St}(((AB)C)) = \text{St}((A(BC)))$ , we need also to specify, by choosing an associator  $\alpha$  as in Eq. (25), the map identifying the linearly independent vectors in  $\text{St}_{\mathbb{R}}(((AB)C))$  with those in  $\text{St}_{\mathbb{R}}((A(BC)))$ . Also, we specify the action of all the local transformations of the form

$(a|_A \boxtimes \mathcal{S}_E \in \text{Transf}_{\mathbb{R}}(AE \rightarrow E))$  on all  $\Sigma \in \text{St}(AE)$ , namely, a rule allowing one to construct the conditioning of bipartite states:

$$\frac{\Sigma}{\text{---}} \frac{A}{\text{---}} \boxed{a} \frac{E}{\text{---}} \in \text{St}_{\mathbb{R}}(E).$$

This defines how the local generalized effects of every  $AB$  embed in  $\text{Eff}_{\mathbb{R}}(AB)$ , but not yet the class of effects, nor that of the observation instruments, of  $\Theta$ . A very common choice throughout the literature is to assume the no-restriction hypothesis in the form “all the positive functionals on states are physical” (see Remark 4).

Then, we can define the action of arbitrary  $\mathcal{T} \boxtimes \mathcal{S}_E \in \text{Transf}(AE \rightarrow BE)$  on  $\text{St}(AE)$  for all systems  $A, B, E \neq I$ , and thus set a braiding  $\mathbf{S}$  for the theory. Finally, we are left with specifying the instruments of  $\Theta$ . One possibility is to choose the preparation instruments of the theory, and then postulate that the class  $\text{Instr}(\Theta)$  includes a collection  $T_X^{A \rightarrow B}$  of transformations if and only if  $T_X^{A \rightarrow B} \boxtimes I_*^{E \rightarrow E}$  maps preparation instruments of  $AE$  to preparation instruments  $BE$  for all  $E \in \text{Sys}(\Theta)$ . Notice that such a requirement does not amount to postulating the no-restriction hypothesis (see Property 2 and Remark 4).

We are now in position to show that the above construction is sufficient to check whether  $\Theta$  is in fact a consistent OPT.

### C. Checking the consistency

By having set the postulates, we are now provided with the action of the associator  $\alpha$  and of all transformations, including the braiding  $\mathcal{S}$ . Accordingly, we can first derive some fundamental expressions which are needed to check whether  $\Theta$  is in fact a consistent OPT.

Sequential composition distributes over sums and is defined in the following way:

$$\frac{A}{\text{---}} \boxed{\mathcal{T}_2 \mathcal{T}_1} \frac{C}{\text{---}} \frac{E}{\text{---}} := \frac{A}{\text{---}} \boxed{\mathcal{T}_1} \frac{B}{\text{---}} \boxed{\mathcal{T}_2} \frac{C}{\text{---}} \frac{E}{\text{---}}, \quad (29)$$

for all systems  $A, B, C, E$  and all  $\mathcal{T}_1 \in \text{Transf}(A \rightarrow B)$ ,  $\mathcal{T}_2 \in \text{Transf}(B \rightarrow C)$ . The rule for extending local transformations can be derived by resorting to the action of the associator  $\alpha$ , namely, via the following identification:

$$\frac{AE}{\text{---}} \boxed{\mathcal{T} \boxtimes \mathcal{S}_E} \frac{BE}{\text{---}} \frac{D}{\text{---}} := \frac{AE}{\text{---}} \boxed{\alpha} \frac{A}{\text{---}} \boxed{\mathcal{T}} \frac{B}{\text{---}} \boxed{\alpha^{-1}} \frac{BE}{\text{---}} \frac{D}{\text{---}} \frac{ED}{\text{---}}. \quad (30)$$

Notice that the above expression is a particular instance of Eq. (24). The action of the local transformations of the form  $\mathcal{S}_E \boxtimes \mathcal{T} \in \text{Transf}(EA \rightarrow EB)$  can be derived by posing the following expression:

$$\frac{E}{\text{---}} \frac{A}{\text{---}} \boxed{\mathcal{T}} \frac{B}{\text{---}} \frac{E}{\text{---}} := \frac{E}{\text{---}} \frac{A}{\text{---}} \frac{B}{\text{---}} \frac{E}{\text{---}} \frac{E}{\text{---}} \frac{B}{\text{---}} \frac{E}{\text{---}} \frac{A}{\text{---}} \boxed{\mathcal{T}} \frac{B}{\text{---}} \frac{E}{\text{---}}. \quad (31)$$

We notice that the latter is a particular instance of Eq. (4). Parallel composition of transformations can be derived by exploiting Eqs. (29)–(31), and then posing the following expression:

$$\begin{array}{c}
 \text{AC} \quad \text{BD} \\
 \boxed{\mathcal{T} \boxtimes \mathcal{T}'} := \begin{array}{c} \text{A} \quad \text{B} \\ \boxed{\mathcal{T}} \\ \text{C} \quad \text{D} \\ \boxed{\mathcal{T}'} \end{array} \quad (32)
 \end{array}$$

We can now proceed by verifying whether a theory  $\Theta$  is a consistent OPT.

*Identity process.* As long as  $\mathcal{I} \subseteq \text{RevTransf}(\Theta)$ , in the light of Eqs. (31) and (32), also Eq. (2) is always satisfied.

*Equations (3) and (4).* Using Eq. (31), one can check whether the following hold:

$$\begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}} \\
 \text{C} \quad \text{D} \\
 \boxed{\mathcal{T}'}
 \end{array} = \begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}} \\
 \text{C} \quad \text{D} \\
 \boxed{\mathcal{T}'}
 \end{array}, \quad (33)$$

$$\begin{array}{c}
 \text{E} \quad \text{A} \quad \text{B} \quad \text{E} \\
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}} \\
 \text{E} \\
 \text{A} \quad \text{B}
 \end{array} = \begin{array}{c}
 \text{E} \\
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}}
 \end{array}. \quad (34)$$

Thus, Eqs. (3) and (4) straightforwardly follow from Eqs. (33) and (34), using Eqs. (2), (29), (30), (31), and (32), along with the associativity of sequential composition. When the OPT is symmetric, Eq. (34) does not need to be checked, being equivalent to Eq. (31).

*Associativity.* After verifying whether the associator  $\alpha$  is invertible, we can check the pentagon identity (26). This is done by using the associator  $\alpha$  as the rule for identifying the linearly independent vectors in  $\text{St}_{\mathbb{R}}(((AB)C)D)$  with those in  $\text{St}_{\mathbb{R}}((A(BC))D)$ . Since the OPT is required to be associative, the following must hold [see Eq. (24)]:

$$\begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}_1} \\
 \text{C} \quad \text{D} \\
 \boxed{\mathcal{T}_2} \\
 \text{E} \quad \text{F} \\
 \boxed{\mathcal{T}_3}
 \end{array} = \begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}_1} \\
 \text{C} \quad \text{D} \\
 \boxed{\mathcal{T}_2} \\
 \text{E} \quad \text{F} \\
 \boxed{\mathcal{T}_3}
 \end{array}, \quad (35)$$

for all  $A, B, C, D, E, F \in \text{Sys}(\Theta)$  and all  $\mathcal{T}_1 \in \text{Transf}(A \rightarrow B)$ ,  $\mathcal{T}_2 \in \text{Transf}(C \rightarrow D)$ ,  $\mathcal{T}_3 \in \text{Transf}(E \rightarrow F)$ . In fact, using Eqs. (2), (25) and (29)–(32), one verifies that Eq. (35) holds by construction.

*Braiding.* One can verify whether the transformations associated with the braiding  $\mathbf{S}$  are invertible and satisfy the two hexagon identities (27) and (28). When the OPT is symmetric, the two hexagon identities (27) and (28) are equivalent, and then just one of them needs to be checked.

*Compatibility with the probabilistic structure.* Finally, one can check whether (i) states separate effects and vice versa, (ii) the instruments  $P_X \in \text{Instr}(I \rightarrow I)$  are probability distributions, (iii) the null transformation  $\varepsilon_{A \rightarrow B}$  is included in

$\text{Transf}(A \rightarrow B)$  for every  $A, B$ , (iv) the coarse-graining operation is well posed for every instrument, and (v) every instrument  $T_X^{A \rightarrow B} \boxtimes |_{\star}^{E \rightarrow E}$  maps preparation instruments of AE to preparation instruments of BE for every system  $A, B, E$ .

Once we have done the above checks, we are done. Indeed, the above conditions are exhaustive, being compliant with the definitions and the coherence results given in Secs. III and VA. We remind at this point that all of the above consistency checks need to be performed extending both sides of every equation with the identity wire of an arbitrary system (see Remark 1). In the next section, we shall first introduce a theory  $\Theta$  in the axiomatic way described in Sec. VB. Subsequently, we show that  $\Theta$  is indeed a consistent OPT by resorting to the above-described consistency checks.

### VI. BILOCAL CLASSICAL THEORY

We now present a classical theory  $\tilde{\Theta}$  with entanglement, which we call bilocal classical theory (BCT).

#### A. Postulates

*Postulate 1 (Classicality, convexity, and types of systems).* The theory  $\tilde{\Theta}$  is classical and convex. For every integer  $D > 1$ ,  $\text{Sys}(\tilde{\Theta})$  contains a type of system having dimension  $D$ .

Classical theories have been defined in Sec. IV C. For every system  $A$ , the elements of  $\text{ExtSt}_0(A)$  are  $D_A + 1$  vertices of a simplex. Moreover, the set of the non-null extremal states  $\text{ExtSt}(A)$  coincides with  $\text{PurSt}(A) = \{|i\rangle_A\}_{i=1}^{D_A}$ . The pure states of every system are jointly perfectly discriminable (see Property 5). In addition, by convexity, the set of deterministic states  $\text{St}_1(A)$  is the convex hull of the pure states.

*Postulate 2 (Parallel composition of systems and states, associator).* For any two systems  $A, B \in \text{Sys}(\tilde{\Theta})$ , the dimension of the composite system  $AB$  is given by the following rule:

$$D_{AB} = D_{BA} = \begin{cases} 2D_A D_B, & \text{if } A \neq I \neq B \\ D_A, & \text{if } B = I. \end{cases} \quad (36)$$

Let  $I \neq A, B, C \in \text{Sys}(\tilde{\Theta})$ . Denoting the pure states of any composite system  $AB$  as  $\text{PurSt}(AB) = \{|(ij)_{-}\rangle_{AB}, |(ij)_{+}\rangle_{AB} | 1 \leq i \leq D_A, 1 \leq j \leq D_B\}$ , for all states  $|i\rangle_A \in \text{PurSt}(A)$ ,  $|j\rangle_B \in \text{PurSt}(B)$  the following parallel composition rule holds:

$$\begin{array}{c}
 \text{B} \\
 i \\
 \text{A} \\
 j
 \end{array} = \frac{1}{2} \sum_{s=-,+} \begin{array}{c} \text{A} \\ (ij)_s \\ \text{B} \end{array}. \quad (37)$$

The associator map is given by the following identification:

$$((ij)_{s_1} k)_{s_2} = (i(jk)_{s_1 s_2})_{s_1} \quad (38)$$

for all local indices  $i, j, k$  and signs  $s_1, s_2$ .

Notice that Postulate 2 complies with the classification of the sets of states in simplicial theories with  $n$ -local discriminability provided in Ref. [8]. In the light of Postulate 2, one sees that the information carriers of the theory, despite being classical, compose differently from the ones of CT. In Sec. VII we will discuss some consequences of this fact. Accordingly,





identity transformation of an arbitrary system, namely, Eq. (9) holds.

By Eq. (38), the associator of the theory is invertible. We now verify the pentagon identity (26) for state composition via consecutive applications of Eq. (38) on the pure states of a pentapartite system. On the one hand, one has

$$\begin{aligned} (((ij)_{s_1} k)_{s_2} l)_{s_3} m)_{s_4} &= (((ij)_{s_1} (kl)_{s_2 s_3})_{s_2} m)_{s_4} \\ &= ((i(j(kl)_{s_2 s_3})_{s_1 s_2})_{s_1} m)_{s_4}. \end{aligned}$$

On the other hand, one also has

$$\begin{aligned} (((ij)_{s_1} k)_{s_2} l)_{s_3} m)_{s_4} &= (((i(jk)_{s_1 s_2})_{s_1} l)_{s_3} m)_{s_4} \\ &= ((i((jk)_{s_1 s_2} l)_{s_1 s_3})_{s_1} m)_{s_4} \\ &= ((i(j(kl)_{s_2 s_3})_{s_1 s_2})_{s_1} m)_{s_4}. \end{aligned}$$

In the case of states and effects, Eq. (33) can be verified using Eqs. (42) and (43). In the case of arbitrary transformations, Eq. (33) can be easily verified just for the atomic transformations (44), and then extended by linearity in the light of Proposition 3. The family of transformations  $\mathcal{S}$  defined in Eq. (41) is manifestly invertible. Moreover, since the theory is symmetric, Eq. (34) is equivalent to Eq. (31), and then it does not need to be checked. Finally, for the same reason, one can verify just one of the hexagon identities, e.g., Eq. (27). This is simply done by an iterative application of the associator (38) and of the braiding (41) on the tetrapartite pure states of the theory.

By Postulates 1 and 6, and by implications (ii)  $\Leftrightarrow$  (iii) and (ii)  $\Rightarrow$  (i) in Proposition 4, the following final requirements hold: (i) states separate effects and vice versa, (ii) the instruments  $P_X \in \text{Instr}(I \rightarrow I)$  are probability distributions, (iii) the null transformation  $\varepsilon_{A \rightarrow B}$  is included in  $\text{Transf}(A \rightarrow B)$  for every  $A, B$ , (iv) the coarse-graining operation is well posed, and (v) every instrument  $T_X^{A \rightarrow B} \boxtimes I_*^{E \rightarrow E}$  maps preparation instruments of AE to preparation instruments of BE for every system  $A, B, E$ .

Accordingly, the postulates of the theory, along with the classification of the atomic transformations, lead to a straightforward check of its consistency.

## VII. FEATURES OF THE THEORY

### A. Entanglement is independent of complementarity

As a consequence of Proposition 3 and Property 7, BCT satisfies the *full information without disturbance principle* (FIWD) [29], i.e., every test can be simulated via a nondisturbing test. This happens in spite of the presence of entanglement. In a theory with FIWD the identity transformation cannot be atomic [29]. Moreover, a theory satisfies FIWD only if the pure states of every system are jointly perfectly discriminable (see Property 5) [29]. Accordingly, BCT does not admit the existence of incompatible observables or, in other words, the theory does not satisfy the *principle of complementarity*. As a first consequence, in BCT it is clearly impossible to violate Bell's inequalities. However, the theory admits of entangled states: every pure state of a composite system is in fact entangled. The above two features allow one to show that complementarity and entanglement are two independent

properties. On the one hand, BCT provides the explicit example of a theory without complementarity, but endowed with entangled states. Conversely, take a modified version of QT whose parallel composition is given by the minimal tensor product on both states and effects, i.e., the theory where every system is quantum, except that the set of states for every composite system is the convex hull of the product states only. The latter is an example of a theory with complementarity but without entanglement (indeed, complementarity is a single-system property). This shows that complementarity and entanglement are in fact two independent properties in a probabilistic theory. Furthermore, the existence of a unique joint probability distribution for the outcomes of every possible set of measurements implies that BCT is also *noncontextual*.<sup>5</sup> It is indeed already known that noncontextuality, being a single-system property, is decoupled from entanglement. On the one hand, Spekkens toy model [31] is noncontextual, but has entanglement. On the other hand, the above-mentioned version of QT with minimal tensor product is contextual, but has no entanglement. However, we remind that Spekkens toy model is not a simplicial theory.

### B. Violation of purity of state composition and of purification, absence of superpositions

In Ref. [8] it is proven that, if a simplicial theory violates local discriminability but  $n$ -local discriminability holds for some  $n$ , then the theory violates atomicity of parallel composition of states, as well. In BCT, this fact is manifest from the state-composition rule (37), where one sees that the parallel composition of any pair of pure states is not pure, namely, also the principle of *purity of parallel composition of states* is violated. Notice that the principle called *purity preservation* in Ref. [23] coincides in fact with atomicity of both parallel and sequential composition (see Remark 3). Moreover, state-composition rule (37) is clearly a particular instance of the classification of composite states for simplicial theories with  $n$ -local discriminability provided in Ref. [8]. The atomic transformations of BCT (see Proposition 3) are the bilocal-tomographic counterpart of the ones of CT. However, resorting to Proposition 3, in particular using expression (44), one can easily verify that atomicity of parallel composition in BCT is not even satisfied for arbitrary transformations. Interestingly, using again Proposition 3, it is also easy to verify that, differently for the case of states, purity of parallel composition of channels is satisfied. Finally, Proposition 3 guarantees that in BCT atomicity and purity are both preserved under *sequential* composition.

Moreover, in BCT no mixed state has a purification, and superpositions of states (in an operational sense) are not

<sup>5</sup>For the definition of a generalized noncontextual ontological model, see Ref. [9]. Therein, the existence of such a model, using a frame representation which is not overcomplete [30], is proved, under the hypotheses of convexity, causality, and local discriminability, to be equivalent to a property called *simplex-embeddability*, and to the existence of a non-negative quasiprobabilistic model. In fact, the authors also exhibit a counterexample showing that in general (in particular, without local discriminability) the result does not hold.

admitted. The latter features are a consequence of the no-go results proven in Ref. [8] for general simplicial theories.<sup>6</sup> Interestingly, this means that entanglement implies neither the purification nor the superposition principles (nor weaker formulations of these, holding just for a finite number of states). However, BCT satisfies the (*essential*) *uniqueness of purification principle* [22].

*Property 8 (Essential uniqueness of purification).* Let  $\Theta$  be an OPT, and  $A, B \in \text{Sys}(\Theta)$ . If there exist  $|\Sigma_1\rangle_{AB}, |\Sigma_2\rangle_{AB} \in \text{PurSt}(AB)$  and  $(\tilde{e}_1|_B, (\tilde{e}_2|_B \in \text{Eff}_1(B)$  such that

$$\begin{array}{c} \text{A} \\ \text{---} \\ \Sigma_1 \\ \text{---} \\ \text{B} \\ \text{---} \\ \tilde{e}_1 \end{array} = \begin{array}{c} \text{A} \\ \text{---} \\ \Sigma_2 \\ \text{---} \\ \text{B} \\ \text{---} \\ \tilde{e}_2 \end{array},$$

then there exists  $\mathcal{R} \in \text{RevTransf}(B \rightarrow B)$  such that

$$\begin{array}{c} \text{A} \\ \text{---} \\ \Sigma_2 \\ \text{---} \\ \text{B} \end{array} = \begin{array}{c} \text{A} \\ \text{---} \\ \Sigma_1 \\ \text{---} \\ \text{B} \\ \text{---} \\ \mathcal{R} \\ \text{---} \\ \text{B} \end{array}.$$

In other words, *if there exists a purification*  $|\Sigma\rangle_{AB}$  of some given deterministic state, then such  $|\Sigma\rangle_{AB}$  is *essentially unique* in  $AB$ , namely, unique up to reversible local transformations of  $B$ . Notice that Property 8, choosing  $A = I$ , implies *transitivity of reversible channels on pure states*. We observe that Property 8 is independent of the purification principle since, e.g., given any OPT, one can always restrict the class of transformations to the identity  $\mathcal{I}$  and the braiding  $\mathcal{S}$  families in a consistent way, ending up with a mere prepare-and-measure scenario possibly satisfying purification, but not Property 8.

### C. Dense coding and additivity of classical capacities

An important communication task that can be performed in BCT is the *dense coding* [33,34]. In quantum dense coding, Alice and Bob share a maximally entangled state. Bob detains his qubit and Alice, after performing on her qubit some local operations in order to encode a two-bit message, sends her local qubit to Bob. The decoding part is then performed by Bob performing a joint measurement in a suitable basis. Via this protocol, Bob manages to decode the message and gains two classical bits of information receiving just one qubit, whose classical capacity, according to Holevo’s bound, is one bit. The protocol in BCT retraces the quantum one in the following way. Alice and Bob share a known entangled state, say  $|(0b)_-\rangle_{AB}$ , where  $D_A = D_B = 2$  and  $b$  is an arbitrary local state of the bibit detained by Bob. Now, Alice can encode her two-bit message. She performs the following encoding via entanglement-preserving local operations [recall the reversible transformations in Eq. (39)]:

$$\begin{aligned} 00 &\mapsto |(0b)_-\rangle_{AB}, & 01 &\mapsto |(0b)_+\rangle_{AB}, \\ 10 &\mapsto |(1b)_-\rangle_{AB}, & 11 &\mapsto |(1b)_+\rangle_{AB}. \end{aligned}$$

<sup>6</sup>However, notice that, in principle, the existence of simplicial theories where at least some mixed states can be purified [32], or satisfying an operational formulation of the superposition principle given in Ref. [8], cannot be excluded.

Then, Alice sends her local bibit to Bob, who can measure the global state, and thus directly decode Alice’s message. Notice that, as far as the success of the protocol is concerned, the specific value  $b$  of Bob’s local state is not relevant. In our BCT dense coding protocol, Alice and Bob share an entangled state, but differently from the quantum case, their local marginal states are always pure. Then, the two-bit message is encoded by Alice into her local bit and into the global degree of freedom (the sign). In this way, by sending one bibit, Alice is able to communicate two bits of information to Bob. We now have a look at the case when dense coding is realized optimally, i.e., with minimal resources. In QT, when  $\log_2 M$  qubits, with  $M = d$ , are sent from Alice to Bob, a bipartite system with local dimensions  $D_A = D_B = d^2$  is optimal, and the latter case is called *tight* [34].

In BCT, regardless of Bob’s local system’s size, the maximum attainable number of distinguished signals is always given by  $M = 2^{2n-1}$ , where  $n$  is the number of bibits that Alice sends to Bob. Accordingly, by receiving  $n$  elementary information carriers, Bob is able to distinguish  $2n$  bits, both in the QT and in the BCT protocol. This means that BCT achieves the same performances as QT in dense coding. We remark that, apparently, BCT seems to exhibit superadditivity of classical capacities since a bipartite system consisting of two bibits, each of which, alone, carries at most one bit, is able to carry three bits. However, if one sticks to the asymptotic definition of classical capacity, every bibit has classical capacity of 2 bits. Indeed, the bits carried by a system of  $n$  bibits are  $2n - 1$ , and thus the asymptotic capacity is 2 bits per bibit. Moreover, the same analysis reveals that classical capacity of BCT systems is additive.

The above analysis shows that BCT does not exhibit *hyperdense coding*, i.e., a coding protocol able to overcome the dense coding limit given by QT. Indeed, the possibility of performing a hyperdense coding would imply *superadditive classical capacities* [35].

### D. Violation of entanglement monogamy and of the no-hypersignaling principle

A quite peculiar feature is that entanglement, in BCT, is not *monogamous*, namely, BCT violates *monogamy of entanglement*. This means that there exist (maximally entangled) states  $|\Sigma\rangle_{ABC}$  such that the marginal states  $(e|_B|\Sigma\rangle_{ABC}, (e|_C|\Sigma\rangle_{ABC}$  are both maximally entangled. Equivalently, a system can be entangled with more than one other system at the same time. On the contrary, in QT entanglement is monogamous. A violation of entanglement monogamy is known to hold in other theories, such as real quantum theory and fermionic quantum theory [36]. However, in the latter cases those states which violate entanglement monogamy are mixed, while in BCT the violation holds *for every m-partite pure state with m ≥ 3*.

In Ref. [10], the *no-hypersignaling principle* in probabilistic theories is introduced and analyzed. Roughly speaking, the principle states that any input-output correlation which can be obtained by transmitting a composite system should also be obtainable by independently transmitting its constituents. For instance, hypersignaling is exhibited if a probabilistic theory, while not contradicting CT and QT at the level of spacelike correlations, displays an anomalous behavior in its timelike

correlations. Both CT and QT satisfy the no-hypersignaling principle. In the case of simplicial theories, a theory  $\Theta$  is hypersignaling if and only if there exist  $A, B \in \text{Sys}(\Theta)$  such that  $D_{AB} > D_A D_B$  (the technical definition of hypersignaling can be found in Ref. [10]). Namely, a simplicial theory is hypersignaling if and only if it has entanglement. BCT, to the best of our knowledge, is the first example of a complete hypersignaling theory (Ref. [10] analyzes the correlations of a model consisting of a two-system scenario). Differently from the model presented in Ref. [10], in BCT the existence of anomalous behaviors in timelike correlations requires the action of bipartite effects on entangled states. We notice that the no-hypersignaling principle is neither sufficient nor necessary for local discriminability [10].

**E. Entanglement swapping, cloning and teleportation, and non-null discord states**

In BCT it is possible to perform *entanglement swapping* [37], namely, the task of transferring entanglement to two remote systems, which are initially uncorrelated, *without interaction*. It is easy to see how this can be done using Eqs. (31), (37), and (42). Let us pick an arbitrary pure entangled state  $|(ij)_s\rangle_{AB}$ . We want to transfer the entanglement from system B to a remote, uncorrelated system D. Now, fix a chosen pure entangled state  $|(kl)_t\rangle_{CD}$  shared by system D and an ancillary system C, and perform the following instrument:

$$\begin{aligned}
 &= \left\{ \frac{1}{2} \sum_{s'=+,-} \delta_{r,ss'} \left( (il)_{s't} \right)^{AD} \right\}_{r=+,-} = \\
 &= \left\{ \frac{1}{2} \left( (il)_{rst} \right)^{AD} \right\}_{r=+,-}, \tag{47}
 \end{aligned}$$

where the possible outcomes of the instrument are  $j, k$  with probability 1, and  $r = +, -$ , each with probability  $\frac{1}{2}$ . As a result, system A and system D become entangled. This entanglement swapping protocol in BCT clearly retraces the quantum one.

Despite the existence of an entanglement-swapping protocol, the analysis of the teleportation scenario arising from it would make poor sense. Indeed, by the FIWD principle and Property 4 it is possible to arbitrarily clone every unknown state, namely, teleportation can be reduced to making a copy of a state and sending it to the receiver (e.g., using a measure-and-prepare channel). For an in-depth discussion about the dependence of teleportation on other properties (such as nonlocality) of a probabilistic theory, we refer the reader to Refs. [38,39].

Finally, the mere existence of entangled states in BCT implies that the theory contains states having *non-null discord* [40] (in an operational sense). However, one may be wondering whether BCT has also some non-null discord separable states. This is actually not the case, due to the FIWD principle and to the absence of delocalized information in any separable state.

**F. Programming and information-theoretically insecure cryptography**

In CT it is possible to perform the task of *programming* any desired channel from a system to any another, due to the existence of a *universal processor*. More precisely, CT satisfies the following property.

*Property 9 (Programming).* Let  $\Theta$  be an OPT. For every pair of systems  $A, B \in \text{Sys}(\Theta)$ , there exist a system P and a channel  $\mathcal{P}_{A,B} \in \text{Transf}_1(PA \rightarrow PB)$  such that the following holds. For every target channel  $\mathcal{C} \in \text{Transf}_1(A \rightarrow B)$ , there exists a program state  $|\sigma\rangle_P \in \text{St}_1(P)$  such that

$$\mathcal{C} = \mathcal{P}_{A,B} \left( |\sigma\rangle_P \right) \tag{48}$$

As a straightforward corollary of Proposition 4, in particular, see the dilation scheme (46) in 4, in BCT the task of programming is possible as well. In QT, despite the *no-programming theorem*, the task of *probabilistic programming* is possible, where the error probability can be made arbitrarily small, provided that  $D_P$  [see Eq. (48)] becomes arbitrarily large [41–43].

Finally, one may be wondering whether entanglement in BCT grants information-theoretical security, in a cryptographic scenario, against the attacks of malicious adversaries. For instance, one can think of the possibility to implement protocols such as secure key generation or distribution. It is easy to see why this is not the case. Despite the existence of entangled states with global information, which remains inaccessible unless all the parties collaborate, such protocols in BCT are intrinsically not secure due to the FIWD principle.

**VIII. DISCUSSION**

Table I provides a survey of the operational features of BCT, along with a list of tasks which can and cannot be performed in the theory.

We adopted the notion of classicality defined from the perspective of states: a theory is classical if and only if the sets of states for every system are those of the systems of CT, namely, simplicial sets with jointly perfectly discriminable pure states. However, the theory presented also abides by the notion of classicality proposed in Ref. [9], admitting of a noncontextual ontological model. In a recent work [44], it is proved the nontrivial result that, under the hypothesis of local tomography, given two sets of states  $\text{St}(A)$  and  $\text{St}(B)$ , the composite space  $\text{St}(AB)$  can admit of entangled states if and only if neither  $\text{St}(A)$  nor  $\text{St}(B)$  is a simplex. On the other hand, a theory without local tomography necessarily features entangled states [8]. BCT provides the concrete example that indeed two classical systems can give rise to entangled states

TABLE I. Survey of the operational properties of BCT.

BCT features $\checkmark$	BCT does not feature $\times$
Causality	Local tomography
Atomicity and purity of sequential composition	Atomicity of parallel composition
Purity of parallel composition for channels	Purity of parallel composition for states
Essential uniqueness of purification	Purification
Bilocal tomography	Complementarity
Noncontextuality	Monogamy of entanglement
Full information without disturbance	No hypersignaling
Non-null discord states	Secure key generation and distribution
Dense coding	Hyperdense coding
Entanglement swapping and teleportation	Superpositions
Universal programmability	No cloning
Transitivity of local reversible channels	Superadditive classical capacity
Joint perfect discriminability of pure states	
No-restriction hypothesis	
Strong self-duality	
Pure conditionalization	

(at the expense of local tomography). In principle, it is not obvious that this could be done in a consistent way. Our result shows that entanglement and incompatibility of measurements (complementarity) are two independent properties in a theory. Interestingly, BCT thus also shows that a violation of local discriminability does not necessarily imply nonlocality in a Bell-type scenario, namely, stronger-than-classical spacelike correlations. As we will discuss in Sec. IX, however, it is not clear at the moment whether BCT is local, i.e., admits of a local ontological model.

We have seen that in BCT entanglement is not monogamous. The same happens in other nonlocal-tomographic theories, such as real quantum theory and fermionic quantum theory [36,45]. One may conjecture this property to be a consequence of the violation of local tomography. Another common trait shared by BCT and fermionic quantum theory is the possibility of activating local discrimination of states with entangled ancillary systems [36,46], as can be easily derived from Eq. (47).

From an axiomatic viewpoint, the existence of BCT allows one to draw some interesting consequences. One can verify that atomicity of parallel composition, being violated by BCT, is independent from causality, ideal compression, perfect discriminability, and bilocal discriminability [22]. The same holds for purity of parallel composition of states since its violation, in this case, is equivalent to the violation of atomicity of state composition. Remarkably, purity of parallel composition is satisfied for channels and effects (i.e., excluding states), while atomicity of parallel composition is violated for any kind of transformation. This fact provides a concrete example motivating the distinction between the two notions of purity and atomicity (see Remark 3). In passing, we notice that the principle of *pure conditionalization* [47] is satisfied by BCT. As a consequence, pure conditionalization implies neither atomicity nor purity of parallel state composition. Moreover, atomicity (or purity) of parallel state composition is neither necessary nor sufficient for the FIWD principle. BCT also shows that the presence of entanglement is compatible with the absence of purification for *any* mixed state, or of *any* kind of superposition [8]. Finally, BCT shows that

the presence of entanglement is compatible with the FIWD principle, namely, entanglement does not imply information-theoretically secure key generation or distribution.

Interestingly, looking at BCT as a *process theory* [17], one realizes that the proposed process-theoretic definition of purity [48,49] would imply that the theory has no pure state. One may then argue that the above definition is not tenable as a notion of purity in a probabilistic scope, at least when purity of state composition is violated.

One may be wondering whether, in principle, BCT is the only classical theory satisfying strict bilocal tomography. In Appendix D (Theorem 4 and Corollary 3), we prove that, for any simplicial theory, rules (36) and (37) come as a consequence of the following assumptions: (a) homogeneous strict bilocal tomography: the theory is strictly bilocal tomographic, with no composite system being local tomographic; (b) essential uniqueness of purification (Property 8). In particular, assumption (a) alone is not sufficient to select rule (36) since, as a counterexample, one can verify (using Theorem 2) that also the rule

$$D_{AB} = D_A D_B + (D_A - 1)(D_B - 1)$$

satisfies (a), being also associative. How the reversible dynamics [assumption (b), see Property 8] affects probabilities in a theory has been recently explored in Ref. [50]. However, there is strong evidence that assumptions (a) and (b) may be not sufficient to single out BCT among classical theories. Indeed, the reversible transformations admitted by the simplicial structure are not in principle exhausted by those defined in Postulate 3, and one can actually define disjoint families of transformations which both are reversible and obey to Eq. (33). On the other hand, coherence (see Sec. VA) is also key to single out BCT. For instance, it is easy to verify that some choices of the associator (see Postulate 2), despite being invertible, lead to violations of the pentagon identity. As for the choices of the associator  $\alpha$  and the braiding  $\mathcal{S}$ , modifying Postulate 3 would lead in principle to conceivable choices differing from (38) and (41). Nevertheless, Postulate 3, along with assumptions (a) and (b), would single out precisely those  $\alpha$  and  $\mathcal{S}$  postulated for BCT.

Finally, as well as CT and QT, also BCT satisfies both the no-restriction hypothesis (see Property 2) and strong self-duality [28]. Notice that BCT also satisfies the usual formulation of the no-restriction hypothesis, the one which is valid in a context where local tomography holds (see Sec. IV B), namely, “all the positive functionals on states are effects.”

## IX. CONCLUSIONS AND OUTLOOK

We conclude this work drawing some final remarks and pointing out some open problems.

In Sec. III we provided a broad-scope presentation of the operational probabilistic framework. In Sec. IV A we proved Theorem 1, providing a characterization of the composition rules for system sizes in arbitrary theories, and then we specialized it to strictly bilocal-tomographic theories, proving Theorem 2. In Sec. V, we set the problem of the consistency of an operational probabilistic theory, identifying a set of sufficient conditions to construct a theory, and to verify its well posedness and coherence. Finally, we have presented BCT, a classical theory featuring entanglement and complete with a nontrivial set of transformations, providing an extensive characterization of it.

BCT violates the principle of local tomography. However, the degree of holism required by the theory in the task of *state tomography* is still limited, due to the property of (strict) bilocal tomography. What is more, notice that in BCT, in order to perform *process tomography* of any  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  (Proposition 3), one just needs a limited set of bipartite pure states, namely, a set  $\{|\rho_i\rangle_{AE}\}_{i=1}^{D_A} \subset \text{PurSt}(AE)$  such that the set of reduced states  $\{(e|_E|\rho_i)\rangle_{AE}\}_{i=1}^{D_A} = \text{PurSt}(A)$ , while the ancilla  $E \neq I$  and all reduced states  $\{(e|_A|\rho_i)\rangle_{AE}\}_{i=1}^{D_A}$  can be arbitrarily chosen. One may still argue that the principle of  $n$ -local discriminability (Property 1) does not sufficiently bind the degree of holism of a physical theory. Nevertheless, there are examples of theories which, while violating local discriminability, have a strong physical motivation, such as fermionic quantum theory. In addition, generic nonlocal-tomographic theories provide a sandbox to investigate the logical interdependence of physical properties, which is a key aspect in view of formulating new physical theories.

Indeed, BCT provides two important proofs of concept. First, entanglement and complementarity are decoupled, i.e., they are independent properties of a physical theory. Second, the set of states, or of correlations, is not sufficient to determine the full theory. This is by the way close to the spirit of the no-hypersignaling principle [10]. Interestingly, since BCT enjoys the same sets of states of CT, it looks like it cannot violate any device-independent principle [51,52]. This leads to the following question: Is it even logically possible to formulate device-independent principles ruling out BCT, but not CT?

As the correlations attainable by a physical theory do not determine the theory itself, and, in fact, may even give rise to quite different theories, one may be questioning *what does it mean to be classical* or, more specifically along the lines of Ref. [9], *what does it mean to be classically explainable*. As we extensively discussed in Sec. VIII, we left the problem of an information-theoretic axiomatization of BCT open.

Such an axiomatization may shed some light on a meaningful notion of classicality even in a context allowing for nonlocal tomography. For instance, both real and fermionic quantum theory are superselected versions of QT: Might it be the case that BCT arises as the superselection of CT? We conjecture that this is not the case, as we argue in the following.

By direct inspection of the state-composition rule in BCT [Postulate 2, Eq. (37)], one sees that every product of pure states is the flat statistical mixture of two entangled states carrying a “delocalized” variable, i.e., the sign  $s$  in Eq. (37). One is tempted to assign an *element of reality* to the global degree of freedom  $s$ . However, is this intuition correct? In fact, there are two ontologically distinct, although operationally equivalent, preparation procedures for the parallel composition of preparation tests in BCT: one is to generate the mixture using classical randomness for the composite system; the other one is to simply prepare two local pure states in parallel. Therefore, is it meaningful to say that an entangled classical measurement on a product state reveals a preexisting value  $s$  in both cases? This would amount to say that the product state was entangled in the first place. Surprisingly, such a paradox is reminiscent of the question whether measurements reveal a predefined value, albeit in a measurement context which is, in all respects, classical. Being the value measurable without disturbance, the issue seems to be irrelevant: there is no contradiction in thinking of  $s$  as an “element of reality” in the sense of EPR. However, this position opens a deeper question: Can the global degree of freedom  $s$  be interpreted as a function of suitably defined local (possibly hidden) variables? We have evidence that the primary difficulty in devising such a model is having hidden variables of finite sizes, i.e., hidden variables which can be stored in a limited memory. This would mean to require that each system cannot retain the values of global data for all the possible choices of remote systems and experiments considered. In Ref. [53], a similar situation was investigated as far as QT is concerned. However, in that case, the hidden register would store the values of the intrinsic properties for all possible sequences of measurements that the observer can perform. The motivation to demand such a property is that abandoning it would allow for some form of *superdeterminism*, where, e.g., all that (infinite amount of) information, which is relevant to predict the outcomes of any experiment, is possibly stored in each system. Nevertheless, such a requirement seemingly clashes with the action of the associator and braiding of the theory. Might it be that BCT does not admit of a local-realistic hidden variable model, in spite of having classical correlations?<sup>7</sup>

Finally, this study provides an adequate toolbox for the comprehensive construction of complete and consistent physical theories. In particular, Sec. V sets an exploitable constructive procedure which can be used in a generic context, while in Sec. VI we provided an explicit concrete application of the latter. This paves the way for a direct employment of

<sup>7</sup>On top of this, at this stage it is not clear whether a (noncontextual) ontological model for BCT may possibly satisfy *diagram preservation*, namely, the property of preserving the compositional structure of the OPT, or whether it may preserve the linear (convex and coarse-graining) relations of the OPT [9].

the developed techniques in enhancing the investigations on post-quantum theories.

**ACKNOWLEDGMENTS**

This paper was made possible through the support of a grant from the John Templeton Foundation, Grant No. 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. M.E. wishes to thank A. Tosini and M. Plávala for stimulating and insightful discussions.

**APPENDIX A: CLASSIFICATION OF BCT’S TRANSFORMATIONS (PROOF OF PROPOSITION 3)**

In this Appendix, in order to make sense of expressions of the form  $|(l_1 l_2)_u\rangle_{S_1 S_2}$  with  $S_1 = I$  or  $S_2 = I$ , these can be safely replaced by the expressions  $|(*l_2)_+\rangle_{S_2}$  or  $|(l_1*)_+\rangle_{S_1}$ , respectively. Moreover, we will extensively exploit the classification of BCT’s effects given in Sec. **VIB**.

*Lemma 1.* Let  $I \neq A \in \text{Sys(BCT)}$ ,  $B \in \text{Sys(BCT)}$ , and  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  be defined as

$$\begin{array}{c} A \\ \text{---} \end{array} \boxed{\mathcal{T}} \begin{array}{c} B \\ \text{---} \end{array} := \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{k} \begin{array}{c} B' \\ \text{---} \end{array} \boxed{\mathcal{R}} \begin{array}{c} A' \\ \text{---} \end{array} \boxed{\tilde{k}} \begin{array}{c} B \\ \text{---} \end{array} \quad , \quad (A1)$$

with  $|k\rangle_{B'} \in \text{PurSt}(B')$ ,  $|\tilde{k}\rangle_{A'} \in \text{Eff}(A')$  such that  $\langle k' | \tilde{k} \rangle_{A'} = \delta_{k' \tilde{k}}$  for any  $|k'\rangle_{A'} \in \text{PurSt}(A')$ , and  $\mathcal{R} \in \text{RevTransf}(B'A \rightarrow A'B)$  such that for all  $E \in \text{Sys(BCT)}$ :

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{((ki)_{s_1 j})_{s_2}} \begin{array}{c} B'A \\ \text{---} \end{array} \boxed{\mathcal{R}} \begin{array}{c} A'B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} := \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{((\kappa_{s_1}^{ki} \iota_{s_1}^{ki})_{r_{s_1}^{ki} j})_{t_{s_1}^{ki} s_2}} \begin{array}{c} A'B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} \quad . \quad (A2)$$

[In Eq. (A2), there is no dependence on  $k$  and  $s_1$  in the case  $B' = I$ , while in the case  $A' = I$  or  $B = I$  we just set  $r_{s_1}^{ki} = +$  for all  $k, i, s_1$ ]. Then, for all  $|(ij)_s\rangle_{AE} \in \text{St}(AE)$ , the following holds:

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(ij)_s} \begin{array}{c} A \\ \text{---} \end{array} \boxed{\mathcal{T}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} = \frac{1}{2} \sum_{s'=+,-} \delta_{\tilde{k}, \kappa_{s'}^{ki}} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(\iota_{s'}^{ki} j)_{r_{s'}^{ki} t_{s'}^{ki} s}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} \quad . \quad (A3)$$

*Proof.* By direct computation, using Postulate 2 and formula (42), one has

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(ij)_s} \begin{array}{c} A \\ \text{---} \end{array} \boxed{\mathcal{T}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} = \frac{1}{2} \sum_{s'=+,-} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(k(ij)_s)_{s'}} \begin{array}{c} B' \\ \text{---} \end{array} \boxed{\mathcal{R} \boxtimes \mathcal{I}_E} \begin{array}{c} A' \\ \text{---} \end{array} \boxed{\tilde{k}} \begin{array}{c} BE \\ \text{---} \end{array} \\ = \frac{1}{2} \sum_{s'=+,-} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{((ki)_{s'} j)_{s' s}} \begin{array}{c} B'A \\ \text{---} \end{array} \boxed{\mathcal{R} \boxtimes \mathcal{I}_E} \begin{array}{c} A' \\ \text{---} \end{array} \boxed{\tilde{k}} \begin{array}{c} BE \\ \text{---} \end{array} \\ = \frac{1}{2} \sum_{s'=+,-} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{((\kappa_{s'}^{ki} \iota_{s'}^{ki})_{r_{s'}^{ki} j})_{t_{s'}^{ki} s}} \begin{array}{c} A' \\ \text{---} \end{array} \boxed{\tilde{k}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} \\ = \frac{1}{2} \sum_{s'=+,-} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(\kappa_{s'}^{ki} (\iota_{s'}^{ki} j)_{r_{s'}^{ki} t_{s'}^{ki} s})_{r_{s'}^{ki}}} \begin{array}{c} A' \\ \text{---} \end{array} \boxed{\tilde{k}} \begin{array}{c} BE \\ \text{---} \end{array} \\ = \frac{1}{2} \sum_{s'=+,-} \delta_{\tilde{k}, \kappa_{s'}^{ki}} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(\iota_{s'}^{ki} j)_{r_{s'}^{ki} t_{s'}^{ki} s}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} \quad .$$

From Postulate 4 and Eqs. (A1), (A2), and (A3), posing  $|\Sigma\rangle_{B'} := \sum_{k=1}^{D_{B'}} \alpha_k |k\rangle_{B'}$ ,  $\langle H |_{A'} := \sum_{\tilde{k}=1}^{D_{A'}} \beta_{\tilde{k}} \langle \tilde{k} |_{A'}$ , where for all  $k, \tilde{k}$  one has  $0 \leq \alpha_k, \beta_{\tilde{k}} \leq 1$  such that  $\sum_{k=1}^{D_{B'}} \alpha_k \in [0, 1]$  and  $\sum_{\tilde{k}=1}^{D_{A'}} \beta_{\tilde{k}} \in [0, D_{A'}]$ , the action of an arbitrary transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  with  $A \neq I$  has the following form:

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(ij)_s} \begin{array}{c} A \\ \text{---} \end{array} \boxed{\mathcal{T}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} = \frac{1}{2} \sum_{k=1}^{D_{B'}} \sum_{s'=+,-} \alpha_k \beta_{\kappa_{s'}^{ki}} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(\iota_{s'}^{ki} j)_{r_{s'}^{ki} t_{s'}^{ki} s}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} \quad . \quad (A4)$$

Let  $I \neq A \in \text{Sys(BCT)}$ ,  $B \in \text{Sys(BCT)}$ , and  $\mathcal{A}$  be a map defined by the following action:

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(ij)_s} \begin{array}{c} A \\ \text{---} \end{array} \boxed{\mathcal{A}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} = \lambda \delta_{ii_0} \begin{array}{c} \text{---} \\ \text{---} \end{array} \boxed{(lj)_{\tau s}} \begin{array}{c} B \\ \text{---} \end{array} \begin{array}{c} E \\ \text{---} \end{array} \quad , \quad (A5)$$

for some  $\lambda \in [0, 1]$ ,  $1 \leq i_0 \leq D_A$ ,  $1 \leq l \leq D_B$ , and  $\tau = \pm$ . Notice that such functions  $\mathcal{A}$  map states to states.

*Lemma 2 (Characterization of transformations).* Let  $I \neq A \in \text{Sys(BCT)}$ ,  $B \in \text{Sys(BCT)}$ , and  $\mathcal{A}$  be a map defined as in Eq. (A5). Then, every  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is a conical combination of elements of the form (A5).

*Proof.* Let  $I \neq A \in \text{Sys(BCT)}$  and  $B \in \text{Sys(BCT)}$ . On the one hand, we know that the action of an arbitrary

transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  has the form (A4). On the other hand, the action on  $\text{PurSt}(AE)$  of an arbitrary conical combination of generalized transformations of the form (A5) is given by

$$\begin{aligned} & \left( (ij)_s \right) \begin{array}{c} A \\ \boxed{\sum_n \mathcal{A}_n} \\ E \end{array} B \\ &= \sum_{m=1}^{D_B} \sum_{\tau=+,-} \lambda_{m,\tau}^{(i)} \left( (mj)_{\tau s} \right) \begin{array}{c} B \\ \boxed{\phantom{\sum_n \mathcal{A}_n}} \\ E \end{array}, \end{aligned} \quad (\text{A6})$$

where the  $\lambda_{m,\tau}^{(i)}$  are non-negative real numbers such that  $\sum_{m,\tau} \lambda_{m,\tau}^{(i)} \in [0, 1]$ . Now, for every  $1 \leq i \leq D_A$  let us define a set  $I_{m,\tau}(i)$  such that  $(k, s') \in I_{m,\tau}(i)$  if and only if  $r_s^{ki} = m$  and  $r_{s'}^{ki} = \tau$ .

For all  $i, m, \tau$  one can pose

$$\lambda_{m,\tau}^{(i)} := \frac{1}{2} \sum_{(k,s') \in I_{m,\tau}(i)} \alpha_k \beta_{k s'}.$$

Clearly, if  $I_{m,\tau}(i) = \emptyset$  one has  $\lambda_{m,\tau}^{(i)} = 0$ . This proves that Eq. (A4) can be rewritten in form (A6), namely, every  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is a conical combination of elements of the form (A5). ■

Let  $I \neq A \in \text{Sys}(\text{BCT})$  and  $B \in \text{Sys}(\text{BCT})$ ,  $h$  be any chosen function from  $D_A$  elements to  $D_B$  elements, and  $\xi$  any chosen function from  $D_A$  to  $\{+, -\}$  ( $\xi$  is set to be identically  $+$  if  $B = I$ ). Moreover, let  $B'' \in \text{Sys}(\text{BCT})$  such that  $D_{B''} = 2D_B^{D_A}$  and  $B' := B''B$ ,  $A' := B''A$ . Define  $\tilde{\mathcal{R}}_{A,B} \in \text{RevTransf}(B'A \rightarrow A'B)$  in the following way:

$$\begin{aligned} & \left( (\sigma_{h,\xi k})_{s_1} (ij)_{s_2} s_3 \right) \begin{array}{c} B'' \\ \boxed{\phantom{\tilde{\mathcal{R}}_{A,B}}} \\ A \end{array} B'' \\ &= \left( (\sigma_{h,\xi i})_{s_1} ([h(i) \oplus k]_{\xi(i)})_{s_2} s_3 \right) \begin{array}{c} B'' \\ \boxed{\phantom{\tilde{\mathcal{R}}_{A,B}}} \\ B \\ \phantom{B} \\ E \end{array}, \end{aligned} \quad (\text{A7})$$

where  $\oplus$  denotes the sum modulo  $D_B$ . The states  $\sigma_{h,\tau}$  in Eq. (A7) implement every possible pair of functions  $(h, \xi)$ . It is easy to realize that the transformation  $\tilde{\mathcal{R}}_{A,B}$  complies indeed with Postulate 3. In the remainder of this Appendix and in Appendix B, we will use  $\tilde{\mathcal{R}}_{A,B}$  as a universal processor for BCT's transformations.

**Lemma 3 (Realization of deterministic transformations).** Let  $I \neq A \in \text{Sys}(\text{BCT})$ ,  $B \in \text{Sys}(\text{BCT})$ , and  $\mathcal{D} \in \text{Transf}_{\mathbb{R}}(A \rightarrow B)$ . Then, the following are equivalent:

(A) For every  $i \in \{1, 2, \dots, D_A\}$  there exists a probability distribution  $\{\lambda_{m,\tau}^{(i)}\}_{(m,\tau) \in I}$ , with  $I = \{1, 2, \dots, D_B\} \times \{+, -\}$ ,

such that the following holds for all  $|(ij)_s\rangle_{AE} \in \text{PurSt}(AE)$ :

$$\left( (ij)_s \right) \begin{array}{c} A \\ \boxed{\mathcal{D}} \\ E \end{array} B = \sum_{(m,\tau) \in I} \lambda_{m,\tau}^{(i)} \left( (mj)_{\tau s} \right) \begin{array}{c} B \\ \boxed{\phantom{\mathcal{D}}} \\ E \end{array}. \quad (\text{A8})$$

(B)  $\mathcal{D} \in \text{Transf}_1(A \rightarrow B)$ .

(C) There exists  $|\Sigma\rangle_{B'} \in \text{St}_1(B')$  such that the following holds:

$$\begin{array}{c} A \\ \boxed{\mathcal{D}} \\ B \end{array} = \begin{array}{c} \Sigma \\ \boxed{\tilde{\mathcal{R}}_{A,B}} \\ A \end{array} \begin{array}{c} B' \\ \boxed{e} \\ B \end{array} \quad (\text{A9})$$

*Proof.* (C)  $\Rightarrow$  (B)  $\Rightarrow$  (A). The chain of implications holds by definition and by the characterization of deterministic transformations given in Sec. III B.

(A)  $\Rightarrow$  (C). Suppose that Eq. (A8) holds. We provide below an explicit construction of a family of states  $|\Sigma\rangle_{B'} \in \text{St}_1(B')$  such that Eq. (A9) holds. First, let  $|\Sigma\rangle_{B'} \in \text{St}_1(B')$  be of the form  $\sum_{(h,\xi) \in J} \mu_{h,\xi} |\sigma_{h,\xi}\rangle_{B'} |0\rangle_B$ . We provide a step-by-step construction for suitable families of sets  $J$ ,  $\{\mu_{h,\xi}\}_{(h,\xi) \in J}$  such that Eq. (A9) reads as Eq. (A8). Start by taking the minimum nonvanishing value  $\lambda_0 := \lambda_{m_0,\tau_0}^{(i_0)}$  over all the probability distributions  $\{\lambda_{m,\tau}^{(i)}\}_{(m,\tau) \in I}$  for  $i \in \{1, 2, \dots, D_A\}$ . In the case where the minimum is not unique, just arbitrarily pick one of them. Define  $h_0, \xi_0$  as those (families of) functions such that  $h_0(i_0) = m_0$ ,  $\xi_0(i_0) = \tau_0$  and  $h_0(i) = m_0^{(i)}$ ,  $\xi_0(i) = \tau_0^{(i)}$ , where  $m_0^{(i)}$ ,  $\tau_0^{(i)}$  are any chosen values such that  $\lambda_{m_0^{(i)},\tau_0^{(i)}}^{(i)}$  is nonvanishing for every  $i$ . In the collections  $\{\lambda_{m,\tau}^{(i)}\}_{(m,\tau) \in I}$  for  $i \in \{1, 2, \dots, D_A\}$ , reset the values of those coefficients  $\lambda_{\tilde{m},\tilde{\tau}}^{(i)}$ , where  $(\tilde{m}, \tilde{\tau})$  are in the image of  $(h_0, \xi_0)$ , in the following way:  $\lambda_{\tilde{m},\tilde{\tau}}^{(i)} \mapsto \lambda_{\tilde{m},\tilde{\tau}}^{(i)} - \lambda_0$ . By construction, this operation does not produce any negative value. Finally, set  $\mu_{h_0,\xi_0} := \lambda_0$ . Iterate the previous zeroth step. One realizes that the iteration of the above procedure has a finite number of steps, say  $N + 1$ , and eventually produces some families of sets  $\{(h_n, \xi_n)\}_{n=0}^N$  and  $\{\mu_{h_n,\xi_n}\}_{n=0}^N$ . Choose now one arbitrary set  $\{(h_n^*, \xi_n^*)\}_{n=0}^N$  and pose  $J = \{(h_n^*, \xi_n^*)\}_{n=0}^N$ . Using Eq. (A4) one verifies that, by construction, Eq. (A9) reads as Eq. (A8), namely, (A)  $\Rightarrow$  (C) holds. ■

**Corollary 1 (Deterministic transformations).** Let  $\mathcal{D} \in \text{Transf}(A \rightarrow B)$ . Then  $\mathcal{D} \in \text{Transf}_1(A \rightarrow B)$  if and only if it admits of a reversible dilation for  $\mathcal{D}$  of the form (40) with  $|\Sigma\rangle_{B'} \in \text{St}_1(B')$  and  $(H|_{A'} = |e\rangle_{A'})$ .

**Lemma 4 (Realization of arbitrary transformations).**

Let  $I \neq A \in \text{Sys}(\text{BCT})$ ,  $B \in \text{Sys}(\text{BCT})$ , and  $\mathcal{T} \in \text{Transf}_{\mathbb{R}}(A \rightarrow B)$ . Then, the following are equivalent:

(a) For every  $i \in \{1, 2, \dots, D_A\}$  there exists a set  $\{\gamma_{m,\tau}^{(i)}\}_{(m,\tau) \in I^{(i)}}$ , with  $I^{(i)} \subseteq \{1, 2, \dots, D_B\} \times \{+, -\}$ ,  $\gamma_{m,\tau}^{(i)} > 0$  for all  $(m, \tau) \in I^{(i)}$ , and  $\sum_{(m',\tau') \in I^{(i)}} \gamma_{m',\tau'}^{(i)} \leq 1$ , such the following holds that for all  $|(ij)_s\rangle_{AE} \in \text{PurSt}(AE)$ :

$$\left( (ij)_s \right) \begin{array}{c} A \\ \boxed{\mathcal{T}} \\ E \end{array} B = \sum_{(m,\tau) \in I^{(i)}} \gamma_{m,\tau}^{(i)} \left( (mj)_{\tau s} \right) \begin{array}{c} B \\ \boxed{\phantom{\mathcal{T}}} \\ E \end{array}. \quad (\text{A10})$$

(b)  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$ .

(c) There exists a deterministic state  $|\Sigma\rangle_{B'} \in \text{St}_1(B')$  such that, for every  $|(ij)_s\rangle_{AE} \in \text{PurSt}(AE)$ ,  $\mathcal{T} \boxtimes \mathcal{S}_E|(ij)_s\rangle_{AE}$  is in the refinement set of some  $\mathcal{D} \boxtimes \mathcal{S}_E|(ij)_s\rangle_{AE}$ , for  $\mathcal{D} \in \text{Transf}_1(A \rightarrow B)$  diluted as in Eq. (A9), and an effect  $(a|_{A'} \in \text{Eff}(A')$ , such that the following holds:

$$\begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}} \\
 \hline
 = \begin{array}{c}
 \text{B}' \\
 \Sigma \\
 \text{A} \quad \text{B} \\
 \boxed{\tilde{\mathcal{R}}_{A,B}} \\
 \text{A}' \quad \text{B} \\
 \text{a}
 \end{array}
 \end{array} \quad (\text{A11})$$

*Proof.* (c)  $\Rightarrow$  (b)  $\Rightarrow$  (a). The chain of implications holds by Postulate 4 and Lemma 2.

(a)  $\Rightarrow$  (c). For every map  $\mathcal{T} \in \text{Transf}_{\mathbb{R}}(A \rightarrow A)$  defined as in Eq. (A10), by Lemma 3 there exists a deterministic transformation  $\mathcal{D} \in \text{Transf}_1(A \rightarrow B)$  and a deterministic state  $|\Sigma\rangle_{B'}$  such that  $\mathcal{T} \boxtimes \mathcal{S}_E|(ij)_s\rangle_{AE}$  is in the refinement set of  $\mathcal{D} \boxtimes \mathcal{S}_E|(ij)_s\rangle_{AE}$  for every  $|(ij)_s\rangle_{AE} \in \text{PurSt}(AE)$ . It can be verified by direct computation that such a  $|\Sigma\rangle_{B'}$  is of the form  $\sum_{(h,\xi) \in J} \mu_{h,\xi} |\sigma_{h,\xi}\rangle_{B'} |0\rangle_B$ , and this has been indeed shown in the proof of Lemma 3. Now consider

$$\sum_{(h,\xi) \in J} \mu_{h,\xi} \begin{array}{c}
 \text{B}'' \\
 \sigma_{h,\xi} \\
 \text{B} \\
 0 \\
 \text{A} \\
 (ij)_s \\
 \text{E} \\
 \boxed{\tilde{\mathcal{R}}_{A,B}} \\
 \text{B}'' \\
 \text{A} \\
 \text{B}
 \end{array} \quad (\text{A12})$$

By construction and by direct inspection of expression (A12) via Eq. (A7), for all  $i$  there exists a suitably chosen set of coefficients  $\{\beta_{h,\xi}^{(i)}\}_{(h,\xi) \in J^{(i)}}$ , with  $J^{(i)} \subseteq J$ ,  $\beta_{h,\xi}^{(i)} > 0$ , and  $\sum_{(h',\xi') \in J^{(i)}} \beta_{h',\xi'}^{(i)} \leq 1$ , such that one attains for every  $\tilde{i}$

$$\begin{array}{c}
 \sum_{(\tilde{h},\tilde{\xi}) \in J^{(i)}} \beta_{\tilde{h},\tilde{\xi}}^{(i)} \begin{array}{c}
 \text{B}'' \\
 \Sigma \\
 \text{B} \\
 \text{A} \\
 (ij)_s \\
 \text{E} \\
 \boxed{\tilde{\mathcal{R}}_{A,B}} \\
 \text{B}'' \\
 \text{A} \\
 \text{B} \\
 \sigma_{\tilde{h},\tilde{\xi}} \\
 \tilde{i}
 \end{array} \\
 = \delta_{\tilde{i}} \sum_{(m,\tau) \in I^{(i)}} \gamma_{m,\tau}^{(i)} \begin{array}{c}
 \text{B} \\
 (mj)_{\tau s} \\
 \text{E}
 \end{array}
 \end{array}$$

Define now  $(a|_{A'} := \sum_{i=1}^{D_A} \sum_{h,\xi \in J^{(i)}} \beta_{h,\xi}^{(i)} (\sigma_{h,\xi}|_{B'}(i|_A \in \text{Eff}_{\mathbb{R}}(A')$ ). Then, one obtains, for all  $|(ij)_s\rangle_{AE} \in \text{PurSt}(AE)$ ,

$$\begin{array}{c}
 \text{B}' \\
 \Sigma \\
 \text{A} \quad \text{B} \\
 \boxed{\tilde{\mathcal{R}}_{A,B}} \\
 \text{A}' \quad \text{B} \\
 \text{a}
 \end{array} = \begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{T}} \\
 \text{E} \\
 (ij)_s
 \end{array}$$

Finally, by construction one observes that  $(a|_{A'} \in \text{Eff}(A')$  (see classification of BCT's effects in Sec. VIB), i.e., Eq. (A11) holds, and this concludes the proof.  $\blacksquare$

We stress that the statement of condition (c) requires a unique deterministic state  $|\Sigma\rangle_{B'}$  for every transformation  $\mathcal{T}$  such that  $\mathcal{T} \boxtimes \mathcal{S}_E|(ij)_s\rangle_{AE}$  refine  $\mathcal{D} \boxtimes \mathcal{S}_E|(ij)_s\rangle_{AE}$  for a fixed deterministic  $\mathcal{D} \in \text{Transf}_1(A \rightarrow B)$ . This is true, in particular, if  $\mathcal{T}$  refines  $\mathcal{D}$ . The latter property will be crucially exploited in the proof of Proposition 4 (see Appendix B).

*Corollary 2.* Let  $I \neq A \in \text{Sys}(\text{BCT})$  and  $B \in \text{Sys}(\text{BCT})$ . Then, every conical combination of elements of the form (A5) that maps states to states is a transformation of BCT.

*Proof.* This is a straightforward consequence of the fact that every conical combination of generalized transformations of the form (A5) can be expressed as in Eq. (A10), combined with implication (a)  $\Rightarrow$  (b) in Lemma 4.  $\blacksquare$

Lemma 2 and Corollary 2 provide a classification of BCT's transformations.

*Lemma 5 (Atomic transformations).* Let  $I \neq A \in \text{Sys}(\text{BCT})$  and  $B \in \text{Sys}(\text{BCT})$ . A map  $\mathcal{A} \in \text{Transf}_{\mathbb{R}}(A \rightarrow B)$  is an atomic transformation if and only if  $\mathcal{A} \boxtimes \mathcal{S}_E$  is of the form (A5) for every  $E \in \text{Sys}(\text{BCT})$ .

*Proof.* By Lemma 4, every map of the form (A5) is an admissible transformation of the theory. First, we show that every transformation of  $\mathcal{A} \in \text{Transf}(A \rightarrow B)$  such that  $\mathcal{A} \boxtimes \mathcal{S}_E$ , for every  $E \in \text{Sys}(\text{BCT})$ , is of the form (A5), satisfies the definition of atomicity [see Eq. (13)]. Indeed, we have already proven that the action of an arbitrary transformation  $\mathcal{T} \in \text{Transf}(A \rightarrow B)$  is given by Eq. (A4). Let us then pose

$$\begin{array}{c}
 \begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{A}} \\
 \text{E} \\
 (ij)_s
 \end{array} = \lambda \delta_{i i_0} \begin{array}{c}
 \text{B} \\
 (lj)_{\tau s} \\
 \text{E}
 \end{array} \\
 = \begin{array}{c}
 \text{A} \quad \text{B} \\
 \boxed{\mathcal{A}_1 + \mathcal{A}_2} \\
 \text{E} \\
 (ij)_s
 \end{array} \\
 = \frac{1}{2} \sum_{n=1,2} \sum_{k,s'} \alpha_k^n \beta_{k,n,s'}^n \begin{array}{c}
 \text{B} \\
 (l_{n,s'}^{ki} j)_{r^{ki} t^{ki} s s'} \\
 \text{E}
 \end{array}
 \end{array}$$

Now, since all the coefficients  $\lambda$ ,  $\alpha_k^n$ ,  $\beta_{k,n,s'}^n$  are non-negative, it must be for all  $|(ij)_s\rangle_{AE} \in \text{PurSt}(AE)$ :

$$(\mathcal{A}_n \boxtimes \mathcal{S}_E)|(ij)_s\rangle_{AE} = \begin{cases} 0, & i \neq i_0 \\ \lambda_n |(lj)_{\tau s}\rangle_{BE}, & i = i_0 \end{cases}$$

for  $n = 1, 2$  and non-negative coefficients  $\lambda_n$  such that  $\lambda_1 + \lambda_2 = \lambda$ . It follows that  $\mathcal{A}_1 \propto \mathcal{A}_2$ , namely, those transformations  $\mathcal{A} \in \text{Transf}(A \rightarrow B)$ , such that  $\mathcal{A} \boxtimes \mathcal{S}_E$  for every  $E \in \text{Sys}(\text{BCT})$  is of the form (A5), are atomic. Conversely, let  $\mathcal{A} \boxtimes \mathcal{S}_E$  be a transformation from  $AE$  to  $BE$ . By Lemma 2, we can write, without loss of generality,  $\mathcal{A} = \sum_{n \in N} \mathcal{A}_n$ , where  $\mathcal{A}_n \boxtimes \mathcal{S}_E$  is of the form (A5) for all  $E \in \text{Sys}(\text{BCT})$ ,  $n \in N$  and  $\mathcal{A}_{n_1} \not\propto \mathcal{A}_{n_2}$  for all  $n_1 \neq n_2$ . Accordingly, by Lemma 4, for any  $\tilde{n} \in N$  both  $\mathcal{A}_{\tilde{n}} \boxtimes \mathcal{S}_E$  and  $(\mathcal{A} - \mathcal{A}_{\tilde{n}}) \boxtimes \mathcal{S}_E$  are transformations of BCT. Now, suppose that  $\mathcal{A}$  is atomic, namely,  $\mathcal{A}_{\tilde{n}} \propto (\mathcal{A} - \mathcal{A}_{\tilde{n}})$ . This implies that  $\mathcal{A} \propto \mathcal{A}_{\tilde{n}}$ , i.e.,  $\mathcal{A} \boxtimes \mathcal{S}_E$  is of the form (A5).  $\blacksquare$

Proposition 3 follows combining Lemmas 2 and 5, and Corollaries 1 and 2.

**APPENDIX B: OPERATIONAL REALIZATION SCHEME  
FOR ARBITRARY INSTRUMENTS IN BCT  
(PROOF OF PROPOSITION 4)**

In this Appendix, in order to make sense of expressions of the form  $|(l_1 l_2)_u\rangle_{S_1 S_2}$  with  $S_1 = I$  or  $S_2 = I$ , these can be safely replaced by the expressions  $|(l_1)_+\rangle_{S_2}$  or  $|(l_1)_+\rangle_{S_1}$ , respectively.

*Lemma 6.* Let  $A, B \in \mathbf{Sys}(\mathbf{BCT})$  and  $\mathcal{T} \in \mathbf{Transf}_{\mathbb{R}}(A \rightarrow B)$  such that  $\mathcal{T} \boxtimes \mathcal{S}_E$  maps  $\mathbf{St}(AE)$  to  $\mathbf{St}(BE)$  for all  $E \in \mathbf{Sys}(\mathbf{BCT})$ . Then,  $\mathcal{T} \in \mathbf{Transf}(A \rightarrow B)$ .

*Proof.* Let  $\mathcal{T} \in \mathbf{Transf}_{\mathbb{R}}(A \rightarrow C)$  be a generalized transformation such that  $\mathcal{T} \boxtimes \mathcal{S}_E$  maps  $\mathbf{St}(AE)$  to  $\mathbf{St}(BE)$  for all  $E \in \mathbf{Sys}(\mathbf{BCT})$ . By simpliciality,  $\mathcal{T}$  may take the following general form:

$$\mathcal{T} \boxtimes \mathcal{S}_B |(ij)_s\rangle_{AB} = \sum_{m,l,\tau} \lambda_{m,l,\tau}^{i,j,s} |(ml)_{\tau s}\rangle_{CB}, \quad (\text{B1})$$

with  $\lambda_{m,l,\tau}^{i,j,s} > 0$  for all  $i, j, s, m, l, \tau$  and  $\sum_{m,l,\tau} \lambda_{m,l,\tau}^{i,j,s} \leq 1$ . Let now  $(j'|_B \in \mathbf{Eff}(B)$  be such that  $(j'|_B = \delta_{j'j}$  for all  $|j\rangle_B \in \mathbf{PurSt}(B)$ .

Imposing Eq. (33), namely,

$$[\mathcal{S}_C \boxtimes (j'|_B)](\mathcal{T} \boxtimes \mathcal{S}_B) = (\mathcal{T} \boxtimes \mathcal{S}_B)[\mathcal{S}_A \boxtimes (j'|_B)],$$

and using Eqs. (42), one has for all  $i, j, j', s$

$$\begin{aligned} & [\mathcal{S}_C \boxtimes (j'|_B)](\mathcal{T} \boxtimes \mathcal{S}_B) |(ij)_s\rangle_{AB} \\ &= [\mathcal{S}_C \boxtimes (j'|_B)] \sum_{m,l,\tau} \lambda_{m,l,\tau}^{i,j,s} |(ml)_{\tau s}\rangle_{CB} = \sum_{m,\tau} \lambda_{m,j',\tau}^{i,j,s} |m\rangle_C \\ &= (\mathcal{T} \boxtimes \mathcal{S}_B)[\mathcal{S}_A \boxtimes (j'|_B)] |(ij)_s\rangle_{AB} = \delta_{j'j} \mathcal{T} |i\rangle_A. \end{aligned}$$

The above equation implies that Eq. (B1) takes the form

$$\mathcal{T} \boxtimes \mathcal{S}_B |(ij)_s\rangle_{AB} = \sum_{m,\tau} \lambda_{m,\tau}^{i,j,s} |(mj)_{\tau s}\rangle_{CB}. \quad (\text{B2})$$

Accordingly, imposing now

$$(\mathcal{S}_C \boxtimes \mathcal{R}')(\mathcal{T} \boxtimes \mathcal{S}_B) = (\mathcal{T} \boxtimes \mathcal{S}_B)(\mathcal{S}_A \boxtimes \mathcal{R}')$$

for all  $\mathcal{R}' \in \mathbf{RevTransf}(B \rightarrow B)$  (see Postulate 3), one has for all  $i, j, s, \pi', \sigma'_m$

$$\begin{aligned} & (\mathcal{S}_C \boxtimes \mathcal{R}')(\mathcal{T} \boxtimes \mathcal{S}_B) |(ij)_s\rangle_{AB} \\ &= (\mathcal{S}_C \boxtimes \mathcal{R}') \sum_{m,\tau} \lambda_{m,\tau}^{i,j,s} |(mj)_{\tau s}\rangle_{CB} \\ &= \sum_{m,\tau} \lambda_{m,\tau}^{i,j,s} |(m\pi'(j))_{\sigma'_m \tau s}\rangle_{CB} \\ &= (\mathcal{T} \boxtimes \mathcal{S}_B)(\mathcal{S}_C \boxtimes \mathcal{R}') |(ij)_s\rangle_{AB} \\ &= \sum_{m,\tau} \lambda_{m,\tau}^{i,\pi'(j),\sigma'_m s} |(m\pi'(j))_{\tau \sigma'_m s}\rangle_{CB}. \end{aligned}$$

The above equation implies that the coefficients  $\lambda_{m,\tau}^{i,j,s}$  cannot depend on  $j, s$  for all  $i, m, \tau$ , namely, Eq. (B2) takes the

following form:

$$\mathcal{T} \boxtimes \mathcal{S}_B |(ij)_s\rangle_{AB} = \sum_{m,\tau} \lambda_{m,\tau}^i |(mj)_{\tau s}\rangle_{CB}. \quad (\text{B3})$$

Using Proposition 3, one realizes from Eq. (B3) that  $\mathcal{T} \in \mathbf{Transf}(\mathbf{BCT})$ , and this concludes the proof.  $\blacksquare$

We now use Lemmas 3, 4, and 6 in order to prove Proposition 4.

(a) (iii)  $\Rightarrow$  (ii)  $\Rightarrow$  (i). Recalling Postulate 6, the chain of implications is trivial.

(b) (i)  $\Rightarrow$  (iii). Let  $\mathbb{T}_X^{A \rightarrow B} \subset \mathbf{Transf}_{\mathbb{R}}(A \rightarrow B)$  such that  $\mathbb{T}_X^{A \rightarrow B} \boxtimes \mathbb{I}_X^{E \rightarrow E}$  maps preparation instruments of AE to preparation instruments of BE for all  $E \in \mathbf{Sys}(\mathbf{BCT})$ .

Let us first prove the implication in the case where  $A = I$ . Choosing  $E = I$ , by hypothesis we have that  $\mathbb{T}_X^{A \rightarrow B}$  maps the unity  $1 \in \mathbf{PurSt}(I)$  to a preparation instrument of B. Namely,  $\mathbb{T}_X^{A \rightarrow B}$  is a preparation instrument of BCT. Let us denote  $\rho_X^{I \rightarrow B} := \mathbb{T}_X^{A \rightarrow B}$  and  $\rho = \sum_{i \in I} \lambda_i |i\rangle_{B'} := \sum_{x \in X} \rho_x \in \mathbf{St}_1(B)$  with  $\lambda_i > 0$  for all  $i \in I$  and  $\sum_{i \in I} \lambda_i = 1$ . Then, the set  $\{\rho_x\}_{x \in X}$  is a refinement of  $\rho$ . Take  $B' = BB$ ,  $A' = B$ ,  $\mathcal{R} = \mathcal{S}_{BB} \in \mathbf{RevTransf}(B'A \rightarrow A'B)$ , and  $|\Sigma\rangle_{B'} = \sum_{i \in I} \lambda_i |i\rangle_{B|B}$ , so that  $\rho = [(e|_{A'} \boxtimes \mathcal{S}_B) \mathcal{R}] |\Sigma\rangle_{B'}$ . By simpliciality, every  $\rho_x$  can be rewritten as  $\sum_{i \in I} \gamma_i^{(x)} |i\rangle_B$ , for suitably defined  $\gamma_i^{(x)}$  such that  $0 \leq \gamma_i^{(x)} \leq \lambda_i$  and  $\sum_{x \in X} \gamma_i^{(x)} = \lambda_i$  for every  $i \in I$ . Define now, for every  $x \in X$  and  $i \in I$ ,  $\beta_i^{(x)} := \gamma_i^{(x)} / \lambda_i$ . It is clear that, by definition, it must be  $0 \leq \beta_i^{(x)} \leq 1$  and  $\sum_{x \in X} \beta_i^{(x)} = 1$  for every  $x \in X$  and  $i \in I$ . Now, defining  $(a_x|_{A'} := \sum_i \beta_i^{(x)} (i|_{A'} \in \mathbf{Eff}_{\mathbb{R}}(A')$  for every  $x \in X$ , one has  $\rho_x = [(a_x|_{A'} \boxtimes \mathcal{S}_B) \mathcal{R}] |\Sigma\rangle_{B'}$ . Finally, since by the classification of BCT's effects (see Sec. VIB) one has  $(a_x|_{A'} \in \mathbf{Eff}(A')$  for all  $x \in X$ , and being  $\sum_{x \in X} (a_x|_{A'} = (e|_{A'}$ , by Postulate 6 one can conclude  $\{(a_x|_{A'}\}_{x \in X} \in \mathbf{Instr}(\mathbf{BCT})$ .

Let now be  $A \neq I$  and  $\tilde{\mathcal{R}}_{A,B} \in \mathbf{RevTransf}(B'A \rightarrow A'B)$  defined as in Eq. (A7). By hypothesis,  $\mathcal{T}_x \boxtimes \mathcal{S}_E$  maps  $\mathbf{St}(AE)$  to  $\mathbf{St}(BE)$  for every  $x \in X$  and  $E \in \mathbf{Sys}(\mathbf{BCT})$ . Accordingly, by Lemma 6,  $\mathcal{T}_x \in \mathbf{Transf}(A \rightarrow B)$  for every  $x \in X$ . On the one hand, condition (A) of Lemma 3 holds for  $\mathcal{D} = \sum_{x \in X} \mathcal{T}_x$ . On the other hand, condition (b) of Lemma 4 holds for  $\mathcal{T} = \mathcal{T}_x$ , and for every  $x \in X$ . Thus, by construction, and invoking implications (A)  $\Rightarrow$  (B) in Lemma 3 and (b)  $\Rightarrow$  (c) in Lemma 4, there exists a state  $|\Sigma\rangle_{B'} \in \mathbf{St}_1(B')$  such that the following holds:

$$A \xrightarrow{\quad} \boxed{\sum_{x \in X} \mathcal{T}_x} \xrightarrow{\quad} B = \begin{array}{c} \text{B}' \\ \Sigma \\ \text{A} \end{array} \boxed{\tilde{\mathcal{R}}_{A,B}} \begin{array}{c} \text{A}' \\ e \\ \text{B} \end{array} \xrightarrow{\quad} .$$

We have thus shown that the collection  $\mathbb{T}_X^{A \rightarrow B} \subset \mathbf{Transf}(A \rightarrow B)$  is a refinement for the deterministic transformation  $\mathcal{D} := \sum_{x \in X} \mathcal{T}_x$ . Considering Eq. (A12), we know that  $|\Sigma\rangle_{B'}$  can be taken of the form  $\sum_{(h,\xi) \in J} \mu_{h,\xi} |\sigma_{h,\xi}\rangle_{B'} |0\rangle_B$  with positive coefficients  $\mu_{h,\xi}$ . Moreover,  $\mathcal{D}$  can be decomposed

into conical combinations of atomic maps

$$\begin{aligned}
 & \delta_{(h,\xi)(\tilde{h},\tilde{\xi})} \left( (ij)_s \right) \begin{array}{c} \text{A} \\ \boxed{\mathcal{A}_i^{(h,\xi)}} \\ \text{B} \\ \text{E} \end{array} \\
 & \equiv \begin{array}{c} \sigma_{h,\xi} \text{ B}'' \\ 0 \text{ B} \\ (ij)_s \text{ A} \end{array} \begin{array}{c} \text{B}'' \\ \boxed{\tilde{\mathcal{R}}_{A,B}} \\ \text{A} \\ \text{B} \\ \text{E} \end{array} \begin{array}{c} \sigma_{\tilde{h},\tilde{\xi}} \\ \tilde{i} \end{array} \\
 & = \delta_{(h,\xi)(\tilde{h},\tilde{\xi})} \delta_{\tilde{i}} \left( (h(i)j)_{\xi(i)s} \right) \begin{array}{c} \text{B} \\ \text{E} \end{array} . \tag{B4}
 \end{aligned}$$

In other terms, the following decomposition holds:

$$\mathcal{D} = \sum_{\tilde{i}=1}^{D_A} \sum_{(h,\xi) \in J} \mu_{h,\xi} \mathcal{A}_i^{(h,\xi)} . \tag{B5}$$

Now, in the light of Lemma 5, the decomposition of a transformation into atomic transformations is unique up to trivial refinements, namely, refinements where the elements are proportional to each other. This implies that every refinement of  $\mathcal{D}$  must consist in a trivial refinement and subsequent coarse graining of the decomposition in Eq. (B5). Consequently, for all  $x' \in \mathbf{X}$ ,  $(h', \xi') \in J$ ,  $\tilde{i}' \in \{1, 2, \dots, D_A\}$ , it must be

$$\begin{aligned}
 \mathcal{T}_{x'} &= \sum_{\tilde{i}=1}^{D_A} \sum_{(h,\xi) \in J} v_{(h,\xi),\tilde{i}}^{(x')} \mathcal{A}_i^{(h,\xi)}, \quad v_{(h',\xi'),\tilde{i}'}^{(x')} \in [0, 1], \\
 \sum_{x \in \mathbf{X}} v_{(h',\xi'),\tilde{i}'}^{(x)} &= \mu_{h',\xi'}, \quad \sum_{(h,\xi) \in J} v_{(h,\xi),\tilde{i}}^{(x')} \leq 1.
 \end{aligned}$$

Now, let us define the following collection of coefficients:

$$\zeta_{(h,\xi),\tilde{i}}^{(x)} := \frac{v_{(h,\xi),\tilde{i}}^{(x)}}{\mu_{h,\xi}} \in [0, 1],$$

$$\forall x \in \mathbf{X}, (h, \xi) \in J, \tilde{i} \in \{1, 2, \dots, D_A\}.$$

Accordingly, each  $\mathcal{T}_x$  can be achieved as follows:

$$\begin{aligned}
 (a_x|_{A'}) &:= \sum_{\tilde{i}=1}^{D_A} \sum_{(h,\xi) \in J} \zeta_{(h,\xi),\tilde{i}}^{(x)} (\sigma_{h,\xi}|_{B''}(\tilde{i}|_A), \\
 & \begin{array}{c} \Sigma \text{ B}' \\ \text{A} \end{array} \begin{array}{c} \text{A}' \\ \boxed{\tilde{\mathcal{R}}_{A,B}} \\ \text{B} \end{array} (a_x) \\
 & = \sum_{\tilde{i}=1}^{D_A} \sum_{(h,\xi) \in J} \mu_{h,\xi} \zeta_{(h,\xi),\tilde{i}}^{(x)} \mathcal{A}_i^{(h,\xi)} \equiv \begin{array}{c} \text{A} \\ \boxed{\mathcal{T}_x} \\ \text{B} \end{array} .
 \end{aligned}$$

By construction, and by the classification of BCT's effects given in Sec. VI B,  $\{(a_x|_{A'})_{x \in \mathbf{X}} \subset \text{Eff}(A')$  and  $\sum_{x \in \mathbf{X}} (a_x|_{A'} = \sum_{(\tilde{h},\tilde{\xi}) \in J} (\sigma_{\tilde{h},\tilde{\xi}}|_{B''}(e|_A)$  hold. In the case where

$\sum_{(\tilde{h},\tilde{\xi}) \in J} (\sigma_{\tilde{h},\tilde{\xi}}|_{B''} \neq (e|_{B''})$ , one could complete the collection  $\{(a_x|_{A'})_{x \in \mathbf{X}}$  adding the effect  $\sum_{(\tilde{h},\tilde{\xi}) \in \tilde{J}} (\sigma_{\tilde{h},\tilde{\xi}}|_{B''}(e|_A)$ , where  $\tilde{J}$  collects all the pairs  $(h, \xi) \notin J$ , to any of the effects in the collection, say e.g.,  $(a_{x_0}|_{A'})$ . This simply amounts to adding the associated null transformation to the corresponding transformation  $\mathcal{T}_{x_0}$  since  $(\sigma_{\tilde{h},\tilde{\xi}}|_{B''}(\sigma_{h',\xi'}))_{B''} = 0$  for every  $(\tilde{h}, \tilde{\xi}) \in \tilde{J}$  and  $(h', \xi') \in J$ . Then, by the first part of Postulate 6, the collection of effects  $\{(a_x|_{A'} \boxtimes \mathcal{C})_{x \in \mathbf{X}}$  maps preparation instruments of AC to preparation instruments of C for all  $C \in \text{Sys}(\text{BCT})$ . Finally, by the second part of Postulate 6, one can conclude that  $\{(a_x|_{A'})_{x \in \mathbf{X}} \in \text{Instr}(\text{BCT})$ .

### APPENDIX C: CONDITIONAL INSTRUMENTS IN THEORIES WITH A UNIQUE DETERMINISTIC EFFECT

We characterize those causal theories (see Property 3) which satisfy Postulate 6 and Property 7, proving that they contain every possible conditional instrument, namely, they satisfy Property 4.

*Theorem 3.* Let  $\Theta$  be an OPT satisfying Property 3, Postulate 6, and Property 7. Then, the theory  $\Theta$  also satisfies Property 4.

*Proof.* We denote the unique deterministic effect of each  $A \in \text{Sys}(\Theta)$  by  $(e|_A)$ . Suppose that  $\rho_X^{I \rightarrow \text{AE}}$  and  $\mathbf{A}_Y^{A \rightarrow B}$  are instruments of  $\Theta$ , and let  $\{\mathcal{B}_Z^{(y)}\}_{y \in Y} \subset \text{Instr}(B \rightarrow C)$  a collection of instruments labeled by  $y \in Y$ . Consider now the following collection of transformations:

$$\mathbf{P} := \bigcup_{y \in Y} \{(\mathcal{B}_Z^{(y)} \mathcal{A}_y \boxtimes \mathcal{I}_E) \rho_x\}_{(x,z) \in \mathbf{X} \times \mathbf{Z}^{(y)}} \in \text{Instr}_{\mathbb{R}}(I \rightarrow C).$$

Now,  $(\mathcal{B}_Z^{(y)} \mathcal{A}_y \boxtimes \mathcal{I}_E) \rho_x \in \text{St}(\text{AE})$  for all  $x \in \mathbf{X}, y \in Y, z \in \mathbf{Z}^{(y)}$  by hypothesis. In addition, by Property 3 and using the characterization of deterministic transformations given in Sec. III B, one has

$$\begin{aligned}
 (e|_{\text{CE}} \sum_{z \in \mathbf{Z}^{(y)}} \mathcal{B}_z^{(y)} \boxtimes \mathcal{I}_E) &= (e|_{\text{BE}}, \quad \forall y \in Y \\
 (e|_{\text{BE}} \sum_{y \in Y} \mathcal{A}_y \boxtimes \mathcal{I}_E) &= (e|_{\text{AE}}, \\
 (e|_{\text{AE}} \sum_{x \in \mathbf{X}} \rho_x \boxtimes \mathcal{I}_E) &= 1.
 \end{aligned}$$

Accordingly,

$$(e|_{\text{CE}} \sum_{x \in \mathbf{X}} \sum_{y \in Y} \sum_{z \in \mathbf{Z}^{(y)}} (\mathcal{B}_z^{(y)} \mathcal{A}_y \boxtimes \mathcal{I}_E) \rho_x = 1,$$

and then, by Postulate 6 and Property 7, and recalling again the characterization of deterministic instruments, one concludes that  $\mathbf{P} \in \text{Instr}(I \rightarrow C)$ . Namely, the conditional generalized instrument  $\bigcup_{y \in Y} \{\mathcal{B}_z^{(y)} \mathcal{A}_y\}_{z \in \mathbf{Z}^{(y)}}$  is an instrument of the theory  $\Theta$ . ■

### APPENDIX D: HOMOGENEOUS STRICT BILOCAL DISCRIMINABILITY AND ESSENTIAL UNIQUENESS OF PURIFICATION IMPLY POSTULATES (36) AND (37) IN A SIMPLICIAL THEORY

Let  $\Theta$  be a simplicial theory, and define, for all pure states  $|i\rangle_A \in \text{PurSt}(A)$ ,  $|j\rangle_B \in \text{PurSt}(B)$ , the set  $\text{PurRef}_{ij}(AB) \subset$

$\text{PurSt}(AB)$  collecting those pure states of  $AB$  which convexly refine the product state  $|i\rangle_A|j\rangle_B$ . Moreover, define

$$n_{AB}^{ij} := |\text{PurRef}_{ij}(AB)|,$$

$$\text{PurRef}(AB) := \bigcup_{\substack{1 \leq i \leq D_A \\ 1 \leq j \leq D_B}} \text{PurRef}_{ij}(AB),$$

$$\overline{\text{PurRef}}(AB) := \text{PurSt}(AB) \setminus \text{PurRef}(AB),$$

$$l_{AB} := |\overline{\text{PurRef}}(AB)|.$$

By direct inspection of the proof of Theorem 2 in Ref. [8], one easily verifies that, for all  $(i, j) \neq (i', j')$ ,  $\text{PurRef}_{ij}(AB) \cap \text{PurRef}_{i'j'}(AB) = \emptyset$ . Accordingly, in a simplicial theory  $\Theta$  each  $|\rho\rangle_{AB} \in \text{PurRef}(AB)$  can be unambiguously labeled as follows:

$$|\rho\rangle_{AB} = |(ij)_\sigma\rangle_{AB}, \quad \sigma \in \{1, 2, \dots, n_{AB}^{ij}\}.$$

Then, for all systems  $A, B$  and  $i, j, \sigma$ , the following holds:

$$D_{AB} = \sum_{\substack{1 \leq i \leq D_A \\ 1 \leq j \leq D_B}} n_{AB}^{ij} + l_{AB}, \quad (\text{D1})$$

$$|i\rangle_A|j\rangle_B = \sum_{\sigma'=1}^{n_{AB}^{ij}} p_{\sigma'}^{ij} |(ij)_{\sigma'}\rangle_{AB}, \quad p_{\sigma'}^{ij} > 0, \quad \sum_{\sigma'=1}^{n_{AB}^{ij}} p_{\sigma'}^{ij} = 1. \quad (\text{D2})$$

*Theorem 4.* Let  $\Theta$  be a simplicial theory satisfying Property 8. Then, for all systems  $A, B \in \text{Sys}(\Theta)$ , there exists a positive integer  $n_{AB}$  such that the following holds. For all pure states  $|i\rangle_A \in \text{PurSt}(A)$ ,  $|j\rangle_B \in \text{PurSt}(B)$ , permutation  $\pi$  of  $D_A$  elements, there exists a reversible transformation  $\mathcal{R} \in \text{RevTransf}(A \rightarrow A)$  and a permutation  $\kappa_{ij}$  of  $n_{AB}$  elements, such that

$$n_{AB}^{ij} = n_{AB}, \quad (\text{D3})$$

$$(\mathcal{R} \boxtimes \mathcal{I}_B)|(ij)_\sigma\rangle_{AB} = |(\pi(ij))_{\kappa_{ij}(\sigma)}\rangle_{AB}, \quad (\text{D4})$$

$$D_{AB} = n_{AB}D_AD_B + l_{AB}, \quad (\text{D5})$$

$$|i\rangle_A|j\rangle_B = \frac{1}{n_{AB}} \sum_{\sigma=1}^{n_{AB}} |(ij)_\sigma\rangle_{AB}. \quad (\text{D6})$$

*Proof.* Property 8, choosing  $A = I$ , implies *transitivity of reversible channels on pure states*. That is, for every system  $A$  and every permutation  $\pi$  of  $D_A$  elements, there exists a reversible transformation  $\mathcal{R} \in \text{RevTransf}(A \rightarrow A)$  such that

$\mathcal{R}|i\rangle_A = |\pi(i)\rangle_A$  for all  $|i\rangle_A \in \text{PurSt}(A)$ . Let  $\pi$  and  $\pi'$  denote two permutations of, respectively,  $D_A$  and  $D_B$  elements, and define  $\tilde{i} := \pi(i)$ ,  $\tilde{j} := \pi'(j)$  for all  $i \in \{1, \dots, D_A\}$ ,  $j \in \{1, \dots, D_B\}$ . Then, combining Property 8 with Proposition 1, for all  $\pi, \pi' \in \text{RevTransf}(A \rightarrow A)$ ,  $\mathcal{R}' \in \text{RevTransf}(B \rightarrow B)$ , such that the following holds:

$$|\tilde{i}\rangle_A := \mathcal{R}|i\rangle_A \in \text{PurSt}(A), \quad (\text{D7})$$

$$|\tilde{j}\rangle_B := \mathcal{R}'|j\rangle_B \in \text{PurSt}(B).$$

Still by Proposition 1, we can also denote

$$(\mathcal{R} \boxtimes \mathcal{R}')|(ij)_\sigma\rangle_{AB} = |(i_\sigma^{ij} j_\sigma^{ij})_{\kappa_{ij}^{ij}}\rangle_{AB}. \quad (\text{D8})$$

Recall now Eq. (D2), which holds by simpliciality. Equations (D7) and (D8), combined with Eq. (D2), read as

$$|\tilde{i}\rangle_A|\tilde{j}\rangle_B = \sum_{\sigma=1}^{n_{AB}^{\tilde{i}\tilde{j}}} p_{\sigma}^{\tilde{i}\tilde{j}} |(\tilde{i}\tilde{j})_\sigma\rangle_{AB} = (\mathcal{R} \boxtimes \mathcal{R}')|i\rangle_A|j\rangle_B$$

$$= \sum_{\sigma=1}^{n_{AB}^{ij}} p_{\sigma}^{ij} |(i_\sigma^{ij} j_\sigma^{ij})_{\kappa_{ij}^{ij}}\rangle_{AB}. \quad (\text{D9})$$

First of all, by simpliciality, Eq. (D9) implies  $i_\sigma^{ij} = \tilde{i}$ ,  $j_\sigma^{ij} = \tilde{j}$  for all  $i \in \{1, \dots, D_A\}$ ,  $j \in \{1, \dots, D_B\}$ ,  $\sigma \in \{1, \dots, n_{AB}^{ij}\}$ . Moreover, by Proposition 1, for all  $i, j, \tilde{i}, \tilde{j}, k$  it must be

$$n_{AB}^{ij} = n_{A'B'}^{\tilde{i}\tilde{j}}, \quad \kappa_{ij}^{ij} = \kappa_{ij}(\sigma), \quad p_{\kappa_{ij}(\sigma)}^{\tilde{i}\tilde{j}} = p_{\sigma}^{ij}, \quad (\text{D10})$$

where  $\kappa_{ij}$  is a permutation of  $n_{AB}^{ij}$  elements. This proves Eqs. (D3), (D4), and, recalling (D1), (D5). For all  $i, j, \sigma$ , by simpliciality,  $|(ij)_\sigma\rangle_{AB}$  is a purification of  $|i\rangle_A$  and  $|j\rangle_B$ . Thus, by Property 8, in Eq. (D10) the permutation  $\kappa_{ij}$  in  $\text{PurRef}_{ij}(AB)$  can be arbitrarily chosen when  $i = \tilde{i}$  and  $j = \tilde{j}$ . Since  $p_{\sigma}^{ij} > 0$  and  $\sum_{\sigma'=1}^{n_{AB}^{ij}} p_{\sigma'}^{ij} = 1$  for all  $i, j, \sigma$ , then one has  $p_{\sigma}^{ij} = 1/n_{AB}$  for all  $i, j, \sigma$ . This finally also proves Eq. (D6). ■

*Corollary 3.* Let  $\Theta$  be a simplicial theory satisfying Property 8,  $I \neq A, B \in \text{Sys}(\Theta)$ , and  $i \in \{1, 2, \dots, D_A\}$ ,  $j \in \{1, 2, \dots, D_B\}$ . Then,  $\Theta$  also satisfies strict bilocal discriminability, with  $D_{AB} > D_AD_B$ , if and only if  $l_{AB} = 0$  and  $n_{AB}^{ij} = 2$ .

*Proof.* By Theorem 4 of Ref. [8], which holds for simplicial theories with  $n$ -local discriminability for some positive integer  $n$ , one has that  $l_{AB} = 0$  for all  $A, B \in \text{Sys}(\Theta)$ . Now, it suffices to plug Eq. (D5) into Eq. (23) of Theorem 2. Solving for  $n_{AB}$ , one finds the two solutions  $n_{AB} = 1$  or  $2$ , and then it must be  $n_{AB} = 2$  for every  $I \neq A, B \in \text{Sys}(\Theta)$ . The converse has been proven in Proposition 2. ■

- [1] E. Schrödinger, Discussion of probability relations between separated systems, *Math. Proc. Cambridge Philos. Soc.* **31**, 555 (1935).
- [2] J. Barrett, Information processing in generalized probabilistic theories, *Phys. Rev. A* **75**, 032304 (2007).
- [3] P. Rastall, Locality, Bell's theorem, and quantum mechanics, *Found. Phys.* **15**, 963 (1985).

- [4] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
- [5] 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).
- [6] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Generalized No-Broadcasting Theorem, *Phys. Rev. Lett.* **99**, 240501 (2007).

- [7] C. Pfister and S. Wehner, An information-theoretic principle implies that any discrete physical theory is classical, *Nat. Commun.* **4**, 1851 (2013).
- [8] G. M. D’Ariano, M. Erba, and P. Perinotti, Classical theories with entanglement, *Phys. Rev. A* **101**, 042118 (2020).
- [9] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, A structure theorem for generalized-noncontextual ontological models, [arXiv:2005.07161](https://arxiv.org/abs/2005.07161).
- [10] M. Dall’Arno, S. Brandsen, A. Tosini, F. Buscemi, and V. Vedral, No-Hypersignaling Principle, *Phys. Rev. Lett.* **119**, 020401 (2017).
- [11] S. Mac Lane, *Categories for the Working Mathematician*, Graduate Texts in Mathematics (Springer, New York, 2013).
- [12] 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, Dordrecht, 2016), pp. 171–221.
- [13] B. Coecke and A. Kissinger, *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning* (Cambridge University Press, Cambridge, 2017).
- [14] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Probabilistic theories with purification, *Phys. Rev. A* **81**, 062348 (2010).
- [15] G. Chiribella and C. M. Scandolo, Entanglement and thermodynamics in general probabilistic theories, *New J. Phys.* **17**, 103027 (2015).
- [16] S. Gogioso and C. M. Scandolo, Categorical probabilistic theories, *Electron. Proc. Theor. Comput. Sci.* **266**, 367 (2018).
- [17] J. H. Selby, A process theoretic triptych: two roads to the emergence of classicality, reconstructing quantum theory from diagrams, looking for post-quantum theories, Ph.D. thesis, Imperial College London, 2017.
- [18] S. Tull, Categorical operational physics, Ph.D. thesis, University of Oxford, 2018.
- [19] P. Janotta and H. Hinrichsen, Generalized probability theories: What determines the structure of quantum theory? *J. Phys. A: Math. Theor.* **47**, 323001 (2014).
- [20] B. Coecke and É. Paquette, Categories for the practising physicist, in *New Structures for Physics*, edited by B. Coecke (Springer, Berlin, 2011), pp. 173–286.
- [21] L. Hardy and W. K. Wootters, Limited holism and real-vector-space quantum theory, *Found. Phys.* **42**, 454 (2012).
- [22] G. M. D’Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, Cambridge, 2017).
- [23] G. Chiribella and C. M. Scandolo, Operational axioms for diagonalizing states, *Electron. Proc. Theor. Comput. Sci.* **195**, 96 (2015).
- [24] G. M. D’Ariano, F. Manessi, P. Perinotti, and A. Tosini, The Feynman problem and fermionic entanglement: Fermionic theory versus qubit theory, *Int. J. Mod. Phys. A* **29**, 1430025 (2014).
- [25] P. Janotta and R. Lal, Generalized probabilistic theories without the no-restriction hypothesis, *Phys. Rev. A* **87**, 052131 (2013).
- [26] Y. Guryanova, R. Silva, A. J. Short, P. Skrzypczyk, N. Brunner, and S. Popescu, Exploring the limits of no backwards in time signalling, *Quantum* **3**, 211 (2019).
- [27] P. Perinotti, Cellular automata in operational probabilistic theories, *Quantum* **4**, 294 (2020).
- [28] M. P. Müller and C. Ududec, Structure of Reversible Computation Determines the Self-Duality of Quantum Theory, *Phys. Rev. Lett.* **108**, 130401 (2012).
- [29] 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).
- [30] C. Ferrie and J. Emerson, Framed Hilbert space: hanging the quasi-probability pictures of quantum theory, *New J. Phys.* **11**, 063040 (2009).
- [31] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, *Phys. Rev. A* **75**, 032110 (2007).
- [32] 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).
- [33] C. H. Bennett and S. J. Wiesner, Communication Via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States, *Phys. Rev. Lett.* **69**, 2881 (1992).
- [34] R. F. Werner, All teleportation and dense coding schemes, *J. Phys. A: Math. Gen.* **34**, 7081 (2001).
- [35] S. Massar, S. Pironio, and D. Pitalúa-García, Hyperdense coding and superadditivity of classical capacities in hypersphere theories, *New J. Phys.* **17**, 113002 (2015).
- [36] G. M. D’Ariano, F. Manessi, P. Perinotti, and A. Tosini, Fermionic computation is non-local tomographic and violates monogamy of entanglement, *Europhys. Lett.* **107**, 20009 (2014).
- [37] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, “Event-Ready-Detectors” Bell Experiment Via Entanglement Swapping, *Phys. Rev. Lett.* **71**, 4287 (1993).
- [38] L. Hardy, Disentangling nonlocality and teleportation, [arXiv:quant-ph/9906123](https://arxiv.org/abs/quant-ph/9906123).
- [39] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Teleportation in general probabilistic theories, in *Mathematical Foundations of Information Flow: Proceedings of the Clifford Lectures 2008, Proceedings of Symposia in Applied Mathematics*, Vol. 71 (American Mathematical Society, Providence, RI, 2012), pp. 25–48.
- [40] P. Perinotti, Discord and Nonclassicality in Probabilistic Theories, *Phys. Rev. Lett.* **108**, 120502 (2012).
- [41] A. Y. Kitaev, Quantum computations: Algorithms and error correction, *Russ. Math. Surveys* **52**, 1191 (1997).
- [42] M. A. Nielsen and I. L. Chuang, Programmable Quantum Gate Arrays, *Phys. Rev. Lett.* **79**, 321 (1997).
- [43] G. M. D’Ariano and P. Perinotti, Efficient Universal Programmable Quantum Measurements, *Phys. Rev. Lett.* **94**, 090401 (2005).
- [44] G. Aubrun, L. Lami, C. Palazuelos, and M. Plávala, Entangleability of cones, [arXiv:1911.09663](https://arxiv.org/abs/1911.09663).
- [45] C. M. Caves, C. A. Fuchs, and P. Rungta, Entanglement of Formation of an Arbitrary State of Two Rebits, *Found. Phys. Lett.* **14**, 199 (2001).
- [46] M. Lugli, P. Perinotti, and A. Tosini, Fermionic State Discrimination by LOCC, *Phys. Rev. Lett.* **125**, 110403 (2020).
- [47] A. Wilce, Formalism and interpretation in quantum theory, *Found. Phys.* **40**, 434 (2010).
- [48] G. Chiribella, Distinguishability and copiability of programs in general process theories, *Int. J. Software Informatics* **8**, 209 (2014).

- [49] J. Selby and B. Coecke, Leaks: Quantum, classical, intermediate and more, *Entropy* **19**, 174 (2017).
- [50] T. D. Galley and L. Masanes, How dynamics constrains probabilities in general probabilistic theories, [arXiv:2002.05088](https://arxiv.org/abs/2002.05088).
- [51] V. Scarani, The device-independent outlook on quantum physics, *Acta Phys. Slov.* **62**, 347 (2012).
- [52] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, Almost quantum correlations, *Nat. Commun.* **6**, 6288 (2015).
- [53] A. Cabello, M. Gu, O. Gühne, J.-Å. Larsson, and K. Wiesner, Thermodynamical cost of some interpretations of quantum theory, *Phys. Rev. A* **94**, 052127 (2016).