Quantum theories from principles without assuming a definite causal structure

Ding Jia (贾丁)*

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada and Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada

(Received 30 June 2018; published 19 September 2018)

There has been a body of work deriving the complex Hilbert-space structure of quantum theory from axioms/principles/postulates to deepen our understanding of quantum theory and to reveal ways to go beyond it to resolve foundational issues. Recent progress in incorporating indefinite causal structure into physical theories suggests that a more comprehensive understanding of both quantum theory and the theory beyond it accounts for indefinite causal structure. We formulate a framework of physical theories without assuming definite causal structure and identify postulates that single out the complex Hilbert-space structure. More than one complex Hilbert-space theory is compatible with the postulates, which leaves room for further search for the best among these theories.

DOI: 10.1103/PhysRevA.98.032112

I. INTRODUCTION

Ordinary quantum theory assumes definite causal structure. This assumption is manifested in the existence of a dynamical law that evolves physical states through a definitely ordered sequence of continuous or discrete times and in the definite causal order presumed for the quantum operations.

In recent years it has been realized that to describe nature more comprehensively it is very likely necessary to drop the assumption of definite causal structure and incorporate indefinite causal structure into the theory. Experiments claiming realizations of operations with indefinite causal structure have been reported [1–3], and protocols have been discovered offering a further layer of indefinite-causal-structure-over-definite-causal-structure advantage in information processing (e.g., [4–9]), in addition to the quantum-over-classical advantage for theories with definite causal structure [10]. Moreover, it was pointed out early on that a theory unifying quantum theory and general relativity is expected to have a causal structure that is both dynamical and indefinite [11,12].

While the pioneer work introduces indefinite causal structure to general operational probabilistic theories [11,12], more recent works specialize to construct theories and models with the complex Hilbert-space structure (e.g., [5,13,14]). Ordinary quantum theory [16] based on the complex Hilbert-space structure suffers foundational problems which motivate a search for better alternatives [17]. In particular, there is a body of work that studies alternative operational probabilistic theories (see, e.g., [18] and references therein). Some alternative theories exhibit interesting new features such as larger violations of Bell's inequality than quantum theory [19,20], but none of the alternatives have so far been found to definitively describe nature better than quantum theory. To answer the deep question of what makes quantum theory special in the landscape of possible probabilistic theories, several different sets of axioms/principles/postulates have been identified which single out quantum theory (e.g., [21–30]). These works usually contain two parts, with the first part offering a framework to formulate a family of probabilistic theories and the second part deriving that only quantum theory obeys certain postulates. It is hoped that these axiomatic characterizations of complex Hilbert-space quantum theory will not only tell us what makes quantum theory special, but also guide the continued search for a superior theory that resolves the foundational problems of quantum theory.

The above axiomatic works commonly assume definite causal structure, either at the level of the general framework, so that all theories in the landscape have definite causal structure, or at the level of the postulates, so that the quantum theory that is singled out has definite causal structure. In view of the need to incorporate indefinite causal structure already mentioned, the assumption of definite causal structure appears to be important limitation. For the sake of understanding what makes quantum theory special for describing nature so well, it is preferable not to impose definite causal structure because, as mentioned above, a more comprehensive description of nature likely incorporates indefinite causal structure. For the sake of searching for a theory superior to quantum theory as well, it is preferable not to impose a causal structure because the superior theory may be a theory with indefinite causal structure.

The task of the present work is to find a set of postulates that singles out the complex Hilbert structure within a framework of theories that does not impose definite causal structure.

The framework of theories that do not impose definite causal structure we use is built on a powerful perspective on physical theories offered by Hardy [11,12]:

> A physical theory, whatever else it does, must correlate recorded data.

This motivates us to take *operations* (through which data are recorded) and *correlations* as the basic concepts of the framework, detailed in Sec. II. To give a mathematical structure to the concepts, an important postulate is made so that operations

*ding.jia@uwaterloo.ca

are associated with ordered vector spaces, and correlations are associated with (multi)linear functionals on these spaces. This framework differs from many of the frameworks used in the previous axiomatic works in that correlations, as a concept distinct from operations, play a very important role.

The task of identifying postulates and deriving the complex Hilbert-space structure is made easy by the previous works of Wilce and Barnum [27,28] (see also [31] and references therein for a comprehensive account of the approach and [29] for a related work based on category theories). The original postulates and derivations in their work are for theories with definite causal structure. Yet we show that the same general strategy of using the Jordan algebra structure to arrive at the complex Hilbert space works in a framework with indefinite causal structure. The list of postulates and the derivation of the complex Hilbert-space structure are presented in Sec. III. Some brief concluding remarks are offered in Sec. IV.

II. PHYSICAL THEORIES AS THEORIES OF OPERATIONS AND CORRELATIONS

No matter what else a theory of physics does, it must correlate recorded data [11,12]. Data are recorded through operations. There are other things a theory of physics can do, such as categorizing the constituents of the universe and offering a picture of reality, but at a minimum, it must deal with operations and correlations. In this paper, we focus on probabilistic theories. Some basic structures about probabilistic theories taking operations and correlations as fundamental concepts are presented in this section.

A. Operation

An operation consists of an action and an observation. For example, the game of "throwing the paper ball into the basket" involves an operation that consists of the action of picking up the paper ball and throwing it towards the basket and the observation of seeing whether the paper ball goes into the basket.

Note that the action and observation do not have to occur in a definite sequence. There are operations with the observation preceding the action and others with the action and the observation occurring simultaneously. It is helpful to simplify the situation by introducing the notion of "general action" to unify action and observation. A general action may be an action with a trivial observation (e.g., Alice throws the paper ball towards the basket and looks into the sky without observing whether the ball falls in), a pure observation (e.g., another person, Bob, observes whether Alice's ball falls in), or a combined action-observation (Alice throws the ball and keeps on observing where it flies).

Data are always gathered through the observation part of the general action. The trivial observation with only one possible outcome is still viewed to gather some data, even though this datum offers no nontrivial information with which to distinguish among more than one possibility.

An operation always refers to some physical objects. In the example above the relevant physical objects are the paper ball and the basket. In general, the relevant physical objects for an operation can be more complicated. For example, the operation of taking an orange and producing a cup of orange juice has the relevant physical object, the orange, going through different forms of existence (raw orange and orange juice). To be specific and talk about the different forms of existence, we speak of the relevant *physical system* of an operation. The physical system shows up as part of the mathematical description of an operation to specify what state of affairs is relevant for the operation. In the example above, we may take the operation to have two relevant physical systems: the state of the orange when it is raw and the state of the orange when it becomes juice. The physical system of an operation specifies a condition that enables the operation and/or a condition that checks the validity of an operation. Only when a paper ball and a basket are present can one play the game of throwing, and only when the orange is turned into juice (but not, say, a half-peeled orange) is the operation valid in that context. We note that in some situations the data recorded also invoke physical systems to store the data. For example, in a paper-ball-throwing competition the result of whether Alice's ball falls into the basket may be recorded on a piece of paper for further reference. This datum, either "yes" or "no," is classical. In other cases the data recorded may take the form of a quantum state or states in some types of systems.

To summarize, in a physical theory, a minimal description of an operation consists of a general action, a set of possible data gathered from the general action, and the relevant physical systems for the operation. More generally, there are situations where multiple choices for the operation are available. A general operation consists of a set of possible general actions, each with its own possible data set and its own relevant physical systems. We settle on this characterization of operations.

To symbolize an operation we adopt the following convention. A general action is denoted with capital letters in the form A. A physical system is denoted with lowercase letters in the form a. Sometimes we group systems together into a composite system. If the composite physical system a consists of subsystems $\mathsf{a}_1, \mathsf{a}_2, \ldots, \mathsf{a}_\mathsf{n}$, we write $\mathsf{a} = \mathsf{a}_1\mathsf{a}_2\ldots\mathsf{a}_\mathsf{n}$ and may use either the left side or the right side to refer to the composite system. The set of possible data is enumerated by letters i in a different font. These symbols A, a, i can be combined to make explicit different pieces of information. For example, a general action A with system a is referred to as A_a, and its ith datum may be referred to as $\mathsf{A}_\mathsf{a}[i]$.

In this language, an operation $\mathcal{O}$ is described by an indexed set of objects $\{\mathsf{A}_\mathsf{a}[i]\}_{\mathsf{A},\mathsf{a},i}$, where it is understood that the sets of possible values a and i vary according to the choice of general action A. We write

$$\mathcal{O} = \{\mathsf{A}_\mathsf{a}[i]\}_{\mathsf{A},\mathsf{a},i}. \tag{1}$$

Example 1. A familiar example of an operation is the quantum instrument used in quantum theory [32]. A quantum instrument is a set of completely positive maps $\{\mathcal{E}[i]\}_i$ from some input-state space $L(\mathcal{H}_{\mathsf{a}_1})$ (the space of bounded linear operators on the complex Hilbert space $\mathcal{H}_{\mathsf{a}_1}$) to some output-state space $L(\mathcal{H}_{\mathsf{a}_2})$. The set of maps is required to sum up to a completely positive trace-preserving map (channel). The quantum instrument describes a general action whose possible observational outcomes are i and whose physical system has two subsystems. The input subsystem a_1 is the one associated with the space $L(\mathcal{H}_{\mathsf{a}_1})$ and the output system a_2 is the one associated with the space $L(\mathcal{H}_{\mathsf{a}_2})$. We write the composite system of the operation as $\mathsf{a} = \mathsf{a}_1\mathsf{a}_2$. Then the operation takes the form $\{\mathcal{E}_{\mathsf{a}}[i]\}_{\mathsf{a},i}$, which is a special case of (1) with only one choice for the general action.

B. Correlations and probabilistic theories

The other basic concept of the framework is the correlation. The correlation among data registered from operations may be established through some other operation that interacts with the physical systems of the original operations. In some information-theory-inspired circuit models of operational probabilistic theories this is the only way to establish a correlation. Yet it is also possible that the correlation is not established through other operations conducted by agencies. For example, the global states in quantum field theory establish correlations for operations coupled to the field states, but the global state is not supposed to always be prepared by some other operations. Both kinds of correlations, those established through and not established through operations, can be described in the present framework.

Correlation is a broad term, and in general a theory mentioning the concept of correlation may not refer to probabilities. Yet in this paper we focus on probabilistic theories. In this context the main function of a probabilistic theory is to calculate the probabilities of allowed operations registering certain data. In general, the probabilities to be calculated take the form of conditional probabilities. When a conditional probability is well defined [33], a probabilistic theory is expected to offer a method with which to calculate it.

In general the conditional probabilities are of the form $p(i, j, \dots, k|\text{cond}) \in \mathbb{R}$, where $i, j, \dots, k$ is a possible set of data to be registered from a set of general actions, and cond encodes a prerequisite condition for the probability to make sense. The conditions include the choice of general action for each operation and further conditions to make the probabilities well defined. For example, in a circuit model cond can include the wiring configurations of the devices. In this probabilistic theory setting a correlation specifically refers to a map from a set of data to the set of real numbers, offering information on the conditional probabilities. A central theme of any probabilistic theory is to specify the properties of such maps. A natural structure to be imposed is linearity, which forms the topic of the next subsection.

C. Theory structure regarding probabilities

Conventionally, *absolute probabilities* are used for probabilities. Conditional probabilities of the form $p(i, j, \dots, k|\text{cond}) \in \mathbb{R}$ obey

$$p(i, j, \dots, k|\text{cond}) \geqslant 0, \tag{2}$$

$$\sum_{i,j,\dots,k} p(i, j, \dots, k|\text{cond}) = 1, \tag{3}$$

where the sum is over possible data to be recorded from the set of general actions. These imply that

$$1 \geqslant p(i, j, \dots, k|\text{cond}) \geqslant 0. \tag{4}$$

There is an alternative option of using *probability weights*. Probability weights $w(i, j, \dots, k|\text{cond}) \in \mathbb{R}$ are only required to obey

$$\infty > w(i, j, \dots, k|\text{cond}) \geqslant 0. \tag{5}$$

These probability weights are meaningful in comparison with each other, which eliminates the need for normalization. For any pair $w(i|\text{cond})$ and $w(j|\text{cond})$ of probability weights (Here for simplicity we use one letter, i or j, to represent a list of observational outcomes.), if $w(j|\text{cond}) \neq 0$, then the prediction is that the datum i is $r = w(i|\text{cond})/w(j|\text{cond})$ times as likely to be recorded as j. If $w(j|\text{cond}) = 0$, a comparison of probability weights in terms of the ratio $r = w(i|\text{cond})/w(j|\text{cond})$ should not be made, and the physical meaning is that the datum j is predicted never to be recorded.

When $0 < \sum_{i,j,\dots,k} w(i, j, \dots, k|\text{cond}) < \infty$, where the sum is over all possible outcomes for the set of general actions, normalization can be conducted and the absolute probabilities can be obtained from the relative probabilities as

$$p(i, j, \dots, k|\text{cond}) = \frac{w(i, j, \dots, k|\text{cond})}{\sum_{i,j,\dots,k} w(i, j, \dots, k|\text{cond})}. \tag{6}$$

The case of $0 = \sum_{i,j,\cdots,k} w(i, j, \dots, k|\text{cond})$ should not appear in a physically meaningful setup, since among all possible outcomes some outcome should happen. In a physically meaningful setup and for finitely many outcomes, $0 < \sum_{i,j,\dots,k} w(i, j, \dots, k|\text{cond}) < \infty$ always holds, and the absolute probabilities can always be obtained from the probability weights. Whereas the absolute probabilities are unique, the probability weights may be rescaled by the same factor without changing the physical content. This means that two theories using probability weights may give physically equivalent predictions even when the exact values for the probability weights of the same outcomes do not agree. The case of a diverging $\sum_{i,j,\dots,k} w(i, j, \dots, k|\text{cond})$ may appear when infinitely many outcomes are allowed by a theory. Then one needs to specify a separate rule to convert probability weights to absolute probabilities, if one still wants to do the conversion. As far as the derivation of the complex Hilbert-space structure of this paper goes we do not need to worry about this case, since the number of outcomes is assumed to be finite.

So far we have been talking about operations as an abstract concept without embedding them in a mathematical model. We now introduce a basic postulate to endow the operations (along with correlations) with some additional mathematical structure. Under this postulate, observational data will become vector-space elements, and the map of correlations will become (multi)linear functionals over such vector spaces.

The motivation comes from the probabilistic mixing of general actions. Let $\mathcal{O} = \{\mathsf{A}_{\mathsf{a}}[i]\}_{\mathsf{A},\mathsf{a},i}$ contain A_{a} and B_{a} as two choices for the general action associated with the same physical system a. Provided both general actions distinguish finitely many possible outcomes, without loss of generality we can suppose that they have the same total number of outcomes (adding void outcomes that are never triggered to the general action with the smaller number of outcomes if needed). Suppose a theory predicts that $w(i|\text{cond}, \mathsf{A})$ should A be chosen as the general action to be performed, and $w(i|\text{cond}, \mathsf{B})$ should B be chosen as the general action to be performed. Probabilistically mixing A and B means performing A with probability weight w_{A} and B with probability weight w_{B}. Under this mixing

$\{\mathsf{A}, \mathsf{B}; w_\mathsf{A}, w_\mathsf{B}\}$ the prediction for the outcomes is expected to be

$$w(i|\text{cond}, \{\mathsf{A}, \mathsf{B}; w_\mathsf{A}, w_\mathsf{B}\}) = w_\mathsf{A}\bar{w}_\mathsf{B} w(i|\text{cond}, \mathsf{A}) + w_\mathsf{B}\bar{w}_\mathsf{A} w(i|\text{cond}, \mathsf{B}), \quad (7)$$

where $\bar{w}_\mathsf{A} = \sum_i w(i|\text{cond}, \mathsf{A})$ and $\bar{w}_\mathsf{B} = \sum_i w(i|\text{cond}, \mathsf{B})$. This is analogous to

$$p(i|\text{cond}, \{\mathsf{A}, \mathsf{B}; p_\mathsf{A}, p_\mathsf{B}\}) = p_\mathsf{A} p(i|\text{cond}, \mathsf{A}) + p_\mathsf{B} p(i|\text{cond}, \mathsf{B}) \quad (8)$$

for ordinary probabilities, where A is performed with probability p_A and B is performed with probability p_B. In (7) there are the extra $\bar{w}_\mathsf{A}$ and $\bar{w}_\mathsf{B}$ factors. Analogous factors are not present for (8) because $\bar{p}_\mathsf{A} := \sum_i p(i|\text{cond}, \mathsf{A}) = 1 = \bar{p}_\mathsf{B} := \sum_i p(i|\text{cond}, \mathsf{B})$. Equation (8) can be arrived at from (7) using $p(i|\text{cond}, \mathsf{A}) := w(i|\text{cond}, \mathsf{A})/\bar{w}_\mathsf{A}$ and $p(i|\text{cond}, \mathsf{B}) := w(i|\text{cond}, \mathsf{B})/\bar{w}_\mathsf{B}$, $p_\mathsf{A} := w_\mathsf{A}/(w_\mathsf{A} + w_\mathsf{B})$, $p_\mathsf{B} := w_\mathsf{B}/(w_\mathsf{A} + w_\mathsf{B})$, $p(i|\text{cond}, \{\mathsf{A}, \mathsf{B}; w_\mathsf{A}, w_\mathsf{B}\}) := w(i|\text{cond}, \{\mathsf{A}, \mathsf{B}; w_\mathsf{A}, w_\mathsf{B}\})/\sum_i w(i|\text{cond}, \{\mathsf{A}, \mathsf{B}; w_\mathsf{A}, w_\mathsf{B}\})$, and noting that

$$\sum_i w(i|\text{cond}, \{\mathsf{A}, \mathsf{B}; w_\mathsf{A}, w_\mathsf{B}\}) = \sum_i w_\mathsf{A}\bar{w}_\mathsf{B} w(i|\text{cond}, \mathsf{A}) + w_\mathsf{B}\bar{w}_\mathsf{A} w(i|\text{cond}, \mathsf{B}) = (w_\mathsf{A} + w_\mathsf{B})\bar{w}_\mathsf{A}\bar{w}_\mathsf{B}. \quad (9)$$

Theories in which Eq. (7) holds have a certain linear structure for the correlation as a map from the outcomes to the probability weights. This suggests that the recorded data on the same physical system be represented as elements in a vector space, with real numbers such as $w_\mathsf{A}\bar{w}_\mathsf{B}$ and $w_\mathsf{B}\bar{w}_\mathsf{A}$ forming the field for the vector space, and the correlations as multilinear maps from these vector spaces to the probability weights. We realize this suggestion as a postulate.

Postulate 1 (linearity). Recorded data for general actions with the same relevant physical system are represented as positive cone elements in an ordered vector space with some trivial datum as an order unit. Correlations are represented as positive multilinear functionals in such spaces.

Here an *ordered vector space* is a real vector space V endowed with a convex cone V^+ such that V^+ spans V and that $V^+ \cap -V^+ = \{0\}$. V^+ is called the *positive cone* of V. An *order unit* of an ordered vector space is an element $u \in V^+$ so that for any $v \in V$, there is an $a > 0$ such that $au - v \in V^+$.

The ordered vector space of Postulate 1 is called an *operational space* and is denoted in the form $\mathfrak{O}_\mathsf{a}$, where a is the relevant physical system. The dimension of the space is denoted d_a. The positive cone is denoted $\mathfrak{O}_\mathsf{a}^+$. It contains the elements that represent physical data. Each $\mathsf{A}_\mathsf{a}[i]$ is represented by an element of $\mathfrak{O}_\mathsf{a}^+$. We refer to these vector-space elements using the same symbols $\mathsf{A}_\mathsf{a}[i]$ for the observational outcomes when no ambiguity arises. Vectors of the form $\mathsf{A}_\mathsf{a}[i]$ are viewed as members of the set of data $\{\mathsf{A}_\mathsf{a}[i]\}_i$ for the corresponding operation. The probability weights have meaning only in comparison, and the probability weights assigned by a state to some vectors are only meaningful for comparing when the vectors represent outcomes of the same operation. When it is clear from the context, though, we often suppress the labels $[i]$ and refer to the vector space elements in the form A_a for simplicity.

The correlations as positive multilinear functionals on $\mathfrak{O}_\mathsf{a}$, $\mathfrak{O}_\mathsf{b}, \ldots \mathfrak{O}_\mathsf{c}$ are denoted in the form $\mathsf{D}^{\mathsf{ab}\ldots\mathsf{c}}$, with the physical systems in the superscript to be distinguished from the recorded data with the system in the subscript:

$$\mathsf{D}^{\mathsf{ab}\ldots\mathsf{c}} : \mathfrak{O}_\mathsf{a} \times \mathfrak{O}_\mathsf{b} \times \ldots \times \mathfrak{O}_\mathsf{c} \to \mathbb{R}, \quad (\mathsf{A}_\mathsf{a}[i], \mathsf{B}_\mathsf{b}[j], \ldots, \mathsf{C}_\mathsf{c}[k]) \mapsto w(i, j, \ldots, k|\text{cond}). \quad (10)$$

The vector space generated by the correlations is called a *correlation space* and is denoted $\mathfrak{C}^{\mathsf{ab}\ldots\mathsf{c}}$. The dimension of the correlation space is denoted $c_{\mathsf{ab}\ldots\mathsf{c}}$.

Example 2. An example of an operational probabilistic theory that incorporates indefinite causal structure and uses probability weights is the *modified Oreshkov-Cerf theory*.

The original Oreshkov-Cerf theory is an operational quantum theory without predefined time [14] (See also [47].). A main new feature in comparison to ordinary operational quantum theory is that, in accordance with the absence of a predefined time, the systems associated with an operation/general action are not separated into input and output subsystems.

Using the notations of the original paper, an operation/general action $\{M_i^{AB\cdots}\}_{i\in O}$ consists of a set of possible events/outcomes indexed by the data set element $i \in O$. $A, B, \ldots$ are the physical systems associated with the operation, with corresponding Hilbert spaces $\mathcal{H}^A, \mathcal{H}^B, \ldots$ whose dimensions are $d^A, d^B, \ldots$. The events are represented by positive semidefinite operators $M_i^{AB\cdots}$ on $\mathcal{H}^A \otimes \mathcal{H}^B \otimes \ldots$. Operations come in equivalence classes. Two operations $\{M_i^{AB\cdots}\}_{i\in O}$ and $\{N_i^{AB\cdots}\}_{i\in O}$ that yield the same joint probabilities for all experimental setups (or circuits) belong to the same equivalence class. Similarly events come in equivalence classes. Two events $M_i^{AB\cdots}$ and $N_i^{AB\cdots}$ coming from different operations that yield the same joint probabilities with other events in all experimental setups (or circuits) belong to the same equivalence class.

Events/operations in the same equivalence class have operators that differ by a constant factor. One way to avoid this ambiguity is to represent an equivalence class of events by specifying a pair of operators in the form $(M_i^{AB\cdots}, \overline{M}^{AB\cdots})$, where $\overline{M}^{AB\cdots} := \sum_{i\in O} M_i^{AB\cdots}$, and fixing a normalization convention, such as

$$\text{Tr}\overline{M}^{AB\cdots} = d^A d^B \ldots. \quad (11)$$

The null operation $\{O^{AB\cdots}\}$ with trace zero is treated separately as a singular case.

The normalization requirement, (11), is weaker than what is usually imposed in ordinary quantum theory. Ordinary quantum theory is time asymmetric in the sense that measurement outcomes represented by POVM elements sum up to the identity (or, more generally, outcomes represented by quantum instrument elements sum up to a channel), but states in a preparation are only required to have their traces sum up to 1. In a theory without predefined time this time asymmetry should be absent, and in the Oreshkov-Cerf theory the time asymmetry is eliminated by weakening the requirement on outcomes so that only a sum of trace condition (11) is imposed.

The correlation is encoded in the following formula for joint probabilities:

$$p(i, j, \dots | \{M_i^{\cdots}\}_{i\in O}, \{N_j^{\cdots}\}_{j\in Q}, \dots ;\ \text{network}) = \frac{\mathrm{Tr}[(M_i^{\cdots} \otimes N_j^{\cdots} \otimes \dots) W_{\text{wires}}]}{\mathrm{Tr}[(\overline{M}^{\cdots} \otimes \overline{N}^{\cdots} \otimes \dots) W_{\text{wires}}]}. \tag{12}$$

This is a special case of (6). The condition in the conditional probability specifies the relevant operations and the way they are connected ("network"). The connection can be specified using a graph. The operations are located at the nodes. Each (sub)system of an operation is connected to a (sub)system of another operation with the same dimension using a "wire," which is an edge labeled by the system dimension. A wire tells which systems interact with which and is mathematically described as a pure bipartite entangled state $|\Phi\rangle\langle\Phi|$ whose precise form depends on the symmetry of the system. The operator W_{wires} is the tensor product of all these wire operators. This is the Oreshkov-Cerf theory in a nutshell. Details on the motivations and discussions about causality can be found in the original article [14].

The theory as presented so far does not fit into the present framework. The map $(M_i^{\cdots}, N_j^{\cdots}, \dots) \mapsto p(i, j, \dots | \{M_i^{\cdots}\}_{i\in O}, \{N_j^{\cdots}\}_{j\in Q}, \dots ; \text{network})$ according to (12) is not multilinear because of the division by $\mathrm{Tr}[(\overline{M}^{\cdots} \otimes \overline{N}^{\cdots} \otimes \dots) W_{\text{wires}}]$. To make the map multilinear and fit into the present framework one could use probability weights with the formula

$$w(i, j, \dots | \{M_i^{\cdots}\}_{i\in O}, \{N_j^{\cdots}\}_{j\in Q}, \dots ;\ \text{network}) = \mathrm{Tr}[(M_i^{\cdots} \otimes N_j^{\cdots} \otimes \dots) W_{\text{wires}}]. \tag{13}$$

This map $(M_i^{\cdots}, N_j^{\cdots}, \dots) \mapsto w(i, j, \dots | \{M_i^{\cdots}\}_{i\in O}, \{N_j^{\cdots}\}_{j\in Q}, \dots ;\ \text{network})$ is then multilinear.

In comparison to (12), in (13) the operators with overline no longer show up. By modifying the theory to use probability weights, we depart from describing operations and events in equivalence classes in the form $(M_i^{AB\cdots}, \overline{M}^{AB\cdots})$. There is now a constant multiplicative factor ambiguity in the probability weights, since one is allowed to rescale the operators of the events in the same operation by an arbitrary common positive factor. This ambiguity does not affect the physical predictions, since the probability weights are only meaningful in comparison to each other, specifically through taking ratios.

D. Subsystem structures

As the last part to specify the basic framework for probabilistic theories with operations and correlations, we discuss the subsystem structure for composite physical systems. We assume two very basic properties for the operational spaces of composite systems. A system a with $d_{\mathsf{a}} = \dim \mathfrak{O}_{\mathsf{a}} = 1$ is called a *trivial system*. The space of a trivial system supports only one linearly independent vector, which describes a trivial data. We assume that for a trivial system a, $\mathfrak{O}_{\mathsf{ab}} \cong \mathfrak{O}_{\mathsf{b}}$ as ordered vector spaces for all b.

The second basic property we assume is that any operational space $\mathfrak{O}_{\mathsf{ab}}$ with two subsystems contain all the product elements while preserving linear independence, i.e., if $\mathsf{A_a} \in \mathfrak{O}_{\mathsf{a}}$ and $\mathsf{B_b} \in \mathfrak{O}_{\mathsf{b}}$, then there is an element $\mathsf{A_aB_b} \in \mathfrak{O}_{\mathsf{ab}}$ so that if $\mathsf{A_a}$ and $\mathsf{A'_a}$ are linearly independent in $\mathfrak{O}_{\mathsf{a}}$ and $\mathsf{B_b}$ and $\mathsf{B'_b}$ are linearly independent in $\mathfrak{O}_{\mathsf{b}}$, then $\mathsf{A_aB_b}$, $\mathsf{A'_aB_b}$, $\mathsf{A_aB'_b}$, and $\mathsf{A'_aB'_b}$ are all linearly independent in $\mathfrak{O}_{\mathsf{ab}}$. This implies that $d_{\mathsf{a}}d_{\mathsf{b}} \leqslant d_{\mathsf{ab}}$.

There is a similar basic property we assume for correlations that pertain to two operational spaces. Suppose $\mathsf{C^a}$ is a correlation pertaining to $\mathfrak{O}_{\mathsf{a}}$ itself and $\mathsf{D^b}$ is a correlation pertaining to $\mathfrak{O}_{\mathsf{b}}$. Then we assume that there is a correlation $\mathsf{C^aD^b}$ pertaining to $\mathfrak{O}_{\mathsf{ab}}$ so that $\mathsf{C^aD^b}(\mathsf{A_aB_b}) = \mathsf{C^a}(\mathsf{A_a})\mathsf{D^b}(\mathsf{B_b})$, i.e., the probability weights multiply.

E. Comments on the framework

The framework just presented family-resembles other frameworks used in previous axiomatic works, but with some notable differences. First, no assumption of definite causal structure is imposed on the current framework. Moreover, correlations carrying nontrivial physical information but not generated by operations are allowed in the current framework. This is in contrast with the circuit models [23,25,34], where the operations carry nontrivial physical correlations and the wires do not. Some theories are more naturally described in the current framework. For example, as mentioned, the global state of quantum field theory is not prepared by an operation and is more suitably viewed as encoding the correlation of operations. Another example is the process matrices that allow correlations with indefinite causal structure [13,35,36]. It is found that the process matrices cannot be parallel-composed without constraints [37]. This would appear unnatural if the process matrices are viewed as operations but natural if they are viewed as correlations among operations.

Another difference lies in the graphical representation of using hypergraphs instead of graphs. Graphical reasoning was important in previous axiomatic works and works on operational theories in general (see, e.g., [23,25,34,38,39] and references therein). If one chooses to work with the current framework, the natural pictorial tool is the hypergraph, rather than the graph, which is widely used in other models (e.g., [14,23,25,34,40]). Roughly speaking, a hypergraph is a generalized graph that allows edges to connect to other integer numbers of nodes rather than just two. The generalized edge is called a "hyperedge." We can associate the nodes of a hypergraph with the operations/outcomes and the hyperedges with the correlations, connecting the nodes they correlate. The implications of using hypergraphs instead of graphs for probabilistic theories remain to be explored.

III. THE COMPLEX HILBERT-SPACE STRUCTURE

In this section we record a list of postulates and show that they single out the complex Hilbert-space structure. We restrict our attention to operations with finite-dimensional operational spaces. Technically, the reason is that the derivation of the complex Hilbert-space structure below uses dimension counting arguments and lemmas that work for finite-dimensional spaces. Conceptually, the restriction to working with finite dimensions can be motivated by the constraints of realistic data gathering. Even for theories whose mathematical description uses infinite-dimensional spaces such as quantum field theory, realistic data gathering is subject to the constraints of finite resolution and finite range, which imply a finite data set.

Despite these motivations for working with finite-dimensional spaces, we do hope that some future work finds a derivation of the complex Hilbert-space structure without restriction to finite-dimensional spaces. There are useful theories described with infinite-dimensional spaces (such as quantum field theory) which introduce new features absent in theories with finite-dimensional spaces. It is an open question to what extent the following derivation generalizes to infinite dimensions.

A. Postulates

To state the next postulate, we need to define the notion of transformation. In ordinary quantum theory, a transformation is a trace-nonincreasing [41] and completely positive map. The trace-nonincreasing property is required so that absolute probabilities remain in the interval [0, 1]. The completely positive property is required to ensure that physical states get mapped to physical states even if the transformation acts partially on a subsystem. We want a generalized definition of transformations that applies to all the theories within the current framework. Since the framework uses probability weights instead of absolute probabilities, there is no requirement regarding the kind of trace-nonincreasing property. The following can be viewed as a generalization of completely positive maps.

Fix two arbitrary operational spaces $\mathfrak{O}_{\mathsf{a}}$ and $\mathfrak{O}_{\mathsf{b}}$. We want to define the notion of a-to-b transformation, which maps not only from $\mathfrak{O}_{\mathsf{a}}$ to $\mathfrak{O}_{\mathsf{b}}$, but also from $\mathfrak{O}_{\mathsf{ac}}$ to $\mathfrak{O}_{\mathsf{bc}}$ for arbitrary c. An a-*to*-b *transformation*, denoted $\mathsf{T}_{\mathsf{a,b}}$, is a family $\{T_{\mathsf{ac,bc}}\}_{\mathsf{c}}$ of linear maps for each c,

$$T_{\mathsf{ac,bc}} : \mathfrak{O}_{\mathsf{ac}} \to \mathfrak{O}_{\mathsf{bc}}, \tag{14}$$

so that (i) for arbitrary $\mathsf{A_a} \in \mathfrak{O}_{\mathsf{a}}$ and $\mathsf{B_c} \in \mathfrak{O}_{\mathsf{c}}$, $T_{\mathsf{ac,bc}}(\mathsf{A_a B_c}) = T_{\mathsf{a,b}}(\mathsf{A_a})\mathsf{B_c}$ for $T_{\mathsf{a,b}} : \mathfrak{O}_{\mathsf{a}} \to \mathfrak{O}_{\mathsf{b}}$, and (ii) $T_{\mathsf{ac,bc}}(\mathfrak{O}^+_{\mathsf{ac}}) \subset \mathfrak{O}^+_{\mathsf{bc}}$. Condition (i) ensures that the transformation acts locally on product elements and condition (ii) generalizes complete positivity.

The transformations as linear maps can be summed linearly. Given $\mathsf{T}_{\mathsf{a,b}} = \{T_{\mathsf{ac,bc}}\}_{\mathsf{c}}$ and $\mathsf{S}_{\mathsf{a,b}} = \{S_{\mathsf{ac,bc}}\}_{\mathsf{c}}$, define $p\mathsf{T}_{\mathsf{a,b}} + q\mathsf{S}_{\mathsf{a,b}} = \{pT_{\mathsf{ac,bc}} + qS_{\mathsf{ac,bc}}\}_{\mathsf{c}}$ for $p, q \in \mathbb{R}$. In this way a vector space is generated. As can be checked easily, the set of all transformations $\mathsf{T}_{\mathsf{a,b}}$ forms a convex cone, making the vector space an ordered vector space. Call it a *transformation space* and denote it $\mathfrak{T}_{\mathsf{a,b}}$. Denote the positive cone by $\mathfrak{T}^+_{\mathsf{a,b}}$ and dim $\mathfrak{T}_{\mathsf{a,b}}$ by $t_{\mathsf{a,b}}$.

The above definition of transformations is intended as a mathematical characterization of the, in principle, possible physical transformations. Whether all these mathematically defined transformations are actually realizable and what the physical interpretation is for the transformations are subject to further specifications of particular theories [42].

We can now state the postulate.

Postulate 2 (dimension). An operational space whose physical system has two subsystems has the same dimension as the correlation space over these two systems and as the transformation spaces between these two systems.

Equivalently, Postulate 2 states that for arbitrary $\mathfrak{O}_{\mathsf{a}}$ and $\mathfrak{O}_{\mathsf{b}}$, $d_{\mathsf{ab}} = c_{\mathsf{ab}} = t_{\mathsf{a,b}} = t_{\mathsf{b,a}}$ (recall that $d_{\mathsf{ab}} = \mathfrak{O}_{\mathsf{ab}}$). One can interpret the postulate as offering the operations enough degrees of freedom to potentially realize all two system correlations and mathematically possible transformations. The correlations of two operations include both those arising naturally and those controlled by agents. The latter type of correlation must interact with the two relevant systems and is controlled by the agents through some operations containing the two systems as subsystems. The postulate states that as far as the degrees of freedom of the vector spaces are concerned, the operations have as many degrees of freedom as the set of all possible correlations, including the type arising from nature. Similarly, there are transformations arising from nature and transformations controlled by agents. The agent-controlled transformations between two systems are realized by the agents through some operations pertaining to the two systems as subsystems. The postulate states that as far as the degrees of freedom of the vector spaces are concerned, the operations have as many degrees of freedom as the set of all possible transformations, including the type arising from nature.

Our discussion now moves from operational-space elements' transformations into each other to how they correlate with each other. Without further constraints the framework allows weird theories such as one in which data recorded from any two operations on different systems are not correlated. In a universe described by this theory little inference can be made. To focus attention on more reasonable theories a postulate is needed to offer some regularity in terms of how systems correlate with each other. We adopt the following "pairing" postulate for this purpose.

To state the postulate, first we need the notion of a "copy" of operational spaces. An *order isomorphism* f between ordered vector spaces V and W is a positive, invertible linear map having a positive inverse, where positive means $f(V^+) \subseteq W^+$. If two operational spaces $\mathfrak{O}_{\mathsf{a}}$ and $\mathfrak{O}_{\mathsf{b}}$ share an order isomorphism, we say that they are *copies* of each other. We use primes on physical systems and vectors to signify copies (e.g., $\mathfrak{O}_{\mathsf{a}'}$ for the copy of $\mathfrak{O}_{\mathsf{a}}$ and $\mathsf{A}'_{\mathsf{a}'}$ for the copy of $\mathsf{A_a}$ under the order isomorphism).

An operational space $\mathfrak{O}_{\mathsf{a}}$ is said to have a *pairing* if there is a copy $\mathfrak{O}_{\mathsf{a}'}$ and a correlation $\mathsf{C}^{\mathsf{aa}'}$ on the two spaces so that $\mathsf{C}^{\mathsf{aa}'}(\mathsf{A_a}, \mathsf{A}'_{\mathsf{a}'}) > 0$ for all nonzero $\mathsf{A_a} \in \mathfrak{O}_{\mathsf{a}}$. The pairing is said to be *symmetric* if $\mathsf{C}^{\mathsf{aa}'}(\mathsf{A_a}, \mathsf{B}'_{\mathsf{a}'}) = \mathsf{C}^{\mathsf{aa}'}(\mathsf{B_a}, \mathsf{A}'_{\mathsf{a}'})$ for all $\mathsf{A_a}, \mathsf{B_a} \in \mathfrak{O}_{\mathsf{a}}$. The pairing is said to be *distinguishing* if whenever an operational space element yields only physical (nonnegative) probability weights through the correlation, the element is physical, i.e., whenever $\mathsf{A_a}$ is such that $\mathsf{C}^{\mathsf{aa}'}(\mathsf{A_a}, \mathsf{B}'_{\mathsf{a}'}) \geqslant 0$ for all $\mathsf{B}'_{\mathsf{a}'} \in \mathfrak{O}^+_{\mathsf{a}'}$, $\mathsf{A_a} \in \mathfrak{O}^+_{\mathsf{a}}$. A *factorizably symmetric distinguishing pairing* is such that it factorizes for operational spaces with factors while preserving the symmetric and distinguishing properties, i.e., for $\mathfrak{O}_{\mathsf{a}} = \mathfrak{O}_{\mathsf{a_1a_2}}$, $\mathsf{A_a} = \mathsf{A_{a_1}A_{a_2}}$, and $\mathsf{B_a} = \mathsf{B_{a_1}B_{a_2}}$, $\mathsf{C}^{\mathsf{aa}'}(\mathsf{A_a}, \mathsf{B}'_{\mathsf{a}'}) = \mathsf{C}_1^{\mathsf{a_1a_1'}}(\mathsf{A_{a_1}}, \mathsf{B}'_{\mathsf{a}'_1})\mathsf{C}_2^{\mathsf{a_2a_2'}}(\mathsf{A_{a_2}}, \mathsf{B}'_{\mathsf{a}'_2})$ factorizes into two pairings $\mathsf{C}_1^{\mathsf{a_1a_1'}}$ and $\mathsf{C}_2^{\mathsf{a_2a_2'}}$ such that both are symmetric and distinguishing.

Postulate 3 (pairing). Each operational space has at least one factorizably symmetric distinguishing pairing.

One can interpret the postulate as imposing some regularity on how recorded data correlate. The existence of a pairing offers the possibility of establishing some positive correlations for pairs of data recorded with operations, in particular, for operations conducted on isomorphic operational spaces, the most elementary pair of spaces that positive correlations can be expected on. The strongest form of correlation we can hope for is that from the outcomes of one operation we can infer unambiguously the outcomes of the paired operation. Postulate 3 is a weaker requirement, asking only that paired outcomes appear together with some positive chance (Note that the physical outcomes are elements of the positive cone, so strictly speaking the pairing condition is an extension of the above requirement to all elements of the operational spaces.). The symmetric property appears as a natural assumption for operational spaces that are isomorphic. The distinguishing property assumes that the correlation of the pairing is strong enough to reflect (at the mathematical level) any unphysical correlation if there is any. Finally, the factorizing property is a natural assumption considering the factor structure.

The next postulate is easy to state. An ordered vector space V is *homogeneous* if $\mathrm{Aut}(V)$, the group of order automorphisms on V, acts transitively on the interior of V_+.

Postulate 4 (homogeneity). Operational spaces are homogeneous.

Intuitively, the postulate says that inside an operational space any region looks locally like any other. For example, the qubit space of ordinary quantum theory is homogeneous, as there is no preferred direction or region inside the space.

The previous postulates already offer strong constraints to arrive at self-dual (Theorem 1) and homogeneous spaces, so that only the self-adjoint parts of real, complex, quaternionic, 3×3 octonion matrix algebras, spin factors, and their direct sums are allowed [43–45]. At this stage, a most general theory fulfilling the postulates appears to be the direct sum of the different types of systems listed above. However, in fact as long as a single quantum qubit shows up in the combination, the theory must be exclusively complex Hilbert space quantum (see "Barnum-Wilce theorem," below). The only possibility preventing this is that a qubit does not show up. Therefore to arrive at the complex Hilbert space we assume:

Postulate 5 (qubit). There exists a qubit.

B. Derivation

The derivation of the complex Hilbert-space structure is simplified immensely thanks to the previous works by Barnum and Wilce [46], Koecher [43], Vinberg [44], and Jordan, von Neumann, and Wigner [45]. The relevance of these results is condensed in the "Barnum-Wilce theorem" (below), which directly yields the final result we look for. To connect the above postulates to the Barnum-Wilce theorem, we only need to do two simple proofs (Theorem 1 and Theorem 2).

A finite-dimensional ordered vector space V is *self-dual* if it has an inner product such that a belongs to the positive cone V^+ iff $\langle a,b\rangle \geqslant 0$ for all $b\in V_+$.

Theorem 1. All $\mathfrak{O}_\mathsf{a}$ are self-dual.

Proof. According to Postulate 3, there is a symmetric distinguishing pairing $(\mathfrak{O}_{\mathsf{a}'},\mathsf{C}^{\mathsf{aa}'})$ for $\mathfrak{O}_\mathsf{a}$. We claim that $\langle\cdot,\cdot\rangle:\mathfrak{O}_\mathsf{a}\times\mathfrak{O}_\mathsf{a}\to\mathbb{R}$ defined by $\langle \mathsf{A}_\mathsf{a},\mathsf{B}_\mathsf{a}\rangle=\mathsf{C}^{\mathsf{aa}'}(\mathsf{A}_\mathsf{a},\mathsf{B}'_{\mathsf{a}'})$ is an inner product, i.e., it is bilinear, symmetric, and positive definite. The first property follows from Postulate 1, and the rest from $(\mathfrak{O}_{\mathsf{a}'},\mathsf{C}^{\mathsf{aa}'})$ being a symmetric pairing.

Now we show that $\mathsf{A}_\mathsf{a}\in\mathfrak{O}^+_\mathsf{a}$ iff $\langle\mathsf{A}_\mathsf{a},\mathsf{B}_\mathsf{a}\rangle\geqslant 0$ for all $\mathsf{B}_\mathsf{a}\in\mathfrak{O}^+_\mathsf{a}$. If $\mathsf{A}_\mathsf{a}\in\mathfrak{O}^+_\mathsf{a}$, then $\langle\mathsf{A}_\mathsf{a},\mathsf{B}_\mathsf{a}\rangle=\mathsf{C}^{\mathsf{aa}'}(\mathsf{A}_\mathsf{a},\mathsf{B}'_{\mathsf{a}'})\geqslant 0$ for all $\mathsf{B}_\mathsf{a}\in\mathfrak{O}^+_\mathsf{a}$ because $\mathsf{C}^{\mathsf{aa}'}$ is positive according to Postulate 1. If $\langle\mathsf{A}_\mathsf{a},\mathsf{B}_\mathsf{a}\rangle\geqslant 0$ for all $\mathsf{B}_\mathsf{a}\in\mathfrak{O}^+_\mathsf{a}$, $\mathsf{A}_\mathsf{a}\in\mathfrak{O}^+_\mathsf{a}$ by the distinguishing property of the pairing.

Theorem 2 (tomographic locality). $d_\mathsf{ab}=d_\mathsf{a}d_\mathsf{b}=c_\mathsf{ab}=c_\mathsf{a}c_\mathsf{b}$.

Proof. Let $\mathfrak{O}_\mathsf{a}$ and $\mathfrak{O}_\mathsf{b}$ be arbitrary. We want to count the number of degrees of freedom, $t_{\mathsf{a},\mathsf{b}}$, in defining a transformation $\mathsf{T}_{\mathsf{b},\mathsf{a}}\in\mathfrak{T}_{\mathsf{b},\mathsf{a}}$. These degrees of freedom fix its action on all $\mathsf{A}_\mathsf{ac}\in\mathfrak{O}_\mathsf{ac}$ for arbitrary c. First, let c be trivial. The local action of T_ab on $\mathfrak{O}_\mathsf{a}$ is determined by taking d_a linearly independent vectors from $\mathfrak{O}_\mathsf{a}$ and specifying an image for each. Each image requires d_b real parameters to specify, so $d_\mathsf{a}d_\mathsf{b}$ independent real parameters are needed in total.

Now let $\mathsf{c}=\mathsf{b}$. Condition (ii) (see Sec. III A) in the definition of transformations fixes the action of T_ab on product elements of the form $\mathsf{A}_\mathsf{a}\mathsf{B}_\mathsf{b}$, but the action on the possible additional linearly independent elements is yet unspecified. For each of the $r_\mathsf{ab}:=d_\mathsf{ab}-d_\mathsf{a}d_\mathsf{b}\geqslant 0$ additional linearly independent vectors, d_bb real parameters are needed to determine the image. Hence specifying T_ab requires at least $l_\mathsf{ab}:=d_\mathsf{a}d_\mathsf{b}+r_\mathsf{ab}d_\mathsf{bb}$ independent real parameters, i.e., $t_{\mathsf{b},\mathsf{a}}\geqslant l_\mathsf{ab}$. By Postulate 2, $t_{\mathsf{b},\mathsf{a}}=d_\mathsf{ab}$, so

$$l_\mathsf{ab}-t_{\mathsf{b},\mathsf{a}}=d_\mathsf{a}d_\mathsf{b}+r_\mathsf{ab}d_\mathsf{bb}-d_\mathsf{ab}=r_\mathsf{ab}(d_\mathsf{bb}-1)\leqslant 0. \quad (15)$$

If $d_\mathsf{b}>1, d_\mathsf{bb}\geqslant d_\mathsf{b}^2>1$. By (15), $r_\mathsf{ab}=0$. If otherwise $d_\mathsf{b}=1$, then $r_\mathsf{ab}=d_\mathsf{ab}-d_\mathsf{a}d_\mathsf{b}=d_\mathsf{a}-d_\mathsf{a}=0$. Therefore r_ab is always 0, and $d_\mathsf{ab}=d_\mathsf{a}d_\mathsf{b}$. By Postulate 2, this also equals to c_ab. Taking b to be trivial, we see that $c_\mathsf{a}=d_\mathsf{a}$. Therefore $d_\mathsf{ab}=d_\mathsf{a}d_\mathsf{b}=c_\mathsf{ab}=c_\mathsf{a}c_\mathsf{b}$.

In Proposition 1.1 in [46], Barnum and Wilce proved the following result.

Barnum-Wilce theorem. For a homogeneous and factorizably self-dual probabilistic theory, if it obeys tomographic locality and contains a qubit, then all its systems are self-adjoint parts of complex matrix algebras.

The theorem was originally obtained in the context of no-signaling probabilistic theories with definite causal structure. However, the proof of the theorem also goes through in the present context, as allowing indefinite causal structure does not affect the proof and "no signaling" was only used to prove that maps of the form $(\mathsf{A}_\mathsf{a},\mathsf{B}_\mathsf{b})\mapsto\mathsf{C}^\mathsf{ab}(\mathsf{A}_\mathsf{a},\mathsf{B}_\mathsf{b})$ are bilinear, which holds automatically in our framework. In the theorem, *factorizably self-dual* means that the self-dualizing inner product factors on two subsystems, i.e., $\langle\mathsf{A}_\mathsf{a}\mathsf{B}_\mathsf{b},\mathsf{X}_\mathsf{a}\mathsf{Y}_\mathsf{b}\rangle=\langle\mathsf{A}_\mathsf{a},\mathsf{X}_\mathsf{a}\rangle\langle\mathsf{B}_\mathsf{b},\mathsf{Y}_\mathsf{b}\rangle$. This property holds for the self-dualizing product in Theorem 1 if we pick the pairing to be factorizable, as allowed by Postulate 3. This plus Postulates 4 and 5 and Theorem 2 leads to the following result.

Corollary. All operational-space elements are self-adjoint parts of complex matrices.

IV. CONCLUSION

We have presented a general framework for physical theories that does not assume definite causal structure. This framework takes operations and correlations as the central concepts. We further identified a list of postulates from which finite-dimensional complex Hilbert-space quantum theories can be derived. This may be viewed as an axiomatic formulation of quantum theories without assuming indefinite causal structure.

More than one quantum theory is compatible with the postulates. The compatible theories include both quantum theories with an explicit indefinite causal structure (e.g., [5,13,14,47,48]), and ordinary formulations of quantum theory with definite causal structure (definite causality can be imposed as a further postulate). This leads to the interesting question whether one among these many compatible theories describes nature best.

The framework presented in Sec. II allows infinite-dimensional systems [49] and can, in principle, incorporate infinite-dimensional theories such as quantum field theory. It is an interesting open question to identify postulates that derive infinite-dimensional quantum theory without assuming definite causal structure.

ACKNOWLEDGMENTS

The author thanks Lucien Hardy and Achim Kempf for guidance and support, Alexander Wilce and Matthew Leifer for making suggestions and pointing out flaws in an earlier version of the manuscript, and Ognyan Oreshkov, Carlo Maria Scandolo, Lee Smolin, and Matthew Graydon for discussion.

Research at Perimeter Institute was supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science. This work was partly supported by a grant from the John Templeton Foundation. The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

[1] J.-P. W. MacLean, K. Ried, R. W. Spekkens, and K. J. Resch, Nat. Commun. **8**, 15149 (2017).
[2] L. M. Procopio, A. Moqanaki, M. Araújo, F. Costa, I. A. Calafell, E. G. Dowd, D. R. Hamel, L. A. Rozema, Č. Brukner, and P. Walther, Nat. Commun. **6**, 7913 (2015).
[3] G. Rubino, L. A. Rozema, A. Feix, M. Araújo, J. M. Zeuner, L. M. Procopio, Č. Brukner, and P. Walther, Sci. Adv. **3**, e1602589 (2017).
[4] L. Hardy, *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle* (Springer, Berlin, 2009), pp. 379–401.
[5] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, Phys. Rev. A **88**, 022318 (2013).
[6] G. Chiribella, Phys. Rev. A **86**, 040301 (2012).
[7] M. Araújo, F. Costa, and Č. Brukner, Phys. Rev. Lett. **113**, 250402 (2014).
[8] A. Feix, M. Araújo, and Č. Brukner, Phys. Rev. A **92**, 052326 (2015).
[9] P. A. Guérin, A. Feix, M. Araújo, and Č. Brukner, Phys. Rev. Lett. **117**, 100502 (2016).
[10] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
[11] L. Hardy, arXiv:gr-qc/0509120.
[12] L. Hardy, J. Phys. A: Math. Theor. **40**, 3081 (2007).
[13] O. Oreshkov, F. Costa, and Č. Brukner, Nat. Commun. **3**, 1092 (2012).
[14] O. Oreshkov and N. J. Cerf, New J. Phys. **18**, 073037 (2016).
[15] E. C. Stueckelberg, Helv. Phys. Acta **33**, 458 (1960).
[16] As a note on terminology, we sometimes refer to complex Hilbert-space quantum theory as "quantum theory" for simplicity. This is to be distinguished from quantum theory in other spaces, such as "real Hilbert-space quantum theory" [15]. By "ordinary quantum theory" we mean complex Hilbert-space quantum theory with definite causal structure.
[17] G. Auletta, *Foundations and Interpretation of Quantum Mechanics: In the Light of a Critical-Historical Analysis of the Problems and of a Synthesis of the Results* (World Scientific, Singapore, 2001).
[18] G. Chiribella and R. W. Spekkens (eds.), *Quantum Theory: Informational Foundations and Foils* (Springer, Berlin, 2016).
[19] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[20] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[21] L. Hardy, arXiv:quant-ph/0101012.
[22] H. Barnum, P. Markus, C. Ududec *et al.*, New J. Phys. **16**, 123029 (2014).
[23] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).
[24] B. Dakić and Č. Brukner, in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, UK, 2011), pp. 365–392.
[25] L. Hardy, arXiv:1104.2066.
[26] L. Masanes and M. P. Müller, New J. Phys. **13**, 063001 (2011).
[27] A. Wilce, in *Probability in Physics*, edited by Y. Ben-Menahem and M. Hemmo (Springer, Berlin, Heidelberg, 2012), pp. 281–298.
[28] A. Wilce, arXiv:1206.2897.
[29] J. H. Selby, C. M. Scandolo, and B. Coecke, arXiv:1802.00367.
[30] P. A. Höhn and C. S. P. Wever, Phys. Rev. A **95**, 012102 (2017).
[31] H. Barnum and A. Wilce, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer, Berlin, 2016), pp. 367–420.
[32] E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970).
[33] See [11] and [12] for a discussion of the nontriviality of the requirement that the probabilities are well defined.
[34] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
[35] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, New J. Phys. **17**, 102001 (2015).
[36] O. Oreshkov and C. Giarmatzi, New J. Phys. **18**, 093020 (2016).
[37] D. Jia and N. Sakharwade, Phys. Rev. A **97**, 032110 (2018).
[38] L. Hardy, *Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky* (Springer, Berlin, 2013), pp. 83–106.
[39] B. Coecke and A. Kissinger, *Picturing Quantum Processes* (Cambridge University Press, Cambridge, UK, 2017).
[40] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).
[41] By allowing not just trace-preserving maps we keep the notion of transformation more general. This general notion of transformation is used, for example, in [34].

[42] A commonly seen understanding of a transformation is that it takes states from a previous time to a later time. This understanding is not general enough. For example, a quantum comb-type transformation can transform an operation (which may be a transformation rather than a state) to another operation that extends from an earlier time to a later time [40].
[43] M. Koecher, Math. Ann. **135**, 192 (1958).
[44] E. Vinberg, Dokl. Acad. Nauk. SSSR **141**, 270 (1961) [English translation, Soviet Math. Dokl 2, 1416 (1961)].
[45] P. Jordan, J. v. Neumann, and E. Wigner, Ann. Math. **35**, 29 (1934).
[46] H. Barnum and A. Wilce, Found. Phys. **44**, 192 (2014).
[47] O. Oreshkov and N. J. Cerf, Nat. Phys. **11**, 853 (2015).
[48] R. Silva, Y. Guryanova, A. J. Short, P. Skrzypczyk, N. Brunner, and S. Popescu, New J. Phys. **19**, 103022 (2017).
[49] Finite-dimensional spaces are used to motivate Postulate 1, but as stated the postulate does not impose finite dimensionality.