Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference

M. S. Leifer[1,2,*] and Robert W. Spekkens[2,†]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*

[2]*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5*

(Received 19 June 2012; revised manuscript received 30 October 2013; published 27 November 2013)

Quantum theory can be viewed as a generalization of classical probability theory, but the analogy as it has been developed so far is not complete. Whereas the manner in which inferences are made in classical probability theory is independent of the causal relation that holds between the conditioned variable and the conditioning variable, in the conventional quantum formalism, there is a significant difference between how one treats experiments involving two systems at a single time and those involving a single system at two times. In this article, we develop the formalism of *quantum conditional states*, which provides a unified description of these two sorts of experiment. In addition, concepts that are distinct in the conventional formalism become unified: Channels, sets of states, and positive operator valued measures are all seen to be instances of conditional states; the action of a channel on a state, ensemble averaging, the Born rule, the composition of channels, and nonselective state-update rules are all seen to be instances of belief propagation. Using a quantum generalization of Bayes' theorem and the associated notion of Bayesian conditioning, we also show that the remote steering of quantum states can be described within our formalism as a mere updating of beliefs about one system given new information about another, and retrodictive inferences can be expressed using the same belief propagation rule as is used for predictive inferences. Finally, we show that previous arguments for interpreting the projection postulate as a quantum generalization of Bayesian conditioning are based on a misleading analogy and that it is best understood as a combination of belief propagation (corresponding to the nonselective state-update map) and conditioning on the measurement outcome.

DOI: 10.1103/PhysRevA.88.052130 PACS number(s): 03.65.Ca, 03.65.Ta, 03.67.−a

I. INTRODUCTION

Quantum theory can be understood as a noncommutative generalization of classical probability theory wherein probability measures are replaced by density operators. Much of quantum information theory, especially quantum Shannon theory, can be viewed as the systematic application of this generalization of probability theory to information theory.

However, despite the power of this point of view, the conventional formalism for quantum theory is a poor analog to classical probability theory because, in quantum theory, the appropriate mathematical description of an experiment depends on its causal structure. For example, experiments involving a pair of systems at spacelike separation are described differently from those that involve a single system at two different times. The former are described by a joint state on the tensor product of two Hilbert spaces, and the latter are described by an input state and a dynamical map on a single Hilbert space. Classical probability works at a more abstract level than this. It specifies how to represent uncertainty prior to, and independently of, causal structure. For example, our uncertainty about two random variables is always described by a joint probability distribution, regardless of whether the variables represent two spacelike separated systems or the input and output of a classical channel. Although channels represent time evolution, they are described mathematically by conditional probability distributions. The input state specifies a marginal distribution, and thus we have the ingredients to define a joint probability distribution over the input and output variables. This joint probability distribution could equally well be used to describe two spacelike separated variables. Therefore, we do not need to know how the variables are embedded in space-time in advance in order to apply classical probability theory. This has the advantage that it cleanly separates the concept of correlation from that of causation. The former is the proper subject of probabilistic inference and statistics. Within the subjective Bayesian approach to probability, independence of inference and causality has been emphasized by de Finetti ([1], Preface pp. x–xi):

"Probabilistic reasoning—always to be understood as subjective—merely stems from our being uncertain about something. It makes no difference whether the uncertainty relates to an unforeseeable future, or to an unnoticed past, or to a past doubtfully reported or forgotten; it may even relate to something more or less knowable (by means of a computation, a logical deduction, etc.) but for which we are not willing to make the effort; and so on."

Thus, in order to build a quantum theory of Bayesian inference, we need a formalism that is even handed in its treatment of different causal scenarios. There are some clues that this might be possible. Several authors have noted that there are close connections, and often isomorphisms, between the statistics that can be obtained from quantum experiments with distinct causal arrangements [2–8]. Time-reversal symmetry is an example of this, but it is also possible to relate experiments involving two systems at the same time with those involving a single system at two times. The equivalence [9] of prepare-and-measure [10] and entanglement-based [11] quantum key distribution protocols is an example of this and provides the basis for proofs of the security of the former [12].

*matt@mattleifer.info

†rspekkens@perimeterinstitute.ca

Such equivalences suggest that it may be possible to obtain a causally neutral formalism for quantum theory by describing such isomorphic experiments by similar mathematical objects.

Among the main goals of this work are to provide this unification for the case of experiments involving two distinct quantum systems at one time and those involving a single quantum system at two times and to provide a framework for making probabilistic inferences that is independent of this causal structure. Both types of experiment can be described by operators on a tensor product of Hilbert spaces, differing from one another only by a partial transpose. Probabilistic inference is achieved using a quantum generalization of Bayesian conditioning applied to *quantum conditional states*, which are the main objects of study of this work.

Quantum conditional states are a generalization of classical conditional probability distributions. Conditional probability plays a key role in classical probability theory, not least due to its role in Bayesian inference, and there have been attempts to generalize it to the quantum case. The most relevant to quantum information are perhaps the quantum conditional expectation [13] (see [14,15] for a basic introduction and [16] for a review) and the Cerf-Adami conditional density operator [17–19]. To date, these have not seen widespread application in quantum information, which casts some doubt on whether they are really the most useful generalization of conditional probability from the point of view of practical applications. Quantum conditional states, which have previously appeared in [4,20,21], provide an alternative approach to this problem. We show that they are useful for drawing out the analogies between classical probability and quantum theory, they can be used to describe both spacelike and timelike correlations, and they unify concepts that look distinct in the conventional formalism.

The remainder of the introduction summarizes the contents of this article. It is meant to provide a broad overview of the conditional states formalism, its motivations, and its applications, while introducing only a minimum of the technical details found in the rest of the paper.

A. Irrelevance of causal structure to the rules of inference

Unifying the quantum description of experiments involving two distinct systems at one time with the description of those involving a single system at two distinct times requires some modifications to the way that the Hilbert space formalism of quantum theory is usually set up. Conventionally, a Hilbert space $\mathcal{H}_A$ describes a system, labeled A, that persists through time. Given two such systems, A and B, the joint system is described by the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. In the present work, a Hilbert space and its associated label should rather be thought of as representing a localized region of space-time. Specifically, an *elementary region* is a small space-time region in which an agent might possibly make a single intervention in the course of an experiment, for example by making a measurement or by preparing a specific state. Each elementary region is associated with a label and a Hilbert space, for instance, A and $\mathcal{H}_A$.

Generally, a *region* refers to a collection of elementary regions. A region that is composed of a pair of disjoint regions, labeled A and B, is ascribed the tensor product Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. In contrast to the usual formalism, this applies regardless of whether A and B describe independent systems or the same system at two different times. Because of this, if an experiment involves a system that does persist through time, then a different label is given to each region it inhabits; e.g., the input and output spaces for a quantum channel are assigned different labels.

Although we have motivated our work by the distinction between spatial and temporal separation, in fact it is *not* the spatiotemporal relation between the regions that is relevant for how they ought to be represented in our quantum generalization of probability theory. Rather, it is the *causal* relation that holds between them which is important.

More precisely, what is important is the distinction between two regions that are *causally related*, which is to say that one has a causal influence on the other (perhaps via intermediaries), and two regions that are *acausally related*, which is to say that neither has a causal influence on the other (although they may have a common cause or a common effect or be connected via intermediaries to a common cause or a common effect).

The causal relation between a pair of regions cannot be inferred simply from their spatiotemporal relation. Consider a relativistic quantum theory for instance. Although a pair of regions that are spacelike separated are always acausally related, a pair of regions that are timelike separated can be related causally, for instance if they constitute the input and the output of a channel, or they can be related acausally, for instance if they constitute the input of one channel and the output of another. Although timelike separation implies that a causal connection is *possible*, it is whether such a connection *actually holds* that is relevant in our formalism. The distinction can also be made in nonrelativistic theories and in theories with exotic causal structure. Indeed, causal structure is a more primitive notion than spatiotemporal structure, and it is all that we need here.

Typically, we confine our attention to two paradigmatic examples of causal and acausal separation (which can be formulated in either a relativistic or a nonrelativistic quantum theory). Two distinct regions at the same time, the correlations between which are conventionally described by a bipartite quantum state, are acausally related. The regions at the input and output of a quantum channel, the correlations between which are conventionally described by an input state and a quantum channel, are causally related (although there are exceptions, such as a channel which erases the state of the system and then reprepares it in a fixed state).[1]

[1] Although it is not required here, one can be more precise about this distinction as follows. A causal structure for a set of quantum regions is represented by a directed acyclic graph wherein the nodes are the regions and the directed edges are relations of causal dependence (the restriction to acyclic graphs prohibits causal loops). Two regions are said to be causally related if for all paths connecting one to the other in the graph every edge along the path is directed in the same sense. Two systems are said to be acausally related if for all paths connecting one to the other not every edge along the path is directed in the same sense. When there exist both sorts of paths between a pair of nodes, the associated regions are neither purely causally nor purely acausally related. We do not consider this case in the article.

TABLE I. Analogies between the classical theory of Bayesian inference and the conditional states formalism for quantum theory.

	Classical	Quantum
State	$P(R)$	ρ_A
Joint state	$P(R,S)$	σ_{AB}
Marginalization	$P(S) = \sum_R P(R,S)$	$\rho_B = \text{Tr}_A(\sigma_{AB})$
Conditional state	$P(S\|R)$ $\sum_S P(S\|R) = 1$	$\sigma_{B\|A}$ $\text{Tr}_B(\sigma_{B\|A}) = I_A$
Relation between joint and conditional states	$P(R,S) = P(S\|R)P(R)$ $P(S\|R) = P(R,S)/P(R)$	$\sigma_{AB} = \sigma_{B\|A} \star \rho_A$ $\sigma_{B\|A} = \sigma_{AB} \star \rho_A^{-1}$
Bayes' theorem	$P(R\|S) = P(S\|R)P(R)/P(S)$	$\sigma_{A\|B} = \sigma_{B\|A} \star (\rho_A \rho_B^{-1})$
Belief propagation	$P(S) = \sum_R P(S\|R)P(R)$	$\rho_B = \text{Tr}_A(\sigma_{B\|A}\rho_A)$

We unify the description of Bayesian inference in the two different causal scenarios in the sense that various formulas are shown to have precisely the same form, in particular the relation between joints and conditionals, the formula for Bayesian inversion, and the formula for belief propagation.

B. Basic elements of the formalism

Without providing all the details, we summarize the analogs, within our formalism, of the most basic elements of classical probability theory. These are presented in Table I.

For an elementary region A, the quantum analog of a normalized probability distribution is a conventional quantum state ρ_A, that is, a positive trace-one operator on $\mathcal{H}_A$. For a region AB, composed of two disjoint elementary regions, the analog of a joint probability distribution is a trace-one operator σ_{AB} on $\mathcal{H}_{AB}$. This operator is not always positive (but we nonetheless refer to it as a *state*). The marginalization operation is replaced by the partial trace operation, Tr_A, which corresponds to ignoring region A. The role of the marginal distribution is played by the marginal state $\rho_B = \text{Tr}_A(\sigma_{AB})$.

The quantum analog of a conditional probability is a *conditional state* for region B given region A. This is an operator on $\mathcal{H}_{AB}$, denoted $\sigma_{B|A}$, that satisfies $\text{Tr}_B(\sigma_{B|A}) = I_A$.

The relation between a conditional state and a joint state is $\sigma_{B|A} = \sigma_{AB} \star \rho_A^{-1}$, where the $\star$ product is a particular noncommutative and nonassociative product, defined by $M \star N \equiv N^{1/2} M N^{1/2}$, where we have adopted the convention of dropping identity operators and tensor products, so that $\sigma_{AB} \star \rho_A^{-1}$ is shorthand for $\sigma_{AB} \star (\rho_A^{-1} \otimes I_B) = (\rho_A^{-1/2} \otimes I_B)\sigma_{AB}(\rho_A^{-1/2} \otimes I_B)$.

This relation implies that the quantum analog of Bayes' theorem, relating $\sigma_{B|A}$ and $\sigma_{A|B}$, is $\sigma_{A|B} = \sigma_{B|A} \star (\rho_A \rho_B^{-1})$. A standard example of inference then proceeds as follows. Suppose a conditional state $\sigma_{B|A}$ represents your beliefs about the relation that holds between a pair of elementary regions. In this case, if you represent your beliefs about A by the quantum state ρ_A, then you must represent your beliefs about B by the quantum state ρ_B, where

$$\rho_B = \text{Tr}_A(\sigma_{B|A}\rho_A). \tag{1}$$

We refer to this map from ρ_A to ρ_B as *belief propagation*.[2]

C. Relevance of causal structure to the form of the state

In the case of acausally related regions, it is the joint state that is easily inferred from the conventional formalism, and the conditional state that is derived from the joint. Specifically, if A and B are acausally related, then their joint state, σ_{AB}, is simply the bipartite state that one would assign to them in the conventional formalism. Consequently, σ_{AB} is a positive operator in this case. The conditional state can be inferred from the rule relating joints to conditionals, namely, $\sigma_{B|A} = \sigma_{AB} \star \rho_A$. It follows that $\sigma_{B|A}$ is also a positive operator.

On the other hand, if A and B are causally related, then it is the conditional state that is easily inferred from the conventional formalism and the joint state that is derivative. Specifically, if the regions are related by a quantum operation $\mathcal{E}_{B|A}$, then $\sigma_{B|A}$ is defined as the operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ that is Jamiołkowski isomorphic to $\mathcal{E}_{B|A}$ [22]. The joint state is then inferred from the rule relating joints to conditionals. One can show that both $\sigma_{B|A}$ and σ_{AB} fail to be positive in general, but they have positive partial transpose.

Because of this, the set of permissible joint and conditional states for acausally related regions is different from the set for causally related regions. To distinguish the two cases, we use the notation ρ_{AB} and $\rho_{B|A}$ for the acausal case and ϱ_{AB} and $\varrho_{B|A}$ for the causal case.

It is important to note that in a classical theory of Bayesian inference, the rules of inference are independent of the causal relations that hold among the variables. The causal relations can still be relevant, however, for constraining the probability distribution that is assigned to those variables. For instance, the causal relations among a triple of variables are significant for the sort of probability distribution that can be assigned to them. Specifically, if variable R is a common cause of variables S and T, while there is no direct causal connection between S and T, then S and T should be conditionally independent given R, which is to say that the joint distribution over

[2]Note that the term "belief propagation" has also been used to describe message-passing algorithms for performing inference on Bayesian networks. This is not the intended meaning here.

TABLE II. Translation of concepts and equations from conventional notation to the conditional states formalism.

	Conventional notation	Conditional states formalism
Probability distribution of X	$P(X)$	ρ_X
Probability that $X = x$	$P(X = x)$	$\rho_{X=x}$
Set of states on A	$\{\rho_x^A\}$	$\varrho_{A\|X}$
Individual state on A	ρ_x^A	$\varrho_{A\|X=x}$
POVM on A	$\{E_y^A\}$	$\varrho_{Y\|A}$
Individual effect on A	E_y^A	$\varrho_{Y=y\|A}$
Channel from A to B	$\mathcal{E}_{B\|A}$	$\varrho_{B\|A}$
Instrument	$\{\mathcal{E}_y^{B\|A}\}$	$\varrho_{YB\|A}$
Individual operation	$\mathcal{E}_y^{B\|A}$	$\varrho_{Y=y,B\|A}$
The Born rule	$\forall\, y : P(Y = y) = \mathrm{Tr}_A\left(E_y^A \rho_A\right)$	$\rho_Y = \mathrm{Tr}_A(\varrho_{Y\|A}\rho_A)$
Ensemble averaging	$\rho_A = \sum_x P(X = x)\rho_x^A$	$\rho_A = \mathrm{Tr}_X(\varrho_{A\|X}\rho_X)$
Action of a channel (Schrödinger)	$\rho_B = \mathcal{E}_{B\|A}(\rho_A)$	$\rho_B = \mathrm{Tr}_A(\varrho_{B\|A}\rho_A)$
Composition of channels	$\mathcal{E}_{C\|A} = \mathcal{E}_{C\|B} \circ \mathcal{E}_{B\|A}$	$\varrho_{C\|A} = \mathrm{Tr}_B(\varrho_{C\|B}\varrho_{B\|A})$
Action of a channel (Heisenberg)	$E_y^A = (\mathcal{E}_{B\|A})^\dagger(E_y^B)$	$\varrho_{Y\|A} = \mathrm{Tr}_B(\varrho_{Y\|B}\varrho_{B\|A})$
Nonselective state-update rule	$\forall\, y : P(Y = y)\rho_y^B = \mathcal{E}_y^{B\|A}(\rho_A)$	$\rho_{YB} = \mathrm{Tr}_A(\varrho_{YB\|A}\rho_A)$

these variables is not arbitrary, but has the form $P(R,S,T) = P(S|R)P(T|R)P(R)$.

In the quantum case, the situation is similar. The rules of inference, such as the formula for belief propagation, the formula for Bayesian inversion, and the relation between the joint and the conditional, do not depend on the causal relations between the regions under consideration, but causal relations *do* constrain the set of operators that can describe joint states.

In fact, the dependence is stronger in the quantum case because the set of permissible states depends on the causal relation even for a *pair* of regions. This is not a feature of a classical theory of inference: If we consider all the possible joint distributions over a pair of variables, R and S, we find that the set of possibilities is the same for the case where R and S are *causally* related as it is for the case where R and S are *acausally* related.

To reiterate, the *fact* that the set of possible states that can be assigned to a set of regions is constrained by the causal relation between those regions is *common* to the classical and quantum theories of inference. What is particular to the theory of quantum inference is that even in the case of a *pair* of regions, the causal relation between the regions is relevant for the set of possible states that can be assigned to those regions.[3]

D. Recasting conventional quantum notions in terms of conditional states and belief propagation

The conditional states formalism incorporates the possibility that a given region is associated to a classical variable rather than a quantum system. In this case, the classical variable is represented by a Hilbert space with a preferred basis, where the different elements of the basis correspond to different values of the variable, and any state assigned to that region is diagonal in that basis. Any joint state or conditional state involving this region is also restricted to have this diagonal form. It follows that for a set of regions that are all classical, the formalism reproduces the classical theory of Bayesian inference.

The formalism also yields a new and unified perspective on many notions in quantum theory. To see this, it is useful to recall that measurements, sets of state preparations, and transformations can all be represented by quantum operations, that is, as completely positive trace-preserving (CPT) linear maps. Channels are CPT maps wherein the input and output spaces are both quantum. A positive operator valued measure (POVM) is a CPT map from a quantum input to a classical output (the measurement outcome). A set of states is a CPT map from a classical input (the state index) to a quantum output (the associated state). Finally, a quantum instrument, which is a measurement together with a state-update rule for every outcome, can be represented as a CPT map from a quantum input to a composite output with a quantum part (the updated state) and a classical part (the measurement outcome). Insofar as every CPT map defines a conditional state, each of these notions in quantum theory is an instance of a conditional state in our formalism. This is summarized in the top half of Table II.

It follows that many relations that seem unrelated in the conventional formalism all become instances of the belief propagation rule in our formalism. This includes the Born rule, the formula for calculating the average state for an ensemble, the composition of channels, the state-update rule in a measurement, and the action of a channel in both the Heisenberg and Schrödinger pictures. This is summarized in the bottom half of Table II.

E. Applications of the formalism

The formalism also accommodates forms of belief propagation that do not fit into the standard list of the previous section.

One example is the inference made about one system based on the outcome of a measurement made on another when the two are correlated by virtue of a common cause. This reproduces the remote collapse postulate of quantum theory, which is sometimes called "remote steering" of a quantum state

[3]There does exist a classical analog of this dependence on causal structure for pairs of regions, but it requires considering the case of a classical theory with an epistemic restriction [23]. We do not pursue the analogy here.

and was made famous by the thought experiment of Einstein, Podolsky, and Rosen. It follows that in the conditional states framework, the steering effect is *merely* belief propagation (updating beliefs about one system based on new evidence about another) and does not require any causal influence from one to the other. This interpretation has been advocated previously by Fuchs [24]. Our formalism also provides an elegant derivation of the formula for the set of ensembles to which a remote system may be steered, previously obtained by conventional methods in [25].

Another example of an unconventional form of belief propagation is retrodiction, that is, inferences about a region based on beliefs about another region in its future. We develop a retrodictive formalism using our quantum Bayes' theorem. The latter is a necessary ingredient because the "givens" in a retrodiction problem are typically the descriptions of sets of state preparations, measurements, and channels, each of which corresponds to a conditional wherein the conditioning system is to the past of the conditioned system. We use Bayes' theorem to invert each of these conditionals to ones wherein the conditioning system is to the future of the conditioned system. Then, one can use these conditionals to propagate one's beliefs backwards in time, that is, to update one's beliefs about the past based on new evidence in the present. This application of our formalism is a good example of how one can achieve causal neutrality: Belief propagation backward in time follows the same rules as belief propagation forward in time. The retrodictive formalism we devise coincides with the one introduced in [26–28] in the case of unbiased sources, but differs in the general case, retaining a closer analogy with classical Bayesian inference.

In the case where a quantum system is passed through a channel (possibly noisy), the Bayesian inversion of the conditional associated to this channel, when interpreted as a quantum operation itself, is the Barnum-Knill approximate error correction map [29]. It follows that this error correction scheme is the quantum analog of the following classical error correction scheme: Based on a channel's output, compute a posterior distribution over inputs (i.e., classical retrodiction) and then sample from the latter.

In the case where a quantum system is prepared in one of a set of states, the Bayesian inversion of the conditional associated to this set of states (a "quantum given classical" conditional) is a conditional associated to a measurement (a "classical given quantum" conditional). Indeed, we find that in these contexts, our quantum Bayes' theorem reproduces the well-known rule relating sets of states to POVMs [3,4,30]. The POVM obtained as the Bayesian inversion of an ensemble of states turns out to be the "pretty-good" measurement for distinguishing those states [31–33]. Therefore, the latter, like the Barnum-Knill recovery operation, can be understood as a quantum analog of sampling from the posterior.

Similarly, the Bayesian inversion of the conditional associated with a measurement is a conditional associated with a set of states. For this case, our quantum Bayes' theorem reproduces a rule proposed by Fuchs as a quantum analog of Bayes' theorem [24].

Finally, we show that our notion of conditioning does not include the projection postulate as a special case and that previous arguments to the contrary (i.e., in favor of the projection postulate being viewed as an instance of Bayesian conditioning) [34,35] are based on a misleading analogy. Within the conditional states formalism, the projection postulate is best described as the application of a belief propagation rule (a nonselective update map), followed by conditioning (the selection). This is broadly in line with the treatment of quantum measurements advocated by Ozawa [36,37]. In support of the argument that the projection postulate is not a type of conditioning, we provide a conditional state version of the argument that all informative measurements must be disturbing, which may be of independent interest due to its close relationship to entanglement monogamy.

F. Structure of the paper

The remainder of this paper is structured as follows. The relevant aspects of classical conditional probability are reviewed in Sec. II. Section III introduces quantum conditional states and the basic concepts of quantum Bayesian inference for a pair of regions. The distinction between conditional states for causally related and acausally related regions is discussed here. This section also provides a detailed discussion of the translations from the conventional formalism to the conditional states formalism that are highlighted in Table II.

Section IV introduces our quantum version of Bayes' theorem and discusses its applications, in particular the connection with the update rule proposed by Fuchs, the correspondence between POVMs and ensemble decompositions of a density operator, the pretty-good measurement, and the Barnum-Knill recovery map. In Sec. IV C, we develop the retrodictive formalism for quantum theory and describe how it relates to the one introduced in [26–28]. Finally, in Sec. IV D, the acausal analog of the symmetry between prediction and retrodiction is discussed in the context of remote measurement.

Section V discusses quantum Bayesian conditioning. After a brief discussion of the general problem of conditioning a quantum region on another quantum region, we focus on conditioning a quantum region on a classical variable. This is the correct way to update quantum states in light of classical data, regardless of the causal relationship between the two. Various examples of this are discussed in Sec. V A, including the case of the remote steering phenomenon. Section V B concerns how to understand within our formalism the rules for updating quantum states after a nondestructive quantum measurement, in particular, how to understand the projection postulate.

In Sec. VI, we discuss related work. Quantum conditional states are compared to other proposals for quantum generalizations of conditional probability in Sec. VI A and the conditional states formalism is compared to several recently proposed operational reformulations of quantum theory in Sec. VI B.

Section VII discusses limitations of the conditional states framework. These arise because the classical operation of taking the product of a conditional and a marginal probability distribution to form a joint distribution is replaced by a noncommutative and nonassociative operation on the corresponding operators in the quantum case. Because of this, unlike in classical probability, an equation involving conditional states does not necessarily remain valid when both sides are conditionalized on an additional variable. This is discussed in Sec. VII A. Section VII B discusses the reasons why causal joint states are limited to two elementary regions. Section VII B1 discusses why they cannot be applied to mixed causal scenarios, such as two acausally related regions with a third region causally related to one of the other two, and Sec. VII B2 discusses the difficulties with generalizing the notion to multiple time steps, where we have three or more causally related regions.

Section VIII discusses an open question about when assignments of conditional states are compatible with one another. That not all conditional assignments are compatible can be shown via the monogamy of entanglement. Indeed, this incompatibility seems to be a more basic notion, of which monogamy is a consequence. Finally, we conclude in Sec. IX.

II. CLASSICAL CONDITIONAL PROBABILITY

In this section, the basic definitions and formalism of classical conditional probability are reviewed, with a view to their quantum generalization in Sec. III.

Let R denote a (discrete) random variable, $R = r$ the event that R takes the value r, $P(R = r)$ the probability of event $R = r$, and $P(R)$ the probability that R takes an arbitrary unspecified value. Finally, $\sum_R$ denotes a sum over the possible values of R.

A conditional probability distribution is a function of two random variables $P(S|R)$, such that for each value r of R, $P(S|R = r)$ is a probability distribution over S. Equivalently, it is a positive function of R and S such that

$$\sum_S P(S|R) = 1 \tag{2}$$

independently of the value of R.

Given a probability distribution $P(R)$ and a conditional probability distribution $P(S|R)$, a joint distribution over R and S can be defined via

$$P(R,S) = P(S|R)P(R), \tag{3}$$

where the multiplication is defined elementwise, i.e., for all values r,s of R and S, $P(R = r,S = s) = P(S = s|R = r)$ $P(R = r)$.

Conversely, given a joint distribution $P(R,S)$, the marginal distribution over R is defined as

$$P(R) = \sum_S P(R,S), \tag{4}$$

and the conditional probability of S given R is

$$P(S|R) = \frac{P(R,S)}{P(R)}. \tag{5}$$

Note that Eq. (5) only defines a conditional probability distribution for those values r of R such that $P(R = r) \neq 0$. The conditional probability is undefined for other values of R.

The chain rule for conditional probabilities states that a joint probability over n random variables $R_1,R_2,\ldots,R_n$ can be written as

$$\begin{aligned} &P(R_1,R_2,\ldots,R_n) \\ &\quad = P(R_n|R_1,R_2,\ldots,R_{n-1}) \\ &\qquad \times P(R_{n-1}|R_1,R_2,\ldots,R_{n-2})\ldots P(R_2|R_1)P(R_1). \end{aligned} \tag{6}$$

Finally, note that the process of marginalizing a distribution over a set of variables commutes with the process of conditioning on a disjoint set of variables, as illustrated in the following commutative diagram:

$$\begin{array}{ccc} P(R,S,T) & \xrightarrow{\sum_R} & P(S,T) \\ \Big\downarrow {\scriptstyle \times P(T)^{-1}} & & \Big\downarrow {\scriptstyle \times P(T)^{-1}} \\ P(R,S|T) & \xrightarrow{\sum_R} & P(S|T). \end{array} \tag{7}$$

III. QUANTUM CONDITIONAL STATES

In this section, the quantum analog of conditional probability—a conditional state—is introduced. We also discuss how the states assigned to disjoint regions are related via a quantum analog of the belief propagation rule $P(S) = \sum_R P(S|R)P(R)$. There is a small difference between conditional states for acausally related and causally related regions. The acausal case is discussed in Secs. III A and III B. On the other hand, Secs. III C–III K mainly concern the causal case, wherein we find that quantum dynamics, ensemble averaging, the Born rule, Heisenberg dynamics, and the transition from the initial state to the ensemble of states resulting from a measurement can all be represented as special cases of quantum belief propagation. Acausal analogies of some of these ideas are also developed in these sections.

A. Acausal conditional states

We begin by defining conditional states for acausally related regions. This scenario, and its classical analog, are depicted in Fig. 1. The definition proceeds in analogy with the classical treatment given in Sec. II. The convention of using $A,B,C,\ldots$ to label quantum regions that are analogous to classical variables $R,S,T,\ldots$ is adopted throughout. The labels $X,Y,Z,\ldots$ are reserved for classical variables associated with preparations and measurements, which remain classical when we pass from probability theory to the quantum analog.

The analog of a probability distribution $P(R)$ assigned to a random variable R is a quantum state (density operator)

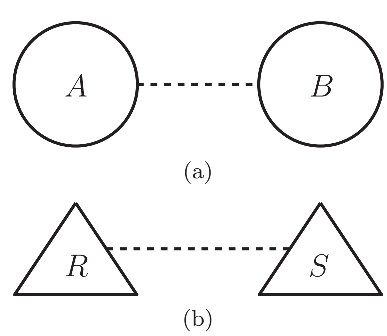


FIG. 1. Acausally related quantum and classical regions. Classical variables are denoted by triangles and quantum regions by circles (this convention is suggested by the shape of the convex set of states in each theory). The dotted line represents acausal correlation. (a) Two quantum regions in an arbitrary joint state (possibly correlated). (b) Two classical variables with an arbitrary joint probability distribution (possibly correlated).

TABLE III. Analogies between classical probability theory for two random variables and quantum theory for two acausally related regions.

Classical probability	Quantum theory
$P(R)$	ρ_A
$P(R,S)$	ρ_{AB}
$P(S) = \sum_R P(R,S)$	$\rho_B = \text{Tr}_A(\rho_{AB})$
$\sum_S P(S\|R) = 1$	$\text{Tr}_B(\rho_{B\|A}) = I_A$
$P(R,S) = P(S\|R)P(R)$	$\rho_{AB} = \rho_{B\|A} \star \rho_A$
$P(S\|R) = P(R,S)/P(R)$	$\rho_{B\|A} = \rho_{AB} \star \rho_A^{-1}$

ρ_A acting on a Hilbert space $\mathcal{H}_A$. When there are two disjoint regions with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ describes the composite region. The quantum analog of a joint distribution $P(R,S)$ is a density operator ρ_{AB} of the composite region, defined on $\mathcal{H}_{AB}$. The analog of marginalization over a variable is the partial trace over a region. These analogies are set out in the top half of Table III.

In analogy to the classical case, where $P(S|R)$ is a positive function that satisfies $\sum_S P(S|R) = 1$, an acausal conditional state for B given A is defined as follows.

Definition 1. An *acausal conditional state* for B given A is a positive operator $\rho_{B|A}$ on $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ that satisfies

$$\text{Tr}_B(\rho_{B|A}) = I_A, \tag{8}$$

where I_A is the identity operator on $\mathcal{H}_A$.

To provide an analogy with Eq. (3), a method of constructing a joint state on $\mathcal{H}_{AB}$ from a reduced state on $\mathcal{H}_A$ and a conditional state on $\mathcal{H}_{AB}$ is required. This is given by

$$\rho_{AB} = \big(\rho_A^{\frac{1}{2}} \otimes I_B\big)\rho_{B|A}\big(\rho_A^{\frac{1}{2}} \otimes I_B\big). \tag{9}$$

Equation (9) involves two constructions that appear repeatedly in what follows. First, the operators $\rho_A^{\frac{1}{2}}$ and $\rho_{B|A}$ are combined via multiplication, but they are defined on different spaces. To solve this problem, $\rho_A^{\frac{1}{2}}$ is expanded to an operator on $\mathcal{H}_{AB}$ by tensoring it with I_B. To simplify notation, the identity operators required to equalize the Hilbert spaces of two operators are left implicit, so that if M_{AB} is an operator on $\mathcal{H}_{AB}$ and N_{BC} is an operator on $\mathcal{H}_{BC}$ then $M_{AB}N_{BC} = (M_{AB} \otimes I_C)(I_A \otimes N_{BC})$ and an equation like $M_{AB} = N_{BC}$ is interpreted as $M_{AB} \otimes I_C = I_A \otimes N_{BC}$. This notation allows us to omit tensor product symbols where convenient, since $M_A \otimes N_B = (M_A \otimes I_B)(I_A \otimes N_B) = M_A N_B$.

Second, rather than simply multiplying $\rho_{B|A}$ with ρ_A in Eq. (9), $\rho_{B|A}$ is conjugated by $\rho_A^{\frac{1}{2}}$. This ensures that the resulting joint operator is positive. To define a notation for this conjugation, let M and N be positive operators on a Hilbert space $\mathcal{H}$. Then define a (nonassociative and noncommutative) product $M \star N$ via

$$M \star N = N^{\frac{1}{2}} M N^{\frac{1}{2}}. \tag{10}$$

With these conventions, Eq. (9) can be rewritten as

$$\rho_{AB} = \rho_{B|A} \star \rho_A, \tag{11}$$

which looks a lot closer to Eq. (3) than Eq. (9) does.

Starting with a joint state ρ_{AB} and its reduced state $\rho_A = \text{Tr}_B(\rho_{AB})$, a conditional state can be defined via

$$\rho_{B|A} = \rho_{AB} \star \rho_A^{-1}, \tag{12}$$

which is the analog of Eq. (5).

As with Eq. (5) there are problems with this formula if ρ_A is not supported on the entire Hilbert space $\mathcal{H}_A$. In that case Eq. (12) is to be understood as an equation on the Hilbert space $\text{supp}(\rho_A) \otimes \mathcal{H}_B$, where $\text{supp}(\rho_A)$ denotes the support of ρ_A (the span of the eigenvectors of ρ_A having nonzero eigenvalues). Because of this, the resulting conditional density operator satisfies $\text{Tr}_B(\rho_{B|A}) = I_{\text{supp}(\rho_A)}$ rather than Eq. (8).

The analogies between the classical and quantum relations between conditionals, marginals, and joints are set out in the bottom half of Table III. As these analogies suggest, the $\star$-product notation allows equations from classical probability to be generalized to quantum theory by replacing functions with operators, products with $\star$ products, and division with $\star$ products with the inverse. However, while this is a useful way of postulating results in the conditional states formalism, one has to take care of the nonassociativity and noncommutativity of the $\star$ product when making such generalizations.

To provide an analogy with the chain rule of Eq. (6) it is helpful to adopt the convention that, in the absence of parentheses, $\star$ products are evaluated right to left. Then, given n disjoint acausally related regions $A_1, A_2, \ldots, A_n$, with Hilbert space $\mathcal{H}_{A_1,A_2,\ldots,A_n} = \bigotimes_{j=1}^n \mathcal{H}_{A_j}$, the joint state can be written as

$$\begin{aligned}\rho_{A_1,A_2,\ldots,A_n} = \rho_{A_n|A_1,A_2,\ldots,A_{n-1}} \star \rho_{A_{n-1}|A_1,A_2,\ldots,A_{n-2}} \\ \star \cdots \star \rho_{A_2|A_1} \star \rho_{A_1}.\end{aligned} \tag{13}$$

Finally, note that the process of marginalizing a conditional state over a region commutes with the process of conditioning on a disjoint region, as illustrated in the following commutative diagram:

$$\begin{tikzcd} \rho_{ABC} \arrow[r, "\text{Tr}_C"] \arrow[d, "\rho_A^{-\frac{1}{2}}(\cdot)\rho_A^{-\frac{1}{2}}"] & \rho_{AB} \arrow[d, "\rho_A^{-\frac{1}{2}}(\cdot)\rho_A^{-\frac{1}{2}}"] \\ \rho_{BC|A} \arrow[r, "\text{Tr}_C"] & \rho_{B|A}. \end{tikzcd} \tag{14}$$

Example 1: Classical states. As one might expect, classical conditional probability is a special case of the quantum constructions outlined above. To see this, the classical variables have to be encoded in quantum regions in some way, and we adopt the convention of using the same letter to denote the classical variable and the corresponding quantum region. Thus, $\mathcal{H}_R, \mathcal{H}_S, \mathcal{H}_T, \ldots$ refer to quantum regions that encode classical random variables $R, S, T, \ldots$, as opposed to $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C, \ldots$, which are general quantum regions.

For classical random variables R and S, pick Hilbert spaces $\mathcal{H}_R$ and $\mathcal{H}_S$ with dimension equal to the number of distinct values of R and S, respectively, and choose orthonormal bases $\{|r\rangle\}$ for $\mathcal{H}_R$ and $\{|s\rangle\}$ for $\mathcal{H}_S$ labeled by the possible values of R and S. Then, joint, marginal, and conditional probability distributions are encoded as operators via

$$\rho_{RS} = \sum_{r,s} P(R=r,S=s)\,|r\rangle\,\langle r|_R \otimes |s\rangle\,\langle s|_S\,, \qquad (15)$$

$$\rho_R = \sum_{r} P(R=r)\,|r\rangle\,\langle r|_R = \mathrm{Tr}_S\,(\rho_{RS})\,, \qquad (16)$$

$$\rho_{S|R} = \sum_{r,s} P(S=s|R=r)\,|r\rangle\,\langle r|_R \otimes |s\rangle\,\langle s|_S\,. \qquad (17)$$

Using Eqs. (2)–(5), it is straightforward to check that these operators satisfy Eqs. (8), (11), and (12).

In order to unify the notation for classical variables and quantum regions, the operators ρ_{RS}, ρ_R, and $\rho_{S|R}$ are often used to directly represent the functions $P(R,S)$, $P(R)$, and $P(S|R)$ without introducing the classical functions explicitly. Whenever states and conditional states have subscripts R,S,T or X,Y,Z, they are implicitly assumed to be of this classical form. If needed, the classical functions can be read off from Eqs. (15)–(17).

Example 2: Pure conditional states. A pure conditional state is one that is of the form $\rho_{B|A} = |\psi\rangle\langle\psi|_{B|A}$ for some vector $|\psi\rangle_{B|A} \in \mathcal{H}_{AB}$. Since $\mathrm{Tr}_B(\rho_{B|A}) = I_A$, and I_A has all eigenvalues equal to 1, the Schmidt decomposition of $|\psi\rangle_{B|A}$ is of the form

$$|\psi\rangle_{B|A} = \sum_k |u_k\rangle_A \otimes |v_k\rangle_B\,, \qquad (18)$$

where $\{|u_k\rangle\}$ is an orthonormal basis for $\mathcal{H}_A$ and $\{|v_k\rangle\}$ is an orthonormal basis for $\mathcal{H}_B$. This implies that a pure conditional state only exists if $\dim(\mathcal{H}_A) \leqslant \dim(\mathcal{H}_B)$ because otherwise there would not be enough orthonormal vectors on the B side to enforce $\mathrm{Tr}_B(\rho_{B|A}) = I_A$.[4]

Since all the Schmidt coefficients are the same, the bases $\{|u_k\rangle\}$ and $\{|v_k\rangle\}$ are highly nonunique. The conditional state $|\psi\rangle_{B|A}$ itself only determines the relationship between the two Schmidt bases; i.e., for any basis in $\mathcal{H}_A$ it determines a corresponding basis in $\mathcal{H}_B$. To see this, fix a reference basis, $\{|j\rangle\}$, for $\mathcal{H}_A$ in order to define a complex conjugation operation. Next, define an isometry $U_{B|A} = \sum_k |v_k\rangle_B\langle u_k^*|_A$, where * denotes complex conjugation in the $\{|j\rangle\}$ basis. Then, if $\{|w_k\rangle_A\}$ is any other basis for $\mathcal{H}_A$, Eq. (18) can be rewritten as

$$|\psi\rangle_{B|A} = \sum_k |w_k\rangle_A \otimes U_{B|A'}|w_k^*\rangle_{A'}, \qquad (19)$$

where A' labels a second copy of $\mathcal{H}_A$. With respect to the reference basis $\{|j\rangle\}$, this simplifies to

$$|\psi\rangle_{B|A} = U_{B|A'}|\Phi^+\rangle_{AA'}, \qquad (20)$$

where $|\Phi^+\rangle_{AA'} = \sum_j |jj\rangle_{AA'}$.

Let ρ_A be an arbitrary density operator on $\mathcal{H}_A$ with eigendecomposition $\rho_A = \sum_k p_k |w_k\rangle\langle w_k|_A$. Combining this with $|\psi\rangle\langle\psi|_{B|A}$ via Eq. (11) in order to define a joint state gives the projector onto the pure state,

$$|\psi\rangle_{AB} = \rho_A^{\frac{1}{2}}|\psi\rangle_{B|A}. \qquad (21)$$

Combining this with Eq. (19) gives the Schmidt decomposition

$$|\psi\rangle_{AB} = \sum_k \sqrt{p_k}|w_k\rangle_A \otimes U_{B|A'}|w_k^*\rangle_{A'}. \qquad (22)$$

Since an arbitrary pure joint state is of this form, this shows that pure conditional states determine pure joint states when combined with arbitrary reduced states and, conversely, the conditional state of a pure joint state is always pure.

Note that using Eq. (20) instead of Eq. (19) gives

$$|\psi\rangle_{AB} = \rho_A^{1/2}U_{B|A'}|\Phi^+\rangle_{AA'}, \qquad (23)$$

which is a well-known canonical decomposition of a bipartite pure state.

B. Acausal belief propagation

Suppose you characterize your beliefs about two classical variables, R and S, by specifying a marginal probability distribution $P(R)$ and a conditional probability distribution $P(S|R)$. Then, you can compute the probability distribution you ought to assign to S via

$$P(S) = \sum_R P(S|R)P(R). \qquad (24)$$

This is called the *classical belief propagation rule* (also known as the law of total probability). It follows from calculating the joint distribution $P(R,S) = P(S|R)P(R)$ and then marginalizing over R.

The belief propagation rule can be thought of as specifying a linear map $\Gamma_{S|R}$ from the space of probability distributions over R to the space of probability distributions over S that preserves positivity and normalization. This is defined as

$$\Gamma_{S|R}(P(R)) \equiv \sum_R P(S|R)P(R). \qquad (25)$$

Propagating beliefs about a quantum region A to an acausally related region B works in a similar way. If you specify a reduced state ρ_A and a conditional state $\rho_{B|A}$ then your state for B is determined by the *acausal quantum belief propagation rule*,

$$\rho_B = \mathrm{Tr}_A(\rho_{B|A}\rho_A), \qquad (26)$$

which follows from the fact that the joint state is $\rho_{AB} = \rho_{B|A} \star \rho_A$, so that $\rho_B = \mathrm{Tr}_A(\rho_{B|A} \star \rho_A)$, and from the cyclic property of the trace.

As in the classical case, acausal belief propagation can also be viewed as a linear map $\mathfrak{E}_{B|A}$ from states on A to states on B that preserves positivity and normalization, defined by

$$\mathfrak{E}_{B|A}(\rho_A) \equiv \mathrm{Tr}_A(\rho_{B|A}\rho_A). \qquad (27)$$

The linear map so defined is clearly positive because it maps states to states. It is not completely positive in general, but its composition with a transpose on A is completely positive. The map $\mathfrak{E}_{B|A}$ is, in fact, identical to the map associated to $\rho_{B|A}$ via the Jamiołkowski isomorphism [22], which is a familiar construction in quantum information theory. These facts are consequences of the following theorem.

Theorem 1: Jamiołkowski isomorphism. Let $\mathfrak{E}_{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ be a linear map and let $M_{AC} \in \mathfrak{L}(\mathcal{H}_{AC})$, where $\mathcal{H}_C$ is a Hilbert space of arbitrary dimension. Then, the

[4] If $|\psi\rangle_{B|A}$ derives from a joint pure state $|\psi\rangle_{AB}$ via Eq. (12), then this only implies that $\dim[\mathrm{supp}(\rho_A)] \leqslant \dim(\mathcal{H}_B)$, which is always true because the ranks of ρ_A and ρ_B are equal.

action of $\mathfrak{E}_{B|A}$ on $\mathfrak{L}(\mathcal{H}_A)$ [tensored with the identity on $\mathfrak{L}(\mathcal{H}_C)$] is given by

$$(\mathfrak{E}_{B|A} \otimes \mathcal{I}_C)(M_{AC}) = \mathrm{Tr}_A(\rho_{B|A} M_{AC}), \tag{28}$$

where $\rho_{B|A} \in \mathfrak{L}(\mathcal{H}_{AB})$ is given by

$$\rho_{B|A} \equiv (\mathfrak{E}_{B|A'} \otimes \mathcal{I}_A)\left(\sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}\right). \tag{29}$$

Here, A' labels a second copy of A, $\mathcal{I}_A$ is the identity superoperator on $\mathfrak{L}(\mathcal{H}_A)$, and $\{|j\rangle\}$ is an orthonormal basis for $\mathcal{H}_A$.

Furthermore, the operator $\rho_{B|A}$ is an acausal conditional state; i.e., it satisfies Definition 1, if and only if $\mathfrak{E}_{B|A} \circ T_A$ is completely positive and trace preserving (CPT), where $T_A : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_A)$ denotes the linear map implementing the partial transpose relative to some basis.

The proof is provided in the Appendix.

C. Causal conditional states

The analogy between conditional probabilities and conditional states presented so far is not complete. In conventional quantum theory, the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ is used to represent a joint system with two subsystems, so that the conditional state $\rho_{B|A}$ refers to the state of two subsystems at a given time. However, for classical conditional probabilities, there is no corresponding requirement that the two random variables R and S appearing in $P(S|R)$ should have any particular causal relation to one another. Indeed, R might equally well represent the input to a classical channel and S the output; i.e., they may be causally related. This is illustrated in Fig. 2(b). If this is indeed the case, then the classical belief propagation rule of Eqs. (24) and (25) can be interpreted as stochastic dynamics.

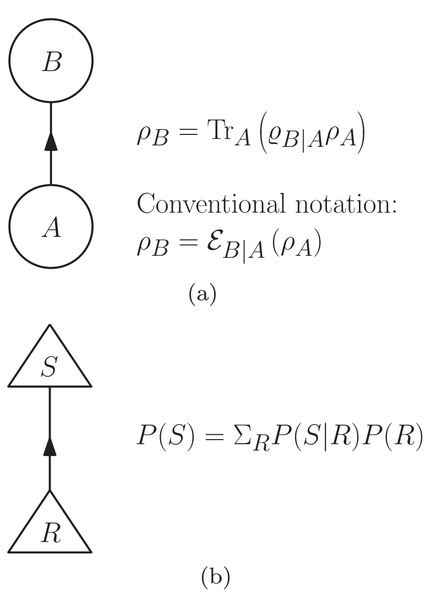


FIG. 2. Causally related quantum and classical regions. The arrows represent the direction of causal influence. (a) General quantum dynamics. A is the input to a CPT map and B is the output. (b) Classical stochastic dynamics. R is the input to a classical channel and S is the output.

In order to formulate quantum theory as a causally neutral theory of Bayesian inference, the same formalism should be used to describe causally related regions as is used to describe acausally related regions. In particular, if A and B are two causally related regions, as depicted in Fig. 2(a), then it ought to be possible to define a quantum conditional state for B given A as an operator on the tensor product $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. Towards this end, we make the following definition.

Definition 2. A *causal conditional state* of B given A is an operator $\varrho_{B|A}$ on $\mathcal{H}_{AB}$ that can be written as

$$\varrho_{B|A} = \rho_{B|A}^{T_A}, \tag{30}$$

for some acausal conditional state $\rho_{B|A}$, where T_A denotes the partial transpose in some basis on $\mathcal{H}_A$.

Thus, the set of causal conditional states is just the image under a partial transpose on the conditioning region of the set of acausal conditional states. Note that, although the partial transpose is basis dependent, its image on the set of acausal conditional states is not and therefore neither is our definition of a causal conditional state. Also, because the set of acausal conditional states is mapped to itself by the full transpose (i.e., the transpose on AB), a partial transpose over the conditioned region B, rather than the conditioning region A, could alternatively have been used to define a causal conditional state. Due to the partial transpose, causal conditional states are not positive operators in general, but they are always *locally* positive; i.e., $\langle\psi|_A \otimes \langle\phi|_B \varrho_{B|A} |\psi\rangle_A \otimes |\phi\rangle_B \geqslant 0$ for all $|\psi\rangle_A \in \mathcal{H}_A, |\phi\rangle_B \in \mathcal{H}_B$.

In this section, we show that defining causal conditional states in this way allows us to implement quantum belief propagation across causally related regions using the same formula as one uses for quantum belief propagation across acausally related regions, namely by a rule of the form $\rho_B = \mathrm{Tr}_A(\varrho_{B|A}\rho_A)$. Belief propagation for dynamics with a quantum input and a quantum output are treated in Sec. III D. Section III I treats belief propagation for causal conditional states themselves, which corresponds to composition of dynamical maps. Section III E introduces the notion of a causal joint state, which is analogous to the joint distribution of input and output variables for a classical channel. Section III F introduces the idea of a quantum-classical hybrid, which is a composite of a quantum region and a classical variable. This allows dynamics with a classical input and quantum output (and vice versa) to be described in terms of causal conditional states. These correspond to ensemble preparation procedures and measurements, as discussed in Secs. III G and III H. In Sec. III J, the Heisenberg picture is translated into the conditional states formalism. In Sec. III K, the most general type of state-update rule that can occur after a measurement—a quantum instrument—is described in terms of causal conditional states. Table II summarizes the translation of these concepts from conventional notation to the conditional states formalism.

D. Quantum channels as causal belief propagation

Conventionally, the transition from a region A to a causally related later region B is described by a dynamical CPT map $\mathcal{E}_{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ such that, if ρ_A is the state of A and ρ_B is the state of B, then $\rho_B = \mathcal{E}_{B|A}(\rho_A)$. However, causal conditional states can provide an alternative representation of quantum dynamics, as we will show. First note the following isomorphism.

Theorem 2. Let $\mathcal{E}_{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ be a linear map and let $\varrho_{B|A} \in \mathfrak{L}(\mathcal{H}_{AB})$ be the Jamiołkowski-isomorphic operator, as defined in Eq. (29). Then, $\varrho_{B|A}$ is a causal conditional state; i.e. it satisfies Definition 2, if and only if $\mathcal{E}_{B|A}$ is CPT.

Proof. Define $\rho_{B|A} \equiv \varrho_{B|A}^{T_A}$ and let $\mathfrak{E}_{B|A}$ be the linear map that is Jamiołkowski isomorphic to $\rho_{B|A}$. It follows that $\mathcal{E}_{B|A} = \mathfrak{E}_{B|A} \circ T_A$. Recalling the relation between causal and acausal conditional states, $\varrho_{B|A}$ is a causal conditional state if and only if $\rho_{B|A}$ is an acausal conditional state. Recalling Theorem 1, $\rho_{B|A}$ is an acausal conditional state if and only if $\mathfrak{E}_{B|A} \circ T_A$ is CPT. It follows that $\varrho_{B|A}$ is a causal conditional state if and only if $\mathcal{E}_{B|A}$ is CPT. ■

Together with Theorem 1, this implies that the action of a CPT map $\mathcal{E}_{B|A}$ on an operator M_{AC} is given by

$$\mathcal{E}_{B|A}(M_{AC}) = \mathrm{Tr}_A(\varrho_{B|A} M_{AC}), \tag{31}$$

where $\varrho_{B|A}$ is the Jamiołkowski isomorphic operator to $\mathcal{E}_{B|A}$.

Quantum dynamics may be represented by causal conditional states as follows.

Proposition 1. Let $\varrho_{B|A}$ be the causal conditional state that is Jamiołkowski-isomorphic to a CPT map $\mathcal{E}_{B|A}$ that describes a quantum dynamics. If the initial state of region A is ρ_A, then the state of B, conventionally written as

$$\rho_B = \mathcal{E}_{B|A}(\rho_A), \tag{32}$$

can be expressed in the conditional states formalism as

$$\rho_B = \mathrm{Tr}_A(\varrho_{B|A}\rho_A), \tag{33}$$

in analogy with the classical belief propagation rule Eq. (24).

We call Eq. (33) the *causal quantum belief propagation rule*. It follows from Eq. (31).

Figure 2 and the fourth and eighth lines of Table II summarize how this representation of quantum dynamics contrasts with the conventional representation, with Fig. 2 emphasizing the analogy between the classical and quantum belief propagation rules.

E. Causal joint states

Section III A showed that a joint state ρ_{AB} of two acausally related regions can be decomposed into a reduced state ρ_A and an acausal conditional state $\rho_{B|A}$. Similarly, two causally related regions can be described by an input state ρ_A and a causal conditional state $\varrho_{B|A}$, but so far there is no causal analog of a joint state. This is addressed by making the following definition, in analogy with Eq. (11).

Definition 3. A *causal joint state* of two causally related regions, A and B, is an operator on $\mathcal{H}_{AB}$ of the form

$$\varrho_{AB} = \varrho_{B|A} \star \rho_A, \tag{34}$$

where ρ_A is a state on $\mathcal{H}_A$ and $\varrho_{B|A}$ is a causal conditional state of B given A.

Note that the reduced state on A of ϱ_{AB} is the initial state (input to the channel) and the reduced state on B is

$$\rho_B = \mathrm{Tr}_A(\varrho_{B|A} \star \rho_A), \tag{35}$$

which, by the cyclic property of the trace and Proposition 1, is the final state (output of the channel).

It is not too difficult to see that a causal joint state ϱ_{AB} is the partial transpose of an acausal joint state on $\mathcal{H}_{AB}$. Specifically, $\varrho_{AB}^{T_A} = \rho_{B|A} \star \rho_A^{T_A}$, where $\rho_{B|A} = \varrho_{B|A}^{T_A}$ is an acausal conditional state and $\rho_A^{T_A}$ is a valid reduced state because the transpose preserves positivity.

Thus, just as a causal conditional state for B given A is an operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ that can be obtained as the partial transpose over A of an acausal conditional state $\rho_{B|A}$, a causal joint state on AB is simply an operator on $\mathcal{H}_A \otimes \mathcal{H}_B$ that can be obtained as the partial transpose over A of an acausal joint state ρ_{AB}.

Example 3: Unitary dynamics. Suppose a region A is assigned the state ρ_A with eigendecomposition $\rho_A = \sum_j p_j |u_j\rangle\langle u_j|$. A is then mapped to a region B, which has a Hilbert space of the same dimension as that of A, by an isometry $U_{B|A} = \sum_j |v_j\rangle_B \langle u_j|_A$. Since the Jamiołkowski isomorphism is basis independent, the causal conditional state associated with the map $\mathcal{E}_{B|A}(\cdot) = U_{B|A}(\cdot)(U_{B|A})^\dagger$ can be written in the eigenbasis of ρ_A as

$$\varrho_{B|A} = \sum_{j,k} |u_j\rangle\langle u_k|_A \otimes U_{B|A'}|u_k\rangle\langle u_j|_{A'}(U_{B|A'})^\dagger \tag{36}$$

$$= \sum_{j,k} |u_j\rangle\langle u_k|_A \otimes |v_k\rangle\langle v_j|. \tag{37}$$

It follows that the causal joint state for AB is

$$\varrho_{AB} = \sum_{j,k} \sqrt{p_j p_k} |u_j\rangle\langle u_k|_A \otimes |v_k\rangle\langle v_j|_B. \tag{38}$$

Note that the pure causal conditional state $\varrho_{B|A}$ is the partial transpose over A of a pure acausal conditional state, as can be seen by comparison with Eq. (18) from Example 2. Also, the pure causal joint state ϱ_{AB} is the partial transpose over A of a pure acausal joint state $\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$, where

$$|\psi\rangle_{AB} = \sum_j \sqrt{p_j} |u_j\rangle_A \otimes |v_j\rangle_B. \tag{39}$$

Causal joint states can be given an operational interpretation similar to that of acausal joint states by specifying a procedure to perform tomography on them (see [3,4]). The motivation for introducing them here is that they allow quantum Bayesian inference to be developed in a way that is blind to the distinction between acausally and causally related regions. This is discussed in Sec. IV.

Note that whereas acausal joint states may involve more than two acausally related regions, causal joint states are thus far only well defined for *two* causally related regions. The reasons for this limitation are discussed in Sec. VII.

F. Quantum-classical hybrid regions

An ensemble preparation procedure can be represented by a CPT map from a classical variable to a quantum region and a measurement can be represented as a CPT map from a quantum region to a classical variable. Therefore, these processes can be represented by conditional states for which either the conditioned or conditioning region is classical as a special case of Proposition 1. In order to compare this to the conventional formalism, we need to describe how composite regions consisting of a classical variable and a quantum region are represented in the conditional states formalism. Such composites are called quantum-classical hybrid regions.

As in Example 1, the classical variable X is associated with a Hilbert space $\mathcal{H}_X$ equipped with a preferred basis $\{|x\rangle\}$ that represents the possible values of X. The quantum region is associated with a Hilbert space $\mathcal{H}_A$ and the joint region with the tensor product $\mathcal{H}_{XA} = \mathcal{H}_X \otimes \mathcal{H}_A$. In order to preserve the classical nature of X, states and conditional states on $\mathcal{H}_{XA}$ are restricted to be of the following form.

Definition 4. A *hybrid operator* on $\mathcal{H}_{XA}$ is an operator of the form

$$M_{XA} = \sum_x |x\rangle\langle x|_X \otimes M_x^A, \tag{40}$$

where $\{|x\rangle\}$ is a preferred basis for $\mathcal{H}_X$ and $\{M_x^A\}$ is a set of operators acting on $\mathcal{H}_A$, labeled by the values of X. The operators M_x^A are referred to as the *components* of M_{XA}.

When M_{XA} is a state, this ensures that the reduced state on X is diagonal in the preferred basis and that there can be no entanglement between the quantum and classical regions.

Although our primary interest is in causal hybrids, since these are relevant to preparations and measurements, one can also have acausal hybrids. As shown below, the set of acausal hybrid conditional states and the set of acausal hybrid joint states are invariant under partial transpose. It follows that the set of acausal hybrid states and the set of causal hybrid states are the same. In particular, this means that, unlike fully quantum causal states, hybrid causal states not only have positive partial transpose but are themselves positive, and unlike fully quantum acausal states, hybrid acausal states not only are positive but also have positive partial transpose.

Therefore, for hybrid states, the notational distinction between ρ and ϱ serves merely as a reminder of the causal arrangement of the regions under consideration. This in contrast with the fully quantum case, where the distinction also has significance for the mathematical properties of the operator. When making claims about the mathematical properties of hybrid conditional states that are independent of causal structure, the notation σ is used. According to these conventions, any formula expressed in terms of σ's will yield a valid formula about hybrid states if the σ's are replaced by either ρ's or ϱ's.

For hybrid regions, there are two possible types of conditional state, depending on whether the conditioning is done on the quantum or the classical region. In the case of conditioning on the classical variable, a hybrid conditional state $\sigma_{A|X}$ is a positive operator on $\mathcal{H}_{XA}$, satisfying

$$\mathrm{Tr}_A(\sigma_{A|X}) = I_X. \tag{41}$$

In the case of conditioning on the quantum region, a hybrid conditional state $\sigma_{X|A}$ is again a positive operator on $\mathcal{H}_{XA}$, but this time satisfying

$$\mathrm{Tr}_X(\sigma_{X|A}) = I_A. \tag{42}$$

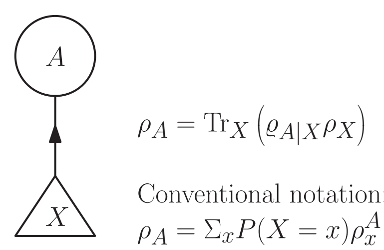
G. Ensemble averaging as belief propagation

A hybrid conditional state of the form $\sigma_{A|X}$ is a quantum state conditioned on a classical variable. The causal interpretation of such states is a process that takes a classical variable as input and outputs a quantum state. This is just an ensemble preparation procedure. In such a preparation procedure, a classical random variable X is sampled from a probability distribution $P(X)$ (by flipping coins, rolling dice, or any other suitable method). Depending on the value x of X thereby obtained, one of a set of quantum states $\{\rho_x^A\}$ is prepared for a quantum region A. If you do not know the value of X, then you should assign the ensemble average state $\rho_A = \sum_x P(X = x)\rho_x^A$ to A. This scenario is depicted in Fig. 3(a). A quantum preparation procedure has an obvious classical analog wherein the quantum region A is replaced by a classical variable R that is prepared in one of a set of probability distributions $P(R|X = x)$ depending on the value x of X. If you do not know the value of X, then the classical belief propagation rule specifies that you should assign the probability distribution $P(R) = \sum_X P(R|X)P(X)$ to R. This case is illustrated in Fig. 3(b). This section shows that a set of density operators can be represented by a hybrid conditional state of the form $\sigma_{A|X}$ and that the formula for the ensemble average state in a preparation procedure is a special case of quantum belief propagation.

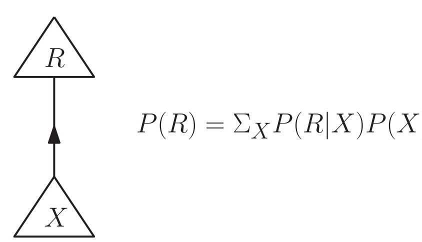
Theorem 3. Let $\sigma_{A|X}$ be a hybrid operator, so that by Eq. (40) it can be written as

$$\sigma_{A|X} = \sum_x \rho_x^A \otimes |x\rangle\langle x|_X, \tag{43}$$

for some set of operators $\{\rho_x^A\}$. Then, $\sigma_{A|X}$ satisfies the definition of both an acausal and a causal conditional state



(a)



(b)

FIG. 3. Quantum and classical preparation procedures. (a) A quantum preparation procedure is a process that takes a classical variable X as input and outputs a quantum region A in one of a set $\{\rho_x^A\}$ of states, depending on the value of X. It is mathematically equivalent to the special case of a CPT map where the input is classical. (b) A classical preparation procedure is a process that takes a variable X as input and outputs one of a set $\{P(R|X = x)\}$ of probability distributions over R, depending on the value of X. It is mathematically equivalent to the stochastic dynamics depicted in Fig. 2(b).

for A given X if and only if each of the components ρ_x^A is a normalized state on $\mathcal{H}_A$.

The proof is provided in the Appendix.

It is often convenient to use the notation $\sigma_{A|X=x} = \langle x|_X \sigma_{A|X} |x\rangle_X = \rho_x^A$ for the components of a conditional state of this form.

If we adopt the convention that the partial transpose on the Hilbert space of a classical variable is performed in its preferred basis then it has no effect on hybrid operators. Thus, acausal conditional states of the form $\rho_{A|X}$ are invariant under partial transpose on X and this is why acausal and causal conditional states that are conditioned on the classical variable have the same form.[5] The remainder of this section concerns the causal interpretation of such states in terms of preparation procedures, so we shift to the notation $\varrho_{A|X}$.

Proposition 2. Let $\varrho_{A|X}$ be the causal hybrid conditional state with components given by a set of states $\{\rho_x^A\}$. The ensemble average state arising from a preparation procedure that samples a value x of a classical variable X from the distribution $P(X)$ and prepares the state ρ_x^A is given by

$$\rho_A = \sum_x P(X=x)\rho_x^A. \tag{44}$$

This can be expressed in the conditional states formalism via the quantum belief propagation rule as

$$\rho_A = \mathrm{Tr}_X(\varrho_{A|X}\rho_X), \tag{45}$$

where $\rho_X = \sum_x P(X=x)|x\rangle\langle x|_X$.

This result follows simply from substituting the definition of ρ_X and $\varrho_{A|X}$ into Eq. (45).

Figure 3 and the second and seventh lines of Table II summarize how the representation of a preparation procedure within the conditional states formalism contrasts with the conventional representation and how the latter generalizes the analogous classical expression.

It should be noted that Theorem 3 can alternatively be derived as a special case of Theorem 2 and Proposition 2 as a special case of Proposition 1. This follows from the fact that a preparation procedure can be represented by a CPT map $\mathcal{E}_{A|X}$ from a classical variable to a quantum region (sometimes called a CQ map). The map is defined on diagonal states ρ_X via

$$\mathcal{E}_{A|X}(\rho_X) = \sum_x \langle x|_X \rho_X |x\rangle_X \rho_x^A. \tag{46}$$

By Proposition 1, the conditional state associated with the preparation is the Jamiołkowski isomorphic operator to this map. Equation (45) is then obtained as a special case of Eq. (33) where the input is classical.

[5]Even if we do not adopt the convention of evaluating partial transposes in the preferred basis, the sets of acausal and causal conditional states are still isomorphic. If $\{|x\rangle_X\}$ is the preferred basis for acausal states, then this amounts to choosing a different preferred basis $\{|x^*\rangle_X\}$ for causal states, where * is complex conjugation in the basis used to define the partial transpose. However, this is an unnecessary complication that is avoided by adopting the recommended convention.

H. The Born rule as belief propagation

A hybrid conditional state of the form $\sigma_{Y|A}$ is a classical probability distribution conditioned on a quantum region. The causal interpretation of such states is as a process that takes a quantum region as input and outputs a classical variable. This is just a measurement. The most general kind of measurement on a quantum region A is conventionally represented by a POVM $\{E_y^A\}$ with the classical variable Y ranging over the possible outcomes. If the state of the region is ρ_A, then the probability of obtaining outcome y is given by the Born rule as $P(Y=y) = \mathrm{Tr}_A(E_y^A\rho_A)$. This scenario is depicted in Fig. 4(a). In the classical analog, the quantum region A is replaced by a classical variable R, the state ρ_A is replaced by a distribution $P(R)$, and the POVM is replaced by a (possibly noisy) classical measurement described by a set of response functions $\{P(Y=y|R)\}$, i.e., a set of functions of R labeled by the values of Y, where $P(Y=y|R=r)$ specifies the probability of obtaining the outcome y given that $R=r$. The overall probability of obtaining the outcome y is given by $P(Y=y) = \sum_R P(Y=y|R)P(R)$, which is just another instance of belief propagation. This case is illustrated in Fig. 4(b). In analogy to this, the remainder of this section shows that the components of a conditional state $\sigma_{Y|A}$ form a POVM and that the Born rule can be written as quantum belief propagation with respect to a causal conditional state of this form.

Theorem 4. Let $\sigma_{Y|A}$ be a hybrid operator so that, by Eq. (40), it can be written in the form

$$\sigma_{Y|A} = \sum_y |y\rangle\langle y|_Y \otimes E_y^A, \tag{47}$$

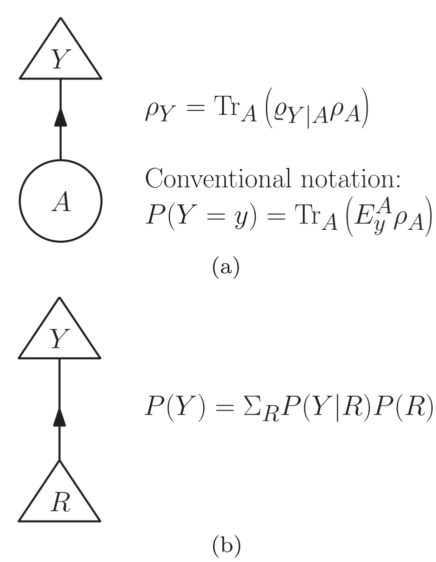


FIG. 4. Quantum and classical measurements. (a) A quantum measurement is a process that takes a quantum region A as input and outputs a classical variable Y. It is mathematically equivalent to the special case of a CPT map where the output is classical. (b) A classical (noisy) measurement is a process that takes a variable R as input and outputs a variable Y that depends on R, possibly in a coarse-grained or nondeterministic way. It is mathematically equivalent to the stochastic dynamics depicted in Fig. 2(b).

for some set of operators $\{E_y^A\}$. Then $\sigma_{Y|A}$ satisfies the definition of both an acausal and a causal conditional state for Y given A if and only if the set $\{E_y^A\}$ is a POVM on $\mathcal{H}_A$, i.e., each E_y^A is positive and $\sum_y E_y^A = I_A$.

The proof is given in the Appendix.

It is sometimes useful to use the notation $\sigma_{Y=y|A} = \langle y|_Y \sigma_{Y|A} |y\rangle_Y = E_y^A$ for the components of conditional states of this form.

Unlike the case of hybrid states that are conditioned on a classical variable, conditional states of the form $\sigma_{Y|A}$ are not invariant under partial transpose on the conditioning region. However, taking the partial transpose over A of $\rho_{Y|A}$ yields another valid acausal conditional state, because $\{(E_y^A)^{T_A}\}$ is a POVM if and only if $\{E_y^A\}$ is. The remainder of this section concerns the causal interpretation of such states, so the notation $\varrho_{Y|A}$ is adopted.

Proposition 3. Consider a measurement of a POVM $\{E_y^A\}$ on a quantum region A in state ρ_A. Let $\varrho_{Y|A}$ be the causal conditional state with components E_y^A. The Born rule,

$$P(Y=y) = \mathrm{Tr}_A\left(E_y^A \rho_A\right), \tag{48}$$

can then be expressed in the conditional states formalism as the quantum belief propagation rule

$$\rho_Y = \mathrm{Tr}_A(\varrho_{Y|A}\rho_A), \tag{49}$$

where $\rho_Y = \sum_y P(Y=y)|y\rangle\langle y|_Y$.

This is easily verified by substituting the definition of $\varrho_{Y|A}$ from Eq. (47) into Eq. (49).

This representation of a measurement as a causal conditional state and of the Born rule as an instance of belief propagation is summarized in Fig. 4 and the third and sixth lines of Table II.

Once again, these results can be understood as a special case of Theorem 2 and Proposition 1 by recognizing that a POVM may be represented as a map from a quantum region to a classical variable (sometimes called a QC map). Specifically, if the probability distribution $P(Y)$ is represented by a diagonal state ρ_Y, then the measurement can be represented by the CPT map $\mathcal{E}_{Y|A}$ defined by

$$\rho_Y = \mathcal{E}_{Y|A}(\rho_A) = \sum_y \mathrm{Tr}_A\left(E_y^A \rho_A\right)|y\rangle\langle y|_Y. \tag{50}$$

The causal conditional state $\varrho_{Y|A}$ appearing in Eq. (49) is simply the Jamiołkowski isomorphic operator to this map.

I. Belief propagation of conditional states

Consider three causally related regions A, B, and C, such that B is in the future of A and C is in the future of B. If the dynamics is Markovian then it can be described by first applying a CPT map $\mathcal{E}_{B|A}$ to A followed by a CPT map $\mathcal{E}_{C|B}$ to B. This scenario is illustrated in Fig. 5(a). If we are only interested in regions A and C, then region B can be eliminated from the description by composing the two maps to obtain $\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}$, where $\mathcal{E} \circ \mathcal{F}(\cdot) \equiv \mathcal{E}(\mathcal{F}(\cdot))$.

In the conditional states formalism the CPT maps are replaced by the Jamiołkowski isomorphic causal conditional states $\varrho_{B|A}$, $\varrho_{C|B}$, and $\varrho_{C|A}$, and thus $\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}$ should be replaced by a formula for $\varrho_{C|A}$ in terms of $\varrho_{B|A}$ and $\varrho_{C|B}$.

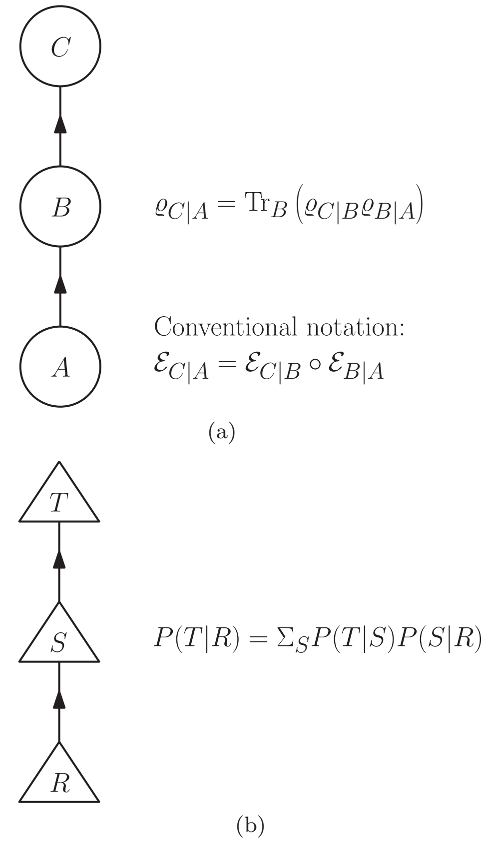


FIG. 5. Propagating causal conditional states and conditional probability distributions. (a) Quantum case. (b) Classical case.

As an aid to intuition, consider the classical analog of this scenario as depicted in Fig. 5(b). Here, the variable S is in the future of R, and T is in the future of S. The three variables are related by a Markovian dynamics, described by the conditional probability distributions $P(S|R)$ and $P(T|S)$. An initial probability distribution $P(R)$ can be propagated into the future to obtain $P(T)$ in two steps. First we propagate from R to S to obtain $P(S) = \sum_R P(S|R)P(R)$ and then from S to T to obtain $P(T) = \sum_S P(T|S)P(S)$. Combining these steps gives $P(T) = \sum_{R,S} P(T|S)P(S|R)P(R)$, so defining the conditional probability distribution,

$$P(T|R) = \sum_S P(T|S)P(S|R), \tag{51}$$

allows the belief propagation from R to T to be performed in a single step via $P(T) = \sum_R P(T|R)P(R)$.

The quantum analog of this is given by the following theorem.

Theorem 5. Let $\mathcal{E}_{B|A}$, $\mathcal{E}_{C|B}$, and $\mathcal{E}_{C|A}$ be linear maps such that $\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}$. Then the Jamiołkowski isomorphic operators, $\varrho_{B|A}$, $\varrho_{C|B}$, and $\varrho_{C|A}$, satisfy

$$\varrho_{C|A} = \mathrm{Tr}_B(\varrho_{C|B}\varrho_{B|A}). \tag{52}$$

Conversely, if three operators satisfy Eq. (52), then the Jamiołkowski isomorphic maps satisfy $\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}$.

The proof is provided in the Appendix.

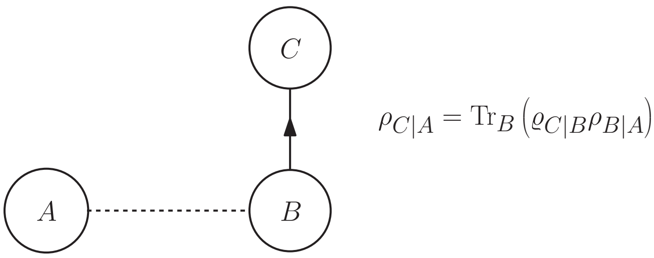


FIG. 6. Propagating an acausal conditional state through a causal conditional state.

Equation (52) can be regarded as a belief propagation rule for causal conditional states. It propagates beliefs about B, conditional on A, into the future to obtain beliefs about C, conditional on A.

A similar formalism can be developed for the propagation of acausal conditional states across acausally related regions. However, for present purposes, it is more interesting to consider a situation of mixed causality, wherein a causal conditional state is propagated across two acausally related regions. This is used in the application to steering developed in Sec. V A3.

Consider the scenario depicted in Fig. 6. Initially, a state ρ_{AB} is assigned to regions A and B, which are acausally related. A CPT map $\mathcal{E}_{C|B}$ (alternatively represented by a causal conditional state $\varrho_{C|B}$) is then applied to region B to obtain the state of AC. In this scenario, C is causally related to B, but acausally related to A. By Theorem 1, the state of AC is given by

$$\rho_{AC} = \mathrm{Tr}_B(\varrho_{C|B}\rho_{AB}). \tag{53}$$

Now $\rho_{AB} = \rho_{B|A} \star \rho_A$ and $\rho_{AC} = \rho_{C|A} \star \rho_A$, so we have

$$\rho_{C|A} \star \rho_A = \mathrm{Tr}_B(\varrho_{C|B}\rho_{B|A}) \star \rho_A, \tag{54}$$

where we have used the fact that ρ_A commutes with $\varrho_{C|B}$. Taking the $\star$ product of this equation with ρ_A^{-1} then gives

$$\rho_{C|A} = \mathrm{Tr}_B(\varrho_{C|B}\rho_{B|A}), \tag{55}$$

in analogy with Eq. (52). This can again be viewed as a belief propagation rule for conditional states, but this time an acausal conditional state is being propagated through a causal conditional state.

J. The Heisenberg picture

In Sec. III D, quantum evolution from an early region A to a late region B was described by a map from states on $\mathcal{H}_A$ to states on $\mathcal{H}_B$. However, dynamics can alternatively be represented in terms of observables rather than states. This is simply the textbook distinction between the Schrödinger picture and the Heisenberg picture. In the Heisenberg picture, a temporal evolution is described by a map from the space of observables on the late region B to the space of observables on the early region A. The observables are usually represented by self-adjoint operators and the dynamics by unitary operations, but this can be generalized to take account of generalized measurements and CPT dynamics. In this generalization, Heisenberg dynamics consists of a map from POVM elements (also known as effects) on B to POVM elements on A. In other words, effects are evolved backwards in time in the Heisenberg picture.

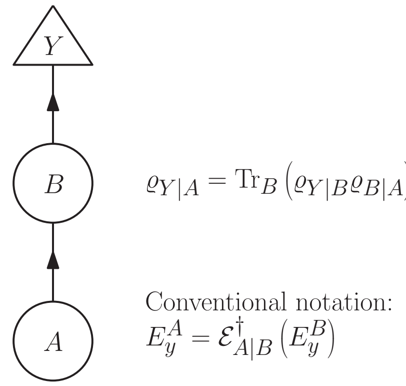


FIG. 7. Dynamics in the Heisenberg picture.

In order to describe the Heisenberg picture for CPT maps, it is necessary to define the notion of a dual map.

Definition 5. The *dual map* $(\mathcal{E}_{B|A})^\dagger : \mathfrak{L}(\mathcal{H}_B) \to \mathfrak{L}(\mathcal{H}_A)$ of a linear map $\mathcal{E}_{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ is the unique map that satisfies

$$\mathrm{Tr}_A[(\mathcal{E}_{B|A})^\dagger(N_B)M_A] = \mathrm{Tr}_B(N_B\mathcal{E}_{B|A}(M_A)) \tag{56}$$

for all $M_A \in \mathfrak{L}(\mathcal{H}_A), N_B \in \mathfrak{L}(\mathcal{H}_B)$.

Note that the input space for the dual map is the output space for the original map and vice versa. The notational convention

$$\mathcal{E}_{A|B}^\dagger \equiv (\mathcal{E}_{B|A})^\dagger \tag{57}$$

is adopted in order to make this clear.

If $\mathcal{E}_{B|A}$ is the map describing time evolution in the Schrödinger picture through the formula $\rho_B = \mathcal{E}_{B|A}(\rho_A)$, then the same evolution is described in the Heisenberg picture by the dual map $\mathcal{E}_{A|B}^\dagger$ through the formula $E_y^A = \mathcal{E}_{A|B}^\dagger(E_y^B)$, where $\{E_y^A\}$ and $\{E_y^B\}$ are POVMs. This follows from the condition that the two pictures should be operationally equivalent; i.e., they should assign the same probabilities. To see this, imagine that the evolution from region A to region B is followed by a measurement on B yielding an outcome Y. This scenario is depicted in Fig. 7. The probability of observing the effect E_y^B after a preparation of ρ_A followed by an evolution $\mathcal{E}_{B|A}$ is expressed in the Schrödinger picture as $\mathrm{Tr}_B(E_y^B\rho_B)$, where $\rho_B = \mathcal{E}_{B|A}(\rho_A)$, while it is expressed in the Heisenberg picture as $\mathrm{Tr}_A(E_y^A\rho_A)$, where $E_y^A = \mathcal{E}_{A|B}^\dagger(E_y^B)$. The definition of the dual map ensures that the two expressions for the probability are equivalent, i.e., $\mathrm{Tr}_B(E_y^B\rho_B) = \mathrm{Tr}_A(E_y^A\rho_A)$.

A CPT map $\mathcal{E}_{B|A}$ can always be written in a Kraus decomposition

$$\mathcal{E}_{B|A}(\cdot) = \sum_\mu K_\mu(\cdot)K_\mu^\dagger, \tag{58}$$

where $K_\mu : \mathcal{H}_A \to \mathcal{H}_B$. The dual map $\mathcal{E}_{A|B}^\dagger$ can then be obtained by taking the adjoint of the operators in the Kraus decomposition, i.e.,

$$\mathcal{E}_{A|B}^\dagger(\cdot) = \sum_\mu K_\mu^\dagger(\cdot)K_\mu. \tag{59}$$

Thus, if $\mathcal{H}_A$ and $\mathcal{H}_B$ are isomorphic and $\mathcal{E}_{B|A}$ is a unitary operation, i.e., $\mathcal{E}_{B|A}(\cdot) = U(\cdot)U^\dagger$ for some unitary operator U, then $\mathcal{E}^\dagger_{A|B}(\cdot) = U^\dagger(\cdot)U$. This is the familiar special case of the Heisenberg picture for unitary dynamics.

In order to translate the Heisenberg picture into the conditional states formalism, first represent the POVM $\{E^B_y\}$ by a conditional state $\varrho_{Y|B}$, the CPT map $\mathcal{E}_{B|A}$ by a causal conditional state $\varrho_{B|A}$, and the POVM $\{E^A_y\}$ by a conditional state $\varrho_{Y|A}$. Second, note that Fig. 7 is just a special case of Fig. 5(a) from Sec. III I in which the final region is classical, so the three conditional states are related by Eq. (52), i.e.,

$$\varrho_{Y|A} = \mathrm{Tr}_B(\varrho_{Y|B}\varrho_{B|A}). \tag{60}$$

In Sec. III I, this was described as a belief propagation formula for causal conditional states because $\varrho_{Y|B}$ was regarded as defining a map from $\varrho_{B|A}$ to $\varrho_{Y|A}$, propagating beliefs about B, conditional on A, into the future. However, in the context of Heisenberg dynamics, we instead regard $\varrho_{B|A}$ as defining a map from $\varrho_{Y|B}$ to $\varrho_{Y|A}$, in the opposite direction to the flow of time. It remains to show that Eq. (60) is equivalent to the conventional description of Heisenberg dynamics in terms of dual maps.

Theorem 6. Let $\varrho_{B|A}$ be the causal joint state that is Jamiołkowski isomorphic to the CPT map $\mathcal{E}_{B|A}$. Then, the action of the dual map $\mathcal{E}^\dagger_{A|B}$ on an operator M_{BC} is given by

$$(\mathcal{E}^\dagger_{A|B} \otimes \mathcal{I}_C)(M_{BC}) = \mathrm{Tr}_B(M_{BC}\varrho_{B|A}). \tag{61}$$

Proof. By Definition 5, the dual map to $\mathcal{E}_{B|A}$ is the unique linear map $\mathcal{E}^\dagger_{A|B}$ that satisfies

$$\mathrm{Tr}_A[\mathcal{E}^\dagger_{A|B}(N_B)M_A] = \mathrm{Tr}_B[N_B\mathcal{E}_{B|A}(M_A)] \tag{62}$$

for all operators M_A and N_B. Using the Jamiołkowski isomorphism and Theorem 2, the right-hand side can be written as

$$\mathrm{Tr}_B[N_B\mathcal{E}_{B|A}(M_A)] = \mathrm{Tr}_{AB}(N_B\varrho_{B|A}M_A) \tag{63}$$

$$= \mathrm{Tr}_A[\mathrm{Tr}_B(N_B\varrho_{B|A})M_A]. \tag{64}$$

The only way this can equal $\mathrm{Tr}_A[\mathcal{E}^\dagger_{A|B}(N_B)M_A]$ for all M_A is if $\mathcal{E}^\dagger_{A|B}(N_B) = \mathrm{Tr}_B(N_B\varrho_{B|A})$. Equation (61) then follows by linear extension to $\mathcal{H}_{BC}$. ■

Combining this with Eq. (60) gives the following proposition.

Proposition 4. Let $\varrho_{B|A}$ be the causal conditional state associated with a quantum evolution described by the CPT map $\mathcal{E}_{B|A}$ and let $\varrho_{Y|A}$ and $\varrho_{Y|B}$ be the hybrid conditional states associated with the POVMs $\{E^A_y\}$ and $\{E^B_y\}$, such that $\{E^A_y\}$ is obtained from $\{E^B_y\}$ by the Heisenberg picture dynamics. The conventional description of evolution in the Heisenberg picture,

$$E^A_y = \mathcal{E}^\dagger_{A|B}\left(E^B_y\right), \tag{65}$$

can be expressed in the conditional states formalism as

$$\varrho_{Y|A} = \mathrm{Tr}_B(\varrho_{Y|B}\varrho_{B|A}). \tag{66}$$

This follows straightforwardly from Theorems 6 and 4.

As in Sec. III I, similar reasoning can be applied to other causal scenarios. Consider the special case of Fig. 6 in which C is replaced by a classical variable Y. This is depicted in

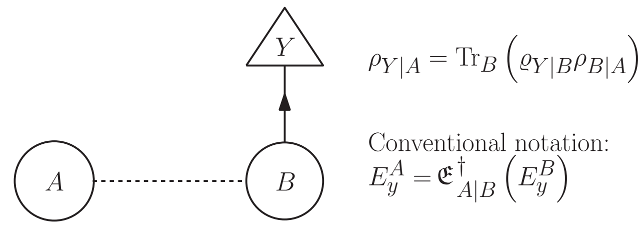


FIG. 8. Heisenberg evolution for a remote measurement.

Fig. 8. Two acausally related regions, A and B, are assigned a state ρ_{AB} and then the POVM $\{E^B_y\}$ (alternatively represented by a conditional state $\varrho_{Y|B}$) is measured on region B. This is the type of scenario that occurs in an EPR experiment. By measuring the region B, information is obtained about the remote region A and we are interested in how the state of A is correlated with the measurement outcome Y.

The belief propagation formula for this scenario, Eq. (55), gives

$$\rho_{Y|A} = \mathrm{Tr}_B(\varrho_{Y|B}\rho_{B|A}). \tag{67}$$

The components of the conditional state $\rho_{Y|A}$ specify a POVM $\{E^A_y\}$. This POVM can be thought of as describing the effective measurement that gets performed on A when we actually measure region B. When combined with ρ_A, the conditional state $\rho_{Y|A}$ specifies the ensemble of states for region A associated with the different measurement outcomes via $\rho_{AY} = \rho_{Y|A} \star \rho_A$. In terms of components, this is

$$\rho_{YA} = \sum_Y |y\rangle\langle y|_Y \otimes \rho_A^{\frac{1}{2}} E^A_y \rho_A^{\frac{1}{2}}, \tag{68}$$

so the un-normalized state of A corresponding to the outcome $Y = y$ is

$$P(Y = y)\rho^A_y = \rho_A^{\frac{1}{2}} E^A_y \rho_A^{\frac{1}{2}}, \tag{69}$$

where $P(Y = y) = \mathrm{Tr}_A(E^A_y\rho_A) = \mathrm{Tr}_B(E^B_y\rho_B)$ is the Born rule probability for the measurement outcome $Y = y$.

If $\rho_{B|A}$ in Eq. (67) is thought of as specifying a map from $\varrho_{Y|B}$ to $\rho_{Y|A}$, then this map is analogous to a Heisenberg picture dynamics, except that the propagation is across acausally related rather than causally related regions. If $\mathfrak{E}_{B|A}$ is the Jamiołkowski isomorphic map to $\rho_{B|A}$, then, by Theorem 6, the POVM elements are related by

$$E^A_y = \mathfrak{E}^\dagger_{A|B}(E^B_y). \tag{70}$$

Mathematically, the only difference between this and a Heisenberg picture map is that $\mathfrak{E}_{B|A} \circ T_A$ is completely positive rather than the map $\mathfrak{E}_{B|A}$ itself. This is just a reflection of the fact that we are propagating across acausally related, rather than causally related, regions.

A similar expression to Eq. (70) has appeared in the context of quantum steering [25], although there it is written in terms of the Choi, rather than Jamiołkowski, isomorphic map so there is a transpose in the expression. We develop this application in Sec. V A3.

K. Quantum instruments as causal belief propagation

Describing a measurement by a POVM is adequate for determining the outcome probabilities of the measurement via the Born rule. However, one might also wish to describe how the state of a postmeasurement region is correlated with the measurement result. In the conventional formalism, the transformative aspect of a measurement is represented by a quantum instrument.

Definition 6. Given quantum regions A and B, and a classical variable Y, a *quantum instrument* is a set $\{\mathcal{E}_y^{B|A}\}$ of CPT maps $\mathcal{E}_y^{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ such that the operators

$$E_y^A = \mathcal{E}_y^{\dagger A|B}(I_B) \tag{71}$$

form a POVM.

If A and B represent causally related regions before and after a measurement and Y represents the measurement outcome, then a quantum instrument can be used to determine the subnormalized state $P(Y = y)\rho_y^B$ of B when the outcome is known via

$$P(Y = y)\rho_y^B = \mathcal{E}_y^{B|A}(\rho_A). \tag{72}$$

It can also be used to compute the outcome probabilities for the measurement by simply tracing over B in Eq. (72) to obtain

$$P(Y = y) = \mathrm{Tr}_B\big[\mathcal{E}_y^{B|A}(\rho_A)\big]. \tag{73}$$

Using Eq. (71), this can be written as

$$P(Y = y) = \mathrm{Tr}_B\big[\mathcal{E}_y^{B|A}(\rho_A)\big] \tag{74}$$

$$= \mathrm{Tr}_B\big[I_B\mathcal{E}_y^{B|A}(\rho_A)\big] \tag{75}$$

$$= \mathrm{Tr}_A\big[\mathcal{E}_y^{\dagger A|B}(I_B)\rho_A\big] \tag{76}$$

$$= \mathrm{Tr}_A\big[E_y^A\rho_A\big], \tag{77}$$

which is just the Born rule with respect to the POVM defined by the instrument.

While an instrument defines a unique POVM, each POVM corresponds to more than one quantum instrument. When performing a measurement of a particular POVM, *any* of the quantum instruments that correspond to it via Eq. (71) may be obtained, depending on how the measurement is implemented. Conversely, the set of instruments corresponding to a given POVM exhaust the possible postmeasurement transformations. This includes, for example, the situation in which the system being measured is absorbed by the detector, which corresponds to choosing the trivial Hilbert space for B, i.e., $\mathcal{H}_B = \mathbb{C}$.

Despite the freedom in choosing a quantum instrument, certain kinds of instrument are usually considered particularly important. For measurements associated with a *projector*-valued measure $\{\Pi_y^A\}$, the possible quantum instruments include the Lüders-von Neumann projection postulate as a special case by taking $\mathcal{H}_B$ to have the same dimension as $\mathcal{H}_A$ and $\mathcal{E}_y^{B|A}(\rho_A) = \mathcal{I}_{B|A}(\Pi_y^A\rho_A\Pi_y^A)$, where $\mathcal{I}_{B|A}$ is an isometry between $\mathcal{H}_A$ and $\mathcal{H}_B$. For general POVMs the rule $\mathcal{E}_y^{B|A}(\rho_A) = \mathcal{I}_{B|A}((E_y^A)^{\frac{1}{2}}\rho_A(E_y^A)^{\frac{1}{2}})$, which is sometimes taken as a natural generalization of the projection postulate, is also included as a special case.

A general measurement procedure, where there is an initial quantum region, a classical outcome of the measurement, and a quantum region after the measurement, is depicted in Fig. 9(a).

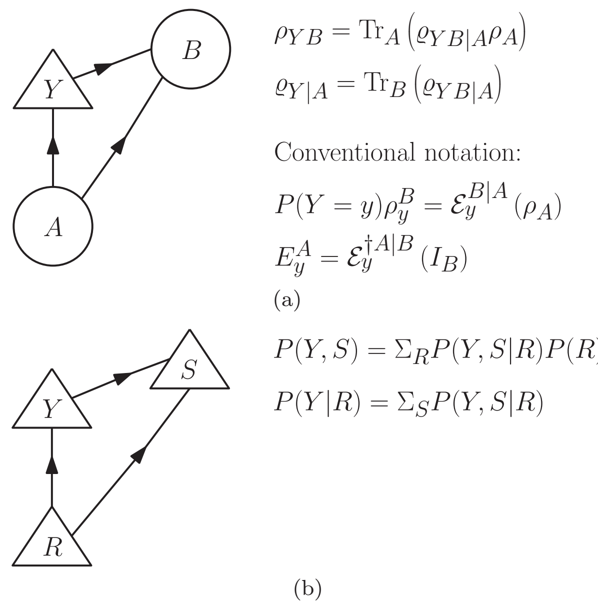


FIG. 9. Measurements and their associated state-update rules. (a) A quantum instrument, representing how the state of a quantum persistent system changes after a measurement. (b) The classical analog of a quantum instrument, representing how the state of a classical persistent system changes after a general (possibly disturbing) measurement.

Note that the final quantum region B may depend causally both on the initial quantum region A and on the outcome Y.

In order to understand how quantum instruments are represented in the conditional states formalism, it is helpful to first look at the classical analog. This is a scenario wherein a classical measurement is made upon a classical system, which persists after the measurement, and in general the measurement procedure is permitted to disturb the state of the system. The variable R describes the system before the measurement and the variable S describes the system after the measurement. This is in line with the quantum treatment, in which distinct regions are given distinct labels. The outcome of the measurement is again denoted by Y. This scenario is depicted in Fig. 9(b). The measurement is then described by a conditional probability $P(Y,S|R)$. This can equivalently be thought of as a set of subnormalized conditional probabilities for S given R, $\{P(Y = y,S|R)\}$, one for each outcome y, which is the analog of a quantum instrument. The joint distribution over Y and S, when the input distribution is $P(R)$, is then given by

$$P(Y,S) = \sum_R P(Y,S|R)P(R). \tag{78}$$

Furthermore, the set of response functions $\{P(Y = y|R)\}$ associated with such a measurement is easily computed from $P(Y,S|R)$ by marginalizing over S,

$$P(Y|R) = \sum_S P(Y,S|R). \tag{79}$$

In the conditional states formalism, Eqs. (72) and (71) are replaced by straightforward analogs of Eqs. (78) and (79) for a causal hybrid state $\varrho_{YB|A}$ of a classical variable Y and quantum region B, conditioned on a quantum region A.

Theorem 7. Let $\varrho_{YB|A}$ be an operator on $\mathcal{H}_{YAB}$ of the form

$$\varrho_{YB|A} \equiv \sum_{y} |y\rangle\langle y|_Y \otimes \varrho_{Y=y,B|A}, \tag{80}$$

where

$$\varrho_{Y=y,B|A} \equiv \mathcal{E}_y^{B|A'}\left(\sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}\right) \tag{81}$$

are the Jamiołkowski isomorphic operators to maps $\mathcal{E}_y^{B|A} : \mathfrak{L}(\mathcal{H}_A) \rightarrow \mathfrak{L}(\mathcal{H}_B)$. Then $\varrho_{YB|A}$ is a causal conditional state if and only if $\{\mathcal{E}_y^{B|A}\}$ is a quantum instrument.

The proof is similar to those of Theorems 2, 3, and 4 and is left to the reader.

Theorem 7 allows the transformative aspect of a quantum measurement to be represented by conditional states as follows.

Proposition 5. Let $\varrho_{YB|A}$ be the causal conditional state associated with the instrument $\{\mathcal{E}_y^{B|A}\}$. Then, when the measurement corresponding to this instrument is made with input state ρ_A, the state-update rule in conventional notation is given by

$$P(Y = y)\rho_y^B = \mathcal{E}_y^{B|A}(\rho_A). \tag{82}$$

In analogy with the classical expression in Eq. (78), this can be expressed in the conditional states framework as

$$\rho_{YB} = \mathrm{Tr}_A(\varrho_{YB|A}\rho_A), \tag{83}$$

where $\rho_{YB} = \sum_y P(Y = y)|y\rangle\langle y|_Y \otimes \rho_y^B$. Furthermore, the conventional expression for the relation between a POVM and a quantum instrument,

$$E_y^A = (\mathcal{E}_y^{\dagger})^{A|B}(I_B), \tag{84}$$

can be expressed simply as

$$\varrho_{Y|A} = \mathrm{Tr}_B(\varrho_{YB|A}), \tag{85}$$

in analogy with the classical expression in Eq. (79).

The proof of Eq. (83) consists of applying Proposition 1, in particular Eq. (33), to Eq. (82) for every value of Y. Equation (85) follows from applying Theorem 6 to each element of the instrument.

Note that, from the perspective of the conditional states framework, the fact that there are many quantum instruments consistent with a given POVM is no more surprising than the fact that in classical probability theory there are many joint distributions consistent with a given marginal distribution.

Finally, note that for a quantum instrument, the map $\mathcal{E}_{B|A}(\rho_A) = \sum_y \mathcal{E}_y^{B|A}(\rho_A)$ is CPT and represents the *nonselective* state-update rule, i.e., the one that you should apply if you know that the measurement has been made but do not know its outcome. In the conditional states framework, if you know that a measurement associated with the causal conditional state $\varrho_{YB|A}$ has been performed, but you do not know the outcome, then you simply marginalize over Y to obtain the causal conditional state $\varrho_{B|A}$. Quantum belief propagation from A to B using $\varrho_{B|A}$ is the nonselective state-update rule.

Table II provides a summary of how dynamics, ensemble preparations, and measurements are represented in the conditional states formalism as compared to the conventional formalism.

IV. QUANTUM BAYES' THEOREM

This section develops a quantum generalization of Bayes' theorem that relates the conditional states $\rho_{B|A}$ ($\varrho_{B|A}$) and $\rho_{A|B}$ ($\varrho_{A|B}$). Formally, the quantum Bayes' theorem is the same for acausal and causal conditional states, so this represents a success in our project to develop a causally neutral theory of quantum Bayesian inference. In Sec. IV A, the quantum Bayes' theorem is introduced for two quantum regions. When written in terms of conventional notation, it reproduces the Barnum-Knill approximate error correction map [29]. Section IV B specializes to the hybrid case, which provides a rule for relating sets of states to POVMs. In conventional notation, this reproduces the definition of the pretty-good measurement [31–33] and a quantum analog of Bayes' theorem previously advocated by Fuchs [24]. As an application of the quantum Bayes' theorem, we develop a retrodictive formalism for quantum theory in Sec. IV C in which states are evolved backwards in time. This demonstrates that the conditional states formalism is causally neutral with respect to the direction of time. Finally, in Sec. IV D, the acausal analog of the symmetry between prediction and retrodiction is discussed in the context of remote measurement.

A. General quantum Bayes' theorem

Recall that the classical Bayes' theorem is

$$P(R|S) = \frac{P(S|R)P(R)}{P(S)}, \tag{86}$$

which is derived by noting two expressions for the joint probability in terms of conditionals and marginals

$$P(R,S) = P(R|S)P(S) \tag{87}$$

$$= P(S|R)P(R). \tag{88}$$

Quantum conditional states can be used to derive a quantum analog of Bayes' theorem. For acausal conditional states, the two analogous expressions to Eqs. (87) and (88) are

$$\rho_{AB} = \rho_{A|B} \star \rho_B \tag{89}$$

$$= \rho_{B|A} \star \rho_A. \tag{90}$$

Combining these gives

$$\rho_{A|B} = \rho_{B|A} \star \left(\rho_A \rho_B^{-1}\right), \tag{91}$$

which is a quantum analog of Bayes' theorem for acausal conditional states.

Classically, the distribution $P(S)$ that appears in the denominator of Bayes' theorem is usually computed via belief propagation as $P(S) = \sum_R P(S|R)P(R)$. This gives the alternative form of Bayes' theorem

$$P(R|S) = \frac{P(S|R)P(R)}{\sum_R P(S|R)P(R)}. \tag{92}$$

Similarly, noting that $\rho_B = \mathrm{Tr}_A(\rho_{B|A} \star \rho_A) = \mathrm{Tr}_A(\rho_{B|A}\rho_A)$, the quantum Bayes' theorem for acausal conditionals can be written as

$$\rho_{A|B} = \rho_{B|A} \star \{\rho_A[\mathrm{Tr}_A(\rho_{B|A}\rho_A)]^{-1}\}. \tag{93}$$

Now consider the case of two causally related regions. Suppose that a region A, described by the state ρ_A, is mapped to B by a CPT map $\mathcal{E}_{B|A}$ that is Jamiołkowski-isomorphic to the causal conditional state $\varrho_{B|A}$. Recall from Sec. III E that the conditional state and input state can be used to define a causal joint state $\varrho_{AB} = \varrho_{B|A} \star \rho_A$. Now we can try to define a new causal conditional state $\varrho_{A|B}$ via an analogous decomposition of the causal joint state, namely, $\varrho_{AB} = \varrho_{A|B} \star \rho_B$, where $\rho_B = \mathrm{Tr}_A(\varrho_{B|A}\rho_A)$ is the output state of the channel. Equating the two expressions for the joint state ϱ_{AB}, we obtain an expression for $\varrho_{A|B}$, which can be regarded as the causal version of the quantum Bayes' theorem,

$$\varrho_{A|B} = \varrho_{B|A} \star \left(\rho_A \rho_B^{-1}\right), \tag{94}$$

and which can also be written as

$$\varrho_{A|B} = \varrho_{B|A} \star \{\rho_A[\mathrm{Tr}_A(\varrho_{B|A}\rho_A)]^{-1}\}. \tag{95}$$

In order for this to make sense, it must be checked that $\varrho_{A|B}$ is indeed a valid causal conditional state. Taking the partial transpose over B of Eq. (94), we have $\varrho_{A|B}^{T_B} = \varrho_{B|A}^{T_B} \star [\rho_A(\rho_B^{T_B})^{-1}]$. Given that ρ_B is a valid state (positive and normalized) and given that the set of such states is mapped to itself by the transpose, $\rho_B^{T_B}$ is also a valid state. Furthermore, given that $\varrho_{B|A}$ is a valid causal conditional state and the fact that the set of such states are mapped to the set of valid acausal conditional states by the partial transpose, it follows that $\varrho_{B|A}^{T_B}$ is a valid acausal conditional state. However, then, by the acausal quantum Bayes' theorem given in Eq. (91), $\varrho_{A|B}^{T_B}$ is a valid acausal conditional state, which implies that $\varrho_{B|A}$ is a valid causal conditional state.

It is instructive to see how the causal and acausal versions of Bayes' theorem appear in conventional notation.

For the causal version, suppose that the causal conditional state $\varrho_{B|A}$ is associated, via the Jamiołkowski isomorphism, with a quantum channel $\mathcal{E}_{B|A}$ and its Bayesian inversion $\varrho_{A|B}$ is associated with a quantum channel $\mathcal{F}_{A|B}$. Then, Eq. (94) is equivalent to

$$\mathcal{F}_{A|B}(\cdot) = \rho_A^{\frac{1}{2}}\left\{\mathcal{E}_{A|B}^{\dagger}\left[\rho_B^{-\frac{1}{2}}(\cdot)\rho_B^{-\frac{1}{2}}\right]\right\}\rho_A^{\frac{1}{2}}, \tag{96}$$

where

$$\rho_B = \mathcal{E}_{B|A}(\rho_A). \tag{97}$$

The converse of this relation, wherein $\mathcal{E}_{B|A}$ is expressed in terms of $\mathcal{F}_{A|B}$, is obtained by simply exchanging A and B as well as $\mathcal{E}$ and $\mathcal{F}$.

The proof that Eq. (94) translates into Eq. (96) is straightforward. From the Jamiołkowski isomorphism (Theorem 1), $\mathcal{F}_{A|B}(\cdot) = \mathrm{Tr}_B[\varrho_{A|B}(\cdot)]$, which implies that

$$\mathcal{F}_{A|B}(\cdot) = \mathrm{Tr}_B\left[\rho_A^{\frac{1}{2}}\rho_B^{-\frac{1}{2}}\varrho_{B|A}\rho_B^{-\frac{1}{2}}\rho_A^{\frac{1}{2}}(\cdot)\right] \tag{98}$$

$$= \rho_A^{\frac{1}{2}}\mathrm{Tr}_B\left[\rho_B^{-\frac{1}{2}}(\cdot)\rho_B^{-\frac{1}{2}}\varrho_{B|A}\rho_A^{\frac{1}{2}}\right], \tag{99}$$

where the first step follows from Eq. (94) and expanding the $\star$ product, and the second step uses the cyclic property of the trace. Equation (96) then follows from the representation of dual maps in terms of conditional states given in Theorem 6.

The map $\mathcal{F}_{A|B}$ is recognizable as the Barnum-Knill recovery map for the channel $\mathcal{E}_{B|A}$ [29]. This map is known to achieve near-optimal quantum error correction in situations where the input state and channel are known. Its connection with Bayesian inversion suggests that the best way of thinking about $\mathcal{F}_{A|B}$ is not as an error correction map, but rather as a means for accurately capturing your beliefs about region A given your beliefs about region B.

A similar result holds for the acausal case. Suppose that the linear map that is Jamiołkowski-isomorphic to the acausal conditional state $\rho_{B|A}$ is $\mathfrak{E}_{B|A}$ and the one associated to its Bayesian inversion $\rho_{A|B}$ is $\mathfrak{F}_{A|B}$ (recall that only $\mathfrak{E}_{B|A} \circ T_A$ and $\mathfrak{F}_{A|B} \circ T_A$ are CPT maps). Following the same reasoning used above, Eq. (91) can be rewritten as

$$\mathfrak{F}_{A|B}(\cdot) = \rho_A^{\frac{1}{2}}\left\{\mathfrak{E}_{A|B}^{\dagger}\left[\rho_B^{-\frac{1}{2}}(\cdot)\rho_B^{-\frac{1}{2}}\right]\right\}\rho_A^{\frac{1}{2}}, \tag{100}$$

where

$$\rho_B = \mathfrak{E}_{B|A}(\rho_A). \tag{101}$$

B. Bayes' theorem for quantum-classical hybrids

For quantum-classical hybrids, there are two versions of Bayes' theorem, depending on whether it is the conditioned region or the conditioning region that is classical. Recall that the mathematical form of hybrid conditional states does not depend on whether they are causal or acausal, and σ is the notation used for results that do not depend on the causal interpretation. The two versions of Bayes' theorem are then

$$\sigma_{X|A} = \sigma_{A|X} \star \left(\rho_X \rho_A^{-1}\right) \tag{102}$$

and

$$\sigma_{A|X} = \sigma_{X|A} \star \left(\rho_A \rho_X^{-1}\right), \tag{103}$$

where $\rho_A = \mathrm{Tr}_X(\sigma_{A|X}\rho_X)$ and $\rho_X = \mathrm{Tr}_A(\sigma_{X|A}\rho_A)$.

A hybrid joint state σ_{XA} may be decomposed into a hybrid conditional state and a reduced state via Eq. (11) in two distinct ways: either in terms of a classical reduced state and a conditional state that is conditioned on the classical part,

$$\sigma_{XA} = \sigma_{A|X} \star \rho_X, \tag{104}$$

or in terms of a quantum reduced state and a conditional state that is conditioned on the quantum part,

$$\sigma_{XA} = \sigma_{X|A} \star \rho_A. \tag{105}$$

The hybrid Bayes' theorems of Eqs. (102) and (103) give the rules for converting between these two decompositions.

To see how these two decompositions appear in conventional notation, recall from Theorem 3 that a general hybrid conditional state $\sigma_{A|X}$ is of the form

$$\sigma_{A|X} = \sum_x |x\rangle\langle x|_X \otimes \rho_x^A, \tag{106}$$

where each ρ_x^A is a normalized density operator on $\mathcal{H}_A$, and from Theorem 4 that $\sigma_{X|A}$ is of the form

$$\sigma_{X|A} = \sum_x |x\rangle\langle x|_X \otimes E_x^A, \tag{107}$$

where $\{E_x^A\}$ is a POVM on A. Finally, recall that the classical state ρ_X is of the form $\rho_X = \sum_x P(X=x)|x\rangle\langle x|_X$ where $P(X)$ is a classical probability distribution.

Equation (104) therefore gives a decomposition of a joint state in terms of a set of states and a probability distribution via

$$\sigma_{XA} = \sum_x P(X=x)|x\rangle\langle x|_X \otimes \rho_x^A, \tag{108}$$

and Eq. (105) gives a decomposition in terms of a POVM and a state for A via

$$\sigma_{XA} = \sum_x |x\rangle\langle x|_X \otimes \rho_A^{\frac{1}{2}} E_x^A \rho_A^{\frac{1}{2}}. \tag{109}$$

In terms of components, the Bayes' theorem of Eq. (102) is a rule for determining a POVM from a probability distribution and a set of states, while Eq. (103) is a rule for determining a set of states from a POVM and a state on A. These rules are

$$\rho_x^A = \frac{\rho_A^{\frac{1}{2}} E_x^A \rho_A^{\frac{1}{2}}}{\mathrm{Tr}_A\left(E_x^A \rho_A\right)} \tag{110}$$

and

$$E_x^A = P(X=x)\rho_A^{-\frac{1}{2}} \rho_x^A \rho_A^{-\frac{1}{2}}, \tag{111}$$

where $\rho_A = \sum_x P(X=x)\rho_x^A$.

These rules have appeared numerous times in the literature, e.g., [3,4,30]. In the context of distinguishing the states in an ensemble, the POVM defined by Eq. (111) is known as the pretty-good measurement [31–33]. Equation (110) is a rule previously advocated by Fuchs as a quantum analog of Bayes' theorem [24]. The fact that these rules are all special cases of a more general quantum Bayes' theorem goes some way to explaining their utility.

C. Retrodiction and time symmetry

As an application of the quantum Bayes' theorem, we use it to develop a retrodictive formalism for quantum theory, in which states are propagated backwards in time from late regions to early regions. This is operationally equivalent to the usual predictive formalism, in which states are propagated forward in time from early regions to late regions. The retrodictive description is particularly useful if you acquire new information about the late region and wish to update your beliefs about the early region, for instance, when you learn about the output of a noisy channel and wish to make inferences about its input. This situation is considered in Sec. V A.

Barnett *et al.* have previously proposed a formalism for retrodictive inference in quantum theory [26–28]. For unbiased sources—for which the ensemble average of the prepared states is the maximally mixed state—their formalism is identical to the one presented here, and the quantum Bayes' theorem provides it with an intuitive derivation. For biased sources, their formalism differs from ours. The one we propose has the advantage that it can be derived as a special case of our general formalism for quantum Bayesian inference and thereby retains a closer analogy with classical Bayesian inference.

As emphasized by the de Finetti quote in the Introduction, the rules for making classical probabilistic inferences about the past are the same as those for making inferences about the future. By analogy, in the conditional states formalism, we would expect to be able to propagate quantum states from future regions to past regions via the same rules used to propagate them from past regions to future regions.

If the state ρ_A of an early region is mapped to the state ρ_B of a later region by a CPT map $\mathcal{E}_{B|A}$, then

$$\rho_B = \mathrm{Tr}_A(\varrho_{B|A}\rho_A), \tag{112}$$

where $\varrho_{B|A}$ is the Jamiołkowski isomorphic causal conditional state to $\mathcal{E}_{B|A}$. By construction, the causal conditional state $\varrho_{A|B}$ defined by Bayes' theorem in Eq. (94) satisfies $\varrho_{AB} = \varrho_{B|A} \star \rho_A = \varrho_{A|B} \star \rho_B$, and this causal joint state has ρ_A and ρ_B as its marginals, so we have

$$\rho_A = \mathrm{Tr}_B(\varrho_{A|B}\rho_B). \tag{113}$$

In conventional notation, this is equivalent to

$$\rho_A = \mathcal{E}_{A|B}^{\mathrm{retr}}(\rho_B), \tag{114}$$

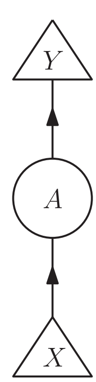
where

$$\mathcal{E}_{A|B}^{\mathrm{retr}}(\cdot) \equiv \rho_A^{\frac{1}{2}} \mathcal{E}_{A|B}^{\dagger}\left(\rho_B^{-\frac{1}{2}}(\cdot)\rho_B^{-\frac{1}{2}}\right)\rho_A^{\frac{1}{2}} \tag{115}$$

is the map that is Jamiołkowski isomorphic to $\varrho_{A|B}$ and the superscript "retr" is used to indicate that this is a retrodictive map that propagates states from future to past regions.

For comparison with the retrodictive formalism of [26], consider a simple prepare-and-measure experiment, as depicted in Fig. 10.

In the predictive description, the preparation procedure is characterized by a probability distribution $P(X)$, which can



Predictive expression:

$$\varrho_{XY} = \mathrm{Tr}_A\left(\varrho_{Y|A}\varrho_{A|X}\right) \star \rho_X$$

Conventional notation:

$$p_{x,y} = \mathrm{Tr}_A\left(E_y^A \rho_x^A\right) P(X=x)$$

Retrodictive expression:

$$\varrho_{XY} = \mathrm{Tr}_A\left(\varrho_{X|A}\varrho_{A|Y}\right) \star \rho_Y$$

Conventional notation:

$$p_{x,y} = \mathrm{Tr}_A\left(E_x^{A,\mathrm{retr}} \rho_y^{A,\mathrm{retr}}\right) P(Y=y)$$

FIG. 10. A prepare-and-measure experiment in which a preparation procedure is followed by a measurement. We are interested in computing the joint probability distribution $P(X,Y)$ of the preparation variable and the measurement outcome. For compactness, $p_{x,y} = P(X=x, Y=y)$ is used for the expressions in conventional notation.

alternatively be represented by a diagonal state ρ_X, and by a set of states ρ_x^A, which can alternatively be represented by a causal conditional state $\varrho_{A|X}$. The measurement is characterized by a POVM $\{E_y^A\}$ or, alternatively, by a causal conditional state $\varrho_{Y|A}$. This is predictive because the conditional states, $\varrho_{A|X}$ and $\varrho_{Y|A}$, are conditioned on regions in their immediate past.

In order to calculate the joint distribution $P(X,Y)$, or equivalently the causal joint state ϱ_{XY}, we need to use the belief propagation formula for causal conditional states as given in Eq. (52). This proceeds as follows.

(1) Propagate the causal conditional state of A given X into the future to obtain $\varrho_{Y|X} = \mathrm{Tr}_A(\varrho_{Y|A}\varrho_{A|X})$.

(2) Combine the causal conditional state for Y given X with the state for X to obtain $\varrho_{XY} = \rho_{Y|X} \star \rho_X$.

Combining these steps, the predictive expression for ϱ_{XY} is

$$\varrho_{XY} = \mathrm{Tr}_A(\varrho_{Y|A}\varrho_{A|X}) \star \rho_X. \tag{116}$$

In conventional notation, if the states $\{\rho_x^A\}$ are the components of $\varrho_{A|X}$ and the elements of the POVM $\{E_y^A\}$ are the components of $\varrho_{Y|A}$, then Eq. (116) is equivalent to

$$P(X=x,Y=y) = \mathrm{Tr}_A\left(E_y^A\rho_x^A\right)P(X=x). \tag{117}$$

A retrodictive description of the same experiment can be given, involving states of regions conditioned on regions in their immediate future, i.e., ρ_Y, $\varrho_{A|Y}$, and $\varrho_{X|A}$. These correspond, respectively, to a probability distribution $P(Y)$, a set of states $\{\rho_y^{A,\mathrm{retr}}\}$, and a POVM $\{E_x^{A,\mathrm{retr}}\}$. Note that, in contrast to the predictive description, the measurement is now being described by a classical probability distribution and a set of states, which we call retrodictive states, while the preparation is being described by a POVM, which we call the retrodictive POVM. The retrodictive calculation of ϱ_{XY} proceeds as follows.

(1) Propagate the causal conditional state of A given Y into the past to obtain $\varrho_{X|Y} = \mathrm{Tr}_A(\varrho_{X|A}\varrho_{A|Y})$.

(2) Combine the causal conditional state for X given Y with the state for Y to obtain $\varrho_{XY} = \rho_{X|Y} \star \rho_Y$.

Combining these steps, the retrodictive expression for ϱ_{XY} is

$$\varrho_{XY} = \mathrm{Tr}_A(\varrho_{X|A}\varrho_{A|Y}) \star \rho_Y. \tag{118}$$

In conventional notation, this is equivalent to

$$P(X=x,Y=y) = \mathrm{Tr}_A\left(E_x^{A,\mathrm{retr}}\rho_y^{A,\mathrm{retr}}\right)P(Y=y). \tag{119}$$

The retrodictive and predictive descriptions of the experiment are related by the quantum Bayes' and quantum belief propagation via

$$\varrho_{X|A} = \varrho_{A|X} \star \left(\rho_X\rho_A^{-1}\right), \tag{120}$$
$$\text{where} \quad \rho_A = \mathrm{Tr}_X(\rho_{A|X}\rho_X), \tag{121}$$

and

$$\varrho_{A|Y} = \varrho_{Y|A} \star \left(\rho_A\rho_Y^{-1}\right), \tag{122}$$
$$\text{where} \quad \rho_Y = \mathrm{Tr}_A(\rho_{Y|A}\rho_A). \tag{123}$$

In conventional notation, these equations are equivalent to

$$E_x^{A,\mathrm{retr}} = P(X=x)\rho_A^{-\frac{1}{2}}\rho_x^A\rho_A^{-\frac{1}{2}}, \tag{124}$$
$$\text{where} \quad \rho_A = \sum_x P(X=x)\rho_x^A, \tag{125}$$

and

$$\rho_y^{A,\mathrm{retr}} = \frac{\rho_A^{\frac{1}{2}}E_y^A\rho_A^{\frac{1}{2}}}{P(Y=y)}, \tag{126}$$
$$\text{where} \quad P(Y=y) = \mathrm{Tr}_B\left(E_y^A\rho_A\right). \tag{127}$$

Equations (120) and (122) can be used to prove that the predictive and retrodictive expressions for ϱ_{XY} do indeed give the same causal joint state. Starting from Eq. (118), we have

$$\begin{aligned}
\varrho_{XY} &= \mathrm{Tr}_A(\varrho_{X|A}\varrho_{A|Y}) \star \rho_Y \qquad (128)\\
&= \mathrm{Tr}_A\Big\{\left(\varrho_{A|X} \star \left[\rho_X\rho_A^{-1}\right]\right)\\
&\quad \left(\varrho_{Y|A} \star \left[\rho_A\rho_Y^{-1}\right]\right)\Big\} \star \rho_Y \qquad (129)\\
&= \mathrm{Tr}_A\Big(\rho_Y^{\frac{1}{2}}\rho_X^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\varrho_{A|X}\rho_X^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\\
&\quad \rho_A^{\frac{1}{2}}\rho_Y^{-\frac{1}{2}}\varrho_{Y|A}\rho_A^{\frac{1}{2}}\rho_Y^{-\frac{1}{2}}\rho_Y^{\frac{1}{2}}\Big) \qquad (130)\\
&= \mathrm{Tr}_A\Big(\rho_Y^{\frac{1}{2}}\rho_X^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\varrho_{A|X}\rho_X^{\frac{1}{2}}\rho_Y^{-\frac{1}{2}}\varrho_{Y|A}\rho_A^{\frac{1}{2}}\Big). \qquad (131)
\end{aligned}$$

Since ρ_Y commutes with ρ_X, ρ_A, and $\varrho_{A|X}$, the $\rho_Y^{\frac{1}{2}}$ term can be moved forward to cancel with $\rho_Y^{-\frac{1}{2}}$ term. The $\rho_A^{-\frac{1}{2}}$ term can be made to cancel with the last $\rho_A^{\frac{1}{2}}$ term via the cyclic property of the trace. This yields

$$\begin{aligned}
\varrho_{XY} &= \mathrm{Tr}_A\left(\rho_X^{\frac{1}{2}}\varrho_{A|X}\rho_X^{\frac{1}{2}}\varrho_{Y|A}\right) \qquad (132)\\
&= \rho_X^{\frac{1}{2}}\mathrm{Tr}_A(\varrho_{A|X}\varrho_{Y|A})\rho_X^{\frac{1}{2}} \qquad (133)\\
&= \mathrm{Tr}_A(\varrho_{A|X}\varrho_{Y|A}) \star \rho_X, \qquad (134)
\end{aligned}$$

where, in the second line, we have used the fact that ρ_X commutes with $\rho_{Y|A}$. Finally, applying the cyclic property of the trace gives

$$\varrho_{XY} = \mathrm{Tr}_A(\varrho_{Y|A}\varrho_{A|X}) \star \rho_X, \tag{135}$$

which is the predictive expression for ϱ_{XY}.

We are now in a position to show that our formalism coincides with that of Ref. [26] for the case of unbiased sources, where $\rho_A = I_A/d$ with d the dimension of $\mathcal{H}_A$. In this case, the definitions of retrodictive states and retrodictive POVMs in [26] were

$$E_x^{A,\mathrm{retr}} = dP(X=x)\rho_x^A, \tag{136}$$

and

$$\rho_y^{A,\mathrm{retr}} = \frac{E_y^A}{dP(Y=y)}, \tag{137}$$

which are easily seen to be special cases of Eqs. (124) and (126).

Finally, note that the analysis above can be extended to deal with the scenario depicted in Fig. 11, in which there is an intervening channel $\mathcal{E}_{B|A}$ between the preparation

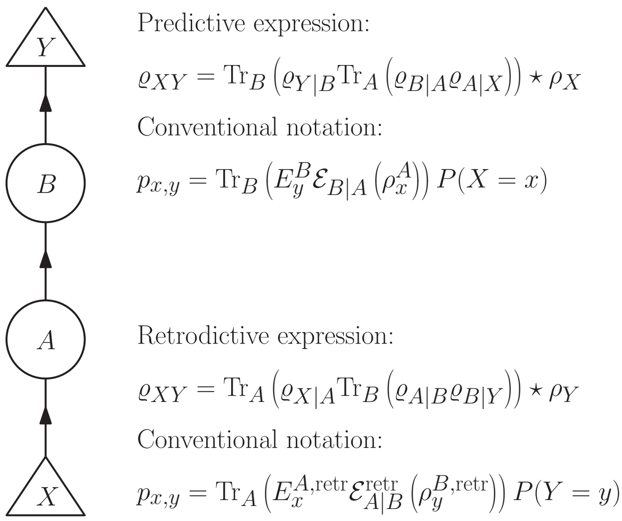


FIG. 11. A prepare-and-measure experiment with an intervening channel. We are interested in computing the joint probability distribution $P(X,Y)$ of the preparation variable and the measurement outcome. For compactness, $p_{x,y} = P(X=x,Y=y)$ is used for the expressions in conventional notation.

and measurement so that the measurement is now made on region B. In fact, by making use of the rule for propagating conditional states given in Eq. (52) [or equivalently the Heisenberg dynamics given in Eq. (60)], the extra region B can be eliminated from the description by defining

$$\varrho_{Y|A} = \mathrm{Tr}_A(\varrho_{Y|B}\varrho_{B|A}), \tag{138}$$

which determines the effective measurement on A that is performed by actually measuring region B. The three operators ρ_X, $\varrho_{A|X}$, and $\varrho_{Y|A}$ then provide a predictive description of a simple prepare-and-measure experiment, and so the previous analysis applies.

Specifically, substituting Eq. (138) into Eq. (116) gives the predictive expression for ϱ_{XY} as

$$\varrho_{XY} = \mathrm{Tr}_{AB}(\varrho_{Y|B}\varrho_{B|A}\varrho_{A|X}) \star \rho_X, \tag{139}$$

or in conventional notation

$$P(X=x,Y=y) = \mathrm{Tr}_B\left[E^B_y \mathcal{E}_{B|A}\left(\rho^A_x\right)\right] P(X=x). \tag{140}$$

Similarly, the retrodictive expression is obtained from Eq. (118) by substituting $\varrho_{A|Y} = \mathrm{Tr}_B(\varrho_{A|B}\varrho_{B|Y})$, where $\varrho_{A|B}$ and $\varrho_{B|Y}$ are obtained from $\varrho_{B|A}$ and $\varrho_{Y|B}$, respectively, by Bayes' theorem. This gives

$$\varrho_{XY} = \mathrm{Tr}_{AB}(\varrho_{X|A}\varrho_{A|B}\varrho_{B|Y}) \star \rho_Y, \tag{141}$$

or in conventional notation

$$P(X=x,Y=y) = \mathrm{Tr}_A\left[E^{A,\mathrm{retr}}_x \mathcal{E}^{\mathrm{retr}}_{A|B}\left(\rho^{B,\mathrm{retr}}_y\right)\right] P(Y=y). \tag{142}$$

To sum up, the formalism of quantum conditional states shows that, just as in the classical case, the rules of quantum Bayesian inference do not discriminate between prediction and retrodiction. Specifically, the rules are the same regardless of whether the propagation is in the same or the opposite direction to the causal arrows. This reveals an important kind of time symmetry that is not apparent in the normal quantum formalism. More importantly, it shows that a formalism for quantum Bayesian inference can be found that is blind to at least this aspect of the causal structure.

D. Remote measurements and spatial symmetry

Many of the novel features of quantum theory exhibit themselves in the correlations that can be obtained between local measurements on a pair of acausally related regions. These include Einstein-Podolsky-Rosen correlations and Bell correlations. Inferences about such measurements can be treated using a formalism that is almost identical to the predictive and retrodictive expressions for prepare-and-measure experiments given above. One simply has to substitute the formula for propagating a causal conditional state across acausally related regions [Eq. (55)] for the formula for propagating them across causally related regions [Eq. (52)] used above.

For two acausally related regions, A and B, it is self-evident that there is complete symmetry between propagation from one region to another and back again, i.e., if

$$\rho_B = \mathrm{Tr}_A(\rho_{B|A}\rho_A), \tag{143}$$

then

$$\rho_A = \mathrm{Tr}_B(\rho_{A|B}\rho_B). \tag{144}$$

The two conditional states, $\rho_{B|A}$ and $\rho_{A|B}$, are related by the quantum Bayes' theorem, so Bayes' theorem allows the direction of belief propagation to be reversed.

To see this spatial symmetry at work, consider the scenario of measurements being implemented on a pair of acausally related regions, A and B, as depicted in Fig. 12.

In the conditional states formalism, the region AB is assigned a joint state ρ_{AB}, and the measurements on A and B are represented by causal conditional states $\varrho_{X|A}$ and $\varrho_{Y|B}$. The joint distribution $P(X,Y)$ over outcomes is given by the components of the acausal joint state ρ_{XY} via

$$\rho_{XY} = \mathrm{Tr}_{AB}[(\varrho_{X|A}\varrho_{Y|B})\rho_{AB}]. \tag{145}$$

These correlations can alternatively be calculated by propagating beliefs about A conditioned on X to beliefs about Y conditioned on X, and also by propagating beliefs about B conditioned on Y to beliefs about X conditioned on Y. These representations are analogous to the predictive and retrodictive expressions for prepare-and-measure experiments discussed in the previous section. We refer to these as *rightward* and *leftward* belief propagation, respectively.

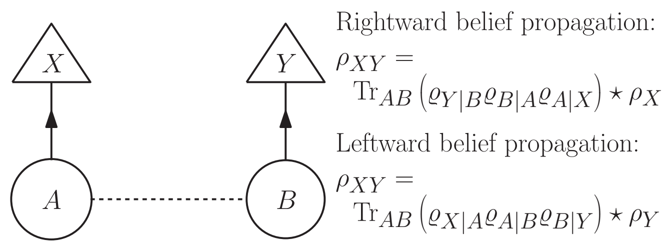


FIG. 12. Measurements on a pair of acausally related regions, rightward and leftward belief propagation.

It may seem convoluted to calculate ρ_{XY} via rightward or leftward belief propagation, when Eq. (145) already gives a simple expression for it. However, it is important to understand how to propagate beliefs across acausally related regions in order to deal with the situation in which you obtain new information about one region and wish to make inferences about the other. This is exactly what happens in the analysis of an EPR experiment. This problem is known as quantum steering and is considered in Sec. V A3.

First, consider rightward belief propagation. The aim is to rewrite Eq. (145) in terms of $\rho_{B|A}$, the state of B conditioned on the region to its left. For greater symmetry with the prepare-and-measure case, we also use ρ_X and $\varrho_{A|X}$ to describe the left-hand wing of the experiment, while retaining $\rho_{Y|B}$ for the right. Then, we write $\rho_{AB} = \rho_{B|A} \star \rho_A$ and note that, by Bayes' theorem,

$$\varrho_{X|A} = \varrho_{A|X} \star \left(\rho_X \rho_A^{-1}\right). \tag{146}$$

Substituting these into Eq. (145) gives

$$\rho_{XY} = \mathrm{Tr}_{AB}\left\{\left[\varrho_{A|X} \star \left(\rho_X \rho_A^{-1}\right)\right]\varrho_{Y|B}[\rho_{B|A} \star \rho_A]\right\}. \tag{147}$$

Expanding the $\star$ products gives

$$\rho_{XY} = \mathrm{Tr}_{AB}\left(\rho_X^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\varrho_{A|X}\rho_X^{\frac{1}{2}}\rho_A^{-\frac{1}{2}}\varrho_{Y|B}\rho_A^{\frac{1}{2}}\rho_{B|A}\rho_A^{\frac{1}{2}}\right). \tag{148}$$

All of the ρ_A terms can be canceled by using the fact that they commute with $\varrho_{Y|B}$ and ρ_X and by using the cyclic property of the trace. Then we have

$$\rho_{XY} = \mathrm{Tr}_{AB}\left(\rho_X^{\frac{1}{2}}\varrho_{A|X}\rho_X^{\frac{1}{2}}\varrho_{Y|B}\rho_{B|A}\right) \tag{149}$$

$$= \mathrm{Tr}_{AB}\left(\varrho_{Y|B}\rho_{B|A}\rho_X^{\frac{1}{2}}\varrho_{A|X}\rho_X^{\frac{1}{2}}\right) \tag{150}$$

$$= \mathrm{Tr}_{AB}(\varrho_{Y|B}\rho_{B|A}\varrho_{A|X}) \star \rho_X, \tag{151}$$

where we have used the cyclic property of the trace and the fact that ρ_X commutes with $\rho_{B|A}$ and $\varrho_{Y|B}$. Equation (151) has the same form as the predictive expression for a prepare-and-measure experiment with an intervening channel given in Eq. (139), except that, in this case, $\rho_{B|A}$ is acausal. We can also use Eq. (55) to define $\rho_{Y|A} = \mathrm{Tr}_B(\varrho_{Y|B}\rho_{B|A})$, which represents the effective measurement on region A that is made by measuring region B. Using this, region B can be eliminated from Eq. (151) to obtain

$$\rho_{XY} = \mathrm{Tr}_{AB}(\rho_{Y|A}\varrho_{A|X}) \star \rho_X. \tag{152}$$

This is similar to Eq. (116), and, in fact, is mathematically identical to it because $\rho_{Y|A}$ is a hybrid conditional state, so its mathematical form does not depend on whether it is acausal or causal. Equation (152) is a useful form to use when we want to consider the effect of measuring B on the remote region A, as in quantum steering.

For leftward belief propagation, a similar analysis gives the expressions

$$\rho_{XY} = \mathrm{Tr}_{AB}(\varrho_{X|A}\rho_{A|B}\varrho_{B|Y}) \star \rho_Y, \tag{153}$$

which is analogous to Eq. (141) and

$$\rho_{XY} = \mathrm{Tr}_A(\rho_{X|A}\varrho_{B|Y}) \star \rho_Y, \tag{154}$$

which is analogous to Eq. (118).

As with prediction and retrodiction, there is complete symmetry between leftward and rightward belief propagation, and there is a strong symmetry between causal and acausal belief propagation in general. This represents progress towards a theory of quantum Bayesian inference that is completely independent of causal structure.

V. QUANTUM BAYESIAN CONDITIONING

Classically, Bayesian conditioning is used to update probabilities when new data are acquired. Specifically, if you are interested in a variable R, and you learn that some correlated variable X takes the value x, then the theory of Bayesian inference recommends that you should update your probability distribution for R from the prior $P(R)$ to the posterior $P_x(R) = P(R|X = x)$.[6]

Bayesian conditioning can be viewed as a two-step process. First, the observation that $X = x$ causes you to update your probability distribution for X from $P(X)$ to $P_x(X)$, where

$$P_x(X = x') = \delta_{x,x'}. \tag{155}$$

Second, assuming that the observation of X does not cause you to change your conditional probabilities $P(R|X)$, the new probability distribution for R is obtained via belief propagation as

$$P_x(R) = \sum_X P(R|X)P_x(X) \tag{156}$$

$$= P(R|X = x). \tag{157}$$

This two-step decomposition of conditioning has been emphasized by Jeffrey [38,39], who showed that whenever an observation causes the value of a random variable to become certain, and that all probabilities conditioned on that random variable are unchanged by the observation, then the change in the probability distribution can be represented as Bayesian conditioning.

The reason for emphasizing this decomposition is that Jeffrey was interested in situations in which an observation does not cause you to believe that some variable takes a precise value. As an example of this, adapted from [40], suppose that X is the color of a jellybean, which has possible values "red," "green," and "yellow," and that R is the flavor, which has possible values "cherry," "strawberry," "lime," and "lemon." Suppose that initially, your probability distribution for X, $P(X)$, assigns a probability 1/3 to each color and that your observation consists of viewing the jellybean in the light of a dim candle. This might not be enough for you to become certain about the color of the jellybean, but it may reduce your uncertainty somewhat. For example, you may now think it reasonable to assign a probability distribution $P_{\text{post}}(X)$ that gives probability 2/3 to $X =$ "red" and 1/6 each to $X =$ "green" and $X =$ "yellow." Assuming that the observation does not cause your conditional probabilities $P(R|X)$ to change, Jeffrey

[6]Strictly speaking, this is only a special case of Bayesian conditioning, which can be formulated more generally for arbitrary events on a sample space rather than just for random variables. We restrict attention to the special case of conditioning one random variable upon another for ease of comparison with quantum theory.

shows that your posterior probability distribution for R is obtained by belief propagation via

$$P_{\text{post}}(R) = \sum_X P(R|X)P_{\text{post}}(X), \tag{158}$$

which is known as Jeffrey conditioning.

Many orthodox Bayesians reject the generalization to Jeffrey conditioning and maintain that the rational way to update probabilities in the light of data is always via Bayesian conditioning, not least because most of the apparatus of Bayesian statistics depends on this. This position can be defended by insisting that the sort of situations described above should really be handled by expanding the sample space to include statements about your perceptions. One can show that Jeffrey conditioning can always be represented as Bayesian conditioning on a larger space in this way. Counter to this, Jeffrey argues that it is not realistic to construct such a space, since you do not actually have precise descriptions of your perceptions. This argument has been eloquently put by Diaconis and Zabell [41].

"For example, suppose we are about to hear one of two recordings of Shakespeare on the radio, to be read by either Olivier or Gielgud, but are unsure of which, and have a prior with mass 1/2 on Olivier, 1/2 on Gielgud. After hearing the recording, one might judge it fairly likely, but by no means certain, to be by Olivier. The change in belief takes place by direct recognition of the voice; all the integration of sensory stimuli has already taken place at a subconscious level. To demand a list of objective vocal features that we condition on in order to affect the change would be a logician's parody of a complex psychological process."

The debate over whether Jeffrey conditioning should be subsumed into Bayesian conditioning is somewhat analogous to a similar argument in quantum theory about whether POVMs should be regarded as fundamental, since they can always be represented by projective measurements on a larger Hilbert space via Naimark extension [42].

In the conditional states formalism, the quantum analog of Jeffrey conditioning is straightforward. If an observation causes you to change the state you assign to a region A from ρ_A to ρ_A^{post}, and if your conditional state for another region B, given A (which will either be an acausal state $\rho_{B|A}$ or a causal state $\varrho_{B|A}$) is unchanged, then your posterior state for B is determined by belief propagation via either

$$\rho_B^{\text{post}} = \text{Tr}_A\left(\rho_{B|A}\rho_A^{\text{post}}\right) \tag{159}$$

or

$$\rho_B^{\text{post}} = \text{Tr}_A\left(\varrho_{B|A}\rho_A^{\text{post}}\right), \tag{160}$$

depending on whether A and B are acausally or causally related.

The question of whether there is a quantum analog of Bayesian conditioning is more subtle, as it depends on whether there is a posterior quantum state for A that is analogous to having certainty about the value of a classical variable, i.e., the point measure $P_x(X = x') = \delta_{x,x'}$. Arguably, a pure state could play this role, since it represents the smallest amount of uncertainty that one can have about a quantum region. However, unlike classical point measures, pure states still assign probabilities other than 0 and 1 to fine-grained measurements, e.g., measurements in a complimentary basis, and there are good reasons to believe that, even if they represent maximal knowledge, that knowledge is still incomplete [43].

We do not pursue this question further here, but instead focus on the case of a hybrid region XA. If the the data is the classical variable (so that one can indeed learn its value), then Bayesian conditioning has a straightforward generalization. Upon learning that $X = x$, the state of A should be updated via

$$\rho_A \rightarrow \sigma_{A|X=x}. \tag{161}$$

Recall that the general form of a state conditioned on a classical variable is $\sigma_{A|X} = \sum_x \rho_x^A \otimes |x\rangle\langle x|_X$, where $\{\rho_x^A\}$ is a set of normalized density operators and $\sigma_{A|X=x}$ is simply our notation for ρ_x^A. The elements of this set are called the components of $\sigma_{A|X}$, so we may describe Bayesian conditioning as replacing σ_A with one of the components of $\sigma_{A|X}$.

Note that, as in the classical case, conditioning can be viewed as a two-step process, wherein first the state of X is updated to the diagonal density operator for the point measure, $\rho_x^X = |x\rangle\langle x|_X$, and then belief propagation is used to determine the posterior state of A,

$$\rho_x^A = \text{Tr}_X\left(\sigma_{A|X}\rho_x^X\right), \tag{162}$$

$$= \langle x|_X\sigma_{A|X}|x\rangle_X, \tag{163}$$

$$= \sigma_{A|X=x}, \tag{164}$$

where we make use of the fact that $\sigma_{A|X} = \sum_x \sigma_{A|X=x} \otimes |x\rangle\langle x|$. Note that this holds regardless of the causal relation between A and X because a hybrid conditional such as $\sigma_{A|X}$ does not distinguish between these causal possibilities.

Recall that the rule for propagating unconditional beliefs about X to beliefs about A is $\rho_A = \text{Tr}_X(\sigma_{A|X}\rho_X)$. In conventional notation, this translates to $\rho_A = \sum_X P(X = x)\sigma_{A|X=x}$. Therefore, Bayesian conditioning is a process by which a state ρ_A is updated to an element $\sigma_{A|X=x}$ within a convex decomposition of ρ_A.

A. Examples of quantum Bayesian conditioning

In this section, we consider several examples of Bayesian conditioning, based on the different causal scenarios discussed in Secs. III and IV. In all these cases, conditioning the state of a quantum region on a classical variable is the correct thing to do in order to update the predictions or retrodictions that can be made about other classical variables correlated with the region. In particular, in Sec. V A3, we develop the application to quantum steering, showing that the set of states of a region that can be steered to by making remote measurements can be expressed compactly in terms of conditioning and belief propagation.

1. Conditioning on a preparation variable

Consider again the preparation scenario depicted in Fig. 3(a), wherein a quantum region A is prepared in one of a set of states $\{\varrho_{A|X=x}\}$ depending on the value of a classical variable X with prior probability distribution $P(X)$. This can alternatively be described by a conditional state $\varrho_{A|X}$ and a diagonal state ρ_X. In this case, Bayesian conditioning of A on

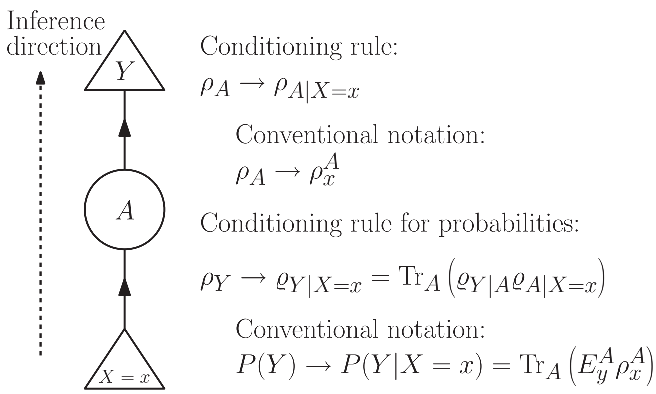


FIG. 13. Predictive inference in a prepare-and-measure experiment. We are interested in inferring the probability of the measurement outcome given knowledge of the preparation variable.

the value x of X corresponds to updating from the ensemble average state $\rho_A = \sum_x P(X=x)\varrho_{A|X=x}$ to the particular state $\varrho_{A|X=x}$, corresponding to the value x of X that actually obtains, which is clearly a reasonable thing to do.

The operational significance of this conditioning becomes apparent by considering a measurement made on region A, described by a conditional state $\varrho_{Y|A}$. We now have a prepare-and-measure experiment, where we are interested in making a predictive inference about the measurement outcome from knowledge of the preparation variable, as depicted in Fig. 13.

Equation (116) gives the predictive expression for the joint probability distribution for this experiment as

$$\varrho_{XY} = \mathrm{Tr}_A(\varrho_{Y|A}\varrho_{A|X}) \star \rho_X. \tag{165}$$

From this, we can compute the marginal probability for Y as

$$\rho_Y = \mathrm{Tr}_A(\varrho_{Y|A}\rho_A), \tag{166}$$

where $\rho_A = \mathrm{Tr}_X(\varrho_{A|X}\rho_X)$ is the ensemble average state. The conditional $\varrho_{Y|X} = \varrho_{XY} \star \rho_X^{-1}$ is given by

$$\rho_{Y|X} = \mathrm{Tr}_A(\varrho_{Y|A}\varrho_{A|X}), \tag{167}$$

from which it follows that

$$\rho_{Y|X=x} = \mathrm{Tr}_A(\varrho_{Y|A}\varrho_{A|X=x}). \tag{168}$$

The transition from Eqs. (166) to (168) is just classical Bayesian conditioning of the probability for Y on the value of X. Both expressions are representations of the Born rule in terms of belief propagation, and Eq. (168) can be obtained from Eq. (166) by replacing ρ_A with $\varrho_{A|X=x}$, which is just quantum Bayesian conditioning. Thus, quantum Bayesian conditioning on a preparation variable can be used as an intermediate step in updating the probability distribution for a measurement outcome by classical Bayesian conditioning.

Nothing changes if we consider the slightly more complicated scenario where there is an intermediate channel between the preparation and measurement, as depicted in Fig. 11. This is because, as shown in Sec. IV C, the joint probability is still given by Eq. (165), where now $\varrho_{Y|A} = \mathrm{Tr}_B(\varrho_{Y|B}\varrho_{B|A})$ describes the effective measurement on A that corresponds to the actual measurement made on the later region B. By similar reasoning, there could be an arbitrary number of time steps between the preparation and measurement and conditioning the state of A on the preparation variable would still be the correct way to update the Born rule probabilities for Y.

2. Conditioning on the outcome of a measurement

When conditioning on a measurement outcome, it is important to recall that the conditional states formalism assigns states to regions rather than to persistent systems. Therefore, when we update the state of a region by conditioning on a measurement outcome, the resulting conditionalized state is assigned to the very same region that we started with. This is a different concept from the usual state-update rules that occur in the standard quantum formalism, such as the projection postulate. These standard rules apply to persistent systems when we are interested in how the state of a system in a region before the measurement gets mapped to its state in a region after the measurement. Because the updated state belongs to a different region than the initial state, this is not an example of pure conditioning in our framework. Therefore, one should not think that conditioning on a measurement outcome must reproduce the projection postulate. The way in which this kind of state-update rule is handled in the conditional states framework is discussed in Sec. V B.

However, there are several other types of inference for which pure conditioning on a measurement outcome is the correct update rule to use. In particular, conditioning can be used for making retrodictions about classical variables in the past of the region of interest. For instance, in a quantum communication scheme, registering the outcome of a measurement on the output of the channel leads us to infer something about which of a set of classical messages was encoded in the quantum state of the channel as input. The use of conditioning for this sort of inference is the topic of this section.

Recall the measurement scenario depicted in Fig. 4(a). A measurement with outcomes labeled by the classical variable Y is implemented upon a quantum region A. In the predictive formalism, this experiment is described by an input state ρ_A and a causal conditional state $\varrho_{Y|A}$ (describing the measured POVM). To condition A on a value y of Y, Bayes' theorem must be applied to determine $\varrho_{A|Y}$, and then the component $\varrho_{A|Y=y}$ gets picked out by conditioning. This gives

$$\rho_A \to \varrho_{A|Y=y} = \varrho_{Y=y|A} \star \left(\rho_A \rho_{Y=y}^{-1}\right), \tag{169}$$

or in conventional notation

$$\rho_A \to \rho_y^{A,\mathrm{retr}} = \frac{\rho_A^{\frac{1}{2}} E_y^A \rho_A^{\frac{1}{2}}}{\mathrm{Tr}_A\left(E_y^A \rho_A\right)}. \tag{170}$$

The state $\varrho_{A|Y=y}$ represents the state of a region A *prior to the measurement*, given the outcome of the measurement; i.e., it is a retrodictive state. Its operational significance is that it allows one to make inferences about variables involved in the preparation of A.

To see this, consider again the prepare-and-measure experiment, where now we are interested in making a retrodictive inference from the measurement outcome to the preparation variable, as depicted in Fig. 14.

Recall from Sec. IV C that there is complete symmetry between the predictive and retrodictive expressions for a prepare-and-measure experiment under exchange of the

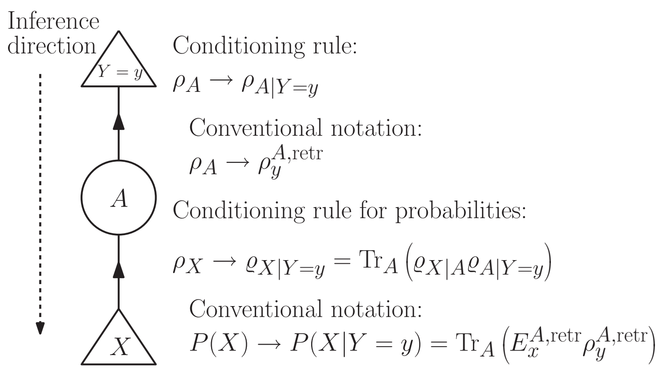


FIG. 14. Retrodictive inference in a prepare-and-measure experiment. We are interested in inferring the probability of the preparation variable given knowledge of the measurement outcome.

preparation variable X with the measurement variable Y. Thus, everything that was said regarding the probability distribution for the measurement outcome in the previous example, applies here to the probability distribution for the preparation variable. In particular, the marginal probability distribution for X is

$$\rho_X = \text{Tr}_A(\varrho_{X|A}\rho_A), \tag{171}$$

and conditioning this on $Y = y$ gives

$$\rho_{X|Y=y} = \text{Tr}_A(\varrho_{X|A}\varrho_{A|Y=y}). \tag{172}$$

Both of these expressions are belief propagation representations of the Born rule, with respect to the retrodictive POVM for X. Thus, conditioning A on the outcome of the measurement can be used as an intermediate step in updating the probability distribution for the preparation variable in light of the measurement outcome. The retrodictive state appearing in the retrodictive Born rule simply gets updated by Bayesian conditioning.

As in the previous example, nothing changes if there are intermediate channels between the preparation and the measurement. We can simply use conditional belief propagation to eliminate the additional regions.

3. Conditioning on the outcome of a remote measurement: Quantum steering

Finally, consider the case of a measurement made on a region B that is acausally related to a region A, as depicted in Fig. 8. This experiment is described by an acausal joint state ρ_{AB}, and a conditional state $\varrho_{Y|B}$, corresponding to the POVM measured on B. We are interested in how the state of the remote region A is updated when we learn the outcome of the measurement made on B. The updated state of A could then be used to predict the outcome of a measurement made on A, corresponding to a causal conditional state $\varrho_{X|A}$. This scenario is depicted in Fig. 12. The results of Sec. IV D establish that the mathematical description of this experiment is formally equivalent to a prepare-and-measure experiment with an intervening channel, the only difference being that the causal conditional state $\varrho_{B|A}$ is replaced by an acausal conditional state $\rho_{B|A}$. This symmetry is enough to establish that, as in the previous example, conditioning the state of A on $Y = y$ must be the correct way of updating the Born rule probabilities for X.

However, it is worth developing this example in a little more detail, since the "remote collapse" at A that occurs upon measuring B is at the core of the EPR argument. This has led to a study of the ensembles of states for A that can be obtained by measuring B, a problem known as quantum steering [25,30,44–49]. The conditional states formalism provides an elegant approach to this problem.

Recall from Sec. IV D that the joint probability for X and Y can be computed via leftward belief propagation, which yields Eq. (154), i.e.,

$$\rho_{XY} = \text{Tr}_A(\varrho_{X|A}\rho_{A|Y}) \star \rho_Y. \tag{173}$$

This formula is obtained by first applying Bayesian inversion to $\varrho_{Y|B}$ to determine

$$\varrho_{B|Y} = \varrho_{Y|B} \star \left(\rho_B\rho_Y^{-1}\right), \tag{174}$$

where $\rho_Y = \text{Tr}_B(\varrho_{Y|B}\rho_B)$ is the Born rule probability distribution for the measurement outcome. Then, conditional belief propagation is used to obtain $\rho_{A|Y} = \text{Tr}_B(\rho_{A|B}\varrho_{B|Y})$.

Equation (173) is formally equivalent to the retrodictive expression for a prepare-and-measure experiment, so the rationale for conditioning is exactly the same as in the previous example. Specifically, the marginal probability distribution for X is

$$\rho_X = \text{Tr}_A(\varrho_{X|A}\rho_A), \tag{175}$$

and conditioning on $Y = y$ gives

$$\rho_{X|Y=y} = \text{Tr}_A(\varrho_{X|A}\rho_{A|Y=y}). \tag{176}$$

Thus, conditioning the state of A on $Y = y$ is an appropriate intermediate step for updating the Born rule probabilities for X.

The following proposition summarizes this result and translates it into conventional notation.

Proposition 6. Let ρ_{AB} be the joint state of two acausally related regions. Suppose that the POVM corresponding to the conditional state $\varrho_{Y|B}$ is measured on B and the outcome $Y = y$ is obtained. Then, the state of region A should be updated from the prior $\rho_A = \text{Tr}_B(\rho_{AB})$ to $\rho_{A|Y=y}$, where

$$\rho_{A|Y} = \text{Tr}_B(\rho_{A|B}\varrho_{B|Y}), \tag{177}$$

$\rho_{A|B} = \rho_{AB} \star \rho_B^{-1}$, and

$$\varrho_{B|Y} = \varrho_{Y|B} \star \left(\rho_B\rho_Y^{-1}\right). \tag{178}$$

Let the components of the conditional state $\varrho_{Y|B}$ be the elements of the POVM $\{E_y^B\}$ and let $\mathfrak{E}_{A|B}$ be the map that is Jamiołkowski-isomorphic to $\rho_{A|B}$. Then the updated state of A is

$$\rho_y^A = \mathfrak{E}_{A|B}\left(\rho_y^B\right), \tag{179}$$

where

$$\rho_y^B = \frac{\rho_B^{1/2}E_y^B\rho_B^{1/2}}{\text{Tr}_B\left(E_y^B\rho_B\right)}. \tag{180}$$

In the above analysis, the method used to compute $\rho_{A|Y=y}$ is to first apply Bayes' theorem to $\varrho_{Y|B}$ and then apply conditional belief propagation. This is the acausal analog of performing the calculation in the retrodictive formalism. However, we could equally well apply belief propagation first to obtain

$$\rho_{Y|A} = \mathrm{Tr}_B(\varrho_{Y|B}\rho_{B|A}) \tag{181}$$

and then apply Bayes' theorem to obtain

$$\rho_{A|Y=y} = \rho_{Y=y|A} \star \left(\rho_A \rho_Y^{-1}\right). \tag{182}$$

This is the acausal analog of performing the calculation in the Heisenberg picture, and it is straightforward to check that it gives the same result.

In conventional notation, this amounts to first determining the effective POVM on A that is performed when B is measured via

$$E_y^A = \mathfrak{E}^{\dagger}_{A|B}\left(E_y^B\right) \tag{183}$$

and then determining the updated state via

$$\rho_y^A = \frac{\rho_A^{\frac{1}{2}} E_y^A \rho_A^{\frac{1}{2}}}{\mathrm{Tr}_B\left(E_y^A \rho_A\right)}. \tag{184}$$

Expressions equivalent to these conventional formulas have appeared previously in [25], the only difference being the appearance of a transpose, due to the use of the Choi isomorphism rather than the Jamiołkowski isomorphism in [25].

We have seen that the remote collapse $\rho_A \to \rho_{A|Y=y}$ is an instance of quantum Bayesian conditioning. Within interpretations of quantum theory wherein quantum states are considered to describe reality completely, the change of the state of A as a result of a distant measurement upon B is sometimes considered to be an instance of action at a distance. Indeed, Einstein criticized the Copenhagen interpretation on exactly these grounds. On the other hand, the analysis above shows clearly that if one views quantum theory as a theory of Bayesian inference, then upon learning the outcome of a measurement on B, all that occurs is that one's beliefs about A are updated. No change to the physical state of A is required within such an approach. This interpretation is bolstered by the formal equivalence to the case of conditioning a region on the outcome of a subsequent measurement, which does not seem to imply retrocausal influences (although for realist interpretations it has been argued that a imposing a particular symmetry principle does imply retrocausality in this scenario [5,6]).

Einstein anticipated such an epistemic interpretation of remote collapse in his writing, as is argued in [50]. However, he most likely thought that the probabilities represented by quantum states could be probabilities for the values of physical variables (possibly hidden) and that these could satisfy classical probability theory. However, by virtue of Bell's theorem [51], such an interpretation of remote collapse is not possible within the standard framework for hidden variable theories. We evade this no-go result here by understanding remote collapse as Bayesian updating within a noncommutative probability theory rather than within classical probability theory. In itself, this does not provide a viable realist interpretation of quantum theory, but it does suggest that an acceptable ontology for quantum theory ought not to include the quantum state. Finally, note that our interpretation of remote steering is broadly in harmony with that of Fuchs [24]. However, the view of quantum theory presented here differs from that of Fuchs in that he views quantum theory as a restriction upon classical probability theory, whereas we consider it to be a generalization thereof.

B. Why the projection postulate is not an instance of Bayesian conditioning

Finally, in this section, we deal with the elephant in the room: the projection postulate. Some authors have argued that the projection postulate is a quantum analog of Bayesian conditioning (see, e.g., [34,35]). However, the projection postulate is not an instance of quantum Bayesian conditioning as defined above. In this section, we discuss the relationship between quantum Bayesian conditioning and the projection postulate (and quantum instruments in general) at some length, in order to dispel the misconception that projection is a kind of conditioning. After pointing out the formal differences between the two, we explain the different types of update rule that are associated with a measurement in both classical probability theory and quantum theory, pointing out where conditioning and projection fit into this picture. Then, we explain how, in the conditional states formalism, the projection postulate should be thought of as a composite operation, consisting of belief propagation to a later region followed by Bayesian conditioning. This is broadly in line with the treatment of quantum measurements advocated by Ozawa [36,37]. Finally, we deal with the possible objection that, although the classical analog of the projection postulate is not just Bayesian conditioning, it can be thought of as conditioning combined with a simple relabeling of the system variable. This argument does not apply in quantum theory because, unlike in a classical theory, any informative measurement necessarily disturbs the system being measured. Although this is well known, we present a formulation of information disturbance in terms of conditional states, which makes it clear that no quantum instrument can be thought of as conditioning combined with relabeling. This may be of interest in its own right, as it emphasizes the similarity between information disturbance and other trade-offs in quantum theory, such as the monogamy of entanglement.

In the conventional formalism, when a projective measurement $\{\Pi_y^A\}$ is made on a system A, the Lüders-von Neumann projection postulate says that, upon learning the outcome y, the initial state ρ_A should be updated via

$$\rho_A \to \frac{\Pi_y^A \rho_A \Pi_y^A}{\mathrm{Tr}_A\left(\Pi_y^A \rho_A\right)}. \tag{185}$$

For a general POVM $\{E_y^A\}$, there is a natural generalization of the projection postulate, given by

$$\rho_A \to \frac{\left(E_y^A\right)^{\frac{1}{2}} \rho_A \left(E_y^A\right)^{\frac{1}{2}}}{\mathrm{Tr}_A\left(E_y^A \rho_A\right)}. \tag{186}$$

This generalization has also been proposed as a quantum analog of Bayesian conditioning [52,53].

On the other hand, in Sec. V A2 the rule for conditioning a state on the outcome of a measurement was found to be

$$\rho_A \to \frac{\rho_A^{\frac{1}{2}} E_y^A \rho_A^{\frac{1}{2}}}{\mathrm{Tr}_A\left(E_y^A \rho_A\right)}. \tag{187}$$

This is distinct from Eq. (186) because the roles of ρ_A and E_y^A have been interchanged. Furthermore, Eq. (186) is not equivalent to an equation of the form of Eq. (187) (even allowing a different POVM to appear therein) because the map $\rho_A \to (E_y^A)^{\frac{1}{2}} \rho_A (E_y^A)^{\frac{1}{2}}$, like all quantum instruments, is linear on the set of states on $\mathcal{H}_A$, whereas the map $\rho \to \rho_A^{\frac{1}{2}} E_y^A \rho_A^{\frac{1}{2}}$ is nonlinear on the state space.

If you are inclined to view the projection postulate as the correct quantum generalization of Bayesian conditioning, then you could take this as evidence *against* the idea that the conditional states formalism provides an adequate theory of quantum Bayesian inference. We therefore go to some length to defend the claim that neither the projection postulate nor any other quantum instrument is an analog of classical Bayesian conditioning.

Consider the scenario depicted in Fig. 15(a). A quantum system located in region A and described by the state ρ_A is subjected to a measurement with outcomes labeled by Y, and the system persists after the measurement. At this later time, it is represented by a region A' which, as always in the present formalism, we distinguish from region A, but which is associated with a Hilbert space of the same dimension. As discussed in Sec. III K, a quantum instrument, $\{\mathcal{E}_y^{A'|A}\}$, determines how the state of A' is related to the state of A, where $\mathcal{E}_y^{A'|A}(\rho_A)$ is the un-normalized state of A' obtained when $Y = y$.

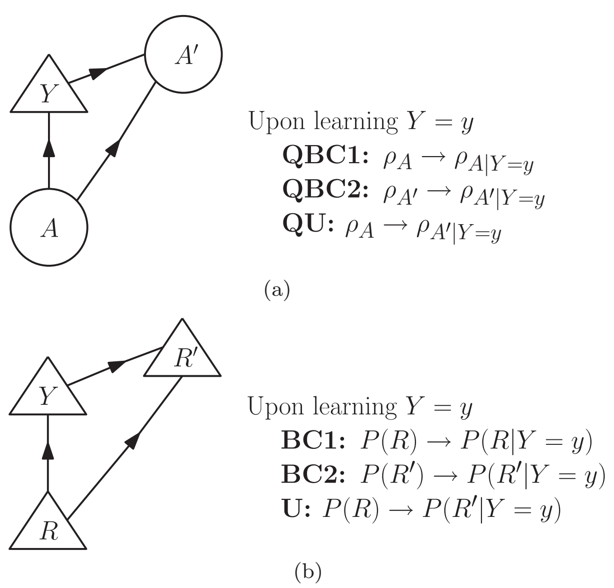


FIG. 15. Causal diagrams for the state-update rules associated with quantum and classical measurements. (a) A quantum instrument, representing how the state of a quantum persistent system changes after a measurement. (b) The classical analog of a quantum instrument, representing how the state of a classical persistent system changes after a measurement.

Now consider the classical analog of this scenario, depicted in Fig. 15(b). A classical system described by the variable R, and assigned a distribution $P(R)$, is subjected to a (possibly noisy) measurement, resulting in the outcome Y, which is a random variable that depends on R through a conditional probability distribution $P(Y|R)$. The system persists after the measurement, where it is described by a random variable R'. The value of R' is presumed to depend probabilistically on R, and the nature of this dependence may vary with the outcome Y. This is captured by a conditional probability $P(Y,R'|R)$, which is the classical analog of a quantum instrument.

It is useful to distinguish three kinds of update rules that might be considered in this scenario, as defined in Fig. 15(b). To describe the difference between these rules, it is useful to introduce some terminology. A distribution over R or R' is said to be a *prior* distribution if it is not conditioned on the value of Y, and it is a *posterior* distribution if it is conditioned on the value of Y. The temporal ordering that is implicit in this prior-posterior terminology specifies whether the distribution characterizes the knowledge you have *before* learning the value of Y or the knowledge you have *after* learning the value of Y. In other words, it refers to the time in your epistemological history, relative to the event of learning Y. On the other hand, the system's configuration at the time before the occurrence of the measurement is called its *initial* configuration and its configuration at the time after the occurrence of the measurement is called its *final* configuration. R is the initial configuration and R' is the final configuration. The temporal ordering implicit in this latter distinction refers to the time in the system's ontological history.

Strictly speaking, Bayesian conditioning is a rule that updates what one knows about *one and the same* variable upon the acquisition of new information. In other words, it maps a prior distribution about some variable to a posterior distribution for that same variable. Consequently, rule **BC1**, which maps the prior distribution of the initial configuration of the system to the posterior distribution of the initial configuration of the system, is an instance of Bayesian conditioning (it is analogous to updating a retrodictive state as discussed in Sec. V A2). Rule **BC2** is also an instance of Bayesian conditioning: It maps the prior distribution of the final configuration of the system to the posterior distribution of the final configuration of the system. However, the rule **U** is *not* an instance of Bayesian conditioning because it maps the prior distribution of the *initial* configuration of the system to the posterior distribution of the *final* configuration. In other words, if one considers R and R' to be distinct variables, then any map from a distribution over one of them to a distribution over the other cannot be an instance of Bayesian conditioning.

We now return to the quantum scenario, this time using the conditional states formalism and paying attention to the analogy with the classical case. The measurement is associated with the causal conditional state $\varrho_{YA'|A}$ that corresponds to the quantum instrument $\{\mathcal{E}_y^{A'|A}\}$. In the quantum conditional states formalism, there are analogs of each of the three rules we considered above. These are specified in Fig. 15(a). The rule **QBC1** corresponds to updating a retrodictive state, as considered in Sec. V A2. The von Neumann-Lüders projection postulate is clearly an instance of rule **QU**. If one stipulates that quantum Bayesian conditioning is a rule that updates the quantum description of *one and the same* region upon acquiring new information, i.e., that it maps a prior state for a region to a posterior state for the same region, then **QBC1** and **QBC2** are instances of quantum Bayesian conditioning, but **QU** is not.

1. State-update rules as a combination of belief propagation and Bayesian conditioning

If **QU** is not an instance of Bayesian conditioning, then what is its status within our framework? We now show that it is a composite of two operations: belief propagation followed by Bayesian conditioning.

First consider the classical analog. The analog of the projection postulate (or any state-update rule arising from an instrument) is given by rule **U**: $P(R) \to P(R'|Y=y)$. This can be obtained by combining an instance of belief propagation, namely

$$P(R) \to P(R') = \sum_R P(R'|R)P(R), \tag{188}$$

followed by the rule **BC2**: $P(R') \to P(R'|Y=y)$, which is an instance of Bayesian conditioning. It is useful to express both these steps in terms of the quantities that are given in the problem, namely, the conditional $P(Y,R'|R)$ and the prior over R, $P(R)$. The conditional probability distribution $P(R'|R)$ used in Eq. (188) is simply the marginal of $P(Y,R'|R)$, i.e., $P(R'|R) = \sum_Y P(Y,R'|R)$. Meanwhile, the expression for the conditional is $P(R'|Y) = P(R',Y)/P(Y)$, where $P(R',Y) = \sum_R P(Y,R'|R)P(R)$. Setting $Y=y$ gives the Bayesian conditioning step as

$$P(R') \to P(R'|Y=y) = \frac{\sum_R P(Y=y,R'|R)P(R)}{P(Y=y)}. \tag{189}$$

The quantum analog of this is straightforward. Quantum state-update rules, such as the projection postulate, are of the form **QU**: $\rho_A \to \varrho_{A'|Y=y}$. This is simply a sequential combination of quantum belief propagation,

$$\rho_A \to \rho_{A'} = \mathrm{Tr}_A(\varrho_{A'|A}\rho_A), \tag{190}$$

with quantum Bayesian conditioning via **QBC2**: $\rho_{A'} \to \varrho_{A'|Y=y}$.

Again, it is useful to express each of these steps in terms of the quantities that are given in the problem: the causal conditional state $\varrho_{YA'|A}$ and the prior ρ_A. The causal conditional state $\varrho_{A'|A}$ used in Eq. (190) is simply a reduced state of $\varrho_{YA'|A}$, i.e., $\varrho_{A'|A} = \mathrm{Tr}_Y(\varrho_{YA'|A})$. To gain some intuition for this step, we translate it into conventional notation. If the quantum instrument associated with $\varrho_{YA'|A}$ is denoted $\{\mathcal{E}_y^{A'|A}\}$, then Eq. (190) becomes

$$\rho_A \to \rho_{A'} = \mathcal{E}_{A'|A}(\rho_A), \tag{191}$$

where

$$\mathcal{E}_{A'|A} = \sum_y \mathcal{E}_y^{A'|A}. \tag{192}$$

The map $\mathcal{E}_{A'|A}$ is the *nonselective* update map. It is the appropriate map to apply when you know that the measurement has been performed, but you do not know which outcome occurred. The standard update map, which is appropriate when one also knows the outcome, is the *selective* update map. The projection postulate and its generalization to POVMs are instances of selective update maps. Applying the nonselective update map is just an instance of quantum belief propagation.

Turning to the Bayesian conditioning step, we have $\varrho_{A'|Y} = \rho_{A'Y} \star \rho_Y^{-1}$, where $\rho_{A'Y} = \mathrm{Tr}_A(\varrho_{YA'|A}\rho_A)$. Combining these and setting $Y=y$ gives

$$\rho_{A'} \to \varrho_{A'|Y=y} = \mathrm{Tr}_A(\varrho_{Y=y,A'|A}\rho_A) \star \rho_{Y=y}^{-1}. \tag{193}$$

In conventional notation, this translates into

$$\rho_{A'} \to \rho_y^{A'} = \frac{\mathcal{E}_y^{A'|A}(\rho_A)}{P(Y=y)}. \tag{194}$$

Given the expression for $\rho_{A'}$ in Eq. (191), we see that **QBC2** is simply a transition from your prior about the system output by the measurement, the result of applying the *nonselective* update map, to your posterior about the system output by the measurement, the result of applying the *selective* update map and normalizing. That this transition from nonselective to selective updates should be regarded as analogous to Bayesian conditioning has previously been argued by Ozawa [36,37].

In fact, the rule **QBC2** is a particular example of the kind of Bayesian conditioning considered in Sec. V A1. Every measurement with output region A' can be considered to define a preparation of A' for every outcome (assuming a fixed input state). The set of states prepared is given by the components of $\varrho_{A'|Y}$, which we can compute from the causal joint state $\varrho_{YA'A} = \varrho_{YA'|A} \star \rho_A$ by tracing over A and conditioning on Y. The rule **QBC2** is then just Bayesian conditioning on Y, thought of as a classical preparation variable.

2. No information gain without disturbance

We have argued that neither the classical rule **U** nor the quantum state-update rule **QU** are instances of Bayesian conditioning. However, a skeptic might counter that our argument is an artifact of our insistence that the system before and after the measurement should be given different labels. If the conditional distribution $P(R'|R)$ in the belief propagation rule of Eq. (188) has the form

$$P(R'=r'|R=R) = \delta_{r',r}, \tag{195}$$

where $\delta_{r',r}$ is the Kronecker-δ function, then R and R' are perfectly correlated and consequently $P(R'|Y=y)$ has precisely the same functional form as $P(R|Y=y)$. In this case, one could say that the rule **U** is *effectively* just Bayesian conditioning.

If $P(R'|R)$ is just a δ function, then we say that the measurement is *nondisturbing*. Recall that for every $P(Y|R)$ that characterizes the outcome probabilities for a measurement, there are many conditionals (i.e., classical instruments) $P(Y,R'|R)$ that might characterize its transformative aspect and are consistent with $P(Y|R)$. It is not difficult to see that, among all such conditionals, there is always one for which the measurement is nondisturbing, namely,

$$P(Y,R'=r'|R=r) = P(Y|R=r)\delta_{r',r}. \tag{196}$$

Of course, there are also many ways of implementing a measurement of $P(Y|R)$ such that it *is* disturbing. Therefore, while the update rule **U** is not an instance of Bayesian conditioning for every possible implementation of the measurement, there is always at least one implementation such that it is effectively just Bayesian conditioning.

The obvious question to ask at this point is whether it is possible to implement a *quantum* measurement in a nondisturbing way, such that the associated quantum update rule **QU** (perhaps the projection postulate, perhaps some other rule) is effectively just quantum Bayesian conditioning. For the measurement to be nondisturbing, the causal conditional state $\varrho_{A'|A}$ in the belief propagation rule of Eq. (190) would have to be of the form

$$\varrho_{A'|A} = \sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}, \tag{197}$$

i.e., it would need to be the partial transpose of the (unnormalized) maximally entangled state. This corresponds to perfect correlation between A and A' because it is Jamiołkowski-isomorphic to the identity channel. If the belief propagation step in the rule **QU** were of this form, then $\varrho_{A'|Y=y}$ would have the same functional form as $\varrho_{A|Y=y}$, and the overall quantum update rule **QU** would be effectively just Bayesian conditioning.

Of course, we are only interested in the case where the measurement is nontrivial. If the measurement gives *no* information about A, then the posterior is the same as the prior and no real conditioning has occurred. Therefore, we restrict our attention to the case where some information is gained. In the language of conditional states, the only kind of measurement that yields no information about the input state is one associated with a causal conditional state that factorizes, that is, one of the form $\varrho_{Y|A} = \rho_Y$ (recall that, in our notation there is an implicit $\otimes I_A$ on the right-hand side of this equation). In conventional notation, this corresponds to a POVM of the form $\{P(Y = y)I_A\}$, which generates a random outcome $Y = y$ from the distribution $P(Y)$ regardless of the state of A. We are interested in nontrivial measurements for which $\varrho_{Y|A}$ does not factorize in this way.

With these definitions in hand, the answer to our question is a resounding "no"; a quantum state-update rule of the form **QU** can *never* be effectively just Bayesian conditioning because, unlike the classical case, in quantum theory information gain necessarily implies a disturbance. This prevents any **QU** rule, such as the projection postulate, from being pure Bayesian conditioning. While this fact is well known, it is instructive to prove it in the conditional states formalism.

Theorem 8: No information gain without disturbance. Consider a measurement described by an instrument associated with the causal conditional state $\varrho_{YA'|A}$. It is impossible for this measurement to be both informative about A ($\varrho_{Y|A} \neq \rho_Y$) and nondisturbing ($\varrho_{A'|A} = \sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}$).

The proof is a causal analog of the monogamy of entanglement (see [3] for related ideas).

Proof. The operator $\varrho_{YA'|A}$ is the partial transpose over A of an acausal conditional state $\rho_{YA'|A}$. Combining this with the maximally mixed state gives a valid tripartite acausal state via

$$\rho_{YA'A} = \rho_{YA'|A} \star I_A/d = \rho_{YA'|A}/d, \tag{198}$$

where d is the dimension of $\mathcal{H}_A$. The condition that the measurement be nondisturbing is equivalent to $\rho_{A'|A} = \sum_{j,k} |j\rangle\langle k|_A \otimes |j\rangle\langle k|_{A'}$, which implies that the tripartite state $\rho_{YA'A}$ should have a reduced state on $A'A$ that is maximally entangled. Meanwhile, the condition that the measurement be informative is equivalent to $\rho_{Y|A} \neq \rho_Y$, which implies that the reduced state on YA of $\rho_{YA'A}$ should not be a product state. However, by the monogamy of entanglement, any tripartite state $\rho_{YA'A}$ for which $\rho_{AA'}$ is maximally entangled must have a product state for its reduced state ρ_{YA}. Hence, both conditions cannot be satisfied simultaneously. ■

We end with another, less obvious, disanalogy between the quantum and classical cases that bears on the question of how to interpret the quantum collapse rule in our framework. In the previous section, we showed that the classical update rule **U** could be decomposed into belief propagation followed by Bayesian conditioning. We could just as well have decomposed it in the opposite order: Bayesian conditioning followed by belief propagation. Specifically, we could first apply **BC1**: $P(R) \to P(R|Y = y)$, and then propagate the conditioned state via $P(R|Y = y) \to P(R'|Y = y) = \sum_R P(R'|R,Y = y)P(R|Y = y)$.

It is natural to ask whether such a reverse-order decomposition is possible in the quantum case. That is, can **QU**: $\rho_A \to \varrho_{A'|Y=y}$ be decomposed into **QBC1**: $\rho_A \to \varrho_{A|Y=y}$, followed by belief propagation $\varrho_{A|Y=y} \to \varrho_{A'|Y=y}$? Perhaps surprisingly, this cannot be done. The belief propagation would have to have the form $\varrho_{A'|Y=y} = \mathrm{Tr}_A(\varrho_{A'|A,Y=y}\varrho_{A|Y=y})$. However, to compute $\varrho_{A'|A,Y=y}$ from $\varrho_{Y=y,A'|A}$ we need to move Y from the left of the conditional to the right while keeping A on the right. Classical analogy suggests that this could be done using a conditionalized form of the quantum Bayes' theorem, i.e., $\varrho_{A'|A,Y=y} = \varrho_{Y=y,A'|A} \star \varrho^{-1}_{Y=y|A}$. Unfortunately, in the conditional states formalism, valid equations do not necessarily remain valid when we conditionalize each term (this is discussed further in Sec. VII). In particular, the conditionalized form of the quantum Bayes' theorem is not valid. Not every intuition from classical Bayesian inference carries over into the conditional states formalism.

VI. RELATED WORK

In this section, quantum conditional states are compared to other proposals for quantum generalizations of conditional probability and the conditional states formalism is compared to several recently proposed operational reformulations of quantum theory.

A. Comparison to other quantum generalizations of conditional probability

Several quantum generalizations of conditional probability have been proposed in the literature, so it is worth comparing their relative merits to the conditional state formalism developed here.

First, Cerf and Adami have proposed an alternative definition of an acausal conditional state [17–19] (their definition does not extend to the causal case). For a bipartite state ρ_{AB}, the Cerf-Adami conditional state is

$$\rho^{(\infty)}_{B|A} = \exp(\ln \rho_{AB} - \ln \rho_A \otimes I_B). \tag{199}$$

This proposal has a close connection to the calculus of quantum entropies, since the conditional von Neumann entropy of a state ρ_{AB} can be succinctly written as

$$S(B|A) = -\mathrm{Tr}_{AB}\left(\rho_{AB}\log_2 \rho^{(\infty)}_{B|A}\right), \tag{200}$$

which is analogous to the classical formula for the conditional Shannon entropy

$$H(S|R) = -\sum_{R,S} P(R,S)\log_2 P(S|R). \tag{201}$$

Similar compact formulas hold for other information theoretic quantities, such as the quantum mutual information and conditional mutual information.

In [54], Leifer and Poulin introduced a family of conditional states, again restricted to the acausal case, indexed by a positive integer n and given by

$$\rho^{(n)}_{B|A} = \left(\rho_{AB}^{\frac{1}{n}} \star \rho_A^{-\frac{1}{n}}\right)^n. \tag{202}$$

This unifies the Cerf-Adami conditional state with the definition used in the present work in the sense that $\rho_{B|A} = \rho^{(1)}_{B|A}$ and $\rho^{(\infty)}_{B|A} = \lim_{n\to\infty}\rho^{(n)}_{B|A}$.

The main concern of [54] was the generalization of graphical models and belief propagation algorithms to quantum theory and their use in simulating many-body quantum systems and decoding quantum error correction codes. In this context, the $n = 1$ and $n \to \infty$ cases are particularly interesting. The $n \to \infty$ case is the natural one to use for simulating many-body systems and it allows for a simple generalization of one direction of the classical Hammersley-Clifford theorem, which characterizes the states on Markov networks. On the other hand, the $n = 1$ case is more useful for decoding quantum error correction codes and, as outlined in the present paper, it extends to the causal case and has close connections to quantum preparations, measurements and dynamics that are lacking for other values of n. Given that different applications work better with different definitions of the conditional state, it is probably fair to say that there is no uniquely compelling quantum generalization of conditional probability.

With this in mind, note that Coecke and Spekkens have outlined a broad framework for generalizations of conditional probability within the category theoretic approach to quantum theory [55]. Encouragingly, it is possible to derive generalizations of Bayes' theorem and conditioning abstractly within this framework, but it is not yet clear what axioms within this framework are sufficient to capture all the important aspects of conditioning.

The final quantum generalization of conditional probability to be considered here is the *quantum conditional expectation* (see [13] for the original paper and [14–16] for reviews). This was proposed in the context of quantum probability theory, which is a noncommutative generalization of classical measure-theoretic probability within the framework of operator algebras. As such, it is well defined for systems with infinite-dimensional Hilbert spaces as well as for systems with an infinite number of degrees of freedom, for which there is more than one unitarily inequivalent Hilbert space representation. Also, it describes conditioning on an arbitrary algebra of observables, rather than just on a tensor factor, as has been considered here. Importantly for the present work, Rédei has proposed an argument based on quantum conditional expectations purporting to show that quantum theory cannot be understood as a theory of Bayesian inference [56] (see also [57]).

However, quantum conditional expectations have a major flaw that is not shared by the conditional states formalism presented here. Fortunately, the full operator-algebraic machinery is not necessary to make the point; the case of finite-dimensional Hilbert spaces and conditioning on a tensor factor suffices. Further details of the general case and how the formalism used below follows from it can be found in [14].

Classically, for a pair of random variables, R and S, a conditional expectation of S given R is a positive map $\Phi_{R|S,R}$ from functions of R and S to functions of R that satisfies

$$\Phi_{R|R,S}(f(R)) = f(R) \tag{203}$$

for all functions $f(R)$ that are independent of S. Any such map is explicitly given by

$$\Phi_{R|R,S}(f(R,S)) = \sum_S P(S|R)f(R,S), \tag{204}$$

where $P(S|R)$ is a conditional probability distribution.

Starting from a joint state $P(R,S)$, one can obtain a conditional expectation by plugging the associated conditional probability $P(S|R)$ into Eq. (204). The main point of this is that it allows the expectation value of any function $f(R,S)$ to be computed from the marginal probability distribution $P(R)$ via

$$\sum_R \Phi_{R|R,S}(f(R,S))P(R) = \sum_{R,S} f(R,S)P(R,S). \tag{205}$$

The set of functions on R and S can be thought of as the dual space to the set of probability distributions on R and S, where the linear functional $\hat{f}$ associated with $f(R,S)$ is given by

$$\hat{f}(P(R,S)) = \sum_{R,S} P(R,S)f(R,S); \tag{206}$$

i.e., it is the functional that maps the probability distribution $P(R,S)$ to the expectation value of $f(R,S)$ with respect to $P(R,S)$. With respect to this identification, a conditional expectation $\Phi_{R|R,S}$ has a dual map $\mathcal{E}_{R,S|R}$ that maps the space of probability distributions over R to the space of probability distributions over R and S. This is given by

$$\mathcal{E}_{R,S|R}(P(R)) = P(S|R)P(R) \tag{207}$$

and is called a *state extension* because every probability distribution $P(R)$ gets mapped to a valid probability distribution $P(R,S) = P(S|R)P(R)$ on a larger space. In addition, state extensions that are dual to conditional expectations satisfy

$$\sum_S \mathcal{E}_{R,S|R}(Q(R)) = Q(R) \tag{208}$$

for every input distribution $Q(R)$.

In the finite-dimensional, tensor factor case, the quantum conditional expectation of B given A is a completely positive[7]

[7]The definition only calls for positivity, but it is a theorem that all quantum conditional expectations are completely positive (see [14]).

linear map $\Phi_{A|AB} : \mathfrak{L}(\mathcal{H}_{AB}) \to \mathfrak{L}(\mathcal{H}_A)$ that acts on the set of observables on $\mathcal{H}_{AB}$ and satisfies

$$\Phi_{A|AB}(M_A \otimes I_B) = M_A \tag{209}$$

for all operators $M_A \in \mathfrak{L}(\mathcal{H}_A)$. This is analogous to the condition given in Eq. (203). The dual map $\mathcal{E}_{AB|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_{AB})$ acts on states and is a state extension, which means that

$$\tau_{AB} = \mathcal{E}_{AB|A}(\tau_A) \tag{210}$$

is a valid state for any state τ_A on $\mathcal{H}_A$. In addition, the state extensions that are dual to conditional expectations satisfy

$$\mathrm{Tr}_B[\mathcal{E}_{AB|A}(\tau_A)] = \tau_A \tag{211}$$

for every state τ_A on $\mathcal{H}_A$, which is analogous to Eq. (208).

As in the classical case, one would like to associate every joint state ρ_{AB} with a conditional expectation, such that the dual state extension $\mathcal{E}_{AB|A}$ satisfies

$$\mathcal{E}_{AB|A}(\rho_A) = \rho_{AB}; \tag{212}$$

i.e., it should give back the state that you started with when you input the reduced state. The analogous requirement is a key property of classical conditional probability as it is what allows an arbitrary joint state to be broken up into a marginal and a conditional that are independent of one another. Unfortunately, the fact that Eq. (211) holds for every input state means that this requirement can only be met for product states, i.e., states of the form $\rho_{AB} = \rho_A \otimes \rho_B$. This severely restricts the applicability of quantum conditional expectations for describing the correlations present in quantum states. Indeed, they are incapable of representing any correlations at all. This fact is known in the quantum probability literature (see [14], Example 9.6), but here is an elementary proof.

Ironically, the easiest way to show that ρ_{AB} has to be a product state is to use the conditional states formalism as outlined in this paper. The state extension $\mathcal{E}_{AB|A}$ is Jamiołkowski isomorphic to a causal conditional state $\varrho_{AB|A'}$, where A' has the same Hilbert space as A and the $'$ is just used to distinguish the input and output spaces. Then, Eq. (212) can be rewritten as

$$\rho_{AB} = \mathrm{Tr}_{A'}(\varrho_{AB|A'}\rho_{A'}), \tag{213}$$

where $\rho_{A'}$ is the same state as ρ_A. Similarly, Eq. (211) can be rewritten as

$$\tau_A = \mathrm{Tr}_{A'B}(\varrho_{AB|A'}\tau_{A'}) \tag{214}$$

$$= \mathrm{Tr}_{A'}(\varrho_{A|A'}\tau_{A'}), \tag{215}$$

for all τ_A, where $\tau_{A'}$ is the same state as τ_A and $\varrho_{A|A'} = \mathrm{Tr}_B(\varrho_{AB|A'})$. Since, Eq. (215) has to hold for every input state, $\varrho_{A|A'}$ has to be Jamiołkowski isomorphic to the identity superoperator, so $\varrho_{A|A'} = |\Phi^+\rangle\langle\Phi^+|^{T_{A'}}_{A|A'}$. Since this is pure, monogamy of entanglement[8] entails that $\varrho_{AB|A'}$ must be of the form $\varrho_{AB|A'} = |\Phi^+\rangle\langle\Phi^+|^{T_{A'}}_{A|A'} \otimes M_B$ for some operator M_B on $\mathcal{H}_B$. Substituting this into Eq. (213) gives

$$\rho_{AB} = \rho_A \otimes M_B, \tag{216}$$

which shows that ρ_{AB} must be a product state and $M_B = \rho_B = \mathrm{Tr}_A(\rho_{AB})$. Therefore, conditional expectations associated with joint states only exist for product states $\rho_{AB} = \rho_A \otimes \rho_B$.

It is unclear why quantum probabilists have not regarded this as a fatal flaw in their definition of quantum conditional expectation. However, despite this problem, quantum conditional expectations are still worthy objects of study as they come up in a variety of contexts. For example, the projection onto the fixed point set of a completely positive map is a quantum conditional expectation. The point is just that the terminology "quantum conditional expectation" is an inaccurate way of describing the way that these maps relate to quantum states.

For comparison, the acausal conditional state defined in the present work can also be described as a map $\mathcal{F}_{AB|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_{AB})$, similar to a state extension, with the crucial difference that $\mathcal{F}_{AB|A}$ is not linear. For a conditional state $\rho_{B|A}$, the map $\mathcal{F}_{AB|A}$ is defined as

$$\mathcal{F}_{AB|A}(\rho_A) = \rho_A^{\frac{1}{2}}\rho_{B|A}\rho_A^{\frac{1}{2}}. \tag{217}$$

The nonlinearity allows this map to satisfy Eqs. (211) and (212) without running into trouble with the monogamy of entanglement. Crucially though, because this map is nonlinear, it does not have a dual map that could be regarded as a conditional expectation. Therefore, the close connection between conditional probabilities and conditional expectations breaks down within this formalism.

Rédei's argument against a Bayesian interpretation of quantum probabilities is based on a variant of Jeffrey conditioning with respect to a quantum conditional expectation. In the present context, if the state of a region A is updated from ρ_A to ρ_A^{post}, then Rédei asserts that the state for AB should be updated to $\rho_{AB}^{\mathrm{post}} = \mathcal{E}_{AB|A}(\rho_A^{\mathrm{post}})$, where $\mathcal{E}_{AB|A}$ is a state extension derived from the prior state ρ_{AB}. He then shows that this update rule fails to satisfy an important stability criterion in the infinite-dimensional case. Since his argument is crucially based on the linearity of conditional expectations and their duality with state extensions, it does not apply to the nonlinear maps $\mathcal{F}_{AB|A}$ associated with quantum conditional states. However, since the failure only occurs for infinite-dimensional systems, a full refutation will have to wait until the conditional states formalism has been extended beyond the finite-dimensional case treated here.

To reiterate, it seems unlikely that there is a unique quantum generalization of conditional probability that has properties analogous to every single property of classical conditional probability that is traditionally regarded as important. For this reason, it is important to keep applications in mind when defining quantum conditionals, rather than working in a formal mathematical vacuum.

B. Operational formalisms for quantum theory

Recent efforts to replace the conventional formalism of quantum theory with a new operational formulation—typically in an effort to provide an axiomatic derivation of quantum

[8]The monogamy of entanglement is well known for positive operators. The fact that it also applies to locally positive operators follows one of the results of [58], which shows that monogamy applies to more general probabilistic theories, including one in which states are locally positive operators.

theory—have much in common with the work presented here. Particularly cognate to our approach is the work of Hardy [59–61], the Pavia group [62,63], Oreshkov *et al.* [64], and Coecke's group [65,66].

The reformulation of quantum theory presented by the Pavia group makes heavy use of the Choi isomorphism between quantum operations and bipartite states and leverages this to represent quantum operations by operators rather than maps. Mathematically, this is also how we achieve a unification of the treatment of acausally related and causally related regions. In particular, our Proposition 1, which specifies how to represent a quantum channel in terms of conditional states, is the counterpart to the expression of the action of a quantum channel in terms of the *link product*. Note that the Pavia group uses the Choi isomorphism, whereas we use the Jamiołkowski-isomorphism. The latter has the advantage of being basis independent, so that the partial transposes that appear in the link product are absent in our approach.

Hardy's latest work on reformulating quantum theory, using the *duotensor framework* [61], also represents quantum operations as operators in much the same way as is done in the quantum combs framework and our own. Furthermore, Hardy's notion of *circuit trace* (an example of the causaloid product introduced in [59,61]) provides a unified way of representing a composition of maps as well as tensor products of system states, which is to say a unified way of representing correlations between acausally related and causally related regions. The motivation for Hardy's work on the causaloid product is very similar to the motivation for our own, namely, to formulate quantum theory in a manner that is even handed with regard to possible causal structures.

Recent work by Oreshkov, Costa, and Brukner [64] also represents correlations between acausally related and causally related regions is a uniform manner by appealing to the Choi isomorphism.

One notable way in which we depart from all of these approaches is that we treat classical systems internally to the formalism, on a par with quantum systems, rather than as indices on operators representing preparations and measurements.

Finally, we compare our approach to the categorical approach of Coecke, where much of quantum theory (in particular the nonmetrical parts) is reformulated using the mathematical framework of symmetric monoidal categories and the graphical calculus that can be defined for these [65,66]. Systems are the objects of the category, while quantum states and quantum operations are the morphisms. The isomorphism between bipartite states and operations also features prominently in this approach and arises from having a *compact structure* in the category. Furthermore, classical systems can be treated internally within this framework. It should also be noted that although Coecke's categorical framework is typically used to formulate quantum theory as a theory of physical processes, it can also be used to express our formulation of quantum theory as a theory of inference. For inference among acausally related regions, this was done in [55]. An extension of this work to the case of inference among causally related regions should be instructive.

The mathematics of all of these approaches and our own are quite analogous. It is in the interpretational aspect that the greatest differences are to be found. In the reformulations considered above, quantum theory is given a rather minimalist interpretation; it is viewed as a framework for making predictions about the outcomes of certain measurements given certain preparations. Quantum states—however they are reformulated mathematically—are taken to be representations of preparation procedures, while quantum operations are taken to be representations of transformation procedures. These approaches follow the interpretational tradition of operationalism. By contrast, our work takes quantum states to represent the beliefs of an agent about a spatiotemporal region and takes quantum operations to represent belief propagation; it has an *epistemological* flavor rather than an *operational* one. For instance, the notions that we deem to be most promising for making sense of the quantum formalism are those one finds in textbooks on statistics and inductive inference, such as Bayes' theorem, conditional probabilities, statistical independence, conditional independence, and sufficient statistics and not the notions that are common to the operational approaches, such as measurements, transformations, and preparations. In this sense, our approach is more closely aligned in its philosophical starting point with *quantum Bayesianism*, the view developed by Caves, Fuchs, and Schack [24,67–70].[9]

The particular merit of our epistemological approach is the strong analogy that it affords between quantum inference and classical probabilistic inference. It makes the conceptual content of various quantum expressions more transparent than they would otherwise be. It also bolsters the view that quantum states ought to be interpreted as states of knowledge.

Our take on how to incorporate causal assumptions into quantum theory also has a rather different starting point than the works described above. The relation we posit between the notions of causation and correlation is most informed by the work on causal networks (also known as Bayesian networks), summarized in the textbooks of Pearl [71] and of Glymour, Scheines, and Spirtes [72]. We ultimately hope to generalize this analysis of causality by replacing classical probability theory with quantum theory, understood as a theory of Bayesian inference [54]. If the goal is to develop a formalism for quantum theory that is causally neutral, then we argue that the causal network approach holds an advantage over the operational approach. Specifically, the problem with taking experimental operations as a primitive notion is that they already have a notion of causal structure built into them insofar as the output of an operation is causally influenced by the input, but not vice versa (similarly, if the operation is a measurement, then the outcome is causally influenced by the input but not vice versa). On the other hand, elementary regions—the primitive notion of the causal network approach—are precisely the sorts of thing that can enter into arbitrary causal relations with one another, while not having any intrinsic causal structure themselves.

[9]Unlike the quantum Bayesians, however, we are not committed to the notion that the beliefs represented by quantum states concern the outcomes of future experiments. Rather, the picture we have in mind is of the quantum state for a region representing beliefs about the physical state of the region, even though we do not yet have a model to propose for the underlying physical states.

VII. LIMITATIONS OF THE $\star$ PRODUCT

Using the $\star$ product makes the conditional states formalism look very similar to classical probability theory. Often, by replacing probabilities with operators and ordinary products with $\star$ products, one can obtain an equation that is valid in the conditional states formalism from one that is valid for classical conditional probabilities. However, because the $\star$ product is nonassociative and noncommutative, this does not always happen. In this section, the limitations of the $\star$ product and the disanalogies between classical conditional probabilities and quantum conditional states are discussed.

A. Conditionalized equations

In classical probability theory, if one takes a universally valid[10] equation relating conditional, joint, and marginal probability distributions over a set of variables $R,S,\ldots$ and conditionalizes every term on a disjoint variable T, then the equation that results is still universally valid. For example, the equation

$$P(R,S) = P(S|R)P(R) \tag{218}$$

generalizes to

$$P(R,S|T) = P(S|R,T)P(R|T), \tag{219}$$

where a T has been placed on the right of the | in each term.

Unfortunately, the analogous property does not hold in the conditional states formalism. For example, for acausal states the analog of Eq. (219) would be

$$\rho_{AB|C} = \rho_{B|AC} \star \rho_{A|C}. \tag{220}$$

Writing this out explicitly, the left-hand side is

$$\rho_C^{-\frac{1}{2}} \rho_{ABC} \rho_C^{-\frac{1}{2}}, \tag{221}$$

whereas the right-hand side is

$$\left(\rho_C^{-\frac{1}{2}} \rho_{AC} \rho_C^{-\frac{1}{2}}\right)^{\frac{1}{2}} \rho_{AC}^{-\frac{1}{2}} \rho_{ABC} \rho_{AC}^{-\frac{1}{2}} \left(\rho_C^{-\frac{1}{2}} \rho_{AC} \rho_C^{-\frac{1}{2}}\right)^{\frac{1}{2}}. \tag{222}$$

Because the $\star$ product is noncommutative, the terms involving ρ_{AC} and ρ_{AC}^{-1} cannot be brought together and made to cancel as they would in the classical case. Therefore, counterexamples to this rule can occur when ρ_{AC} and ρ_C do not commute. For example, it is straightforward to verify that a generalized W state of the form

$$|\psi\rangle_{ABC} = \frac{1}{2}(|001\rangle_{ABC} + |010\rangle_{ABC}) + \frac{1}{\sqrt{2}}|100\rangle_{ABC} \tag{223}$$

does not satisfy this rule. The calculation is not especially instructive, so it is omitted.

Note that a universally valid equation relating quantum conditional, joint, and marginal states is still universally valid if one conditionalizes every term on a classical variable. This follows from the equality of the expressions for the left-hand and right-hand sides of Eq. (220) when C is replaced with a classical variable T.

B. Limitations of causal joint states

In Sec. III E, a causal joint state was defined as an operator of the form $\varrho_{B|A} \star \rho_A$. This representation of two causally related regions highlights the symmetry with the acausal case, since, up to a partial transpose, it is the same sort of operator that would be used to represent two acausally related regions. Based on this, the quantum Bayes' theorem was developed in a way that is formally equivalent for acausally and causally related regions. For this to work, we only needed causal joint states for two causally related regions. However, since acausal states are not limited to just two regions, it is natural to ask whether causal joint states can be defined for more than two regions. Unfortunately, the naive generalization does not work for scenarios with mixed causal structure, e.g., two causally related regions that are both acausally related to a third regions, and it also does not work for multiple time steps.

In the remainder of this section, these limitations are discussed and a different definition of a causal joint state is suggested, which works more generally, but does not exhibit the symmetry between the acausal and causal scenarios for two regions.

1. *Mixed causal scenarios*

If region B is in the causal future of region A, then they can be assigned a causal joint state $\varrho_{AB} = \varrho_{B|A} \star \rho_A$. The causal conditional state is Jamiołkowski isomorphic to the dynamical CPT map $\mathcal{E}_{B|A}$ and ρ_A is the input state. Now suppose that there is a third region, C, that is acausally related to both A and B; i.e., we start with a joint state ρ_{AC} of A and C and apply $\mathcal{E}_{B|A}$ to region A, while doing nothing to C, to obtain B and C in a joint state $\rho_{BC} = (\mathcal{E}_{B|A} \otimes \mathcal{I}_C)(\rho_{AC})$.

One might think that a joint state of ABC could be defined via $\varrho_{ABC} = \rho_{B|A} \star \rho_{AC}$. We would like this state to have the correct input and output states as marginals, so, in particular, it should satisfy

$$\rho_{BC} = \mathcal{E}_{B|A}(\rho_{AC}) = \mathrm{Tr}_A(\varrho_{ABC}). \tag{224}$$

Unfortunately, this fails because Theorem 1 implies that

$$\rho_{BC} = \mathrm{Tr}_A(\varrho_{B|A}\rho_{AC}), \tag{225}$$

whereas

$$\mathrm{Tr}_A(\varrho_{ABC}) = \mathrm{Tr}_A(\varrho_{B|A} \star \rho_{AC}). \tag{226}$$

Because there is no trace over C, the cyclic property of the trace cannot be used to equate these two expressions.

In fact, in addition to not having ρ_{BC} as a reduced state, ϱ_{ABC} fails to correctly represent the correlations between the causally related regions A and B as well, i.e., $\mathrm{Tr}_C(\varrho_{ABC}) \neq \varrho_{AB}$.

To see the failure of both these conditions explicitly, consider an example in which $\mathcal{H}_A$, $\mathcal{H}_B$, and $\mathcal{H}_C$ are all of dimension d, the input state $\rho_{AC} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{AC}$ is a maximally entangled state, and $\mathcal{E}_{B|A}$ is the identity superoperator. Then, the output state is $\rho_{BC} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{BC}$, and the causal joint state is $\varrho_{AB} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{AB}^{T_A}$.

[10]Universal validity means that the equation holds for the joint, marginal, and conditional probability distributions derived from *any* joint probability distribution $P(R,S,\ldots)$, rather than holding only in special cases, e.g., due to symmetries or degeneracies of a particular distribution.

In this case, explicitly calculating the operator $\varrho_{ABC} = \varrho_{B|A} \star \rho_{AC}$ gives

$$\varrho_{ABC} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{AC} \otimes \frac{I_B}{d}. \tag{227}$$

While the reduced state ρ_{AC} gives the correct input state, ρ_{BC} is not the output state and ϱ_{AB} is not the causal joint state. In fact, no operator of the form $\varrho_{ABC} = \varrho_{B|A} \star \rho_{AC}$ can ever satisfy all three conditions $\rho_{AC} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{AC}$, $\varrho_{AB} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{AB}^{T_A}$ and $\rho_{BC} = \frac{1}{d}|\Phi^+\rangle\langle\Phi^+|_{BC}$ simultaneously. This is because ϱ_{ABC} is a locally positive operator, and hence it must satisfy the monogamy of entanglement, but requiring all three bipartite reduced states to be maximally entangled would violate monogamy and, in fact, only one of the three conditions can be satisfied. Because of this, we have to move beyond locally positive operators in order to faithfully represent all the correlations.

The validity of Eq. (225) suggests that using an ordinary product instead of a $\star$ product could be a better way of defining a joint state in this scenario. Indeed, the operator $\varrho_{B|A}\rho_{AC}$ does have all the correct bipartite marginals. For example, in our example of maximally entangled input and identity map dynamics, explicit calculation gives

$$\varrho_{B|A}\rho_{AC} = \frac{1}{d}\sum_{j,k,m=1}^{d} |j\rangle\langle k|_A \otimes |m\rangle\langle j|_B \otimes |m\rangle\langle k|_C. \tag{228}$$

This violates the monogamy of entanglement, which is allowed because it is not a locally positive operator. In fact, it is not even Hermitian.

While this may turn out to be a useful representation, it has a number of disadvantages compared to the $\star$ product. First of all, it is nonunique because the operator $\rho_{AC}\varrho_{B|A}$ gives an equally good account of all the correlations. One could even take combinations of the two operators, such as

$$\tfrac{1}{2}(\varrho_{B|A}\rho_{AC} + \rho_{AC}\varrho_{B|A}), \tag{229}$$

which might be useful because it is Hermitian.

Second, given an arbitrary operator M_{AB}, it is not clear how to check whether it is of the form $\varrho_{B|A}\rho_A$ without running over all possible input states ρ_A and checking whether $M_{AB}\rho_A^{-1}$ is a valid causal conditional state. In contrast, when using the $\star$ product, we know that an operator is of the form $\varrho_{B|A} \star \rho_A$ if and only if it is the partial transpose of a valid acausal state on AB, which is a straightforward condition to check.

Finally, a related issue is that, when using the ordinary product instead of the $\star$ product, the set of possible causal joint states depends on the causal direction. For an evolution from A to B, a causal joint state would be of the form $\varrho_{B|A}\rho_A$, but for an evolution from B to A it would be of the form $\varrho_{A|B}\rho_B$. These define two different sets of operators, so we lose the symmetry that was used to obtain Bayes' theorem in the causal case.

2. *Multiple time steps*

For three spacelike separated regions, a Markovian joint state ρ_{ABC}, where A and C are conditionally independent given B, can always be decomposed via the chain rule into

$$\rho_{ABC} = \rho_{C|B} \star (\rho_{B|A} \star \rho_A). \tag{230}$$

For three timelike separated regions, the analog of Markovianity is a two time-step dynamics where the first CPT map is $\mathcal{E}_{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ and the second is $\mathcal{E}_{C|B} : \mathfrak{L}(\mathcal{H}_B) \to \mathfrak{L}(\mathcal{H}_C)$; i.e., it has no direct dependence on A. It is natural to ask whether this situation can be represented by a tripartite causal joint state that can be decomposed in a manner similar to Eq. (230).

Suppose that the causal conditional states isomorphic to $\mathcal{E}_{B|A}$ and $\mathcal{E}_{C|B}$ are $\varrho_{B|A}$ and $\varrho_{C|B}$. If the input state is ρ_A then the first time step is represented as

$$\rho_B = \mathcal{E}_{B|A}(\rho_A) \tag{231}$$

$$= \mathrm{Tr}_A(\varrho_{B|A}\rho_A) \tag{232}$$

and the second time step is represented as

$$\rho_C = \mathcal{E}_{C|B}(\rho_B) \tag{233}$$

$$= \mathrm{Tr}_B(\varrho_{C|B}\rho_B). \tag{234}$$

It follows that

$$\rho_C = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}(\rho_A) \tag{235}$$

$$= \mathrm{Tr}_B[\varrho_{C|B}\mathrm{Tr}_A(\varrho_{B|A}\rho_A)], \tag{236}$$

The question is whether the complete dynamics might also be representable as

$$\rho_C = \mathrm{Tr}_{AB}[\varrho_{C|B} \star (\varrho_{B|A} \star \rho_A)], \tag{237}$$

which in turn would suggest that $\varrho_{ABC} = \varrho_{C|B} \star (\varrho_{B|A} \star \rho_A)$ might be a good candidate for a tripartite causal joint state.

This fails because the expression $\varrho_{ABC} = \varrho_{C|B} \star (\varrho_{B|A} \star \rho_A)$ is not well defined. To see this, expand the first $\star$ product to obtain

$$\varrho_{ABC} = \sqrt{\varrho_{B|A} \star \rho_A}\,\varrho_{C|B}\sqrt{\varrho_{B|A} \star \rho_A}. \tag{238}$$

The term $\sqrt{\varrho_{B|A} \star \rho_A}$ is not well defined because $\varrho_{B|A} \star \rho_A$ is not a positive operator, but only locally positive, so it may have negative eigenvalues. Hence, it does not have a unique square root. This could be remedied by adopting a convention for square roots of Hermitian operators, such as demanding that the square root of each negative eigenvalue has positive imaginary part. The resulting tripartite operator ϱ_{ABC} would then have the correct reduced states, ρ_A, ρ_B, and ρ_C, representing the state of the system at each time step. However, it is not related to a tripartite state of three acausally related regions via partial transposes, so the symmetry that motivates the use of the $\star$ product representation is lost. This loss of symmetry resonates with previous work showing that tripartite "entanglement in time" is not isomorphic to ordinary tripartite entanglement [2].

As with the evolution of a subsystem, these problems can be remedied by using the ordinary product instead of the $\star$ product to represent time evolutions, but it is subject to the same disadvantages that were discussed in that context.

VIII. OPEN QUESTIONS

A. The quantum conditionals problem

Monogamy of entanglement is a key feature that distinguishes classical from quantum information. There is a closely related difference between classical conditional probability distributions and acausal quantum conditional states that deserves further investigation. Classically, if $P(R)$ is a probability distribution and $P(S|R)$ and $P(T|R,S)$ are conditional probability distributions, then

$$P(R,S,T) = P(T|R,S)P(S|R)P(R) \tag{239}$$

is always a valid probability distribution. Furthermore, the distribution so defined has the correct marginal and conditional states in the sense that

$$\sum_{S,T} P(R,S,T) = P(R), \tag{240}$$

$$\frac{\sum_T P(R,S,T)}{\sum_{S,T} P(R,S,T)} = P(S|R), \tag{241}$$

$$\frac{P(R,S,T)}{\sum_T P(R,S,T)} = P(T|R,S), \tag{242}$$

whenever the left-hand sides are well defined.

In the quantum case, the analogous properties do not hold. Although a tripartite state ρ_{ABC} can always be decomposed as

$$\rho_{ABC} = \rho_{C|AB} \star \rho_{B|A} \star \rho_A \tag{243}$$

via the chain rule, one cannot start with an arbitrary reduced state ρ_A and two arbitrary conditional states, $\rho_{B|A}$ and $\rho_{C|AB}$, and expect there to be a joint state ρ_{ABC} that has these conditional and reduced states.

For example, suppose that B and C are conditionally independent of A, i.e., $\rho_{C|AB} = \rho_{C|A}$. Now, suppose that ρ_A is chosen to have more than one nonzero eigenvalue and $\rho_{B|A}$ is chosen to be maximally entangled, e.g., $\rho_{B|A} = |\Phi^+\rangle\langle\Phi^+|_{B|A}$. This implies that the reduced state $\rho_{AB} = \rho_{B|A} \star \rho_A$ is pure and entangled. If, in addition, $\rho_{C|A}$ is chosen to be $\rho_{C|A} = |\Phi^+\rangle\langle\Phi^+|_{C|A}$, then the reduced state $\rho_{AC} = \rho_{C|A} \star \rho_A$ should also be pure and entangled. However, monogamy of entanglement says that this is impossible, so these choices of conditional state are not compatible. Determining the full set of constraints on coexistent conditional states for three acausally related regions would be an interesting problem, as would determining the computational complexity of the n-party generalization.

IX. CONCLUSIONS

The formalism of quantum conditional states presented in this paper provides a step towards a formalism for quantum theory that is independent of causal structure, as a theory of probabilistic inference ought to be, and provides a closer analogy between quantum theory and classical probability theory. There is significant potential to use these results to simplify and generalize existing approaches to problems in quantum information theory. As an example of this, in a companion paper [73] we provide an approach to the problems of compatibility and pooling of quantum states that is based on a principled application of Bayesian conditioning and is a direct generalization of existing approaches to the classical versions of these problems. It seems unlikely that the possibility of this approach would have been noticed within the conventional quantum formalism. However, this is only the beginning and we anticipate applications to quantum estimation theory and to quantum cryptography, such as studying the relationship between cryptography protocols that employ different causal arrangements to achieve the same task. As it stands, the formalism is limited to two disjoint elementary quantum regions and the most pressing problem is to generalize it to arbitrary causal scenarios. This is a topic of ongoing work.

ACKNOWLEDGMENTS

We thank David Poulin for useful discussions about conditional states. M.L. would like to thank the UCL Department of Physics and Astronomy for its hospitality. Part of this work was completed while M.L. was a postdoctoral fellow at the Institute for Quantum Computing, University of Waterloo, Canada, where he was supported in part by MITACS and ORDCF. M.L. also acknowledges support from The Foundational Questions Institute through Grant No. RFP1-06-006. Research at the Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

APPENDIX: PROOFS OF THEOREMS

Theorem 1: Jamiołkowski isomorphism. Let $\mathfrak{E}_{B|A} : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_B)$ be a linear map and let $M_{AC} \in \mathfrak{L}(\mathcal{H}_{AC})$ be a linear operator, where $\mathcal{H}_C$ is a Hilbert space of arbitrary dimension. Then the action of $\mathfrak{E}_{B|A}$ on $\mathfrak{L}(\mathcal{H}_A)$ [tensored with the identity on $\mathfrak{L}(\mathcal{H}_C)$] is given by

$$(\mathfrak{E}_{B|A} \otimes \mathcal{I}_C)(M_{AC}) = \mathrm{Tr}_A(\rho_{B|A} M_{AC}), \tag{A1}$$

where $\rho_{B|A} \in \mathfrak{L}(\mathcal{H}_{AB})$ is given by

$$\rho_{B|A} \equiv (\mathfrak{E}_{B|A'} \otimes \mathcal{I}_A)\left(\sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}\right). \tag{A2}$$

Here A' labels a second copy of A, $\mathcal{I}_A$ is the identity superoperator on $\mathfrak{L}(\mathcal{H}_A)$, and $\{|j\rangle\}$ is an orthonormal basis for $\mathcal{H}_A$.

Furthermore, the operator $\rho_{B|A}$ is an acausal conditional state; i.e., it satisfies Definition 1, if and only if $\mathfrak{E}_{B|A} \circ T_A$ is CPT, where $T_A : \mathfrak{L}(\mathcal{H}_A) \to \mathfrak{L}(\mathcal{H}_A)$ denotes the linear map implementing the partial transpose relative to some basis.

To prove this theorem, it is useful to make use of the connection between the Choi and the Jamiołkowski isomorphisms. The map that is Choi-isomorphic to an operator $\rho_{B|A}$ is given by

$$(\mathcal{E}_{B|A} \otimes \mathcal{I}_C)(M_{AC}) = \langle\Phi^+|_{AA'} \rho_{B|A'} M_{AC} |\Phi^+\rangle_{AA'}, \tag{A3}$$

where $|\Phi^+\rangle_{AA'} = \sum_j |jj\rangle_{AA'}$ is a canonical maximally entangled state defined with respect to a preferred basis $\{|j\rangle\}$ for $\mathcal{H}_A$.

The operator is recovered from the map via

$$\rho_{B|A} \equiv (\mathcal{E}_{B|A'} \otimes \mathcal{I}_A)(|\Phi^+\rangle\langle\Phi^+|_{AA'}). \tag{A4}$$

Because

$$\sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'} = (|\Phi^+\rangle\langle\Phi^+|_{AA'})^{T_A}, \tag{A5}$$

Eqs. (A4) and (A2) differ only by whether one uses the projector onto the maximally entangled state (Choi) or the partial transpose thereof (Jamiołkowski) and the two isomorphic maps to $\rho_{B|A}$ are related by $\mathfrak{E}_{B|A} = \mathcal{E}_{B|A} \circ T_A$, where T_A is the partial transpose operation with respect to the basis used to define the Choi isomorphism.

The equivalence of Eqs. (A3) and (A1) is established as follows:

$$\begin{aligned}
&(\mathcal{E}_{B|A} \otimes \mathcal{I}_C)(M_{AC}) \\
&= \langle\Phi^+|_{AA'}\rho_{B|A'}M_{AC}|\Phi^+\rangle_{AA'} \\
&= \sum_{j,k}\langle jj|_{AA'}\rho_{B|A'}M_{AC}|kk\rangle_{AA'} \\
&= \sum_{j,k}\langle j|_{A'}\rho_{B|A'}|k\rangle_{A'}\langle j|_A M_{AC}|k\rangle_A \\
&= \sum_{j,k}\langle j|_A\rho_{B|A}|k\rangle_A\langle k|_A M_{AC}^{T_A}|j\rangle_A \\
&= \mathrm{Tr}_A\big(\rho_{B|A}M_{AC}^{T_A}\big) \\
&= ([\mathfrak{E}_{B|A} \circ T_A] \otimes \mathcal{I}_C)(M_{AC}).
\end{aligned} \tag{A6}$$

Proof of Theorem 1. Equation (A1) is derived from Eq. (A2) as follows:

$$\begin{aligned}
&(\mathfrak{E}_{B|A} \otimes \mathcal{I}_C)(M_{AC}) \\
&= (\mathfrak{E}_{B|A'} \otimes \mathcal{I}_C)\left(\left[\sum_k |k\rangle\langle k|_{A'}\right]M_{A'C}\left[\sum_j |j\rangle\langle j|_{A'}\right]\right) \\
&= (\mathfrak{E}_{B|A'} \otimes \mathcal{I}_C)\left(\sum_{j,k}\langle k|_{A'}M_{A'C}|j\rangle_{A'}|k\rangle\langle j|_{A'}\right) \\
&= (\mathfrak{E}_{B|A'} \otimes \mathcal{I}_C)\left(\sum_{j,k}\mathrm{Tr}_A(|j\rangle\langle k|_A M_{AC})|k\rangle\langle j|_{A'}\right) \\
&= \mathrm{Tr}_A\left[(\mathfrak{E}_{B|A'} \otimes \mathcal{I}_C)\left(\sum_{j,k}|j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}\right)M_{AC}\right] \\
&= \mathrm{Tr}_A(\rho_{B|A}M_{AC}).
\end{aligned} \tag{A7}$$

Now suppose that $\rho_{B|A}$ is an acausal conditional state; i.e., it is positive and $\mathrm{Tr}_B(\rho_{B|A}) = I_A$. To show that the Jamiołkowski-isomorphic map composed with a partial transpose, $\mathfrak{E}_{B|A} \circ T_A$, is trace preserving, note that T_A is trace preserving, so it suffices to show that $\mathfrak{E}_{B|A}$ is trace preserving. This proceeds as follows:

$$\mathrm{Tr}_B[\mathfrak{E}_{B|A}(M_A)] = \mathrm{Tr}_B[\mathrm{Tr}_A(\rho_{B|A}M_A)] \tag{A8}$$

$$= \mathrm{Tr}_A[\mathrm{Tr}_B(\rho_{B|A})M_A] \tag{A9}$$

$$= \mathrm{Tr}_A(I_A M_A) \tag{A10}$$

$$= \mathrm{Tr}_A(M_A). \tag{A11}$$

To show that $\mathfrak{E}_{B|A} \circ T_A$ is completely positive, note that it is equal to the Choi-isomorphic map $\mathcal{E}_{B|A}$, so it suffices to show that the latter is completely positive. By definition,

$$\mathcal{E}_{B|A} \otimes \mathcal{I}_C(\rho_{AC}) = \langle\Phi^+|_{AA'}\rho_{B|A'} \otimes \rho_{AC}|\Phi^+\rangle_{AA'} \tag{A12}$$

and this is a positive operator for arbitrary positive ρ_{AC}, where $\mathcal{H}_C$ can have any dimension.

Conversely, suppose $\mathfrak{E}_{B|A} \circ T_A$ is CPT. Then $\mathfrak{E}_{B|A}$ is also trace preserving, so

$$\mathrm{Tr}_B(\rho_{B|A}) = \mathrm{Tr}_B\left[\sum_{j,k}|j\rangle\langle k|_A \otimes \mathfrak{E}_{B|A}(|k\rangle\langle j|_{A'})\right] \tag{A13}$$

$$= \sum_{j,k}|j\rangle\langle k| \otimes \mathrm{Tr}_B[\mathfrak{E}_{B|A}(|k\rangle\langle j|_{A'})] \tag{A14}$$

$$= \sum_{j,k}|j\rangle\langle k|_A \otimes \mathrm{Tr}_{A'}(|k\rangle\langle j|_{A'}) \tag{A15}$$

$$= \sum_{j,k}|j\rangle\langle k|_A\delta_{j,k} \tag{A16}$$

$$= \sum_j |j\rangle\langle j|_A \tag{A17}$$

$$= I_A. \tag{A18}$$

Also,

$$\rho_{B|A} = (\mathcal{E}_{B|A'} \otimes \mathcal{I}_A)(|\Phi^+\rangle\langle\Phi^+|_{AA'}), \tag{A19}$$

and this is a CPT map acting on a positive operator, so $\rho_{B|A}$ is positive. ■

Theorem 3. Let $\sigma_{A|X}$ be a hybrid operator, so that by Eq. (40) it can be written as

$$\sigma_{A|X} = \sum_x \rho_x^A \otimes |x\rangle\langle x|_X \tag{A20}$$

for some set of operators $\{\rho_x^A\}$. Then, $\sigma_{A|X}$ satisfies the definition of both an acausal and a causal conditional state for A given X, if and only if each of the components ρ_x^A is a normalized state on $\mathcal{H}_A$.

Proof. Suppose $\sigma_{A|X}$ has the form of Eq. (A20) for a set of normalized states $\{\rho_x^A\}$. Then it is clearly positive and satisfies $\mathrm{Tr}_A(\sigma_{A|X}) = I_X$ because $\mathrm{Tr}_A(\rho_x^A) = 1$ for every x. Therefore, $\sigma_{A|X}$ is an acausal conditional state. On the other hand, $\sigma_{A|X}$ is invariant under partial transpose on X, so it is also a causal conditional state.

Conversely, suppose that $\sigma_{A|X}$ is an acausal conditional state. This means that it is positive and satisfies $\mathrm{Tr}_A(\sigma_{A|X}) = I_X$. Positivity means that $\langle\psi|_{AX}\sigma_{A|X}|\psi\rangle_{AX} \geqslant 0$ for all $|\psi\rangle_{AX} \in \mathcal{H}_{AX}$. If $\sigma_{A|X}$ has the form of Eq. (A20), then taking $|\psi\rangle_{AX} = |\phi\rangle_A \otimes |x\rangle_A$ gives $\langle\phi|_A\rho_x^A|\phi\rangle_A \geqslant 0$. By varying over all $|\phi\rangle_A \in \mathcal{H}_A$, this implies that each ρ_x^A is a positive operator. To prove that these operators are normalized, note that

$$\mathrm{Tr}_A(\sigma_{A|X}) = \sum_x \mathrm{Tr}_A\big(\rho_x^A\big)|x\rangle\langle x|_X. \tag{A21}$$

This is an eigendecomposition of $\mathrm{Tr}_A(\sigma_{A|X})$ with eigenvalues $\mathrm{Tr}_A(\rho_x^A)$ and if this is the identity operator then each of these eigenvalues must be 1.

On the other hand, if $\sigma_{A|X}$ is a causal conditional state, then its partial transpose over X must be positive and satisfy $\mathrm{Tr}_A(\varrho_{A|X}^{T_X}) = I_X$. However, operators of the form of Eq. (A20) are invariant under partial transpose on X, so the same argument applies. ■

Theorem 4. Let $\sigma_{Y|A}$ be a hybrid operator so that it can be written in the form

$$\sigma_{Y|A} = \sum_y |y\rangle\langle y|_Y \otimes E_y^A \tag{A22}$$

for some set of operators $\{E_y^A\}$. Then $\sigma_{Y|A}$ satisfies the definition of both an acausal and a causal conditional state for Y given A if and only if the components E_y^A form a POVM on $\mathcal{H}_A$; i.e., each E_y^A is positive and $\sum_y E_y^A = I_A$.

Proof. Suppose $\sigma_{Y|A}$ has the form of Eq. (A22) for a POVM $\{E_y^A\}$. Then it is clearly positive and satisfies $\mathrm{Tr}_Y(\sigma_{Y|A}) = \sum_y E_y^A = I_A$. Therefore, $\sigma_{Y|A}$ is an acausal conditional state. On the other hand,

$$\sigma_{Y|A}^{T_A} = \sum_y |y\rangle\langle y|_Y \otimes \left(E_y^A\right)^{T_A} \tag{A23}$$

is also positive because the positive operators E_y^A remain positive under the transpose. Also, $\mathrm{Tr}_Y(\sigma_{Y|A}^{T_A}) = \sum_y (E_y^A)^{T_A} = I_A^{T_A} = I_A$. Therefore, $\sigma_{Y|A}$ is also a causal conditional state.

Conversely, suppose that $\sigma_{Y|A}$ is an acausal conditional state. This means that it is positive and satisfies $\mathrm{Tr}_Y(\sigma_{Y|A}) = I_A$. By the same argument used in the proof of Theorem 3, positivity implies that, if $\sigma_{Y|A}$ is of the form of Eq. (A22), then each of the components E_y^A must be positive. Since $\mathrm{Tr}_Y(\sigma_{Y|A}) = \sum_y E_y^A$, the components must form a POVM.

On the other hand, if $\sigma_{Y|A}$ is an acausal conditional state then, by the argument just given, its partial transpose over A must be of the form of Eq. (A22) for some POVM $\{E_y^A\}$. This means that $\sigma_{Y|A}$ itself can be written as

$$\sigma_{Y|A} = \sum_y |y\rangle\langle y|_Y \otimes \left(E_y^A\right)^{T_A}, \tag{A24}$$

but since the operators $(E_y^A)^{T_A}$ form a POVM whenever $\{E_y^A\}$ is a POVM, $\sigma_{Y|A}$ is of the required form. ■

Theorem 5. Let $\mathcal{E}_{B|A}$, $\mathcal{E}_{C|B}$, and $\mathcal{E}_{C|A}$ be linear maps such that $\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}$. Then the Jamiołkowski isomorphic operators, $\varrho_{B|A}$, $\varrho_{C|B}$, and $\varrho_{C|A}$, satisfy

$$\varrho_{C|A} = \mathrm{Tr}_B(\varrho_{C|B}\varrho_{B|A}). \tag{A25}$$

Conversely, if three operators satisfy Eq. (A25), then the Jamiołkowski isomorphic maps satisfy $\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}$.

Proof. By definition, the Jamiołkowski isomorphic operator to $\mathcal{E}_{C|A}$ is

$$\varrho_{C|A} = (\mathcal{E}_{C|A'} \otimes \mathcal{I}_A)\left(\sum_{j,k} |j\rangle\langle k|_A \otimes |k\rangle\langle j|_{A'}\right) \tag{A26}$$

$$= \sum_{j,k} |j\rangle\langle k| \otimes [\mathcal{E}_{C|B} \circ \mathcal{E}_{B|A'}(|k\rangle\langle j|_{A'})]. \tag{A27}$$

Applying Theorem 1 to $\mathcal{E}_{B|A'}$ gives

$$\varrho_{C|A} = \sum_{j,k} |j\rangle\langle k|_A \otimes \mathcal{E}_{C|B}[\mathrm{Tr}_{A'}(\varrho_{B|A'}|k\rangle\langle j|_{A'})]$$
$$= \sum_{j,k} |j\rangle\langle k|_A \otimes \mathcal{E}_{C|B}(\langle j|_{A'}\varrho_{B|A'}|k\rangle_{A'}), \tag{A28}$$

and then applying the same theorem to $\mathcal{E}_{C|B}$ gives

$$\varrho_{C|A} = \sum_{j,k} |j\rangle\langle k|_A \otimes \langle j|_{A'}\mathrm{Tr}_B(\varrho_{C|B}\varrho_{B|A'})|k\rangle_{A'}. \tag{A29}$$

Since A' is a dummy label in this expression, it can be changed to A and then

$$\varrho_{C|A} = \sum_{j,k} |j\rangle\,\langle j|_A\, \mathrm{Tr}_B(\varrho_{C|B}\varrho_{B|A})\,|k\rangle\,\langle k|_A \tag{A30}$$
$$= \mathrm{Tr}_B(\varrho_{C|B}\varrho_{B|A}). \tag{A31}$$

For the converse direction, we have

$$\mathcal{E}_{C|A}(M_A) = \mathrm{Tr}_A(\varrho_{C|A}M_A) \tag{A32}$$
$$= \mathrm{Tr}_A[\mathrm{Tr}_B(\varrho_{C|B}\varrho_{B|A})M_A] \tag{A33}$$
$$= \mathrm{Tr}_B[\varrho_{C|B}\mathrm{Tr}_A(\varrho_{B|A}M_A)] \tag{A34}$$
$$= \mathrm{Tr}_B[\varrho_{C|B}\mathcal{E}_{B|A}(M_A)] \tag{A35}$$
$$= \mathcal{E}_{C|B}[\mathcal{E}_{B|A}(M_A)] \tag{A36}$$
$$= \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}(M_A). \tag{A37}$$

■

[1] B. de Finetti, *Theory of Probability*, Vol. 1 (Wiley, New York, 1974).
[2] S. Taylor, S. Cheung, C. Brukner, and V. Vedral, in *Proceedings of Quantum Communication, Measurement and Computing*, AIP Conf. Proc. No. 734 (AIP, New York, 2004), pp. 281–284.
[3] M. S. Leifer, Phys. Rev. A **74**, 042310 (2006).
[4] M. S. Leifer, in *Foundations of Probability and Physics—4*, edited by G. Adenier, C. A. Fuchs, and A. Y. Khrennikov, AIP Conf. Proc. No. 889 (AIP, New York, 2007), pp. 172–186.
[5] P. W. Evans, H. Price, and K. B. Wharton, Br. J. Philos. Sci. **64**, 297 (2013).
[6] K. B. Wharton, D. J. Miller, and H. Price, Symmetry **3**, 524 (2011).
[7] S. Marcovitch and B. Reznik, arXiv:1103.2557.
[8] S. Marcovitch and B. Reznik, arXiv:1107.2186.
[9] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[10] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), p. 175.
[11] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[12] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).
[13] H. Umegaki, Kodai Math. Sem. Rep. **14**, 59 (1962).
[14] D. Petz, *Quantum Information Theory and Quantum Statistics* (Springer, Berlin, 2008), Chap. 9.2, pp. 123–130.

[15] M. Rédei and S. J. Summers, Stud. Hist. Philos. Mod. Phys. **38**, 390 (2007).
[16] D. Petz, in *Quantum Probability and Applications III*, edited by L. Accardi and W. Waldenfels, Lecture Notes in Mathematics, Vol. 1303 (Springer, Berlin, Heidelberg, 1988), pp. 251–260.
[17] N. J. Cerf and C. Adami, Phys. Rev. Lett. **79**, 5194 (1997).
[18] N. J. Cerf and C. Adami, in *Fundamental Theories of Physics*, edited by M. Ferrero and A. Merwe, Vol. 81 (Kluwer Academic, Dordrecht, 1997).
[19] N. J. Cerf and C. Adami, Physica D **120**, 62 (1998).
[20] M. Asorey, A. Kossakowski, G. Marmo, and E. C. G. Sudarshan, Open Syst. Inf. Dyn. **12**, 319 (2005).
[21] Z. Zhang and S. Luo, Phys. Rev. A **75**, 032312 (2007).
[22] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).
[23] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **86**, 012103 (2012).
[24] C. A. Fuchs, J. Mod. Opt. **50**, 987 (2003).
[25] F. Verstraete, Ph.D. thesis, University of Leuven, 2002.
[26] S. M. Barnett, D. T. Pegg, and J. Jeffers, J. Mod. Opt. **47**, 1779 (2000).
[27] D. T. Pegg, S. Barnett, and J. Jeffers, J. Mod. Opt. **49**, 913 (2002).
[28] D. T. Pegg, S. M. Barnett, and J. Jeffers, Phys. Rev. A **66**, 022106 (2002).
[29] H. Barnum and E. Knill, J. Math. Phys. **43**, 2097 (2002).
[30] L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A **183**, 14 (1993).
[31] P. Hausladen and W. K. Wootters, J. Mod. Opt. **41**, 2385 (1994).
[32] V. P. Belavkin, Radio Eng. Electron. Phys. **20**, 39 (1975).
[33] V. P. Belavkin, Stochastics **1**, 315 (1975).
[34] J. Bub, in *The Logico-algebraic Approach to Quantum Mechanics*, edited by C. A. Hooker, Vol. II (Reidel, Dordrecht, 1978), pp. 209–226.
[35] J. Bub, Stud. Hist. Philos. Mod. Phys. **38**, 232 (2007).
[36] M. Ozawa, in *Quantum Communication, Computing and Measurement*, edited by O. Hirota, A. S. Holevo, and C. M. Caves (Kluwer/Plenum, New York, 1997), pp. 233–241.
[37] M. Ozawa, Fortschr. Phys. **46**, 615 (1998).
[38] R. C. Jeffrey, *The Logic of Decision*, 2nd ed. (University of Chicago Press, Chicago, 1983).
[39] R. C. Jeffrey, *Subjective Probability: The Real Thing* (Cambridge University Press, Cambridge, UK, 2004).
[40] J. Earman, in *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory* (MIT Press, Cambridge, MA, 1992), Chap. 2, pp. 34–35.
[41] P. Diaconis and S. L. Zabell, J. Am. Stat. Assoc. **77**, 822 (1982).
[42] M. A. Naimark, Izv. Akad. Nauk SSSR Ser. Mat. **4**, 277 (1940).
[43] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).
[44] E. Schrödinger, Proc. Cambridge Philos. Soc. **31**, 555 (1935).
[45] E. Schrödinger, Proc. Cambridge Philos. Soc. **32**, 446 (1936).
[46] M. Vujičić and F. Herbut, J. Phys. A **21**, 2931 (1988).
[47] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007).
[48] S. J. Jones, H. M. Wiseman, and A. C. Doherty, Phys. Rev. A **76**, 052116 (2007).
[49] F. Herbut, J. Phys. A **42**, 085202 (2009).
[50] R. W. Spekkens and N. Harrigan, Found. Phys. **40**, 125 (2010).
[51] J. S. Bell, Physics **1**, 195 (1964).
[52] K. Jacobs, Quantum Inf. Proc. **1**, 73 (2002).
[53] K. Jacobs, Phys. Rev. A **72**, 044101 (2005).
[54] M. S. Leifer and D. Poulin, Ann. Phys. **323**, 1899 (2008).
[55] B. Coecke and R. W. Spekkens, Synthese **186**, 651 (2012).
[56] M. Rédei, Int. Stud. Philos. Sci. **6**, 129 (1992).
[57] G. Valente, Stud. Hist. Philos. Mod. Phys. **38**, 832 (2007).
[58] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, arXiv:quant-ph/0611295.
[59] L. Hardy, arXiv:gr-qc/0509120.
[60] L. Hardy, J. Phys. A **40**, 3081 (2007).
[61] L. Hardy, arXiv:1104.2066.
[62] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).
[63] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).
[64] O. Oreshkov, F. Costa, and C. Brukner, Nat. Commun. **3**, 1092 (2012).
[65] B. Coecke and E. O. Paquette, in *New Structures for Physics*, edited by B. Coecke, Lecture Notes in Physics (Springer, Berlin, 2009), pp. 173–286.
[66] B. Coecke, E. O. Paquette, and D. Pavlovic, in *Sematic Techniques in Quantum Computation*, edited by S. Gay and I. Mackie (Cambridge University Press, Cambridge, UK, 2010), pp. 29–69.
[67] C. M. Caves, C. A. Fuchs, and R. Schack, Phys. Rev. A **65**, 022305 (2002).
[68] C. M. Caves, C. A. Fuchs, and R. Schack, Studies in History Philos. Sci. Part B: Studies in History Philos. Mod. Phys. **38**, 255 (2007).
[69] C. A. Fuchs, arXiv:1003.5209.
[70] C. A. Fuchs, Physics in Canada **66**, 77 (2010).
[71] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. (Cambridge University Press, Cambridge, U.K., 2009).
[72] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. (MIT Press, Cambridge, MA, 2001).
[73] M. S. Leifer and R. W. Spekkens, arXiv:1110.1085.