Quantum theory as a Causally Neutral Theory of Bayesian Inference

M. S. Leifer

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

Robert W. Spekkens

Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, Ontario, Canada, N2L 2Y5

(Dated: June 17, 2012)

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. Classical probability theory is independent of causal structure, whereas the conventional quantum formalism requires causal structure to be fixed in advance. In this paper, we develop the formalism of quantum conditional states, which unifies the description of experiments involving two systems at a single time with the description of those involving a single system at two times. The analogies between quantum theory and classical probability theory are expressed succinctly within the formalism and it unifies the mathematical description of distinct concepts, such as ensemble preparation procedures, measurements, and quantum dynamics. We introduce a quantum generalization of Bayes’ theorem and the associated notion of Bayesian conditioning. Conditioning a quantum state on a classical variable is the correct rule for updating quantum states in light of classical data, regardless of the causal relationship between the classical variable and the quantum degrees of freedom, but it does not include the projection postulate as a special case. We show that previous arguments that projection is the quantum generalization of conditioning are based on misleading analogies. Since our formalism is causally neutral, conditioning provides a unification of the predictive and retrodictive formalisms for prepare-and-measure experiments and leads to an elegant derivation of the set of states that a system can be “steered” to by making measurements on a remote system.

PACS numbers: 03.65.Ca, 03.65.Ta, 03.67.-a

Keywords: quantum conditional probability, quantum dynamics, quantum measurement, retrodiction, steering

Contents

I. Introduction 2

A. Causal Neutrality 3

II. Classical Conditional Probability 4

III. Quantum Conditional States 5

A. Acausal Conditional States 5
B. Acausal Belief Propagation 6
C. Causal Conditional States 7
D. Quantum Channels as Causal Belief Propagation 8
E. Causal Joint States 9
F. Quantum-Classical Hybrid Regions 10
G. Ensemble Averaging as Belief Propagation 11
H. The Born Rule as Belief Propagation 12
I. Belief Propagation of Conditional States 13
J. The Heisenberg Picture 14
K. Quantum Instruments as Causal Belief Propagation 16

IV. Quantum Bayes’ Theorem 17

A. General Quantum Bayes’ Theorem 18
B. Bayes’ Theorem for Quantum-Classical Hybrids 19
C. Retrodiction and Time Symmetry 20
D. Remote Measurements and Spatial Symmetry 22

V. Quantum Bayesian Conditioning 23

A. Examples of Quantum Bayesian Conditioning 25
1. Conditioning on a Preparation Variable 25
2. Conditioning on the Outcome of a Measurement 25
3. Conditioning on the Outcome of a Remote Measurement: Quantum Steering 26
B. Why the Projection Postulate is not an Instance of Bayesian Conditioning 27
1. State-Update Rules as a Combination of Belief Propagation and Bayesian Conditioning 29
2. No Information Gain Without Disturbance 30

VI. Related Work 31

A. Comparison to Other Quantum Generalizations of Conditional Probability 31
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 \[12\]. 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.

One of the main goals of this work is 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 \[14\] for a basic introduction and \[15\] for a review) and the Cerf-Adami conditional density operator...
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 [23, 24], 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 space-like and time-like correlations, and they unify concepts that look distinct in the conventional formalism. For example, the descriptions of the preparation of an ensemble of quantum states, the probabilities for the outcomes of measurements, and quantum dynamics can all be written as a generalization of the classical belief propagation rule (also called the law of total probability) [24], which is given by

\[ P(S) = \sum_R P(S|R)P(R). \]  

(1)

The three cases differ only in the choice of which variables: \( R \), \( S \) or neither; remain classical in the generalization. We also show that the ensemble of states generated by the most general update rule for a quantum system after a measurement — a quantum instrument — can be described by belief propagation with respect to a quantum conditional state.

We introduce a quantum version of Bayes’ theorem, which generalizes a rule previously advocated by Fuchs [22]. Several well-known constructions in quantum information are instances of this theorem, including the correspondence between Positive Operator Valued Measures (POVMs) and ensemble decompositions of a density operator [3, 4, 22], the “pretty good” measurement [24, 26], and the Barnum-Knill recovery map for approximate error correction [27].

Finally, we discuss conditioning a quantum state 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 classical variable and the quantum degrees of freedom. The causal neutrality of our formalism unifies the treatment of predictions with that of retrodictions (inferences about the past), in analogy with the unification found in classical Bayesian inference. The retrodictive formalism we devise coincides with the one introduced in [28–30] in the case of unbiased sources, but differs in the general case, retaining a closer analogy with classical Bayesian inference. The formalism also describes the case of conditioning on the outcome of a remote measurement, such as in the EPR experiment or more generally in “quantum steering”. Although our notion of conditioning does not include the projection postulate as a special case, we argue that it is nevertheless the correct way to update states in the light of measurement results, and that the assertion that the projection postulate is analogous to Bayesian conditioning [51–52] is based on a misleading analogy. The latter is best described as the application of a belief propagation rule (a non-selective update map), followed by conditioning (the selection). The remote measurement case 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 [33].

### A. Causal Neutrality

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, \( A \), and a Hilbert space \( \mathcal{H}_A \).

Generally, a region will refer 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.

As discussed in the introduction, we will make use of a conditional quantum state to achieve a unified description of the spatial and temporal scenarios. In fact, although we motivate our work by the distinction between spatial and temporal separation, we find that it is not the spatio-temporal 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 spatio-temporal relation. Consider a relativistic quantum theory for instance.
though a pair of regions that are space-like separated are always acausally related, a pair of regions that are time-like 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 time-like 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 spatio-temporal structure, and it is all that we need here.

Typically, we shall confine our attention to two paradigmatic examples of causal and acausal separation (which can be formulated in either a relativistic or a non-relativistic 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 re-prepares it in a fixed state) [75].

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.

Nonetheless, the different causal scenarios continue to be distinguished insofar as the set of operators that can represent a possible state in one scenario is different from the set that does so in the other scenario. This latter fact does not constitute a failure to achieve causal neutrality in the formalism for Bayesian inference because a similar phenomenon occurs classically. 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 these variables is not arbitrary, but has the form

\[
P(R, S, T) = P(S|R)P(T|R)P(R)\]

In our framework, the operator describing the space of some set of regions may also depend on the causal relations among those regions.

There is, however, one sense in which the formalism for quantum Bayesian inference that we develop here is more sensitive to the causal structure than the formalism for classical Bayesian inference. In the latter, 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 sum up, 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 [76].

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 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

\[
\sum_S P(S|R) = 1
\]

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

\[
P(R, S) = P(S|R)P(R),
\]

where the multiplication is defined element-wise, 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

\[
P(R) = \sum_S P(R, S),
\]

and the conditional probability of \( S \) given \( R \) is

\[
P(S|R) = \frac{P(R, S)}{P(R)}.
\]

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

\[
P(R_1, R_2, \ldots, R_n) = P(R_n|R_1, R_2, \ldots, R_{n-1})
\times P(R_{n-1}|R_1, R_2, \ldots, R_{n-2}) \cdots P(R_2|R_1)P(R_1).
\]
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.

\[
P(R, S, T) \xrightarrow{\sum_R} P(S, T) \\
\downarrow \times P(T)^{-1} \quad \downarrow \times P(T)^{-1}
\]

\[
P(R, S | T) \xrightarrow{\sum_R} P(S | T)
\]

### 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 §III A — 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. However, acausal analogs of some of these ideas are also developed in these sections.

#### A. Acausal Conditional States

In this section, the quantum analog of a conditional probability distribution, a conditional state, is developed as it applies to acausally related regions. This causal scenario, and its classical analog, are depicted in fig. 1. The definition proceeds in analogy with the classical treatment given in §II A. 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) \( \rho_A \) 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 \( \rho_{AB} \) 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 I.

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 III.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

\[
\text{Tr}_B (\rho_{B|A}) = I_A,
\]

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

\[
\rho_{AB} = (\rho_A^\perp \otimes I_B) \rho_{B|A} (\rho_A^\perp \otimes I_B).
\]

Eq. (9) involves two constructions that appear repeatedly in what follows. Firstly, the operators \( \rho_A^\perp \) and \( \rho_{B|A} \) are combined via multiplication, but they are defined on different spaces. To solve this problem, \( \rho_A^\perp \) 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 will be left implicit, so

\[
\text{TABLE I: Analogies between classical probability theory for two random variables and quantum theory for two acausally related regions.}
\]

| Classical Probability | Quantum Theory |
|-----------------------|----------------|
| \( P(R) \)           | \( \rho_A \)   |
| \( P(R, S) \)        | \( \rho_{AB} \) |
| \( P(S) = \sum_R P(R, S) \) | \( \text{Tr}_B (\rho_{B|A}) = I_A \) |
| \( \sum_S P(S | R) = 1 \) | \( \rho_{AB} = \rho_{B|A} \ast \rho_A \) |
| \( P(S | R) = P(R, S) / P(R) \) | \( \rho_{B|A} = \rho_{AB} \ast \rho_A^{-1} \) |

\[
\text{FIG. 1: Acausally related quantum and classical regions. Classical variables are denoted by triangles and quantum regions by circles (this convention is suggested by the shape of the convex set of states in each theory). The dotted line represents acausal correlation. (a) Two quantum regions in an arbitrary joint state (possibly correlated). (b) Two classical variables with an arbitrary joint probability distribution (possibly correlated).}
\]
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 $M_{A} \otimes N_{B} = (M_{A} \otimes I_B) (I_A \otimes N_B) = M_{A} N_{B}$.

Secondly, rather than simply multiplying $\rho_{B|A}$ with $\rho_A$ in eq. (9), $\rho_{B|A}$ is conjugated by $\rho_A^\dagger$. 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 (non-associative and non-commutative) product $M \ast N$ via
\[ M \ast N = N^\dagger M N^\dagger. \tag{10} \]

With these conventions, eq. (9) can be rewritten as
\[ \rho_{AB} = \rho_{B|A} \ast \rho_A, \tag{11} \]
which looks a lot closer to eq. (3) than eq. (9) does.

Starting with a joint state $\rho_{AB}$ and its reduced state $\rho_A = Tr_B (\rho_{AB})$, a conditional state can be defined via
\[ \rho_{B|A} = \rho_{AB} \ast \rho_A^{-1}, \tag{12} \]
which is the analog of eq. (4).

As with eq. (5), there are problems with this formula if $\rho_A$ 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 $supp(\rho_A) \otimes \mathcal{H}_B$, where $supp(\rho_A)$ denotes the support of $\rho_A$ (the span of the eigenvectors of $\rho_A$ having nonzero eigenvalues). Because of this, the resulting conditional density operator satisfies $Tr_B (\rho_{B|A}) = I_{supp(\rho_A)}$ 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 I. As these analogies suggest, the $\ast$-product notation allows equations from classical probability to be generalized to quantum theory by replacing functions by operators, products by $\ast$-products, and division by $\ast$-products with the inverse. However, whilst 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 $\ast$-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, $\ast$-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
\[ \rho_{A_1 A_2 \ldots A_n} = \rho_{A_n | A_1 A_2 \ldots A_{n-1}} \ast \rho_{A_{n-1} | A_1 A_2 \ldots A_{n-2}} \ast \ldots \ast \rho_{A_2 | A_1} \ast \rho_{A_1}. \tag{13} \]

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:
\[
\begin{align*}
\rho_{ABC} & \xrightarrow{\text{Tr}_C} \rho_{AB} \\
\rho_{BC|A} & \xrightarrow{\text{Tr}_C} \rho_{B|A}
\end{align*}
\]
\[ \text{(14)} \]

Example III.2 (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
\[ \rho_{RS} = \sum_{r,s} P(R = r, S = s) |r\rangle \langle r| \otimes |s\rangle \langle s|, \tag{15} \]
\[ \rho_R = \sum_r P(R = r) |r\rangle \langle r| = Tr_S (\rho_{RS}) \tag{16} \]
\[ \rho_{S|R} = \sum_{r,s} P(S = s | R = r) |r\rangle \langle r| \otimes |s\rangle \langle s|. \tag{17} \]

Using eqs. (2-3), 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 $\rho_{RS}, \rho_R$ 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. (13-17).

Example III.3 (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 $Tr_B (\rho_{B|A}) = I_A$, and $I_A$ has all eigenvalues equal to 1, the Schmidt decomposition of $|\psi\rangle_B|A$ is of the form
\[ |\psi\rangle_B|A = \sum_k |u_k\rangle_A \otimes |v_k\rangle_B, \tag{18} \]
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 $\dim(\mathcal{H}_A) \leq \dim(\mathcal{H}_B)$ because otherwise there would not be enough orthonormal vectors on the $B$ side to enforce $Tr_B (\rho_{B|A}) = I_A$. [7]
Since all the Schmidt coefficients are the same, the bases \{\{u_k\}\} and \{\{v_k\}\} are highly non-unique. The conditional state \(\psi_{B|A}\) itself only determines the relationship between the two Schmidt bases, i.e. for any basis in \(H_A\) it determines a corresponding basis in \(H_B\). To see this, fix a reference basis, \(\{\{j\}\}\), for \(H_A\) 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 \(\dagger\) denotes complex conjugation in the \(\{\{j\}\}\) basis. Then, if \(\{\{|w_k\rangle_A\}\) is any other basis for \(H_A\), eq. (18) can be rewritten as
\[
|\psi\rangle_{B|A} = \sum_k |w_k\rangle_A \otimes U_{B|A'} |u_k\rangle_{A'} ,
\]
(19)
where \(A'\) labels a second copy of \(H_A\). With respect to the reference basis \(\{\{j\}\}\), this simplifies to
\[
|\psi\rangle_{B|A} = U_{B|A'} |\Phi^+\rangle_{AA'} ,
\]
(20)
where \(|\Phi^+\rangle_{AA'} = \sum_j |\langle j\rangle_{AA'}\).

Let \(\rho_A\) be an arbitrary density operator on \(H_A\) with eigendecomposition \(\rho_A = \sum_k p_k |w_k\rangle_A \langle w_k|_A\). 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
\[
|\psi\rangle_{AB} = \rho_A^{\frac{1}{2}} |\psi\rangle_{B|A} .
\]
(21)
Combining this with eq. (10) gives the Schmidt decomposition
\[
|\psi\rangle_{AB} = \sum_k \sqrt{p_k} |w_k\rangle_A \otimes U_{B|A'} |u_k\rangle_{A'} .
\]
(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
\[
|\psi\rangle_{AB} = \rho_A^{1/2} U_{B|A'} |\Phi^+\rangle_{AA'} .
\]
(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
\[
P(S) = \sum_R P(S|R) P(R).
\]
(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 \(\Gamma_{SR}\) 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
\[
\Gamma_{SR}(P(R)) = \sum_R P(S|R)P(R).
\]
(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 \(\rho_A\) and a conditional state \(\rho_{B|A}\) then your state for \(B\) is determined by the acausal quantum belief propagation rule
\[
\rho_B = \text{Tr}_A (\rho_{B|A} \rho_A)
\]
(26)
which follows from the fact that the joint state is \(\rho_{AB} = \rho_{B|A} \rho_A\) so that \(\rho_B = \text{Tr}_A (\rho_{B|A} \rho_A)\), and from the cyclic property of the trace.

As in the classical case, acausal belief propagation can also be viewed as a linear map \(\mathcal{E}_{B|A}\) from states on \(A\) to states on \(B\) that preserves positivity and normalization, defined by
\[
\mathcal{E}_{B|A}(\rho_A) = \text{Tr}_A (\rho_{B|A} \rho_A).
\]
(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 \(\mathcal{E}_{B|A}\) is in fact identical to the map associated to \(\rho_{B|A}\) via the Jamiołkowski isomorphism \(\mathcal{J}\), which is a familiar construction in quantum information theory. These facts are consequences of the following theorem.

**Theorem III.4 (Jamiołkowski Isomorphism).** Let \(\mathcal{E}_{B|A} : \mathcal{L}(H_A) \rightarrow \mathcal{L}(H_B)\) be a linear map and let \(M_{AC} \in \mathcal{L}(H_{AC})\), where \(H_C\) is a Hilbert space of arbitrary dimension. Then, the action of \(\mathcal{E}_{B|A}\) on \(\mathcal{L}(H_A)\) (tensored with the identity on \(\mathcal{L}(H_C)\)) is given by
\[
(\mathcal{E}_{B|A} \otimes I_C)(M_{AC}) = \text{Tr}_A (\rho_{B|A} M_{AC}) ,
\]
(28)
where \(\rho_{B|A} \in \mathcal{L}(H_{AB})\) is given by
\[
\rho_{B|A} = (\mathcal{E}_{B|A} \otimes I_A) \left( \sum_{j,k} |j\rangle_A \langle k|_A \otimes |j\rangle_{A'} \langle k|_{A'} \right) .
\]
(29)
Here, \(A'\) labels a second copy of \(A\), \(I_A\) is the identity superoperator on \(\mathcal{L}(H_A)\), and \(\{\{j\}\}\) is an orthonormal basis for \(H_A\).

Furthermore, the operator \(\rho_{B|A}\) is an acausal conditional state, i.e. it satisfies definition (11), if and only if \(\mathcal{E}_{B|A} \otimes T_A\) is CPT, where \(T_A : \mathcal{L}(H_A) \rightarrow \mathcal{L}(H_A)\) denotes the linear map implementing the partial transpose relative to some basis.

The proof is provided in appendix A.
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.

\[
\begin{align*}
\rho_B &= \text{Tr}_A (\varrho_{B|A} \rho_A) \\
\text{Conventional notation:} &
\rho_B = \mathcal{E}_{B|A}(\rho_A) \\

P(S) &= \sum_R P(S|R)P(R)
\end{align*}
\]

(a)

(b)

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 III.5.** A causal conditional state of $B$ given $A$ is an operator $\varrho_{B|A}$ on $\mathcal{H}_{AB}$ that can be written as

\[
\varrho_{B|A} = \rho_{B|A}^{T_A},
\]

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. $\langle \psi | A_0 \otimes \phi | B \varrho_{B|A} | \psi \otimes | \phi | B \geq 0$ for all $| \psi \rangle_A \in \mathcal{H}_A$, $| \phi \rangle_B \in \mathcal{H}_B$.

In this section, we show how causal conditional states can be used for quantum belief propagation across causally related regions in the same way as acausal conditional states are used for acausally related regions, namely by a rule of the form $\rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A)$. Belief propagation for dynamics with a quantum input and a quantum output are treated in III D and 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. III E 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 III G and III H. In III I, the Heisenberg picture is translated into the conditional states formalism. In 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 III 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 $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ such that, if $\rho_A$ is the state of $A$ and $\rho_B$ 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.
Conditional States Formalism

| Dynamics (CPT Map)       | Conventional Notation | Conditional States Formalism |
|--------------------------|-----------------------|-----------------------------|
| \( \mathcal{E}_{B|A} \)  | \( \rho_B = \mathcal{E}_{B|A} (\rho_A) \)                  | \( \varrho_{B|A} \)       |
|                          | \( \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) \)            |                              |

| Preparation (Set of States) |                                                                 |
|-----------------------------|------------------------------------------------------------------|
| \( \rho_A = \sum_x P(X=x) \rho_x^A \) | \( \rho_A \) = \text{Tr}_X (\varrho_{A|X} \rho_X) |

| Measurement (POVM)           |                                                                 |
|-----------------------------|------------------------------------------------------------------|
| \( \mathcal{E}^A_y \)       | \( P(Y=y) = \text{Tr}_A (\mathcal{E}^A_y \rho_A) \)               |
|                              | \( \rho_Y = \text{Tr}_A (\varrho_{Y|A} \rho_A) \) |

| Composition of Dynamics      |                                                                 |
|-----------------------------|------------------------------------------------------------------|
| \( \mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A} \) | \( \varrho_{C|A} = \text{Tr}_B (\varrho_{C|B} \varrho_{B|A}) \) |

| Heisenberg picture (Dual Map) |                                                                 |
|-------------------------------|------------------------------------------------------------------|
| \( \mathcal{E}^\dagger_{A|B} \) | \( \mathcal{E}^\dagger_{C|A} = \mathcal{E}^\dagger_{C|B} \circ \mathcal{E}^\dagger_{B|A} \) |
| \( \mathcal{E}^\dagger_{y|A} \) | \( \varrho_{Y|A} = \text{Tr}_B (\varrho_{Y|B} \varrho_{B|A}) \) |

| State-update (Quantum Instrument) |                                                                 |
|----------------------------------|------------------------------------------------------------------|
| \( \mathcal{E}^B_{y|A} \)       | \( P(Y=y) \rho^B_y = \mathcal{E}^B_{y|A} (\rho_A) \)             |
|                                  | \( \rho_Y = \text{Tr}_A (\varrho_{Y|B} \rho_B) \) |

TABLE II: Comparison of conventional and conditional state notation for CPT maps, POVMs, sets of states, quantum instruments and the Heisenberg picture.

**Theorem III.6.** Let \( \mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B) \) be a linear map and let \( \varrho_{B|A} \in \mathcal{L}(\mathcal{H}_{AB}) \) be the Jamiołkowski isomorphic operator, as defined in eq. (24). Then, \( \varrho_{B|A} \) is a causal conditional state, which satisfies definition III.3, if and only if \( \mathcal{E}_{B|A} \) is CPT.

**Proof.** Theorem III.3 states that the operator \( \rho_{B|A} \) that is Jamiołkowski isomorphic to a linear map \( \mathcal{E}_{B|A} \) is an acausal conditional state iff the linear map \( \mathcal{E}_{B|A} = \mathcal{E}_{B|A} \circ T_A \) is CPT. However, the Jamiołkowski isomorphic operator to \( \mathcal{E}_{B|A} \) is \( \varrho_{B|A} = \rho_{B|A} \circ T_A \), so this is a causal conditional state iff \( \mathcal{E}_{B|A} \) is CPT.

Together with theorem III.3 this implies that the action of a CPT map \( \mathcal{E}_{B|A} \) on an operator \( M_{AC} \) is given by

\[ \mathcal{E}_{B|A}(M_{AC}) = \text{Tr}_A (\varrho_{B|A} M_{AC}) , \]

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 III.7.** Let \( \varrho_{B|A} \) be the Jamiołkowski isomorphic causal conditional to a CPT map \( \mathcal{E}_{B|A} \) that describes a quantum dynamics. If the initial state of region \( A \) is \( \rho_A \), then the state of \( B \), conventionally written as

\[ \rho_B = \mathcal{E}_{B|A} (\rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

in analogy with the classical belief propagation rule eq. (34).

We call eq. (34) the causal quantum belief propagation rule. It follows from eq. (31).

Fig. 2 and the first line 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**

\[ \mathcal{E}_{B|A} \] showed that a joint state \( \rho_{AB} \) of two acausally related regions can be decomposed into a reduced state \( \rho_A \) and an acausal conditional state \( \rho_{B|A} \). Similarly, two causally related regions can be described by an input state \( \rho_A \) 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 III.8.** A causal joint state of two causally related regions, \( A \) and \( B \), is an operator on \( \mathcal{H}_{AB} \) of the form

\[ \varrho_{AB} = \varrho_{B|A} \circ \rho_A , \]

\[ \varrho_{AB} = \varrho_{B|A} \circ \rho_A , \]

\[ \varrho_{AB} = \varrho_{B|A} \circ \rho_A , \]

\[ \varrho_{AB} = \varrho_{B|A} \circ \rho_A , \]

where \( \rho_A \) 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 \( \varrho_{AB} \) is the initial state (input to the channel) and the reduced state on \( B \) is

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

\[ \rho_B = \text{Tr}_A (\varrho_{B|A} \rho_A) , \]

in analogy with the classical belief propagation rule eq. (34).
which, by the cyclic property of the trace and proposition [III.7] is the final state (output of the channel).

It is not too difficult to see that a causal joint state \( \varrho_{AB} \) is the partial transpose of an acausal joint state on \( \mathcal{H}_{AB} \). Specifically, \( \varrho_{AB}^{TA} = \varrho_{B|A} \ast \varrho_{A}^{T^*} \), where \( \varrho_{B|A} = \varrho_{B|A}^{T^*} \) is an acausal conditional state and \( \varrho_{A}^{T^*} \) 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 \( \varrho_{B|A} \), 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 \( \varrho_{AB} \).

Example III.9 (Unitary Dynamics). Suppose a region \( A \) is assigned the state \( \varrho_{A} \) with eigendecomposition \( \varrho_{A} = \sum_{j} p_{j} |u_{j}⟩⟨u_{j}|. \) A is then mapped to a region \( B \), which has a Hilbert space of the same dimension as that of \( A \), by an isometry \( U_{B|A} = \sum_{j} |v_{j}\rangle_{B} |u_{j}\rangle_{A}. \) Since the Jamiołkowski isomorphism is basis independent, the causal conditional state associated with the map \( E_{B|A} \) can be written in the eigenbasis of \( \varrho_{A} \) as

\[
\varrho_{B|A} = \sum_{j} p_{j} |u_{j}\rangle_{A} \otimes U_{B|A} |u_{j}\rangle_{A} \langle U_{B|A}^{\dagger}
\]

(36)

\[
= \sum_{j,k} |u_{j}\rangle ⟨u_{k}|_{A} \otimes |v_{k}\rangle ⟨v_{j}|. \quad (37)
\]

It follows that the causal joint state for \( AB \) is

\[
\varrho_{AB} = \sum_{j,k} \sqrt{p_{j}p_{k}} |u_{j}\rangle_{A} \otimes |v_{k}\rangle \langle u_{j}|_{B}. \quad (38)
\]

Note that \( \varrho_{B|A} \) is the partial transpose over \( A \) of a pure acausal conditional state, as can be seen by comparison with eq. (15) from example III.3. Also, \( \varrho_{AB} \) is the partial transpose over \( A \) of a pure acausal joint state \( \varrho_{AB} = |ψ⟩⟨ψ|_{AB} \), where

\[
|ψ⟩_{AB} = \sum_{j} \sqrt{p_{j}} |u_{j}\rangle_{A} \otimes |v_{j}\rangle_{B}. \quad (39)
\]

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 [4, 11]). 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 [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 [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 [III.7]. 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 [III.2] the classical variable \( X \) is associated with a Hilbert space \( \mathcal{H}_{X} \) equipped with a preferred basis \( \{ |x⟩ \} \) 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 III.10.** A hybrid operator on \( \mathcal{H}_{XA} \) is an operator of the form

\[
M_{XA} = \sum_{x} |x⟩⟨x|_{X} \otimes M^{A}_{x},
\]

(40)

where \( \{ |x⟩ \} \) is a preferred basis for \( \mathcal{H}_{X} \) and \( \{ M^{A}_{x} \} \) is a set of operators acting on \( \mathcal{H}_{A} \) labeled by the values of \( X \). The operators \( M^{A}_{x} \) 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 \( \rho \) and \( \varrho \) 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 \( \sigma \) is used. According to these conventions, any formula expressed in terms of \( \sigma \)'s will yield a valid formula about hybrid states if the \( \sigma \)'s are replaced by either \( \rho \)'s or \( \varrho \)'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

\[
\text{Tr}_A ( \sigma_{A|X} ) = I_X. \tag{41}
\]

Theorem III.11. Let \( \sigma_{A|X} \) be a hybrid operator, so that by eq. (44) 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 for \( A \) given \( X \) iff each of the components \( \rho_x^A \) is a normalized state on \( \mathcal{H}_A \).

The proof is provided in appendix A.

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. 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 III.12.** 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 \( \rho_x^A \), 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

\[
\rho_A = \text{Tr}_X ( \varrho_{A|X} \rho_X ), \tag{45}
\]

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

This result follows simply from substituting the definition of \( \rho_X \) and \( \varrho_{A|X} \) into eq. (43).

Fig. 3 and the second line of table II summarize how the representation of a preparation procedure within the
This follows from the fact that a preparation procedure proposition III.12 as a special case of proposition III.7.

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 $\rho_X$ via

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

By proposition III.7, the conditional state associated with the preparation is the Jamiołkowski isomorphic operator to this map. Eq. (45) is then obtained as a special case of eq. (46) where the input is classical.

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 $\{E^A_y\}$ with the classical variable $Y$ ranging over the possible outcomes. If the state of the region is $\rho_A$, then the probability of obtaining outcome $y$ is given by the Born rule as $P(Y = y) = \text{Tr}_A (E^A_y \rho_A)$. This scenario is depicted in fig. 4(a). In the classical analogue, the quantum region $A$ is replaced by a classical variable $R$, the state $\rho_A$ is replaced by a distribution on $P(R)$ 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 causally ordered 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 III.13.** Let $\sigma_{Y|A}$ be a hybrid operator so that, by eq. (10), it can be written in the form

$$\sigma_{Y|A} = \sum_y |y\rangle_Y \langle y|_Y \otimes E^A_y,$$  (47)

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

![Fig. 4: Quantum and classical measurements.](image)

**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 non-deterministic way. It is mathematically equivalent to the stochastic dynamics depicted in fig. 2(b).

The proof is given in appendix A.

It is sometimes useful to use the notation $\sigma_{Y=y|A} = \langle y|_Y \sigma_{Y|A} |y\rangle_Y = E^A_y$ 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 of $\rho_{Y|A}$ yields another valid acausal conditional state, because $\{(E^A_y)^T\}$ is a POVM iff $\{E^A_y\}$ is. The remainder of this section concerns the causal interpretation of such states, so the notation $\varrho_{Y|A}$ is adopted.

**Proposition III.14.** Consider a measurement of a POVM $\{E^A_y\}$ on a quantum region $A$ in state $\rho_A$. Let $\varrho_{Y|A}$ be the causal conditional state with components $E^A_y$. The Born rule,

$$P(Y = y) = \text{Tr}_A (E^A_y \rho_A),$$  (48)

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

$$\rho_Y = \text{Tr}_A (\varrho_{Y|A} \rho_A),$$  (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 line of Table I.

Once again, these results can be understood as a special case of Theorem III.6 and Proposition III.7 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 \( \rho_Y \), then the measurement can be represented by the CPT map \( \mathcal{E}_{Y|A} \) defined by

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

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 \( \varrho_{C|B} \).

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 III.15.** 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

(a)

Conventional notation:

\[
\mathcal{E}_{C|A} = \mathcal{E}_{C|B} \circ \mathcal{E}_{B|A}.
\]

(b)

\[
P(T|R) = \sum_S P(T|S)P(S|R)
\]

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

Jamiołkowski isomorphic operators, \( \varrho_{B|A}, \varrho_{C|B} \) and \( \varrho_{C|A} \) satisfy

\[
\varrho_{C|A} = \text{Tr}_B \left( \varrho_{C|B} \varrho_{B|A} \right). \tag{52}
\]

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 Appendix A.

Eq. (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 [V.A.3]

Consider the scenario depicted in Fig. 4. Initially, a state \( \rho_{AB} \) 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 III.4, the state of $AC$ is given by

$$\rho_{AC} = \text{Tr}_B \left( \varrho_{C|B} \rho_{AB} \right).$$  \hspace{1cm} (53)

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

$$\rho_{C|A} \ast \rho_A = \text{Tr}_B \left( \varrho_{C|B} \rho_{B|A} \right) \ast \rho_A,$$

(54)

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

$$\rho_{C|A} = \text{Tr}_B \left( \varrho_{C|B} \rho_{B|A} \right),$$

(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.

![FIG. 6: Propagating an acausal conditional state through a causal conditional state.](image)

**J. The Heisenberg Picture**

In [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.

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

**Definition III.16.** The dual map $(\mathcal{E}_{B|A})_{\uparrow} : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_A)$ of a linear map $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ is the unique map that satisfies

$$\text{Tr}_A \left( (\mathcal{E}_{B|A})_{\uparrow} (N_B) M_A \right) = \text{Tr}_B \left( N_B \mathcal{E}_{B|A} (M_A) \right)$$

(56)

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

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

$$\mathcal{E}_{A|B}^\dagger = (\mathcal{E}_{B|A})_{\uparrow},$$

(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 $E_B^y = \mathcal{E}_{A|B}^\dagger(E_y^B)$ 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 $B$ is followed by a measurement on $B$ yielding an outcome $Y$. This scenario is depicted in fig. [II]. The probability of observing the effect $E_y^B$ after a preparation of $\rho_A$ followed by an evolution $\mathcal{E}_{B|A}$ is expressed in the Schrödinger picture as $\text{Tr}_B (E_y^B \rho_B)$ where $\rho_B = \mathcal{E}_{B|A}(\rho_A)$, while it is expressed in the Heisenberg picture as $\text{Tr}_A (E_y^A \rho_A)$ 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. $\text{Tr}_B (E_y^B \rho_B) = \text{Tr}_A (E_y^A \rho_A)$.

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

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

(58)

where $K_{\mu} : \mathcal{H}_A \rightarrow \mathcal{H}_B$. 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.,

$$\mathcal{E}_{A|B}^\dagger (\cdot) = \sum_{\mu} K_{\mu}^\dagger (\cdot) K_{\mu}.$$  \hspace{1cm} (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 $\mathcal{E}_{A|B}^\dagger (\cdot) = U^\dagger (\cdot) U$. 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 $\{E_y^B\}$ by a conditional state $\varrho_{B|A}$, the CPT map $\mathcal{E}_{B|A}$ by a causal conditional state $\varrho_{B|A}$, and the POVM $\{E_y^A\}$ by a conditional state $\varrho_{Y|A}$. Secondly, note that fig. [III] is just a special case of fig. [II] in which the final region is classical, so the three conditional states are related by eq. [52], i.e.

$$\varrho_{Y|A} = \text{Tr}_B \left( \varrho_{Y|B} \varrho_{B|A} \right).$$  \hspace{1cm} (60)

In [III] 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 $\rho_{B|A}$ as defining a map from $\rho_{Y|B}$ to $\rho_{Y|A}$, 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.

$$\rho_{Y|A} = \text{Tr}_B \left( \rho_{Y|B} \rho_{B|A} \right)$$

Conventional notation: $E_y^A = \mathcal{E}_{A|B}^\dagger (E_y^B)$

FIG. 7: Dynamics in the Heisenberg picture.

**Theorem III.17.** Let $\rho_{BC}$ be the causal joint state that is Jamiołkowski isomorphic to the CPT map $\mathcal{E}_{BC}$, Then, the action of the dual map $\mathcal{E}_{A|B}$ on an operator $M_{BC}$ is given by

$$\left( (\mathcal{E}_{A|B})^\dagger \otimes \mathcal{I}_C \right) (M_{BC}) = \text{Tr}_B \left( M_{BC} \rho_{B|A} \right)$$  \hspace{1cm} (61)

Proof. By definition III.18, the dual map to $\mathcal{E}_{BC}$ is the unique linear map $\mathcal{E}_{A|B}$ that satisfies

$$\text{Tr}_A \left( (\mathcal{E}_{A|B})^\dagger (N_B) M_A \right) = \text{Tr}_B \left( N_B \mathcal{E}_{B|A} (M_A) \right)$$  \hspace{1cm} (62)

for all operators $M_A$ and $N_B$. Using the Jamiołkowski isomorphism, theorem III.6 the right hand side can be written as

$$\text{Tr}_B \left( N_B \mathcal{E}_{B|A} (M_A) \right) = \text{Tr}_{AB} (N_B \rho_{AB} M_A)$$  \hspace{1cm} (63)

$$= \text{Tr}_A \left( \text{Tr}_B (N_B \rho_{B|A}) M_A \right).$$  \hspace{1cm} (64)

The only way this can equal $\text{Tr}_A \left( (\mathcal{E}_{A|B})^\dagger (N_B) M_A \right)$ for all $M_A$ is if $\mathcal{E}_{A|B}^\dagger (N_B) = \text{Tr}_B (N_B \rho_{B|A})$. Eq. (61) then follows by linear extension to $\mathcal{H}_{BC}$. \hfill $\Box$

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

**Proposition III.18.** Let $\rho_{BC}$ be the causal conditional state associated with a quantum evolution described by the CPT map $\mathcal{E}_{BC}$, and let $\rho_{Y|A}$ and $\rho_{Y|B}$ be the hybrid conditional states associated with the POVMs $\{E_y^A\}$ and $\{E_y^B\}$, such that $\{E_y^A\}$ is obtained from $\{E_y^B\}$ by the Heisenberg picture dynamics. The conventional description of evolution in the Heisenberg picture,

$$E_y^A = \mathcal{E}_{A|B}^\dagger (E_y^B),$$  \hspace{1cm} (65)

can be expressed in the conditional states formalism as

$$\rho_{Y|A} = \text{Tr}_B \left( \rho_{Y|B} \rho_{B|A} \right).$$  \hspace{1cm} (66)

This follows straightforwardly from theorem III.17 and theorem III.13.

As in §III similar reasoning can be applied to other causal scenarios. Consider the special case of fig. 8 in which $C$ is replaced by a classical variable $Y$. This is depicted in fig. 8. Two causally related regions, $A$ and $B$, are assigned a state $\rho_{AB}$ and then the POVM $\{E_y^B\}$ (alternatively represented by a conditional state $\rho_{Y|B}$) 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 $Y$.

FIG. 8: Heisenberg evolution for a remote measurement.

The belief propagation formula for this scenario, eq. (65), gives

$$\rho_{Y|A} = \text{Tr}_B \left( \rho_{Y|B} \rho_{B|A} \right).$$  \hspace{1cm} (67)

The components of the conditional state $\rho_{Y|A}$ specify a POVM $\{E_y^A\}$. 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 $\rho_A$, the conditional state $\rho_{Y|A}$ specifies the ensemble of states for region $A$ associated with the different measurement outcomes via $\rho_{AY} = \rho_{Y|A} \rho_A$. In terms of components, this is

$$\rho_{Y|A} = \sum_Y |y\rangle \langle y|_Y \otimes \rho_y^A E_y^A \rho_y^A,$$  \hspace{1cm} (68)

so the unnormalized state of $A$ corresponding to the outcome $Y = y$ is

$$P(Y = y) \rho_y^A = \rho_y^A E_y^A \rho_y^A,$$  \hspace{1cm} (69)

where $P(Y = y) = \text{Tr}_A (E_y^A \rho_A) = \text{Tr}_B (E_y^B \rho_B)$ 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 $\rho_{Y|B}$ to $\rho_{Y|A}$, then this map is analogous to a Heisenberg picture dynamics, except that the propagation is across causally related rather than causally related regions. If $\mathcal{E}_{B|A}$ is the Jamiołkowski isomorphic
map to $\rho_{B|A}$ then, by theorem III.17, the POVM elements are related by

$$E^A_y = \mathcal{E}^A_{y|B}(E^B_y).$$  \hspace{1cm} (70)

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

A similar expression to eq. (70) has appeared in the context of quantum steering \cite{33}, 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 \emph{Y.1.3}. To our knowledge, the connection to Heisenberg dynamics has not previously been recognized.

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 post-measurement region is correlated with the measurement result. In the conventional formalism, the transformative aspect of a measurement is represented by a quantum instrument.

**Definition III.19.** Given quantum regions $A$ and $B$, and a classical variable $Y$, a quantum instrument is a set $\{\mathcal{E}^B_{y|A}\}$ of CPT maps $\mathcal{E}^B_{y|A} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$ such that the operators

$$E^A_y = \mathcal{E}^A_{y|B}(I_B)$$  \hspace{1cm} (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 $P(Y = y)\rho^B_y$ of $B$ when the outcome is known via

$$P(Y = y)\rho^B_y = \mathcal{E}^B_{y|A}(\rho_A).$$  \hspace{1cm} (72)

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

$$P(Y = y) = \text{Tr}_B\left(\mathcal{E}^B_{y|A}(\rho_A)\right).$$  \hspace{1cm} (73)

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

$$P(Y = y) = \text{Tr}_B\left(\mathcal{E}^B_{y|A}(\rho_A)\right) = \text{Tr}_B\left(I_B \mathcal{E}^B_{y|A}(\rho_A)\right) = \text{Tr}_A\left(\mathcal{E}^A_{y|B}(I_B) \rho_A\right) = \text{Tr}_A\left(E^A_y \rho_A\right),$$  \hspace{1cm} (77)

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

Whilst 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 post-measurement 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. $\mathcal{H}_B = \{\emptyset\}$.

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 $\{\Pi^A_y\}$, 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 $\mathcal{E}^B_{y|A}(\rho_A) = I_B \rho_A \Pi^A_y$, where $I_B \mathcal{H}_A$ is an isometry between $\mathcal{H}_A$ and $\mathcal{H}_B$. For general POVMs the rule $\mathcal{E}^B_{y|A}(\rho_A) = I_B \rho_A \left((E^A_y)^{1/2} \rho_A (E^A_y)^{1/2}\right)$, 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). 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

$$P(Y, S) = \sum_R P(Y, S|R)P(R).$$  \hspace{1cm} (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 $S$,
\[ P(Y|R) = \sum_S P(Y,S|R). \]  
(79)

In the conditional states formalism, eqs. (72) and (74) are replaced by straightforward analogs of eqs. (75) and (79) for a causal hybrid state $\varrho_{Y|B,A}$ of a classical variable $Y$ and quantum region $B$, conditioned on a quantum region $A$.

Theorem III.20. Let $\varrho_{Y|B,A}$ be an operator on $\mathcal{H}_{Y|AB}$ of the form
\[\varrho_{Y|B,A} = \sum_y \rvert y \rangle \langle y \rvert_Y \otimes \varrho_{Y=y|B,A}, \]  
(80)

where
\[\varrho_{Y=y,B,A} = \mathcal{E}_y^{B|A'} \left( \sum_{i,k} \rvert j_i \rangle \langle k \rangle_{A} \otimes \rvert j \rangle_{A'} \right)\]  
(81)

are the Jamiołkowski isomorphic operators to maps $\mathcal{E}_y^{B|A} : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$. Then $\varrho_{Y|B,A}$ 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 III.6, III.11 and III.13 and is left to the reader.

Proposition III.21. Let $\varrho_{Y|B,A}$ 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 $\rho_A$, the state-update rule in conventional notation is given by
\[ P(Y | y = y) \rho_B = \mathcal{E}_y^{B|A}(\rho_A). \]  
(82)

In analogy with the classical expression in eq. (78), this can be expressed in the conditional states framework as
\[ P(Y | y) \rho_B = Tr_A \left( \varrho_{Y|B,A} \rho_A \right), \]  
(83)

where $\rho_B = \sum_y P(Y = y) |y \rangle \langle y |_Y \otimes \rho_B$. Furthermore, the conventional expression for the relation between a POVM and a quantum instrument,
\[ E_y^A = (\mathcal{E}_y^1)^{A|B}(I_B), \]  
(84)

can be expressed simply as
\[ \varrho_{Y|A} = Tr_B \left( \varrho_{Y|B,A} \right), \]  
(85)

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

The proof of eq. (83) consists of applying proposition III.17 in particular eq. (83), to eq. (82) for every value of $Y$. Eq. (85) follows from applying theorem III.17 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 non-selective 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 $\varrho_{Y|B,A}$ 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 non-selective state-update rule.

Table III 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 $\rho_{B|A}$
gives the alternative form of Bayes’ theorem and the quantum analog of Bayes’ theorem for acausal conditional states. In [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 [24]. This 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 [24, 26] and a quantum analog of Bayes’ theorem previously advocated by Fuchs [22]. As an application of the acausal analog of the symmetry between prediction and interference. In  and a quantum analog of Bayes’ theorem is developed. The converse of this relation, wherein is also a valid state. Furthermore, given that is a valid causal conditional state, and the fact that the set of such states is mapped to itself by the transpose, is also a valid state. 

In order for this to make sense, it must be checked that is indeed a valid causal conditional state. Taking the partial transpose over , we have 

and which can also be written as 

In the converse of this relation, wherein is associated, via the Jamiołkowski isomorphism, with a quantum channel . Then, eq. (93) is equivalent to 

The converse of this relation, wherein is expressed in terms of , is obtained by simply exchanging and as well as and . The proof that eq. (93) translates into eq. (95) is straightforward. From the Jamiołkowski isomorphism
rewritten as

\[ \mathcal{F}_{A|B}(\cdot) = \text{Tr}_B \left( \varrho_{A|B}(\cdot) \right), \]

which implies that

\[ \mathcal{F}_{A|B}(\cdot) = \text{Tr}_B \left( \rho_A^+ \mathcal{E}_{B|A} \rho_B^+ \rho_A^+ \right) \]

where the first step follows from eq. (94) and expanding the \( \star \)-product, and the second step uses the cyclic property of the trace. Eq. (99) then follows from the representation of dual maps in terms of conditional states given in theorem III.17.

The map \( \mathcal{F}_{A|B} \) is recognizable as the Barnum–Knill recovery map for the channel \( \mathcal{E}_{B|A} \) [27]. This map is known to achieve near-optimal quantum error correction in situations where the input state and channel are known. To our knowledge, its connection with Bayesian inversion has not previously been noted. It 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 \( \mathcal{E}_{B|A} \) and the one associated to its Bayesian inversion \( \mathcal{F}_{A|B} \) is \( \mathcal{S}_{A|B} \) (recall that only \( \mathcal{E}_{B|A} \circ T_A \) and \( \mathcal{S}_{A|B} \circ T_A \) are CPT maps). Following the same reasoning as above, eq. (101) can be rewritten as

\[ \mathcal{S}_{A|B}(\cdot) = \rho_A^+ \left( \mathcal{E}_{A|B} \left( \rho_B^+ \cdot \rho_B^+ \right) \right) \rho_A^+, \]

where

\[ \rho_B = \mathcal{E}_{B|A}(\rho_A). \]

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

\[ \sigma_{X|A} = \sigma_{X|A} \cdot \rho_A. \]

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 III.11 that a general hybrid conditional state \( \sigma_{A|X} \) is of the form

\[ \sigma_{A|X} = \sum_x |x\rangle \langle x| \otimes \rho_x^A, \]

where each \( \rho_x^A \) is a normalized density operator on \( \mathcal{H}_A \), and from theorem III.13 that \( \sigma_{X|A} \) is of the form

\[ \sigma_{X|A} = \sum_x |x\rangle \langle x| \otimes E_x^A, \]

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

Eq. (101) therefore gives a decomposition of a joint state in terms of a set of states and a probability distribution via

\[ \sigma_{X|A} = \sum_x P(X = x) |x\rangle \langle x| \otimes \rho_x^A, \]

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

\[ \sigma_{X|A} = \sum_x |x\rangle \langle x| \otimes E_x^A \rho_A. \]

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, whilst eq. (103) is a rule for determining a set of states from a POVM and a state on \( A \). These rules are:

\[ \rho_x^A = \frac{\rho_A E_x^A \rho_A^+}{\text{Tr}_A (E_x^A \rho_A)}, \]

and

\[ E_x^A = P(X = x) \rho_A \frac{1}{\rho_x^A} \rho_x^A \rho_A^+. \]

These rules have appeared numerous times in the literature, e.g. [8, 22]. In the context of distinguishing the states in an ensemble, the POVM defined by eq. (111) is known as the “pretty good” measurement. Eq. (110) is a rule previously advocated by Fuchs as a quantum analog of Bayes’ theorem [22]. 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 will be considered in Appendix A.

Barnett et al. have previously proposed a formalism for retrodictive inference in quantum theory [28–30]. 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 $\rho_A$ of an early region is mapped to the state $\rho_B$ of a later region by a CPT map $\mathcal{E}_{B|A}$ then

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

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 $\rho_A$ and $\rho_B$ as its marginals, so we have

$$\rho_A = \text{Tr}_B \left( \varrho_{A|B} \rho_B \right).$$

In conventional notation, this is equivalent to

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

where

$$\mathcal{E}^{\text{retr}}_{A|B} (\cdot) \equiv \rho_A \mathcal{E}^\dagger_{A|B} \left( \rho_B^{-1} \mathcal{E}^\dagger_{A|B} (\cdot) \right) \rho_B^{-1}.$$

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.

Predictive expression:

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

Conventional notation:

$$p_{x,y} = \text{Tr}_A \left( E^A_y \rho_A^x \right) P(X = x)$$

Retrodictive expression:

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

Conventional notation:

$$p_{x,y} = \text{Tr}_A \left( E^A_x \varrho_{Y|X} \star \rho^y \right) P(Y = y)$$

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.

For comparison with the retrodictive formalism of [28], 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 alternatively be represented by a diagonal state $\rho_X$, and by a set of states $\rho_A^x$, which can alternatively be represented by a causal conditional state $\varrho_{A|X}$. The measurement is characterized by a POVM $\{ E^A_x \}$, or alternatively 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 $\varrho_{XY}$, 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 $\varrho_{Y|X} = \text{Tr}_A \left( \varrho_{Y|A} \varrho_{A|X} \right)$.

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

Combining these steps, the predictive expression for $\rho_{XY}$ is

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

In conventional notation, if the states $\{ \rho_A^x \}$ are the components of $\varrho_{A|X}$ and the elements of the POVM $\{ E^A_x \}$ are the components of $\varrho_{Y|A}$, then eq. (116) is equivalent to

$$P(X = x, Y = y) = \text{Tr}_A \left( E^A_y \rho^x \right) P(X = x).$$

A retrodictive description of the same experiment can be given, involving states of regions conditioned on regions in their immediate future, i.e. $\rho_Y^x$, $\varrho_{A|Y}$ and $\varrho_{X|A}$. 
These correspond, respectively, to a probability distribution $P(Y)$, a set of states $\{\rho^A_y \}$, and a POVM $\{E^A_x, retr\}$. 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, whilst the preparation is being described by a POVM, which we call the retrodictive POVM. The retrodictive calculation of $\varrho_{XY}$ proceeds as follows:

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

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

Combining these steps, the retrodictive expression for $\varrho_{XY}$ is

$$\varrho_{XY} = \text{Tr}_A (\varrho_{X|A} \varrho_{A|Y}) \ast \rho_Y.$$  \hfill (118)

In conventional notation, this is equivalent to

$$P(X = x, Y = y) = \text{Tr}_A (E^A_x^{\text{retr}} \rho^A_y) P(Y = y).$$  \hfill (119)

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

$$\varrho_{X|A} = \varrho_{A|x} \ast (\rho_x \rho_A^{-1}),$$  \hfill (120)

where $\rho_A = \text{Tr}_X (\varrho_{A|X} \rho_X)$, \hfill (121)

and

$$\varrho_{A|Y} = \varrho_{Y|A} \ast (\rho_Y \rho_A^{-1}),$$  \hfill (122)

where $\rho_Y = \text{Tr}_A (\varrho_{Y|A} \rho_A)$. \hfill (123)

In conventional notation, these equations are equivalent to

$$E^A_x^{\text{retr}} = P(X = x) \rho_A^{-\frac{1}{2}} \rho_x \rho_A^{-\frac{1}{2}},$$  \hfill (124)

where $\rho_A = \sum_x P(X = x) \rho_x A$, \hfill (125)

and

$$\rho^A_y^{\text{retr}} = \frac{\rho_A E^A_y \rho_A}{P(Y = y)},$$  \hfill (126)

where $P(Y = y) = \text{Tr}_B (E^A_y \rho_A)$. \hfill (127)

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

$$\varrho_{XY} = \text{Tr}_A (\varrho_{X|A} \varrho_{A|Y}) \ast \rho_Y = \text{Tr}_A \{ (\varrho_{A|x} \ast \rho_x \rho_A^{-1}) \} \ast \rho_Y$$

$$= \text{Tr}_A \left( \frac{1}{2} \rho_A \rho_Y^{-\frac{1}{2}} \varrho_{A|x} \rho_A \rho_Y^{-\frac{1}{2}} \rho_Y A \right)$$

$$= \text{Tr}_A \left( \rho_A \rho_Y^{-\frac{1}{2}} \varrho_{A|x} \rho_A \rho_Y^{-\frac{1}{2}} \rho_Y A \right)$$

Since $\rho_Y$ commutes with $\rho_x$, $\rho_A$ and $\varrho_{A|x}$, the $\rho_Y$ term can be moved forward to cancel with $\rho_Y^\ast$ term. The $\rho_A^\ast$ term can be made to cancel with the last $\rho_A^\ast$ term via the cyclic property of the trace. This yields

$$\varrho_{XY} = \text{Tr}_A \left( \rho_A \rho_Y^\ast \varrho_{A|x} \rho_A \rho_Y^\ast \varrho_{Y|A} \rho_A \right) = \text{Tr}_A (\varrho_{Y|A} \varrho_{A|x}) \ast \rho_X,$$  \hfill (137)

where, in the second line, we have used the fact that $\rho_X$ commutes with $\rho_Y$. Finally, applying the cyclic property of the trace gives

$$\varrho_{XY} = \text{Tr}_A (\varrho_{Y|A} \varrho_{A|x}) \ast \rho_X.$$  \hfill (138)

which is the predictive expression for $\varrho_{XY}$.

We are now in a position to show that our formalism coincides with that of Ref. \cite{28} 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 \cite{28} were

$$E^A_x^{\text{retr}} = dP(X = x) \rho_x A,$$  \hfill (139)

and

$$\rho^A_y^{\text{retr}} = \frac{E^A_y}{P(Y = y)},$$  \hfill (139)

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. \ref{fig11}, in which there is an intervening channel $E_{B|A}$ between the preparation 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. (122) (or equivalently the Heisenberg dynamics given in eq. (100)), the extra region $B$ can be eliminated from the description by defining

$$\varrho_{Y|A} = \text{Tr}_A (\varrho_{Y|B} \varrho_{B|A}),$$  \hfill (140)

which determines the effective measurement on $A$ that is performed by actually measuring region $B$. The three
operators \(\rho_X\), \(\varrho_{A|X}\), 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. (118) into eq. (116) gives the predictive expression for \(\varrho_{XY}\) as

\[
\varrho_{XY} = \text{Tr}_{AB} \left( \varrho_{Y|B}\varrho_{B|A}\varrho_{A|X} \right) * \rho_X,
\]

or in conventional notation

\[
P(X = x, Y = y) = \text{Tr}_B \left( E_y^B \varrho_{B|A} \left( \rho_x^A \right) \right) P(X = x).
\]

Predictive expression:

\[
\varrho_{XY} = \text{Tr}_B \left( \varrho_{Y|B}\text{Tr}_A \left( \varrho_{B|A}\varrho_{A|X} \right) \right) * \rho_X
\]

Conventional notation:

\[
p_{x,y} = \text{Tr}_B \left( E_y^B \varrho_{B|A} \left( \rho_x^A \right) \right) P(X = x)
\]

Retrodictive expression:

\[
\varrho_{XY} = \text{Tr}_A \left( \varrho_{X|A}\text{Tr}_B \left( \varrho_{A|B}\varrho_{B|Y} \right) \right) * \rho_Y
\]

Conventional notation:

\[
p_{x,y} = \text{Tr}_A \left( E_x^A \text{retr} \varrho_{Y|B} \left( \rho_y^B \text{retr} \right) \right) P(Y = y)
\]

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.

Similarly, the retrodictive expression is obtained from eq. (118) by substituting \(\varrho_{A|Y} = \text{Tr}_B \left( \varrho_{A|B}\varrho_{B|Y} \right)\), where \(\varrho_{A|B}\) and \(\varrho_{B|Y}\) are obtained from \(\varrho_{B|A}\) and \(\varrho_{Y|B}\) by Bayes’ theorem. This gives

\[
\varrho_{XY} = \text{Tr}_{AB} \left( \varrho_{X|A}\varrho_{A|B}\varrho_{B|Y} \right) * \rho_Y,
\]

or in conventional notation

\[
P(X = x, Y = y) = \text{Tr}_A \left( E_x^A \text{retr} \varrho_{Y|B} \left( \rho_y^B \text{retr} \right) \right) P(Y = y).
\]

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. (145), for the formula for propagating them across causally related regions, eq. (142), 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

\[
\rho_B = \text{Tr}_A \left( \rho_{B|A}\rho_A \right)
\]

then

\[
\rho_A = \text{Tr}_B \left( \rho_{A|B}\rho_B \right).
\]

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.

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

In the conditional states formalism, the region \(AB\) is assigned a joint state \(\rho_{AB}\), 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 causal joint state \(\rho_{XY}\) via

\[
\rho_{XY} = \text{Tr}_{AB} \left( \left( \varrho_{X|A}\varrho_{Y|B} \right) \rho_{AB} \right).
\]
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 shall refer to these as rightward and leftward belief propagation respectively.

It may seem convoluted to calculate $\rho_{XY}$ via rightward or leftward belief propagation, when eq. (147) 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 will be considered in §V.A.3.

First, consider rightward belief propagation. The aim is to rewrite eq. (147) 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 $\rho_X$ and $\varrho_{A|X}$ to describe the left hand wing of the experiment, whilst retaining $\rho_{Y|B}$ for the right. Then, we write $\rho_{AB} = \rho_{B|A} \rho_A$ and note that, by Bayes’ theorem,

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

Substituting these into eq. (147) gives

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

Expanding the $\ast$-products gives

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

(150)

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

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

(151)

$$= \text{Tr}_{AB} \left( \rho_{Y|B} \rho_{B|A} \rho_{A|X} \rho_X \right)$$

(152)

$$= \text{Tr}_{AB} \left( \rho_{Y|B} \rho_{B|A} \rho_{A|X} \right) \ast \rho_X,$$

(153)

where we have used the cyclic property of the trace and the fact that $\rho_X$ commutes with $\rho_{B|A}$ and $\varrho_{Y|B}$.

Eq. (153) has the same form as the predictive expression for a prepare-and-measure experiment with an intervening channel given in eq. (141), except that, in this case, $\rho_{B|A}$ is acausal. We can also use eq. (150) to define $\rho_{Y|A} = \text{Tr}_B \left( \rho_{Y|B} \rho_{B|A} \right)$, which represents the effective measurement on region $A$ that is made by measuring region $B$. Using this, region $B$ can be eliminated from eq. (153) to obtain

$$\rho_{XY} = \text{Tr}_{AB} \left( \rho_{Y|A} \rho_{A|X} \right) \ast \rho_X.$$  (154)

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. Eq. (153) 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

$$\rho_{XY} = \text{Tr}_{AB} \left( \varrho_{X|A} \rho_{B|Y} \right) \ast \rho_Y,$$

(155)

which is analogous to eq. (143) and

$$\rho_{XY} = \text{Tr}_A \left( \rho_{X|A} \rho_{B|Y} \right) \ast \rho_Y,$$

(156)

which is analogous to eq. (118).

As with prediction and retrodictive, 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 $R$ to the posterior $P_r(R) = P(R | X = x)$.

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

$$P_x(X = x') = \delta_{x,x'}.$$  (157)

Secondly, 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

$$P_x(R) = \sum_X P(R | X) P_x(X)$$

(158)

$$= P(R | X = x).$$  (159)

This two step decomposition of conditioning has been emphasized by Richard Jeffrey [35, 36] 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 [37], 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 \( X = \text{‘red’} \) and 1/6 each to \( X = \text{‘green’} \) and \( X = \text{‘yellow’} \). Assuming that the observation does not cause your conditional probabilities \( P(R|X) \) to change, Jeffrey shows that your posterior probability distribution for \( R \) is obtained by belief propagation via

\[
P_{\text{post}}(R) = \sum_X P(R|X) P_{\text{post}}(X),
\]

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 [38]:

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 [39].

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 \( \rho_A \) to \( \rho_A^{\text{post}} \), 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

\[
\rho_B^{\text{post}} = \text{Tr}_A \left( \rho_{B|A} \rho_A^{\text{post}} \right)
\]

or

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

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 \( P_x(X = x^\prime) = \delta_{x,x^\prime} \). 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 [40].

We will not pursue this question further here, but instead focus on the case of a hybrid region \( XA \). If 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

\[
\rho_A \rightarrow \sigma_{A|X=x}.
\]

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

\[
\rho^A_x = \text{Tr}_X \left( \sigma_{A|X} \rho_X^X \right), \quad (164)
\]

\[
= \langle x \rvert_X \sigma_{A|X} \left| x \right\rangle_X, \quad (165)
\]

\[
= \sigma_{A|X=x}, \quad (166)
\]

where we’ve made use of the fact that \( \sigma_{A|X} = \sum_x \sigma_{A|X=x} \otimes \left| x \right\rangle \left\langle x \right| \). 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 \( \rho_A \) is updated to an element \( \sigma_{A|X=x} \) within a convex component decomposition of \( \rho_A \).

### A. Examples of Quantum Bayesian Conditioning

In this section, we consider several examples of Bayesian conditioning, based on the different causal scenarios discussed in \( \text{III} \) and \( \text{IV} \). 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 \( \text{IV} \) 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 \( \{ \theta_{A|X=x} \} \) depending on the value of a classical variable \( X \) with prior probability distribution \( P(X) \). This can alternatively be described by a conditional state \( \theta_{A|X} \) and a diagonal state \( \rho_X \). In this case, Bayesian conditioning of \( A \) on the value \( x \) of \( X \) corresponds to updating from the ensemble average state \( \rho_A = \sum_x P(X = x) \theta_{A|X=x} \) to the particular state \( \theta_{A|X=x} \) 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 \( \theta_{XY} \). We now have a prepare-and-measure experiment, as depicted in fig. 3 and eq. (159) gives the predictive expression for the joint probability distribution for this experiment as

\[
\theta_{XY} = \text{Tr}_A \left( \theta_{Y|A} \theta_{A|X} \right) \otimes \rho_X. \quad (167)
\]

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

\[
\rho_Y = \text{Tr}_A \left( \theta_{Y|A} \rho_A \right), \quad (168)
\]

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

\[
\rho_{Y|X} = \text{Tr}_A \left( \theta_{Y|A} \theta_{A|X} \right) \rho_X, \quad (169)
\]

from which it follows that

\[
\rho_{Y|X=x} = \text{Tr}_A \left( \theta_{Y|A} \theta_{A|X=x} \right). \quad (170)
\]

The transition from eq. (168) to eq. (170) 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. (170) can be obtained from eq. (168) by replacing \( \theta_A \) with \( \theta_{A|X=x} \), 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 \( \text{IV} \), the joint probability is still given by eq. (167), where now \( \theta_{Y|A} = \text{Tr}_B \left( \theta_{Y|B} \theta_{B|A} \right) \) 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

Recall the measurement scenario depicted in fig. 1(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 \( \rho_A \) and a causal conditional state \( \theta_{Y|A} \) (describing the measured POVM). To condition \( A \) on a value \( y \) of \( Y \), Bayes’ theorem must be applied to determine \( \theta_{A|Y=y} \), and then the component \( \theta_{A|Y=y} \) gets picked out by conditioning. This gives

\[
\rho_A \rightarrow \theta_{A|Y=y} = \theta_{Y=y|A} \ast \left( \rho_A \rho_Y^{-1} \right), \quad (171)
\]

or in conventional notation

\[
\rho_A \rightarrow \rho^{A\text{retr}}_A = \frac{\rho^{\frac{1}{2}}_A E^A_y \rho^{\frac{1}{2}}_A}{\text{Tr}_A \left( E^A_y \rho_A \right)}. \quad (172)
\]

The state \( \theta_{A|Y=y} \) represents the state of a region \( A \) prior to the measurement, given the outcome of the measurement, i.e. it is a retrodiction 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 of fig. [10]. Recall from [IVC] that there is complete symmetry between the predictive and retrodictive expressions for a prepare-and-measure experiment under exchange of the 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

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

(173)

and conditioning this on $Y = y$ gives

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

(174)

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 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. [5]. This experiment is described by an acausal joint state $\rho_{AB}$, 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 [IVD] 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 $\varrho_{Y|B}$ 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 [23, 33, 41–46]. The conditional states formalism provides an elegant approach to this problem.

Recall from [IVD] that the joint probability for $X$ and $Y$ can be computed via leftward belief propagation, which yields eq. (156), i.e.

$$\rho_{XY} = \text{Tr}_A \left( \varrho_{X|A} \rho_A |Y \right) * \rho_Y.$$  

(175)

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

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

(176)

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

Eq. (175) 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

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

(177)

and conditioning on $Y = y$ gives

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

(178)

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 V.1.** Let $\rho_{AB}$ 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 $\rho_A = \text{Tr}_B \left( \rho_{AB} \right)$ to $\rho_{A|Y = y}$, where

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

(179)

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

and

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

(180)

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

$$\rho^A_y = E_{A|B} \left( \rho_y^B \right),$$

(181)

where

$$\rho_y^B = \left( \rho_B \right)^{1/2} E_y \rho_B \left( \rho_B \right)^{-1/2}$$

(182)
In the above analysis, the method used to compute $\rho_{A|Y=y}$ is to first apply Bayes’ theorem to $\rho_{Y|B}$ 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

$$\rho_{Y|A} = \text{Tr}_B \left( \rho_{Y|B} \rho_{B|A} \right),$$  \hspace{1cm} (183)

and then apply Bayes’ theorem to obtain

$$\rho_{A|Y=y} = \rho_{Y=y|A} \ast \left( \rho_A \rho_{Y}^{-1} \right).$$  \hspace{1cm} (184)

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

$$E_{y}^{A} = \mathcal{C}_{A|B}^{dagger} \left( E_{y}^{B} \right),$$  \hspace{1cm} (185)

and then determining the updated state via

$$\rho_{y}^{A} = \frac{\rho_{A}^{\frac{1}{2}} \mathcal{E}^{A}_{y} \rho_{A}^{\frac{1}{2}}}{\text{Tr}_B \left( E_{y}^{B} \rho_B \right)}. \hspace{1cm} (186)$$

Expressions equivalent to these conventional formulas have appeared previously in [33], the only difference being the appearance of a transpose, due to the use of the Choi isomorphism rather than the Jamiołkowski isomorphism in [33].

We have seen that the remote collapse $\rho_A \rightarrow \rho_{A|Y=y}$ 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 [47]. 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 [18], 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 [22]. 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.

\section*{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. [31, 32]). 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 [49, 50]. 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 \( \rho_A \) should be updated via
\[
\rho_A \rightarrow \frac{\Pi_{\rho_A}^A \Pi_{\rho_y}^A}{\text{Tr}_A (\Pi_{\rho_y}^A \rho_A)}.
\] (187)

For a general POVM \( \{ E_y^A \} \), there is a natural generalization of the projection postulate, given by
\[
\rho_A \rightarrow \frac{(E_y^A)\rho_A (E_y^A)^\dagger}{\text{Tr}_A (E_y^A \rho_A)}.
\] (188)

This generalization has also been proposed as a quantum analog of Bayesian conditioning \[51,52\].

On the other hand, in fig. 13(b) the rule for conditioning a state on the outcome of a measurement was found to be
\[
\rho_A \rightarrow \frac{\rho_A^\frac{1}{2} E_y^A \rho_A^\frac{1}{2}}{\text{Tr}_A (E_y^A \rho_A^\frac{1}{2})}.
\] (189)

This is distinct from eq. (188) because the roles of \( \rho_A \) and \( E_y^A \) have been interchanged. Furthermore, eq. (188) is not equivalent to an equation of the form of eq. (189) (even allowing a different POVM to appear therein) because the map \( \rho_A \rightarrow (E_y^A)^\dagger \rho_A (E_y^A)^\dagger \), like all quantum instruments, is linear on the set of states on \( \mathcal{H}_A \), whereas the map \( \rho \rightarrow \rho_A^\frac{1}{2} E_y^A \rho_A^\frac{1}{2} \) 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 shall 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. 13(a). A quantum system located in region \( A \) and described by the state \( \rho_A \) 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 \[III\], a quantum instrument, \( \{ \mathcal{E}_{y'}^{A'} \} \), determines how the state of \( A' \) is related to the state of \( A \), where \( \mathcal{E}_{y'}^{A'} (\rho_A) \) is the unnormalized state of \( A' \) obtained when \( Y = y \).

Now consider the classical analogue of this scenario, depicted in fig. 13(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. 13(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 will be called its initial configuration and its configuration at the time after the occurrence of the measurement its final configuration. \( R \) is the initial configuration and \( R' \) the final. 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 \textit{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 \( \text{BC1} \), which maps the prior distribution of the initial configuration of the system to the posterior distribution

\[
\text{BC1: } P(R) \rightarrow P(R'|R = y)
\]

\[
\text{BC2: } P(R') \rightarrow P(R'|Y = y)
\]

\[
\text{QU: } P(A) \rightarrow P(A'|Y = y)
\]
of the initial configuration of the system, is an instance of Bayesian conditioning (it is analogous to updating a retrodictive state as discussed in \(\{V A 2\}\)). 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 of the system. 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 \(\varrho_{Y, A' | A}\) that corresponds to the quantum instrument \(\{\xi_{y}^{A' | A}\}\).

In the quantum conditional states formalism, there are analogues of each of the three rules we considered above. These are specified in fig. \(\{3(a)\}\). The rule QBC1 corresponds to updating a retrodictive state, as considered in \(\{V A 2\}\). 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: \(P(R) \rightarrow P(R' | Y = y)\). This can be obtained by combining an instance of belief propagation, namely

\[
P(R) \rightarrow P(R') = \sum_{R} P(R' | R) P(R),
\]

followed by the rule BC2: \(P(R') \rightarrow P(R' | Y = y)\), 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. \(190\) 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

\[
P(R') \rightarrow P(R' | Y = y) = \frac{\sum_{R} P(Y = y, R' | R) P(R)}{P(Y = y)}. \tag{191}
\]

The quantum analogue of this is straightforward. Quantum state update rules, such as the projection postulate, are of the form QU: \(\varrho_{A} \rightarrow \varrho_{A' | Y = y}\). This is simply a sequential combination of quantum belief propagation

\[
\varrho_{A} \rightarrow \varrho_{A'} = \text{Tr}_{A} \left( \varrho_{A' | A} \rho_{A} \right), \tag{192}
\]

with quantum Bayesian conditioning via QBC2: \(\varrho_{A} \rightarrow \varrho_{A' | Y = y}\).

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 \(\varrho_{Y, A' | A}\) and the prior \(\varrho_{A}\). The causal conditional state \(\varrho_{Y, A' | A}\) used in eq. \(192\) is simply a reduced state of \(\varrho_{Y, A' | A}\), i.e. \(\varrho_{Y, A' | A} = \text{Tr}_{Y} \left( \varrho_{Y, A' | A} \right)\). To gain some intuition for this step, we translate it into conventional notation. If the quantum instrument associated with \(\varrho_{Y, A' | A}\) is denoted \(\{\xi_{y}^{A' | A}\}\), then eq. \(192\) becomes

\[
\varrho_{A} \rightarrow \varrho_{A'} = \xi_{y}^{A' | A}(\varrho_{A}), \tag{193}
\]

where

\[
\xi_{y}^{A' | A} = \sum_{y} \xi_{y}^{A' | A}, \tag{194}
\]

The map \(\xi_{y}^{A' | A}\) is the non-selective 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 non-selective update map is just an instance of quantum belief propagation.

Turning to the Bayesian conditioning step, we have \(\varrho_{A' | Y = y} = \varrho_{A' | Y } * \rho_{Y}^{-1}\), where \(\varrho_{A' | Y} = \text{Tr}_{A} \left( \varrho_{Y, A' | A} \rho_{A} \right)\). Combining these and setting \(Y = y\) gives

\[
\varrho_{A' \rightarrow \varrho_{A' | Y = y}} = \text{Tr}_{A} \left( \varrho_{y = y, A' | A} \rho_{A} \right) * \rho_{y = y}^{-1}. \tag{195}
\]

In conventional notation, this translates into

\[
\varrho_{A'} \rightarrow \varrho_{y}^{A'} = \frac{\xi_{y}^{A' | A}(\varrho_{A})}{P(Y = y)}. \tag{196}
\]

Given the expression for \(\varrho_{A'}\) in eq. \(193\), we see that QBC2 is simply a transition from your prior about the system output by the measurement, the result of applying the non-selective 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 non-selective to selective updates should be regarded as analogous to Bayesian conditioning has previously been argued by Ozawa. 

In fact, the rule $\text{QBC2}$ is a particular example of the kind of Bayesian conditioning considered in [54, 55]. 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 $\varrho_{Y'A|A} = \varrho_{YA'|A} \otimes \rho_A$ by tracing over $A$ and conditioning on $Y$. The rule $\text{QBC2}$ 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 $\text{U}$ nor the quantum state update rule $\text{QU}$ 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. (190) has the form

$$P(R' = r'|R = r) = \delta_{y', r}, \quad (197)$$

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

If $P(R'|R)$ is just a delta function, then we say that the measurement is non-disturbing. 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 non-disturbing, namely,

$$P(Y, R' = r'|R = r) = P(Y|R = r)\delta_{y', r}. \quad (198)$$

Of course, there are also many ways of implementing a measurement of $P(Y|R)$ such that it is disturbing. Therefore, whilst the update rule $\text{U}$ 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 non-disturbing way, such that the associated quantum update rule $\text{QU}$ (perhaps the projection postulate, perhaps some other rule) is effectively just quantum Bayesian conditioning. For the measurement to be non-disturbing, the causal conditional state $\varrho_{A'|A}$ in the belief propagation rule of eq. (192) would have to be of the form

$$\varrho_{A'|A} = \sum_{j,k} |j\rangle_A \langle k| \otimes |j\rangle_{A'}, \quad (199)$$

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 $\text{QU}$ 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 $\text{QU}$ 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 $\{P(Y = y)|_A\}$, 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 $\text{QU}$ can never be effectively just Bayesian conditioning because, unlike the classical case, in quantum theory information gain necessarily implies a disturbance. This prevents any $\text{QU}$ rule, such as the projection postulate, from being pure Bayesian conditioning. Whilst this fact is well-known, it is instructive to prove it in the conditional states formalism.

**Theorem V.2** (No information gain without disturbance). Consider a measurement described by an instrument associated with the causal conditional state $\varrho_{Y A'|A}$. It is impossible for this measurement to be both informative about $A$ ($\varrho_{Y|A} \neq \rho_Y$) and non-disturbing ($\varrho_{A'|A} = \sum_{j,k} |j\rangle_A \otimes |j\rangle_{A'}$).

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

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

$$\rho_{Y A'|A} = \rho_{Y A'|A} \otimes I_A/d = \rho_{Y A'|A}/d, \quad (200)$$

where $d$ is the dimension of $\mathcal{H}_A$. The condition that the measurement be non-disturbing is equivalent to $\rho_{A'|A} = \sum_{j,k} |j\rangle_A \otimes |j\rangle_{A'}$, which implies that the tripartite...
state $\rho_{Y'A'}$ 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 $\rho_{Y'A'}$ should not be a product state. But, by the monogamy of entanglement, any tripartite state $\rho_{Y'A'}$ for which $\rho_{AA'}$ is maximally entangled must have a product state for its reduced state $\rho_{Y|A}$. 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: $P(R) \rightarrow P(R|Y=y)$, and then propagate the conditioned state via $P(R|Y=y) \rightarrow P(R'|Y=y) = \sum R P(R'|R,Y=y)P(R|Y=y)$.

It is natural to ask whether such a reverse-order decomposition is possible in the quantum case. That is, can QU: $\rho_A \rightarrow \varrho_{A|Y=y}$ be decomposed into QBC1: $\rho_A \rightarrow \varrho_{A|Y=y}$, followed by belief propagation $\varrho_{A|Y=y} \rightarrow \varrho_{A'|Y=y}$? Perhaps surprisingly, this cannot be done. The belief propagation would have to have the form $\varrho_{A'|Y=y} = \text{Tr}_A(\varrho_{A'|A,Y=y}\varrho_{A|Y=y})$. But 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 conditionedized form of the quantum Bayes’ theorem, i.e. $\varrho_{A'|A,Y=y} = \varrho_{Y=y,A'|A} * \varrho_{Y=y|A}$. Unfortunately, in the conditional states formalism, valid equations do not necessarily remain valid when we conditionize each term (this is discussed further in VIII). In particular, the conditionedized 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.

Firstly, Cerf and Adami have proposed an alternative definition of an acausal conditional state $\rho_{(\infty)}^{(B|A)}$ (their definition does not extend to the causal case). For a bipartite state $\rho_{AB}$, the Cerf-Adami conditional state is

$$\rho_{B|A}^{(\infty)} = \exp(\log \rho_{AB} - \log \rho_A \otimes I_B).$$

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

$$S(B|A) = -\text{Tr}_{AB}(\rho_{AB} \log \rho_{B|A}^{(\infty)}),$$

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

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

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

In [53], Leifer and Poulin introduced a family of conditional states, again restricted to the acausal case, indexed by a positive integer $n$ and given by

$$\rho_{B|A}^{(n)} = \left(\rho_{AB}^{*} \rho_{A}^{-1}\right)^{n}.$$

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

The main concern of [53] 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 [54]. 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 [12] for the original paper and [14–16] for reviews). This was proposed in the context of quantum probability theory, which is a non-commutative 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 (see also [5]).

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 $\Phi_{R|S}$ from functions of $R$ and $S$ to functions of $R$ that satisfies

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

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

$$\Phi_{R|S}(f(R,S)) = \sum_{S} P(S|R)f(R,S),$$

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. (200). 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

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

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

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

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 $\Phi_{R|S}$ has a dual map $\mathcal{E}_{R|S}$ that maps the space of probability distributions over $R$ to the space of probability distributions over $R$ and $S$. This is given by

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

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

$$\sum_{S} \mathcal{E}_{R|S}(Q(R)) = Q(R)$$

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 linear map $\Phi_{A|B}: \mathfrak{L}(H_{A}) \rightarrow \mathfrak{L}(H_{A})$ that acts on the set of observables on $H_{AB}$ and satisfies

$$\Phi_{A|B}(M_{A} \otimes I_{B}) = M_{A}$$

for all operators $M_{A} \in \mathfrak{L}(H_{A})$. This is analogous to the condition given in eq. (200). The dual map $\mathcal{E}_{A|B}: \mathfrak{L}(H_{A}) \rightarrow \mathfrak{L}(H_{AB})$ acts on states and is a state extension, which means that

$$\tau_{AB} = \mathcal{E}_{A|B}(\tau_{A})$$

is a valid state for any state $\tau_{A}$ on $H_{A}$. In addition, the state extensions that are dual to conditional expectations satisfy

$$\text{Tr}_{B}(\mathcal{E}_{A|B}(\tau_{A})) = \tau_{A}$$

for every state $\tau_{A}$ on $H_{A}$, which is analogous to eq. (210).

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

$$\mathcal{E}_{A|B}(\rho_{A}) = \rho_{AB},$$

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. (213) 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.

Irrationally, the easiest way to show that $\rho_{AB}$ 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 prime is just used to distinguish the input and output spaces. Then, eq. (214) can be rewritten as

$$\rho_{AB} = \text{Tr}_{A'} \left( \varrho_{AB|A'} \rho_{A'} \right),$$

where $\rho_{A'}$ is the same state as $\rho_A$. Similarly, eq. (215) can be rewritten as

$$\tau_A = \text{Tr}_{A'B} \left( \varrho_{AB|A'} \tau_{A'} \right) \tag{216}$$

$$\tau_A = \text{Tr}_{A'B} \left( \varrho_{AB|A'} \tau_{A'} \right), \tag{217}$$

for all $\tau_A$, where $\tau_{A'}$ is the same state as $\tau_A$ and $\varrho_{A|A'} = \text{Tr}_B \left( \varrho_{AB|A'} \right)$. Since, eq. (217) has to hold for every input state, $\varrho_{A|A'}$ has to be Jamiołkowski isomorphic to the identity superoperator, so $\varrho_{A|A'} = |\Phi^+\rangle \langle \Phi^+ | \tau_{A'}$. Since this is pure, monogamy of entanglement [31] entails that $\varrho_{AB|A'}$ must be of the form $\varrho_{AB|A'} = |\Phi^+\rangle \langle \Phi^+ | \tau_{A'} \otimes M_B$ for some operator $M_B$ on $\mathcal{H}_B$. Substituting this into eq. (215) gives

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

which shows that $\rho_{AB}$ must be a product state and $M_B = \text{Tr}_B (\rho_{AB})$. 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 $\mathcal{F}_{AB|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{AB})$, 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

$$\mathcal{F}_{AB|A} (\rho_A) = \hat{\rho}_A \rho_B \hat{\rho}_A. \tag{219}$$

The nonlinearity allows this map to satisfy eqs. (213) and (214) 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.

Rényi’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 $\rho_A$ to $\rho_A^\text{post}$, then Rédei asserts that the state for $AB$ should be updated to $\rho_{AB}^\text{post} = \mathcal{E}_{AB|A} (\rho_A^\text{post})$, where $\mathcal{E}_{AB|A}$ is a state extension derived from the prior state $\rho_{AB}$. 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 theory — have much in common with the work presented here. Particularly cognate to our approach is the work of Hardy [57–59], the Pavia group [60, 61], Oreshkov et al. [62] and Coecke’s group [63, 64].

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 11.7, 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 [59], 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 [57, 53]) 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 also represents correlations between acausally related and causally related regions in 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 non-metrical parts) is reformulated using the mathematical framework of symmetric monoidal categories and the graphical calculus that can be defined for these systems. Systems are the objects of the category, while quantum states and quantum operations are the morphisms. The isomorphism between bipartite states with regard to possible causal structures.

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 non-metrical parts) is reformulated using the mathematical framework of symmetric monoidal categories and the graphical calculus that can be defined for these systems. 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.

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 non-metrical parts) is reformulated using the mathematical framework of symmetric monoidal categories and the graphical calculus that can be defined for these systems. 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.

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 spatio-temporal 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.

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 and of Glymour, Scheines and Spirtes. We ultimately hope to generalize this analysis of causality by replacing classical probability theory with quantum theory, understood as a theory of Bayesian inference. 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, whilst not having any intrinsic causal structure themselves.

VII. LIMITATIONS OF THE ⋆-PRODUCT

Using the ⋆-product makes the conditional states formalism look very similar to classical probability theory. Often, by replacing probabilities with operators and ordinary products with ⋆-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 ⋆-product is nonassociative and noncommutative, this does not always happen. In this section, the limitations of the ⋆-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 equation relating conditional, joint and marginal probability distributions over a set of variables R, S, . . . and conditionalizes every term on a disjoint variable T, then the equation that results is still universally
valid. For example, the equation
\[ P(R, S) = P(S|R)P(R) \] (220)
generalizes to
\[ P(R, S|T) = P(S|R, T)P(R|T), \] (221)
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. (221) would be
\[ \rho_{AB|C} = \rho_{B|AC} \star \rho_{A|C}. \] (222)
Writing this out explicitly, the left hand side is
\[ \rho_{AC}^{-\frac{1}{2}} \rho_{ABC} \rho_{AC}^{-\frac{1}{2}}, \] (223)
whereas the right hand side is
\[ \left( \rho_{AC}^{-\frac{1}{2}} \rho_{AC} \rho_{AC}^{-\frac{1}{2}} \right)^{\frac{1}{2}} \rho_{AC} \rho_{AC} \rho_{AC}^{-\frac{1}{2}} \left( \rho_{AC}^{-\frac{1}{2}} \rho_{AC} \rho_{AC}^{-\frac{1}{2}} \right)^{\frac{1}{2}}. \] (224)
Because the \( \star \)-product is noncommutative, the terms in-
volving \( \rho_{AC} \) and \( \rho_{AC}^{-1} \) cannot be brought together and
made to cancel as they would in the classical case. Therefore,
counterexamples to this rule can occur when \( \rho_{AC} \)
and \( \rho_{C} \) do not commute. For example, it is straightforward
to verify that a generalized W-state of the form
\[ \lvert \psi \rangle_{ABC} = \frac{1}{2} \left( \lvert 001 \rangle_{ABC} + \lvert 010 \rangle_{ABC} \right) \]
\[ + \frac{1}{\sqrt{2}} \lvert 100 \rangle_{ABC} \] (225)
does not satisfy this rule. The calculation is not especially instructive, so it is omitted.

Note that a universally valid equation relating quan-
tum 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 expres-
sions for the left hand and right hand sides of eq. (222) when \( C \) is replaced by a classical variable \( T \).

**B. Limitations of Causal Joint States**

In \[ \text{1111} \] 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’ the-
orem 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 \( \rho_{A} \) 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 \( \rho_{AC} \) of \( A \) and \( C \) and apply \( \mathcal{E}_{B|A} \) to region \( A \), whilst doing nothing to \( C \), to obtain \( B \) and \( C \) in a joint state \( \rho_{BC} = (\mathcal{E}_{B|A} \otimes I_{C})(\rho_{AC}) \).

One might think that a joint state of \( ABC \) could be defined via \( \varrho_{ABC} = \varrho_{B|A} \star \rho_{AC} \). We would like this state to have the correct input and output states as marginals, so, in particular, it should satisfy
\[ \rho_{BC} = \mathcal{E}_{B|A}(\rho_{AC}) = \text{Tr}_{A}(\varrho_{ABC}). \] (226)
Unfortunately, this fails because theorem \[ \text{III.4} \] implies that
\[ \rho_{BC} = \text{Tr}_{A}(\varrho_{B|A} \star \rho_{AC}), \] (227)
whereas
\[ \text{Tr}_{A}(\varrho_{ABC}) = \text{Tr}_{A}(\varrho_{B|A} \star \rho_{AC}). \] (228)
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 \( \rho_{BC} \) as a reduced state, \( \varrho_{ABC} \) fails to correctly represent the correlations between the causally related regions \( A \) and \( B \) as well, i.e.
\[ \text{Tr}_{C}(\varrho_{ABC}) \neq \varrho_{AB}. \]

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 \( \rho_{AC} = \frac{1}{d} \lvert \Phi^{+} \rangle \langle \Phi^{+} \rvert_{AC} \) is a maximally entangled state, and \( \mathcal{E}_{B|A} \) is the identity superoperator. Then, the output state is \( \rho_{BC} = \frac{1}{d} \lvert \Phi^{+} \rangle \langle \Phi^{+} \rvert_{BC} \) and the causal joint state is \( \varrho_{AB} = \frac{1}{d} \lvert \Phi^{+} \rangle \langle \Phi^{+} \rvert_{AB} \).

In this case, explicitly calculating the operator \( \varrho_{ABC} = \varrho_{B|A} \star \rho_{AC} \) gives
\[ \varrho_{ABC} = \frac{1}{d} \lvert \Phi^{+} \rangle \langle \Phi^{+} \rvert_{AC} \otimes I_{B}. \] (229)

Whilst the reduced state \( \rho_{AC} \) gives the correct input state, \( \rho_{BC} \) is not the output state and \( \varrho_{AB} \) is not the
causal joint state. In fact, no operator of the form \( \varrho_{ABC} = \varrho_{B|A} \ast \rho_{AC} \) can ever satisfy all three conditions \( \rho_{AC} = \frac{1}{2} \langle \Phi^+ | \Lambda \rangle, \varrho_{AB} = \frac{1}{2} \langle \Phi^+ | T_A \rangle \) and \( \rho_{BC} = \frac{1}{2} \langle \Phi^+ | T_B \rangle \) simultaneously. This is because \( \varrho_{ABC} \) 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. (227) suggests that using an ordinary product instead of a \( \ast \)-product could be a better way of defining a joint state in this scenario. Indeed, the operator \( \varrho_{B|A} \ast \rho_{AC} \) does have all the correct bipartite marginals. For example, in our example of maximally entangled input and identity map dynamics, explicit calculation gives

\[
\varrho_{B|A} \ast \rho_{AC} = \frac{1}{d} \sum_{j,k,m=1}^d \langle j | \langle k | \otimes | m \rangle \langle j | B \otimes | m \rangle \langle k | C \rangle.
\]

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

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

\[
\frac{1}{2} \left( \varrho_{B|A} \ast \rho_{AC} + \rho_{AC} \ast \varrho_{B|A} \right),
\]

which might be useful because it is Hermitian.

Secondly, 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 \( \rho_A \) and checking whether \( M_{AB} \rho_A^{-1} \) is a valid causal conditional state. In contrast, when using the \( \ast \)-product, we know that an operator is of the form \( \varrho_{B|A} \ast \rho_A \) 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 \( \ast \)-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 space-like separated regions, a Markovian joint state \( \rho_{ABC} \), where \( A \) and \( C \) are conditionally independent given \( B \), can always be decomposed via the chain rule into

\[
\rho_{ABC} = \rho_{C|B} \ast (\rho_{B|A} \ast \rho_A)
\]

For three time-like separated regions, the analog of Markovianity is a two time-step dynamics where the first CPT map is \( \mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B) \) and the second is \( \mathcal{E}_{C|B} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_C) \), 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. (229).

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 \( \rho_A \) then the first time-step is represented as

\[
\rho_B = \mathcal{E}_{B|A}(\rho_A) = \text{Tr}_A \left( \varrho_{B|A} \rho_A \right)
\]

and the second time-step is represented as

\[
\rho_C = \mathcal{E}_{C|B}(\rho_B) = \text{Tr}_B \left( \varrho_{C|B} \text{Tr}_A \left( \varrho_{B|A} \rho_A \right) \right).
\]

It follows that

\[
\rho_C = \text{Tr}_{AB} \left( \varrho_{C|B} \ast \left( \varrho_{B|A} \ast \rho_A \right) \right),
\]

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

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

\[
\varrho_{ABC} = \sqrt{\varrho_{B|A} \ast \rho_A \ast \varrho_{C|B} \ast \sqrt{\varrho_{B|A} \ast \rho_A}}.
\]

The term \( \sqrt{\varrho_{B|A} \ast \rho_A} \) is not well defined because \( \varrho_{B|A} \ast \rho_A \) 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 \( \varrho_{ABC} \) would then have the correct reduced states, \( \rho_A, \rho_B \) and \( \rho_C \), 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 \( \ast \)-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 [3].
As with the evolution of a subsystem, these problems can be remedied by using the ordinary product instead of the $\ast$-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 $P(S|R)$, $P(T|R,S)$ are conditional probability distributions then

$$P(R,S,T) = P(T|R,S)P(S|R)P(R)$$  \hspace{1cm} (241)

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

$$\sum_{S,T} P(R,S,T) = P(R)$$  \hspace{1cm} (242)

$$\sum_{T} P(R,S,T) = P(S|R)$$  \hspace{1cm} (243)

$$\sum_{S,T} P(R,S,T) = P(T|R,S),$$  \hspace{1cm} (244)

whenever the left hand sides are well defined.

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

$$\rho_{ABC} = \rho_{C|AB} \ast \rho_{B|A} \ast \rho_{A}$$  \hspace{1cm} (245)

via the chain rule, one cannot start with an arbitrary reduced state $\rho_{A}$ and two arbitrary conditional states, $\rho_{B|A}$ and $\rho_{C|AB}$, and expect there to be a joint state $\rho_{ABC}$ 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 $\rho_{A}$ 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 $\rho_{AB} = \rho_{B|A} \ast \rho_{A}$ 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 $\rho_{AC} = \rho_{C|A} \ast \rho_{A}$ 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 [71] 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 UCL Department of Physics and Astronomy and the Perimeter Institute for Theoretical Physics for their hospitality. M.L. is supported by Grant RFP1-06-006 from The Foundational Questions Institute. Part of this work was completed whilst 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, and an associate postdoctoral fellow at the Perimeter Institute for Theoretical Physics, where research is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

Appendix A: Proofs of Theorems

Theorem III.4 (Jamiolkowski Isomorphism). Let $\mathcal{E}_{B|A} : \mathcal{L} (\mathcal{H}_{A}) \rightarrow \mathcal{L} (\mathcal{H}_{B})$ be a linear map and let $M_{AC} \in \mathcal{L} (\mathcal{H}_{AC})$, where $\mathcal{H}_{C}$ is a Hilbert space of arbitrary dimension. Then, the action of $\mathcal{E}_{B|A}$ on $\mathcal{L} (\mathcal{H}_{A})$ (tensored with the identity on $\mathcal{L} (\mathcal{H}_{C})$) is given by

$$(\mathcal{E}_{B|A} \otimes \mathcal{I}_{C}) (M_{AC}) = \text{Tr}_{A} (\rho_{B|A} M_{AC}),$$  \hspace{1cm} (A1)
where \( \rho_{B|A} \in \mathcal{L}(\mathcal{H}_{AB}) \) is given by

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

Here, \( A' \) labels a second copy of \( A \), \( \mathcal{I}_A \) is the identity superoperator on \( \mathcal{L}(\mathcal{H}_A) \), 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 [III.2] if and only if \( \mathcal{E}_{B|A} \circ T_A \) is CPT, where \( T_A : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_A) \) 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 Choi isomorphic map to an operator \( \rho_{B|A} \) is given by

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

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

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

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 \( \mathcal{E}_{B|A} = \mathcal{E}_{B|A} \circ T_A \), where \( T_A \) is the partial transpose operation with respect to the basis used to define the Choi isomorphism.

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

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

\[
= \sum_{j,k} \langle j|_{AA'} \rho_{B|A'} M_{AC} |kk\rangle_{AA'},
\]

\[
= \sum_{j,k} \langle j|_{A'} \rho_{B|A'} |k\rangle_{A'} \langle j|_A M_{AC} |k\rangle_A,
\]

\[
= \sum_{j,k} \langle j|_A \rho_{B|A} |k\rangle_A \langle j|_A M_{AC} |k\rangle_A,
\]

\[
= \sum_{j,k} \langle j|_A \rho_{B|A} M_{AC} |j\rangle_A,
\]

\[
= \operatorname{Tr}_A \left( \rho_{B|A} M_{AC}^{T_A} \right).
\]

Proof of theorem [III.2] Eq. (A1) is derived from eq. (A2) as follows:

\[
(\mathcal{E}_{B|A} \otimes \mathcal{I}_C) (M_{AC}) = (\mathcal{E}_{B|A'} \otimes \mathcal{I}_C) (M_{AC'}),
\]

\[
= (\mathcal{E}_{B|A'} \otimes \mathcal{I}_C) \left( \sum_{j,k} |k\rangle \langle k|_{A'} \right) M_{AC} \left[ \sum_j \langle j| \langle j|_{A'} \right)
\]

\[
= (\mathcal{E}_{B|A'} \otimes \mathcal{I}_C) \left( \sum_{j,k} \langle j|_{A'} M_{AC} |j\rangle_A \langle j|_{A'} \right),
\]

\[
= (\mathcal{E}_{B|A'} \otimes \mathcal{I}_C) \left( \sum_{j,k} \operatorname{Tr}_A (|j\rangle \langle j|_A M_{AC}) |k\rangle \langle j|_{A'} \right),
\]

\[
= \operatorname{Tr}_A \left( (\mathcal{E}_{B|A'} \otimes \mathcal{I}_C) \left( \sum_{j,k} |j\rangle \langle j|_A \otimes |k\rangle \langle j|_{A'} \right) \right) M_{AC},
\]

\[
= \operatorname{Tr}_A (\rho_{B|A} M_{AC}). \tag{A7}
\]

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

\[
\operatorname{Tr}_B (\mathcal{E}_{B|A} (M_A)) = \operatorname{Tr}_B (\operatorname{Tr}_A (\rho_{B|A} M_A)), \tag{A8}
\]

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

\[
= \operatorname{Tr}_A (I_A M_A), \tag{A10}
\]

\[
= \operatorname{Tr}_A (M_A). \tag{A11}
\]
To show that $\mathcal{E}_{B|A} \circ T_A$ 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

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

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

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

$$\text{Tr}_B (\rho_{B|A}) = \text{Tr}_B \left( \sum_{j,k} |j\rangle \langle k| \otimes \mathcal{E}_{B|A} (|k\rangle \langle j|_{A'}) \right)$$

$$= \sum_{j,k} |j\rangle \langle k| \otimes \text{Tr}_B (\mathcal{E}_{B|A} (|k\rangle \langle j|_{A'})) \quad (A13)$$

$$= \sum_{j,k} |j\rangle \langle k| \otimes \text{Tr}_{B'} (\mathcal{E}_{B'|A'} (|k\rangle \langle j|_{A'})) \quad (A14)$$

$$= \sum_{j,k} |j\rangle \langle k| \otimes \text{Tr}_{B'} (\rho_{B'|A'} \delta_{j,k}) \quad (A15)$$

$$= \sum_{j} |j\rangle \langle j|_A = I_A. \quad (A17)$$

Also,

$$\rho_{B|A} = (\mathcal{E}_{B|A'} \otimes I_A) \langle \Phi^+ |_{AA'} \rangle \quad (A19)$$

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

**Theorem III.11.** Let $\sigma_{A|X}$ be a hybrid operator, so that by eq. (A10) it can be written as

$$\sigma_{A|X} = \sum_x \rho^X_x \otimes |x\rangle \langle x|_X, \quad (A20)$$

for some set of operators $\{\rho^X_x\}$. Then, $\sigma_{A|X}$ satisfies the definition of both an acausal and a causal conditional state for $A$ given $X$, iff each of the components $\rho^X_x$ 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 $\{\rho^X_x\}$. Then it is clearly positive and satisfies $\text{Tr}_A (\sigma_{A|X}) = I_X$ because $\text{Tr}_A (\rho^X_x) = 1$ 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 $\text{Tr}_A (\sigma_{A|X}) = I_X$. Positivity means that $\langle \psi|_{AX} \sigma_{A|X} |\psi\rangle_{AX} \geq 0$ for all $|\psi\rangle_{AX} \in \mathcal{H}_{AX}$. If $\sigma_{A|X}$ has the form of eq. (A20) then taking $|\psi\rangle_{AX} = |\phi\rangle_A \otimes |x\rangle_X$ gives $\langle \phi|_A \rho^X_x |\phi\rangle_A \geq 0$. By varying over all $|\phi\rangle_A \in \mathcal{H}_A$, this implies that each $\rho^X_x$ is a positive operator. To prove that these operators are normalized note that

$$\text{Tr}_A (\sigma_{A|X}) = \sum_x \text{Tr}_A (\rho^X_x) = I_X. \quad (A21)$$

This is an eigendecomposition of $\text{Tr}_A (\sigma_{A|X})$ with eigenvalues $\text{Tr}_A (\rho^X_x)$ 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 $\text{Tr}_A (\sigma^X_{A|X}) = I_X$. However, operators of the form of eq. (A20) are invariant under partial transpose on $X$, so the same argument applies. \hfill \Box

**Theorem III.13.** Let $\sigma_{Y|A}$ be a hybrid operator so that it can be written in the form

$$\sigma_{Y|A} = \sum_y |y\rangle \langle y|_Y \otimes E^A_y, \quad (A22)$$

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

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

$$\sigma^{T_A}_{Y|A} = \sum_y |y\rangle \langle y|_Y \otimes (E^A_y)^{T_A}, \quad (A23)$$

is also positive because the positive operators $E^A_y$ remain positive under the transpose. Also, $\text{Tr}_Y (\sigma^{T_A}_{Y|A}) = \sum_y (E^A_y)^{T_A} = I^T_A = I_A$. 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 $\text{Tr}_Y (\sigma_{Y|A}) = I_A$. By the same argument used in the proof of theorem III.11, positivity implies that, if $\sigma_{Y|A}$ is of the form of eq. (A22), then each of the components $E^A_y$ must be positive. Since $\text{Tr}_Y (\sigma_{Y|A}) = \sum_y E^A_y$, 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 $\{E^A_y\}$. This means that $\sigma_{Y|A}$ itself can be written as

$$\sigma^{T_A}_{Y|A} = \sum_y |y\rangle \langle y|_Y \otimes (E^A_y)^{T_A}, \quad (A24)$$

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

**Theorem III.15.** 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

$$\varrho_{C|A} = \text{Tr}_B (\varrho_{C|B} \varrho_{B|A}). \quad (A25)$$
Conversely, if three operators satisfy eq. \[ (A25) \], then the Jamiołkowski isomorphic maps satisfy \[ E_{C|A} = E_{C|B} \circ E_{B|A} \].

**Proof.** By definition, the Jamiołkowski isomorphic operator to \[ E_{C|A} \] is

\[
E_{C|A} = (E_{C|A'} \otimes I_A) \left( \sum_{j,k} |j\rangle \langle k|_A \otimes |j\rangle \langle k|_A' \right) 
= \sum_{j,k} |j\rangle \langle k| \otimes [E_{C|B} \circ E_{B|A'} \langle k| \langle j|_A'] \tag{A26}
\]

Applying theorem III.6 to \[ E_{B|A'} \] gives

\[
E_{C|A} = \sum_{j,k} |j\rangle \langle k|_A \otimes E_{C|B} \left( \text{Tr}_{A'} \left( |B\rangle \langle A| \otimes |j\rangle \langle k|_A' \right) \right) 
= \sum_{j,k} |j\rangle \langle k|_A \otimes E_{C|B} \left( \langle j|_{A'} \otimes |B\rangle \langle A| \otimes |k\rangle \langle j|_A' \right), \tag{A28}
\]

and then applying the same theorem to \[ E_{C|B} \] gives

\[
E_{C|A} = \sum_{j,k} |j\rangle \langle k|_A \otimes \text{Tr}_{A'} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle j|_A' \right), \tag{A29}
\]

Since \[ A' \] is a dummy label in this expression, it can be changed to \[ A \] and then

\[
E_{C|A} = \sum_{j,k} |j\rangle \langle j|_A \text{Tr}_{B} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right) 
= \text{Tr}_{B} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right). \tag{A30}
\]

For the converse direction, we have

\[
E_{C|A} = \sum_{j,k} |j\rangle \langle j|_A \text{Tr}_{A'} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right) \tag{A31}
\]

\[
E_{C|A} = \text{Tr}_{A} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right). \tag{A32}
\]

\[
E_{C|A} = \text{Tr}_{A} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right) \tag{A33}
\]

\[
E_{C|A} = \text{Tr}_{B} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right). \tag{A34}
\]

\[
E_{C|A} = \text{Tr}_{C} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right). \tag{A35}
\]

\[
E_{C|A} = \text{Tr}_{A} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right). \tag{A36}
\]

\[
E_{C|A} = \text{Tr}_{B} \left( E_{C|B} \otimes |B\rangle \langle A| \otimes |k\rangle \langle k|_A \right). \tag{A37}
\]

---

[1] B. de Finetti, *Theory of Probability*, vol. 1 (Wiley, 1974).
[2] S. Taylor, S. Cheung, C. Brukner, and V. Vedral, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (1984), p. 175.
[3] C. H. Bennett and G. Brassard, in *Proceedings of the 67th AIP Conference Proceedings*, pp. 281–284, eprint arXiv:quant-ph/0402127.
[4] M. S. Leifer, Phys. Rev. A 74, 042310 (2006), eprint arXiv:quant-ph/0606022.
[5] M. S. Leifer, in *Foundations of Probability and Physics – 4*, edited by G. Adenier, C. A. Fuchs, and A. Y. Khrennikov (AIP, 2007), vol. 889 of *AIP Conference Proceedings*, pp. 172–186, eprint arXiv:quant-ph/0611233.
[6] P. Evans, H. Price, and K. B. Wharton (2010), to appear in British Jnl. for the Philosophy of Sci., eprint arXiv:1001.5057.
[7] K. B. Wharton, D. J. Miller, and H. Price, Symmetry 3, 524 (2011), eprint arXiv:1103.2492.
[8] S. Marcovitch and B. Reznik (2011), eprint arXiv:1103.2557.
[9] S. Marcovitch and B. Reznik (2011), eprint arXiv:1107.2186.
[10] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (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), eprint arXiv:quant-ph/0003004.
[13] H. Umegaki, Kodai Math. Sem. Rep. 14, 59 (1962).
[14] D. Petz, *Quantum Information Theory and Quantum Statistics* (Springer, 2008), chap. 9.2 Conditional Expectations, pp. 123–130, Theoretical and Mathematical Physics.
[15] M. Rédei and S. J. Summers, Stud. Hist. Philos. M. P. 38, 390 (2007).
[16] D. Petz, in *Quantum probability and applications III*, edited by L. Accardi and W. Waldenfels (1988), vol. 1303 of *Lecture Notes in Mathematics*, pp. 251–260.
[17] N. J. Cerf and C. Adami, Phys. Rev. Lett. 79, 5194 (1997), eprint arXiv:quant-ph/9705202.
[18] N. J. Cerf and C. Adami, in *New Developments on Fundamental Problems in Quantum Physics*, edited by M. Ferrero and A. Merwe (Kluwer, 1997), vol. 81 of *Fundamental Theories of Physics*, eprint arXiv:quant-ph/9610005.
[19] N. J. Cerf and C. Adami, Physica D 120, 62 (1998), eprint arXiv:quant-ph/9605039.
[20] M. Asorey, A. Kossakowski, G. Marmo, and E. C. G. Sudarshan, Open Syst. Inf. Dyn. 12, 319 (2005), eprint arXiv:quant-ph/0602228.
[21] Z. Zhang and S. Luo, Phys. Rev. A 75, 032312 (2007).
[22] C. A. Fuchs, J. Mod. Optic. 50, 987 (2003), subanitally similar to eprint arXiv:quant-ph/0205039.
[23] L. P. Hughston, R. Jozsa, and W. K. Wootters, Phys. Lett. A 183, 14 (1993).
[24] P. Hausladen and W. K. Wootters, J. Mod. Optic. 41,
2385 (1994).

[25] V. P. Belavkin, Radio Engineering and Electronic Physics 20, 39 (1975).

[26] V. P. Belavkin, Stochastics 1, 315 (1975).

[27] H. Barnum and E. Knill, J. Math. Phys. 43, 2097 (2002), eprint arXiv:quant-ph/0004088.

[28] S. M. Barnett, D. T. Pegg, and J. Jeffers, J. Mod. Optic. 47, 1779 (2000).

[29] D. Pegg, S. Barnett, and J. Jeffers, J. Mod. Optic. 49, 913 (2002), eprint arXiv:quant-ph/0207174.

[30] D. Pegg, S. M. Barnett, and J. Jeffers, Phys. Rev. A 66, 022106 (2002), eprint arXiv:quant-ph/0208082.

[31] J. Bub, in The Logico-Algebraic Approach to Quantum Mechanics, edited by C. A. Hooker (Dordrecht: Reidel, 1978), vol. II, pp. 209–226.

[32] J. Bub, Stud. Hist. Philos. M. P. 38, 232 (2007).

[33] F. Verstraete, Ph.D. thesis, University of Leuven (2002).

[34] A. Jamiołkowski, Rep. Math. Phys. 3, 275 (1972).

[35] R. C. Jeffrey, The Logic of Decision (University of Chicago Press, 1983), 2nd ed.

[36] R. C. Jeffrey, Subjective Probability: The Real Thing (Cambridge University Press, 2004).

[37] J. Earman, Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory (MIT Press, 1992), chap. 2, pp. 34–35.

[38] P. Diaconis and S. L. Zabell, J. Am. Stat. Assoc. 77, 822 (1982).

[39] M. A. Naimark, Izv. Akad. Nauk SSSR Ser. Mat. 4, 277 (1940).

[40] R. W. Spekkens, Phys. Rev. A 75, 032110 (2007), eprint arXiv:quant-ph/0401052.

[41] E. Schrödinger, Proc. Camb. Phil. Soc. 31, 555 (1935).

[42] E. Schrödinger, Proc. Camb. Phil. Soc. 32, 446 (1936).

[43] M. Vujčić and F. Herbut, J. Phys. A 21, 2931 (1988).

[44] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. 98, 140402 (2007), eprint arXiv:quant-ph/0612147.

[45] S. J. Jones, H. M. Wiseman, and A. C. Doherty, Phys. Rev. A 76, 052116 (2007), eprint arXiv:0709.0390.

[46] F. Herbut, J. Phys. A 42, 085202 (2009), eprint arXiv:0809.1631.

[47] R. W. Spekkens and N. Harrigan, Found. Phys. 40, 125 (2010), eprint arXiv:0706.2261.

[48] J. S. Bell, Physics 1, 195 (1964).

[49] M. Ozawa, in Quantum Communication, Computing and Measurement, edited by O. Hirota, A. S. Holevo, and C. A. Hooker (Dordrecht: Reidel, 1997), pp. 233–241, eprint arXiv:quant-ph/9705030.

[50] M. Ozawa, Fortschr. Phys. 46, 615 (1998), eprint arXiv:quant-ph/9711006.

[51] K. Jacobs, Quantum Inf. Process. 1, 73 (2002), eprint arXiv:quant-ph/0201096.

[52] K. Jacobs, Phys. Rev. A 72, 044101 (2005), eprint arXiv:quant-ph/0504193.

[53] M. S. Leifer and D. Poulin, Ann. Phys. 323, 1899 (2008), eprint arXiv:0708.1337.

[54] B. Coecke and R. W. Spekkens, Synthese (2011), published online first, eprint arXiv:1102.2368.

[55] M. Rédei, International Studies in the Philosophy of Science 6, 129 (1992).

[56] G. Valente, Stud. Hist. Philos. M. P. 38, 832 (2007).

[57] L. Hardy (2005), eprint arXiv:gr-qc/0509120.

[58] L. Hardy, J. Phys. A 40, 3081 (2007), eprint arXiv:gr-qc/0608043.

[59] L. Hardy (2011), eprint arXiv:1104.2066.

[60] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Phys. Rev. A 80, 022339 (2009), eprint arXiv:0904.4843.

[61] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Phys. Rev. A 84, 012311 (2011), eprint arXiv:1011.6451.

[62] O. Oreshkov, F. Costa, and C. Brukner (2011), eprint arXiv:1105.4646.

[63] B. Coecke and E. O. Paquette, in New Structures for Physics, edited by B. Coecke (Springer, 2009), Lecture Notes in Physics, pp. 173–286, eprint arXiv:0905.3010.

[64] B. Coecke, E. O. Paquette, and D. Pavlovic (2009), to appear in Semantic Techniques for Quantum Computation, I. Mackie and S. Gay (eds.), Cambridge University Press, eprint arXiv:0904.1997.

[65] C. M. Caves, C. A. Fuchs, and R. Schack, Phys. Rev. A 65, 022305 (2002), eprint arXiv:quant-ph/0106133.

[66] C. M. Caves, C. A. Fuchs, and R. Schack (2006), eprint arXiv:quant-ph/0608190.

[67] C. A. Fuchs (2010), eprint arXiv:1003.5209.

[68] C. A. Fuchs, Physics in Canada 66, 77 (2010), eprint arXiv:1003.5182.

[69] J. Pearl, Causality: Models, Reasoning and Inference (Cambridge University Press, 2009), 2nd ed.

[70] P. Spirtes, C. Glymour, and R. Scheines, Causation, Prediction, and Search (MIT Press, 2001), 2nd ed.

[71] M. S. Leifer and R. W. Spekkens (2011), eprint arXiv:1110.1085.

[72] S. D. Bartlett, T. Rudolph, and R. W. Spekkens (2011), to appear in Phys. Rev. A, eprint arXiv:1111.5057.

[73] H. Barnum, J. Barrett, M. Leifer, and A. Wilce (2006), eprint arXiv:quant-ph/0611295.

[74] 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.

[75] 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.

[76] 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[22]. We do not pursue the analogy here.

[77] If \( |\psi\rangle_{BA} \) derives from a joint pure state \( |\psi\rangle_{AB} \) via eq. 12 then this only implies that \( \dim(\text{supp}(\rho_A)) \leq \dim(H_B) \), which is always true because the ranks of \( \rho_A \) and \( \rho_B \) are equal.

[78] 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 \( \{ |\pi\rangle_X \} \)
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.

[79] 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.

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

[81] 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 [73], which shows that monogamy applies to more general probabilistic theories, including one in which states are locally positive operators.

[82] 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.

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