Quantum common causes and quantum causal models

John-Mark A. Allen,1 Jonathan Barrett,1 Dominic C. Horsman,2 Ciarán M. Lee,3 and Robert W. Spekkens4

1Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK
2Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK
3Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
4Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada

Reichenbach’s principle asserts that if two observed variables are found to be correlated, then there should be a causal explanation of these correlations. Furthermore, if the explanation is in terms of a common cause, then the conditional probability distribution over the variables given the complete common cause should factorize. The principle is generalized by the formalism of causal models, in which the causal relationships among variables constrain the form of their joint probability distribution. In the quantum case, however, the observed correlations in Bell experiments cannot be explained in the manner Reichenbach’s principle would seem to demand. Motivated by this, we introduce a quantum counterpart to the principle. We demonstrate that under the assumption that quantum dynamics is fundamentally unitary, if a quantum channel with input A and outputs B and C is compatible with A being a complete common cause of B and C, then it must factorize in a particular way. Finally, we show how to generalize our quantum version of Reichenbach’s principle to a formalism for quantum causal models, and provide examples of how the formalism works.

I. INTRODUCTION

It is a general principle of scientific thought—and indeed of everyday common sense—that if physical variables are found to be statistically correlated, then there ought to be a causal explanation of this fact. If the dog barks every time the telephone rings, we do not ascribe this to coincidence. A likely explanation is that the sound of the telephone ringing is causing the dog to bark. This is a case where one of the variables is a cause of the other. If sales of ice cream are high on the same days of the year that many people get sunburned, a likely explanation is that the sun was shining on these days and that the hot sun causes both sunburns and the desire to have an ice cream. Here the explanation is not that buying ice cream causes people to get sunburned, nor vice versa, but instead that there is a common cause of both: the hot sun.

That the principle is highly natural is most apparent when it is expressed in its contrapositive form: if there is no causal relationship between two variables (i.e. neither is a cause of the other and there is no common cause) then the variables will not be correlated. In particular, without a general commitment to this latter statement, it would be impossible ever to regard two different experiments as independent from one another, or for the results of one scientific team to be regarded as an independent confirmation of the results of another.

This principle of causal explanation was first made explicit by Reichenbach [1]. It is key in scientific investigations which aim to find causal accounts of phenomena from observed statistical correlations. Not only that, but the scientific process of experimentation requires that our tests are not subject to arbitrary causal influences. Without assuming some form of Reichenbach’s principle it becomes impossible to reason about experimental outcomes.

Despite the central role of causal explanations in science, there are significant challenges to providing them for the correlations that are observed in quantum experiments [2]. In a Bell experiment, a pair of systems are prepared together, then taken to space-like separated locations for measurement. The choices of measurement being made at those space-like separated locations also. The natural causal explanation of the correlations that one observes in such experiments is that each measurement outcome is influenced by the local measurement setting as well by a common cause located in the joint past of the two measurement events. But Bell’s theorem famously rules out this possibility: within the standard framework of causal models, if the correlations violate a Bell inequality [3]—as is predicted by quantum theory and verified experimentally [4,5]—then a common cause explanation of the correlations is ruled out. Furthermore, Ref. [2] proves that it is not possible to explain Bell correlations with classical causal models without unwelcome fine-tuning of the parameters. In the study of classical causation, fine-tuning is normally ruled out by assumption [3]. This includes any attempt to explain Bell correlations with exotic causal influences, such as retrocausality and superluminal signalling.

However, the results of Ref. [2] assume our classical models of causation. It has been suggested [2] that an appropriate quantum generalisation of these models might preserve Reichenbach’s principle without fine-tuning. Such a generalisation would satisfactorily explain Bell experiments and all other quantum experiments on a causal level. This article seeks to develop such a model by first suggesting an intrinsically quantum version of Reichenbach’s principle.

Specifically, we consider the case of a quantum system A in the causal past of a bipartite quantum system BC and ask what constraints on the channel from A to BC follows from the assumption that A is the complete common cause of B and C. In this scenario we are able to
find a natural quantum analog to Reichenbach’s principle. This analog can be expressed in several equivalent forms, each of which naturally generalises a corresponding classical expression. In particular, one of these conditions states that \( A \) is a complete common cause of \( BC \) if one can dilate the channel from \( A \) to \( BC \) to a unitary by introducing two ancillary systems, contained in the causal past of \( BC \), such that each ancillary system can influence only one of \( B \) and \( C \). This unitary dilation codifies the causal relationship between \( A \) and \( BC \) and illustrates the fact that no other system can influence both \( B \) and \( C \). Moreover, our quantum Reichenbach’s principle contains the classical version as a special case in the appropriate limit. This suggests that our quantum version is the correct way to generalise Reichenbach’s principle.

The mathematical framework of causal models \([3, 4]\) can be seen as a direct generalisation of Reichenbach’s principle to arbitrary causal structures. By following the classical example we are able to generalise our quantum Reichenbach’s principle to a framework of quantum causal models. In each case, the original Reichenbach’s principle becomes a special case of the framework. Just as with classical causal models, the framework of quantum causal models allows us to analyse the causal structure of arbitrary quantum experiments. It also does so while preserving an appropriate form of Reichenbach’s principle (by construction) and avoiding fine-tuning.

Although our main motivation for developing quantum causal models is the possibility of finding a satisfactory (i.e., non-fine-tuned) causal explanation of Bell inequality violations \([2, 10]\), they are also likely to have practical applications. For instance, finding quantum-classical separations in the correlations achievable in novel causal scenarios might lead to new device-independent protocols \([11]\), such as randomness extraction and secure key distribution. Quantum causal models may also provide novel schemes for simulating many body systems in condensed matter physics \([12]\) and novel means for inferring the underlying causal structure from quantum correlations \([13, 14]\).

The structure of the paper is as follows. Section \( \text{II} \) provides a formal statement of Reichenbach’s principle and shows how it can be rigorously justified under certain philosophical assumptions. The main body of results is in Sec. \( \text{III} \). Here our quantum generalisation of Reichenbach’s principle is presented and justified by reasoning parallel to that of the classical case. This is then fleshed out with alternative characterisations of our quantum version of conditional independence and some specific examples. We return to the classical world in Sec. \( \text{IV} \), discussing classical causal models and providing a rigorous justification of the Markov condition, which plays the role of Reichenbach’s principle for general causal structures. Sec. \( \text{V} \) then generalizes these ideas to the quantum sphere, and presents our proposal for quantum causal models. Finally, in Sec. \( \text{VI} \) we summarize the relationship of our proposal to prior work on quantum causal models, and in Sec. \( \text{VII} \) we summarize and describe some directions for future work.

**II. REICHENBACH’S PRINCIPLE**

**A. Statement**

Reichenbach gave his principle a formal statement in Ref. \( [1] \). Following Ref. \( [15] \), we here distinguish two parts of the formalized principle. First is the *qualitative part* which expresses the intuitions described at the beginning of the introduction. The other is the *quantitative part* which constrains the sorts of probability distributions one should assign in the case of a common cause explanation.

The qualitative part of Reichenbach’s principle may be stated as follows: if two physical variables \( Y \) and \( Z \) are found to be statistically dependent, then there should be a causal explanation of this fact, either:

1. \( Y \) is a cause of \( Z \);
2. \( Z \) is a cause of \( Y \);
3. there is no causal link between \( Y \) and \( Z \), but there is a common cause \( X \) of \( Y \) and \( Z \);
4. \( Y \) is a cause of \( Z \) and there is a common cause \( X \) of \( Y \) and \( Z \); or
5. \( Z \) is a cause of \( Y \) and there is a common cause \( X \) of \( Y \) and \( Z \).

Note that the causal influences considered here may be indirect (mediated by other variables). If none of these causal relations hold between \( Y \) and \( Z \), then we refer to them as *ancestrally independent* (because their respective causal ancestries constitute disjoint sets). Using this terminology, the qualitative part of Reichenbach’s principle can be expressed particularly succinctly in its contrapositive form as: ancestral independence implies statistical independence, i.e., \( P(Y \mid Z) = P(Y)P(Z) \).

The quantitative part of Reichenbach’s principle applies only to the case where the correlation between \( Y \) and \( Z \) is due purely to a common cause (case 3 above). Let \( X \) be a complete common cause for \( Y \) and \( Z \), meaning that \( X \) is the collection of all variables acting as common causes. The quantitative part states that \( Y \) and \( Z \) must be conditionally independent given \( X \), so the joint probability distribution \( P(XYZ) \) satisfies

\[
P(YZ|X) = P(Y|X)P(Z|X).
\]

**B. Justifying the quantitative part of Reichenbach’s principle**

Within the philosophy of causality, providing an adequate justification of Reichenbach’s principle is a delicate issue. It rests on controversy over basic questions, such
as what it means for one variable to have a causal influence on another and what is the correct interpretation of probabilistic statements. In this section, we discuss one way of justifying the principle which provides a clean motivational story with a natural quantum analogue. This involves temporarily making a strong philosophical assumption about determinism. The reader need not accept this assumption as it is only temporarily adopted to motivate the quantitative part of Reichenbach’s principle, which stands apart from the assumption.

Suppose we adopt a Bayesian point of view on probabilities: they are the degrees of belief of a rational agent. Dutch book arguments—based on the principle that a rational agent will never accept a set of bets on which they are certain to lose money—can then be given as to why probabilities should be non-negative, sum to 1 and so forth. But why should an agent who takes a decision ignoring the values of unobserved variables. Under these assumptions, one can show that the qualitative part of Reichenbach’s principle implies the quantitative part.

In general, a classical channel describing the influence of random variable $X$ on $Y$ is given by a conditional probability distribution $P(Y|X)$. Assuming underlying deterministic dynamics, then although the value of the variable $Y$ might not be completely determined by the value of $X$, it must be determined by the value of $X$ along with the values of some extra, unobserved, variables in the past of $Y$ which can collectively be denoted $\lambda$. Any variation in the value of $Y$ for a given value of $X$ is then explained by variation in the value $\lambda$. This can be formalised as follows.

**Definition 1** (Classical dilation). For a classical channel $P(Y|X)$, a causal deterministic dilation is given by some random variable $\lambda$ with probability distribution $P(\lambda)$ and some deterministic function $Y = f(X, \lambda)$ such that

$$P(Y|X) = \sum_{\lambda} \delta(Y, f(X, \lambda))P(\lambda),$$

(2)

where $\delta(Y, X) = 1$ if $X = Y$ and 0 otherwise.

We now apply this to the situation depicted in Fig. 1, where $X$ is the complete common cause of $Y$ and $Z$. The conditional distribution $P(YZ|X)$ admits of a dilation in terms of an ancillary unobserved variable $\lambda$, for some distribution $P(\lambda)$ and a function $f = (f_Y, f_Z)$ from $(\lambda, X)$ to $(Y, Z)$ such that $Y = f_Y(\lambda, X)$ and $Z = f_Z(\lambda, X)$. The assumption that $X$ is the complete common cause of $Y$ and $Z$ implies that the ancillary variable $\lambda$ can be split into a pair of ancestrally independent variables, $\lambda_Y$ and $\lambda_Z$, where $\lambda_Y$ only influences $Y$ and $\lambda_Z$ only influences $Z$. It follows that there must exist $\lambda_Y$ and $\lambda_Z$ that are causally related to $X$, $Y$ and $Z$ as depicted in Fig. 2, where the causal dependences are deterministic and given by a pair of functions $f_Y$ and $f_Z$ such that $Y = f_Y(\lambda_Y, X)$ and $Z = f_Z(\lambda_Z, X)$.

In this case, we have

$$P(YZ|X) = \sum_{\lambda_Y, \lambda_Z} \delta(Y, f_Y(\lambda_Y, X))\delta(Z, f_Z(\lambda_Z, X))P(\lambda_Y, \lambda_Z)$$

(3)

Finally, given the qualitative part of Reichenbach’s principle, the ancestral independence of $\lambda_Y$ and $\lambda_Z$ in the causal structure implies that $P(\lambda_Y, \lambda_Z) = P(\lambda_Y)P(\lambda_Z)$. It then follows that $P(YZ|X) = P(Y|X)P(Z|X)$ admits of a dilation where $X$ is the complete common cause of $Y$ and $Z$.

A well-known converse statement is also worth noting: any classical channel $P(YZ|X)$ satisfying $P(YZ|X) = P(Y|X)P(Z|X)$ admits of a dilation where $X$ is the complete common cause of $Y$ and $Z$ by appealing to the qualitative part of Reichenbach’s principle and fundamental determinism. The definition can be formalized into a mathematical condition as follows:

**Definition 2** (Classical compatibility). $P(YZ|X)$ is said to be compatible with $X$ being the complete common cause of $Y$ and $Z$ if one can find variables $\lambda_Y$ and $\lambda_Z$, distributions $P(\lambda_Y)$ and $P(\lambda_Z)$, a function $f_Y$ from $(\lambda_Y, X)$ to $Y$ and a function $f_Z$ from $(\lambda_Z, X)$ to $Z$, such
that these constitute a dilation of $P(YZ|X)$, that is, such that

$$P(YZ|X) = \sum_{\lambda_Y, \lambda_Z} \delta(Y, f_Y(\lambda_Y, X))\delta(Z, f_Z(\lambda_Z, X))P(\lambda_Y)P(\lambda_Z)$$

(4)

With this definition, we can summarize the result described above as follows.

**Theorem 1.** Given a conditional probability distribution $P(YZ|X)$, the following are equivalent:

1. $P(YZ|X)$ is compatible with $X$ being the complete common cause of $Y$ and $Z$.

2. $P(YZ|X) = P(Y|Z)P(Z|X)$.

The $1 \rightarrow 2$ implication is what establishes that a rational agent should espouse the quantitative part of Reichenbach’s principle if they espouse the qualitative part and fundamental determinism.

The $2 \rightarrow 1$ implication allows one to deduce a possible causal explanation of an observed distribution from a feature of that distribution. However, it is important to stress that it only establishes a possible causal explanation. It does not state that this is the only causal explanation. Indeed, it is possible to satisfy this conditional independence relation within alternative causal structures by fine-tuning the strengths of the causal dependences. However, as noted above, fine-tuned causal explanations are typically rejected as bad explanations in the field of causal inference. Therefore, the best explanation of the conditional independence of $Y$ and $Z$ given $X$ is that $X$ is the complete common cause of $Y$ and $Z$.

III. THE QUANTUM VERSION OF REICHENBACH’S PRINCIPLE

In this section, we introduce our quantum version of Reichenbach’s principle. The definition of a quantum causal model that we provide in Sec. [11] can be seen as generalizing these ideas in much the same way that classical causal models generalize the classical Reichenbach’s principle.

**A. Quantum preliminaries**

We assume throughout that all quantum systems are finite-dimensional for simplicity. Given a quantum system $A$, we will write $\mathcal{H}_A$ for the corresponding Hilbert space, $d_A$ for the dimension of $\mathcal{H}_A$, and $I_A$ for the identity on $\mathcal{H}_A$. We will also write $\mathcal{H}_{A^*}$ for the dual space to $\mathcal{H}_A$, and $I_{A^*}$ for the identity on the dual space. If a quantum system is initially uncorrelated with any other system, then the most general time evolution of the system corresponds to a quantum channel, i.e., a completely postive trace-preserving (CPTP) map. If the system at the initial time is labelled $A$, with Hilbert space $\mathcal{H}_A$, and the system at the later time is labelled $B$, with Hilbert space $\mathcal{H}_B$, then the CPTP map is

$$\mathcal{E}_{B|A} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{H}_B),$$

(5)

where $\mathcal{B}(\mathcal{H})$ is the set of bounded operators on $\mathcal{H}$.

An alternative way to express the channel $\mathcal{E}_{B|A}$ as is an operator, using a variant of the Choi-Jamiołkowski isomorphism [16, 17]:

$$\rho_{B|A} := \sum_{ij} \mathcal{E}_{B|A}(|i\rangle_A\langle j|) \otimes |i\rangle_{A^*}\langle j|.$$  

(6)

Here, the vectors $\{|i\rangle_A\}$ form an orthonormal basis of the Hilbert space $\mathcal{H}_A$. The vectors $\{|i\rangle_{A^*}\}$ form the dual basis, belonging to $\mathcal{H}_{A^*}$. The operator $\rho_{B|A}$ therefore acts on the Hilbert space $\mathcal{H}_B \otimes \mathcal{H}_{A^*}$. Although the expression above involves an arbitrary choice of orthonormal basis, the operator $\rho_{B|A}$ thus defined is independent of the choice of basis. This version of the Choi-Jamiołkowski isomorphism was chosen because it is both basis-independent and a positive operator. Following [18], we have chosen the operator $\rho_{B|A}$ to be normalized in such a way that $Tr_B(\rho_{B|A}) = I_{A^*}$ (in analogy with the normalization condition $\sum_i P(Y|X) = 1$ for a classical channel $P(Y|X)$).

Suppose that $\rho_B = \mathcal{E}_{B|A}(\rho_A)$. Given that the operator $\rho_{B|A}$ contains all of the information about the channel $\mathcal{E}_{B|A}$, the question arises of how one can express $\rho_B$ in terms of $\rho_{B|A}$ and $\rho_A$. Recall that $\rho_{B|A}$ is defined on $\mathcal{H}_B \otimes \mathcal{H}_{A^*}$, while $\rho_A$ is defined on $\mathcal{H}_A$. As we discuss further in Sec. [11] by defining an appropriate “linking operator”

$$\tau_{A^*} := \sum_{lm} |l\rangle_{A^*}\langle m| \otimes |l\rangle_A\langle m|$$

(7)

with respect to dual orthonormal bases on $\mathcal{H}_A$ and $\mathcal{H}_{A^*}$, one can write $\rho_B = Tr_{AA^*}(\rho_{B|A}\tau_{A^*})$, $\rho_{B|A}$. This expression is meant to be reminiscent of the classical formula $P(Y) = \sum_i P(Y|X)P(X)$.

Given an operator $\rho_{AB\cdots C|D\cdots}$ acting on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \cdots \otimes \mathcal{H}_C \otimes \mathcal{H}_D \otimes \cdots$, we will use the same expression with missing indices to denote the result of taking partial traces on the corresponding factor spaces. For example, given a channel $\rho_{BB|AA}$, we write $\rho_{B|A} := Tr_B(\rho_{BB|AA})$.

We will sometimes use a hat to denote an operator renormalized such that the trace is 1. For example, if $\rho_{B|A}$ is the operator representing a channel from $A$ to $B$, then $\hat{\rho}_{B|A} = (1/d_A)\rho_{B|A}$.

When writing products of operators, we will sometimes suppress tensor products with identities. For example, $(\rho_{B|A} \otimes I_C)(\rho_{C|A} \otimes I_B)$ will be written simply as $\rho_{B|A} \rho_{C|A}$.
FIG. 3: A causal structure relating three quantum systems with $A$ the complete common cause of $B$ and $C$.

B. Main result

The qualitative part of Reichenbach’s principle can be applied to quantum theory with almost no change: if quantum systems $B$ and $C$ are correlated then this must have a causal explanation in one of the five forms listed above. Here, for quantum systems to be correlated means there exist measurements upon them for which the outcomes are correlated.

Finding a quantum version of the quantitative part of Reichenbach’s principle is more subtle. If a quantum system $A$ is a complete common cause of $B$ and $C$ (as depicted in Fig. 3), then one expects there to be some constraint analogous to $P(YZ|X) = P(Y|X)P(Z|X)$ in the classical case. If one tries to do this by generalising the joint distribution $P(XYZ)$, then one immediately faces the problem that textbook quantum theory has no analogue of a joint distribution for a collection of quantum systems in which some are causal descendants of others. The situation is improved if one focusses on finding an analogue of $P(YZ|X)$ instead. In standard quantum theory, as long as a system $A$ is initially uncorrelated with its environment, then the evolution from $A$ to $BC$ is described by a channel $\rho_{BC|A}$, and this seems to be a natural analogue of $P(YZ|X)$. However, even in this case it is not obvious what constraint on $\rho_{BC|A}$ should serve as the analogue of the classical constraint $P(YZ|X) = P(Y|X)P(Z|X)$.

The treatment of quantum systems over time is deferred to the full definition of quantum causal models in Sec. V. This section focuses on the case of a channel $\rho_{BC|A}$.

In Sec. II B we demonstrated how to justify the quantitative part of Reichenbach’s principle from the qualitative part in the classical case by temporarily assuming that all dynamics are fundamentally deterministic. We shall now make an analogous argument in the quantum case by temporarily assuming that quantum dynamics are fundamentally unitary. Just as in the classical case, this assumption simply provides a clean way to motivate our result and the reader need not accept it as philosophically accurate since the result can stand apart from this assumption.

In general, a quantum channel from $A$ to $B$ is given by a CPTP map $\mathcal{E}_{B|A}$. Assuming underlying unitary dynamics, then the output state at $B$ must depend unitarily on $A$ along with some extra ancillary system $\lambda$ in the past of $B$. This can be formalised as follows.

**Definition 3** (Unitary dilation). For a quantum channel $\mathcal{E}_{B|A}$ a quantum unitary dilation is given by some ancillary quantum system $\lambda$ with state $\rho_\lambda$ and some unitary $U$ from $\mathcal{H}_A \otimes \mathcal{H}_\lambda$ to $\mathcal{H}_B \otimes \mathcal{H}_B$ such that

$$\mathcal{E}_{B|A}() = \text{Tr}_B (U(\cdot \otimes \rho_\lambda)U^\dagger),$$

where the dimension of $B$ is fixed by unitarity since $d_{AdA} = d_Bd_B$.

If we represent the channels by our variant of the Choi–Jamiołkowski isomorphism, with $\rho_{B|A}$ representing $\mathcal{E}_{B|A}$ and $\rho_{B|A}^U$ representing $U(\cdot)^U$, then the dilation equation has the form

$$\rho_{B|A} = \text{Tr}_{B\lambda^\ast} \left( \rho_{B|A}^{U^\dagger} \tau_{\lambda^\ast}\rho_\lambda \right)$$

where $\tau_{\lambda^\ast}$ is the linking operator defined in Sec. V.

Just as in the classical case, we would like to apply this to the situation depicted in Fig. 3 where $A$ is the complete common cause for $B$ and $C$. This was easy classically as it is clear what it means for a classical variable, $X$, to have no causal influence on another, $Y$, in a deterministic system. Specifically, if the collection of inputs other than $X$ is denoted $\bar{X}$ so there is a deterministic function $f$ such that $Y = f(X, \bar{X})$, then the assumption that $X$ has no causal influence on $Y$ is formalized as $f(X, \bar{X}) = f'(\bar{X})$ for some function $f'$. In unitary quantum theory the corresponding condition is less obvious, so we spell it out explicitly with a definition.

**Definition 4** (No influence). Consider a unitary channel $\rho_{B|A|\bar{A}}$ from $AA$ to $BB$. A has no causal influence on $B$ if and only if for $\rho_{B|A|\bar{A}} := \text{tr}_{B|A} \rho_{B|A|\bar{A}}$, we have $\rho_{B|A|\bar{A}} = I_{A^\ast} \otimes \rho_{B|A^\ast}$. Equivalently to this definition, $A$ has no causal influence on $B$ in some unitary channel if and only if the marginal output state at $B$ is independent of any operations performed on $A$ before the $A$ system enters the channel. There is a rich literature concerning similar properties of unitary operators from various perspectives. In particular, the results of Ref. [19] are very close to ours (where they use the phrase “nonsignalling” rather than “no causal influence”) and Ref. [20, 21] contains similar results (where they say “semi-causal” rather than “no causal influence”).

We can now apply this to the complete common cause situation of Fig. 3. The channel $\mathcal{E}_{B|A}$ admits a unitary dilation in terms of an ancillary system $\lambda$, for some state $\rho_\lambda$ and unitary $U$ from $\lambda A$ to $B\bar{A}$. Here, an ancillary output $\bar{D}$ is generally required so that dimensions of inputs and outputs match, but is not important and will always be traced out. This dilation is such that $\mathcal{E}_{B|A}() = \text{Tr}_{\bar{D}} (U(\cdot \otimes \rho_\lambda)U^\dagger)$.

Just as in Sec. II B the assumption that $A$ is a complete common cause for $B$ and $C$ implies that the ancilla $\lambda$ can
factorise into ancestrally independent $\lambda_B$ and $\lambda_C$ where $\lambda_B$ has no causal influence on $C$ and $\lambda_C$ has no causal influence on $B$. It follows that systems $\lambda_B$ and $\lambda_C$ exist and are causally related to $A$, $B$, and $C$ as depicted in Fig. 4, where the channel is unitary and given by $U$.

The ancestral independence of $\lambda_B$ and $\lambda_C$ implies that the input quantum state factorises $\rho_{AAB} = \rho_A \rho_B \rho_C$, suggesting the following quantum analog to our “classical compatibility” condition.

**Definition 5** (Quantum compatibility). $\rho_{BC|A}$ is said to be compatible with $A$ being a complete common cause of $B$ and $C$, if it is possible to find ancillary quantum systems $\lambda_B$ and $\lambda_C$, states $\rho_{\lambda_B}$ and $\rho_{\lambda_C}$, and a unitary channel where $\lambda_B$ has no causal influence on $C$ and $\lambda_C$ has no causal influence on $B$, such that these constitute a dilation of $\rho_{BC|A}$.

All that remains is to show that this, together with the qualitative part of the quantum Reichenbach’s principle, implies an appropriate quantitative part (generalising Thm. 1).

**Theorem 2.** The following are equivalent:

1. $\rho_{BC|A}$ is compatible with $A$ being the complete common cause of $B$ and $C$.

2. $\rho_{BC|A} = \rho_{B|A \rho_{C|A}}$.

The proof is in Appendix A. Note that there is no ordering ambiguity on the right-hand side of the second condition, because the two terms must commute. This is seen by taking the Hermitian conjugate of both sides of the equation.

The strong analogy that exists between Thms. 1 and 2 suggests the following definition:

**Definition 6** (Quantum conditional independence of outputs given input). Given a quantum channel $\rho_{BC|A}$, the outputs are said to be quantum conditionally independent given the input, if and only if $\rho_{BC|A} = \rho_{B|A \rho_{C|A}}$.

It is easily seen that the quantum definition reduces to the classical definition in the case that the channel $\rho_{BC|A}$ is invariant under the operation of completely dephasing the systems $A$, $B$, and $C$ in some basis. More precisely: if fixed bases are chosen for $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_C$, and the operator $\rho_{BC|A}$ is diagonal when written with respect to the tensor product of these bases, then the outputs are quantum conditionally independent given the input if and only if the classical channel defined by the diagonal elements of the matrix has the property that the outputs are conditionally independent given the input.

With this terminological convention in hand, we can express our quantum version of the quantitative part of Reichenbach’s principle as follows: if a channel $\rho_{BC|A}$ is compatible with $A$ being a complete common cause of $B$ and $C$, then for this channel $B$ and $C$ are quantum conditionally independent given $A$.

The $1 \rightarrow 2$ implication in the theorem is what establishes the quantum version of the quantitative part of Reichenbach’s principle.

The $2 \rightarrow 1$ implication is pertinent to causal inference: analogously to the classical case, if one grants the implausibility of fine-tuning, then one must grant that the most plausible explanation of the quantum conditional independence of outputs $B$ and $C$ given input $A$ is that $A$ is a complete common cause of $B$ and $C$.

Thm. 2 and the surrounding discussion, motivates the definition of quantum causal models given in Sec. V. For the rest of this section we make some further remarks about the quantum version of Reichenbach’s principle.

C. Alternative expressions for quantum conditional independence of outputs given input

Classically, conditional independence of $Y$ and $Z$ given $X$ is standardly expressed as $P(YZ|X) = P(Y|X)P(Z|X)$. However, there are alternative ways of expressing this constraint.

For instance, if one defines the joint distribution over $X, Y, Z$ that one obtains by feeding the uniform distribution on $X$ into the channel $P(YZ|X)$—that is, $P(XYZ) = P(Y|X)P(Z|X)$, where $d_X$ is the cardinality of $X$—then $Y$ and $Z$ being conditionally independent given $X$ in $P(YZ|X)$ can be expressed as the vanishing of the conditional mutual information of $Y$ and $Z$ given $X$ in the distribution $P(XYZ)$.

This is defined as $I(Y:Z|X) := H(Y,X) + H(Z,X) − H(X,Y,Z) − H(X)$, with $H(\cdot)$ denoting the Shannon entropy of the marginal on the subset of variables indicated in its argument. Therefore, the condition is simply $I(Y:Z|X) = 0$.

Similarly, if $Y$ and $Z$ are conditionally independent given $X$ in $P(YZ|X)$, then it is possible to mathematically represent the channel $P(YZ|X)$ as the following sequence of operations: copy $X$, then process one copy into $Y$ via the channel $P(Y|X)$ and process the other into $Z$ via the channel $P(Z|X)$.

We present here the quantum analogues of these alternative expressions. They will be found to be useful for developing intuitions about quantum conditional independence and in the proofs. Recall that the quantum conditional mutual information of $B$ and $C$ given $A$ is defined as $I(B : C|A) = S(B, A) + S(C, A) − S(A, B, C) − S(A)$.
where $S(\cdot)$ denotes the von Neumann entropy of the reduced trace-one quantum state on the subsystem that is specified by its argument.

**Theorem 3.** Given a channel $\rho_{BC|A}$, the following conditions are also equivalent to the conditions 1 and 2 of Thm. 2:

3. $I(B : C|A) = 0$ where $I(B : C|A)$ is the quantum conditional mutual information of $B$ and $C$ given $A$ evaluated on the (positive, trace-one) operator $\hat{\rho}_{BC|A}$.

4. The Hilbert space for the $A$ system can be decomposed as $\mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i^L} \otimes \mathcal{H}_{A_i^R}$ and $\rho_{BC|A} = \sum_i (\rho_{B|A_i^L} \otimes \rho_{C|A_i^R})$, where for each $i$, $\rho_{B|A_i^L}$ represents a CPTP map $B(\mathcal{H}_{A_i^L}) \to B(\mathcal{H}_B)$, and $\rho_{C|A_i^R}$ a CPTP map $B(\mathcal{H}_{A_i^R}) \to B(\mathcal{H}_C)$.

The proof is in Appendix A. The theorem that conditions 3 and 4 are equivalent follows as a corollary of Thm. 6 of Ref. [22], our main contribution is showing that these are also equivalent to conditions 1 and 2 of Thm. 2.

The final condition can be described as follows. First one imagines decomposing the system $A$ into a direct sum of subspaces, each of which is denoted $A_i$. For each $i$, the subspace $A_i$ is split into two factors, denoted $A_i^L$ and $A_i^R$, with one factor evolving via a channel $\rho_{B|A_i^L}$ into system $B$, and the other factor evolving via $\rho_{C|A_i^R}$ into system $C$. In the special case where $i$ is only a single value of $i$, this is simply a factorization of the $A$ system into two parts. In the special case where all of the $A_i^L$ and $A_i^R$ are 1-dimensional Hilbert spaces, it is simply a coherent version of a classical copy operation.

**D. Circuit representations**

It is instructive to summarize the contents of Thms. 1 and 2 using circuit diagrams.

Consider Fig. 5. Equality (4) in the figure asserts that $Y$ and $Z$ are conditionally independent given $X$ in $P(YZ|X)$. Here conditional independence is expressed in the form of the classical analogue of condition 4 in Thm 3 based on making a copy of $X$, as discussed at the beginning of Sec. H. Equality (1) asserts that the conditional probability distribution $P(YZ|X)$ always admits of a classical dilation of the sort described below Definition 1 while equality (3) asserts simply that each of $P(Y|X)$ and $P(Z|X)$ admits of such a classical dilation. Finally, equality (2) asserts that $P(YZ|X)$ is compatible with $X$ being a complete common cause of $Y$ and $Z$.

Analogous circuit diagrams can be provided in the quantum case, as depicted in Fig. 6. In this case, equality (4) asserts that the channel $\rho_{BC|A}$ is such that $B$ and $C$ are quantum conditionally independent given $A$, according to the definition we have proposed. The circuit decomposition follows from the expression for quantum conditional independence described in condition 4 of Thm. 3. Because quantum systems cannot be copied, there is no possibility of an operation sending a copy of $A$ to each of the two gates. Instead, we introduce a new symbol that indicates the decomposition of the Hilbert space $\mathcal{H}_A$ into a direct sum of tensor products. It is a circle decorated with the set $\{\lambda\}$, where the value $\lambda$ indexes the terms in the direct sum. For each value of $\lambda$, the left-
hand wire carries the factor $\mathcal{H}_{A_L}$ and the right-hand wire the factor $\mathcal{H}_{A_R}$. The gate applied to the left-hand wire is decorated by a set of quantum channels, also indexed by $i$, representing the fact such that the gate acts as the $i$th channel on the $i$th factor.

Equality (1) in Fig. 6 simply asserts the fact that $\rho_{BC|A}$ admits a unitary dilation.

Equality (3) in Fig. 6 asserts that for each $i$, the channel $\rho_{B|iA}$ admits a unitary dilation (where the unitary is denoted $V_i$ and acts on $\mathcal{H}_{AB} \otimes \mathcal{H}_{A_L}$) and the channel $\rho_{C|iA}$ admits a unitary dilation (where the unitary is denoted $B_i$ and acts on $\mathcal{H}_{AC} \otimes \mathcal{H}_{A_R}$).

Finally, equality (2) asserts that $\rho_{BC|A}$ is compatible with $A$ being a complete common cause of $B$ and $C$ by depicting conditions under which $\lambda_B$ has no influence on $C$ and $\lambda_C$ has no influence on $B$.

E. Examples

1. A unitary transformation

Consider the case in which inputs $A$ and $D$ evolve, via a generic unitary transformation $U$ into outputs $B$ and $C$. In Fig. 7 we illustrate the circuit and the corresponding causal diagram.

The channel $\rho_{BC|AD}$ which one obtains in this case is compatible with the complete common cause of $B$ and $C$ being the composite system $AD$. This follows from the fact that $\rho_{BC|AD}$ is its own dilation and hence trivially satisfies the condition for compatibility laid out in Def. 5. It follows from Thm. 2 that for such a $\rho_{BC|AD}$, the outputs $B$ and $C$ are quantum conditionally independent given the input $AD$, which means that $\rho_{BC|AD} = \rho_{B|AD}\rho_{C|AD}$, as can also be verified by direct calculation. Similarly, the alternative expressions for this sort of quantum conditional independence, namely, conditions 3 and 4 of Thm. 5 can be verified to hold.

2. Coherent copy vs. incoherent copy

Consider the simple example of a classical channel, taking $X$ to $Y, Z$, where $X, Y, Z$ are bit-valued and

$$0_X \rightarrow 0_Y0_Z, \quad 1_X \rightarrow 1_Y1_Z.$$  \hspace{1cm} (8)

The outputs of the channel are conditionally independent given the input. In particular, if $X$ is some variable that obtains at an earlier time, and $Y, Z$ two variables at a later time, with the channel describing the evolution of $X$ into $Y, Z$, then it is intuitively obvious that variation in $X$ fully explains any correlation between $Y$ and $Z$. Indeed this example may be seen as the paradigmatic case of the explanation of classical correlations via a complete common cause.

One quantum analogue of this channel is the incoherent copy of a qubit: a qubit $A$ is measured in the computational basis; if 0 is obtained, then prepare the state $|00\rangle_{BC}$ and if 1 is obtained, prepare $|11\rangle_{BC}$. The operator representing this channel is

$$\rho_{BC|A} = |000\rangle\langle 000|_{BCA^*} + |111\rangle\langle 111|_{BCA^*}.$$  

It is easily verified that this operator satisfies each of the conditions of Thm. 2. The decomposition of the $A$ Hilbert space implied by Condition 4 is

$$\mathcal{H}_A = (\mathbb{C} \otimes \mathbb{C}) \oplus (\mathbb{C} \otimes \mathbb{C}),$$

where $\mathbb{C}$ is the 1-dimensional complex Hilbert space, i.e., the complex numbers.

The other direct quantum analogue of the classical copy above is the coherent copy of a qubit:

$$\alpha|0\rangle_A + \beta|1\rangle_A \rightarrow \alpha|0\rangle_B|0\rangle_C + \beta|1\rangle_B|1\rangle_C,$$ \hspace{1cm} (9)

with

$$\rho_{BC|A} = (|000\rangle_{BCA^*} + |111\rangle_{BCA^*})\langle 000|_{BCA^*} + (111|_{BCA^*},$$

which corresponds to an unnormalized GHZ state. It can easily be verified that $I(B : C|A) = 1$ for a trace-one version of this state, hence it is not the case that outputs $B$ and $C$ are quantum conditionally independent given the input $A$. There is, then, no way in which this channel can arise as a marginal channel in a situation in which $A$ is the quantum complete common cause of $B$ and $C$.

At first blush, this conclusion may seem surprising. Given the evolution of Eq. (9), where would correlations come from, other than being completely explained by the input $A$?

The puzzle is resolved by considering the dilation of the coherent copy to a unitary transformation, and the interpretation of quantum pure states. Consider Figs. 8 and 9 which show realizations of the classical copy, and the quantum coherent copy, via the classical and quantum CNOT gates respectively.\footnote{[69]}.
In the classical case, there are two reasons why any correlation between \(Y\) and \(Z\) must be entirely explained by statistical variation in the value of \(X\). First, the ancillary variable \(\lambda\) is prepared deterministically with value 0, so there is no possibility that statistical variation in the value of \(\lambda\) underwrites the correlations between \(B\) and \(C\). Second, the classical CNOT gate, 

\[
\begin{align*}
0_X0_\lambda & \rightarrow 0_Y0_Z, \\
0_X1_\lambda & \rightarrow 0_Y1_Z, \\
1_X0_\lambda & \rightarrow 1_Y0_Z, \\
1_X1_\lambda & \rightarrow 1_Y1_Z,
\end{align*}
\]  

(10) (which one easily verifies to reduce to the classical copy of Eq. 8 when one sets \(\lambda\) to 0), has the causal structure depicted in Fig. 8 so that \(\lambda\) does not act as a common cause of \(Y\) and \(Z\) but only a local cause of \(Z\).

In the quantum case, there is no analogue for either of these reasons. Concerning the second reason, the quantum CNOT has the causal structure depicted in Fig. 9: the quantum CNOT is such that not only does \(A\) have a causal influence on \(C\), but \(\lambda\) has a causal influence on \(B\) as well. In other words, unlike the classical CNOT, there is a back action of the target on the control. It follows that in the quantum case, \(\lambda\) can act as a common cause of \(B\) and \(C\). Furthermore, the ancilla is prepared in a quantum pure state \(|\lambda\rangle\). This is dis-analogous to a point distribution on the value 0 for the classical variable \(\lambda\) if one takes the view that a quantum pure state represents maximal but incomplete information about a quantum system [24, 28]. In this case, one must allow for the possibility that some correlation between \(B\) and \(C\) is due to the ancilla, in which case \(A\) is not the complete common cause of \(B\) and \(C\). [70].

F. Generalization to one input, \(k\) outputs

Thms. 2 and 3 which apply to quantum channels with one input and two outputs, can be generalized to the case of one input and \(k\) outputs.

Consider a channel \(\rho_{B_1...B_k|A}\), and let \(\bar{B}_i\) denote the collection of all outputs apart from \(B_i\). The notion of quantum compatibility from Def. 5 generalizes in the obvious way: \(\rho_{B_1...B_k|A}\) is said to be compatible with \(A\) being a complete common cause of \(B_1\) \(\ldots\) \(B_k\) if it is possible to find ancillary quantum systems \(\lambda_1\) \(\ldots\) \(\lambda_k\), states \(\rho_{\lambda_1}\) \(\ldots\) \(\rho_{\lambda_k}\), and a unitary channel where, for each \(i\) \(\lambda_i\) has no causal influence on \(\bar{B}_i\), such that these constitute a dilation of \(\rho_{B_1...B_k|A}\).

The generalization of Thms. 2 and 3 consolidated into a single theorem, is as follows:

**Theorem 4.** The following are equivalent:

1. \(\rho_{B_1...B_k|A}\) is compatible with \(A\) being a complete common cause of \(B_1\) \(\ldots\) \(B_k\).
2. \(\rho_{B_1...B_k|A} = \rho_{A|1}\cdots\rho_{A|k}\), where for all \(i, j\), \([\rho_{B_i|A}, \rho_{B_j|A}] = 0\).
3. For each \(i\), \(I(B_i : \bar{B}_i|A) = 0\) where \(I(B_i : \bar{B}_i|A)\) is the quantum conditional mutual information evaluated on the (positive, trace-one) operator \(\hat{\rho}_{B_1...B_k|A}\).

4. The Hilbert space for the \(A\) system can be decomposed as \(\mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i}^1 \otimes \cdots \otimes \mathcal{H}_{A_i}^k\) such that \(\rho_{B_1...B_k|A} = \sum_i \left(\rho_{B_1|A_i}^1 \otimes \cdots \otimes \rho_{B_k|A_i}^k\right)\), where for each \(i\), and each \(l \in \{1, \ldots, k\}\), \(\rho_{B_i|A_i^l}\) represents a CPTP map \(\mathcal{B}(\mathcal{H}_{A_i}) \rightarrow \mathcal{B}(\mathcal{H}_{B_i})\).

The proof is in Appendix 3. By analogy to the classical case, we refer to conditions 2, 3 and 4 of Thm. 4 as \(B_1\) \(\ldots\) \(B_k\) being quantum conditionally independent given \(A\) for the channel \(\rho_{B_1...B_k|A}\).

IV. CAUSAL MODELS

A. Definitions

Reichenbach’s principle is important because it generalizes to the modern formalism of causal models [8, 9].

A causal model consists of two entities: (i) a causal structure, represented by a directed acyclic graph (DAG) where the nodes represent random variables and the directed edges represent the directed causal influences among these (several examples have already been presented in this article), and (ii) some parameters, which specify the strength of the causal dependences. Some terminology is required to present the formal definitions.

Given a DAG with nodes \(X_1, \ldots, X_n\), let Parents\((i)\) denote the parents of node \(X_i\), that is, the set of nodes...
that have an arrow into \( X_i \), and let \( \text{Children}(i) \) denote the children of node \( X_i \), that is, the set of nodes \( X_j \) such that there is an arrow from \( X_i \) to \( X_j \). The descendants of \( X_i \) are those nodes \( X_j \) such that \( X_i \) has an arrow into \( X_j \). The ancestors of \( X_i \) are those nodes \( X_j \) such that there is a directed path from \( X_i \) to \( X_j \). The ancestors of \( X_i \) are those nodes \( X_j \) such that \( X_i \) is a descendent of \( X_j \).

**Definition 7.** A causal model specifies a DAG, with nodes corresponding to random variables \( X_1, \ldots, X_n \), and a family of conditional probability distributions \( \{ P(X_i|\text{Parents}(i)) \} \), one for each \( i \).

**Definition 8.** Given a DAG, with random variables \( X_1, \ldots, X_n \) for nodes, and given an arbitrary joint distribution \( P(X_1 \ldots X_n) \), the distribution is said to be Markov for the graph if and only if it can be written in the form of

\[
P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i|\text{Parents}(i)).
\]

(Recall that each conditional \( P(X_i|\text{Parents}(i)) \) can be computed from the joint \( P(X_1 \ldots X_n) \).)

The generalization of Reichenbach’s principle that is afforded by the formalism of causal models is this: if there are statistical dependences among variables \( X_1, \ldots, X_n \), expressed in the particular form of the joint distribution \( P(X_1 \ldots X_n) \), then there should be a causal explanation of these dependences in terms of a DAG relative to which the distribution \( P(X_1 \ldots X_n) \) is Markov.

Note that an alternative way of formalizing the Markov property is that \( P(X_1 \ldots X_n) \) is Markov for the graph if and only if, for each \( i \), \( P(X_i|\text{Parents}(i)) = P(X_i|\text{Nondesc}(i)) \), where \( \text{Nondesc}(i) \) is the set of non-descendants of node \( X_i \). The intuitive idea is that the parents of a node screen off that node from the rest of the graph: once the values of the parents are fixed, the values of other non-descendent nodes are irrelevant to the value of \( X_i \).

Note that the qualitative and quantitative parts of Reichenbach’s principle are easily seen to be special cases of the requirement that for a joint distribution to be explainable by the causal structure of some DAG, it must be Markov for that DAG: if two variables, \( Y \) and \( Z \), are ancestrally independent in the graph, then any distribution that is Markov for this graph must factorize on these, \( P(YZ) = P(Y)P(Z) \), which is the qualitative part of Reichenbach’s principle in its contrapositive form; if two variables, \( Y \) and \( Z \), have a variable \( X \) as a complete common cause, as in the DAG of Fig. 1, then any distribution that is Markov for the graph must satisfy \( P(YZ|X) = P(Y|X)P(Z|X) \), which is the quantitative part of Reichenbach’s principle.

**B. Justifying the Markov condition**

Just as we previously asked whether there was some principle that forced a rational agent to assign probability distributions in accordance with the quantitative part of Reichenbach’s principle, we can similarly ask why a rational agent who takes causal relationships to be given by a particular DAG should arrange their beliefs so that the joint distribution is Markov for the DAG.

The justification of the Markov condition parallels the justification of the quantitative part of Reichenbach’s principle that was presented in Sec. 111B. We begin by outlining what the qualitative part of Reichenbach’s principle and the assumption of fundamental determinism imply for any arbitrary causal structure.

**Definition 9** (Classical compatibility with a DAG). \( P(X_1 \ldots X_n) \) is said to be compatible with a DAG \( G \) with nodes \( X_1, \ldots, X_n \) if one can find a DAG \( G' \) that is obtained from \( G \) by adding extra nodes \( \lambda_1, \ldots, \lambda_n \), such that for each \( i \), the node \( \lambda_i \) has a single outgoing arrow, to \( X_i \), and no incoming arrows, and one can find, for each \( i \), a distribution \( P(\lambda_i) \) and a function \( f_i \) from \( (\lambda_i, \text{Parents}(i)) \) to \( X_i \) such that

\[
P(X_1 \ldots X_n) = \sum_{\lambda_1 \ldots \lambda_n} \prod_{i=1}^{n} \delta(X_i, f_i(\lambda_i, \text{Parents}(i)))P(\lambda_i).
\]

**Theorem 5** (Ref. [8]). Given a joint distribution \( P(X_1 \ldots X_n) \) and a DAG \( G \) with nodes \( X_1, \ldots, X_n \), the following are equivalent:

1. \( P(X_1 \ldots X_n) \) is compatible with the causal structure described by the DAG \( G \).
2. \( P(X_1 \ldots X_n) \) is Markov for \( G \), that is,

\[
P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i|\text{Parents}(i)).
\]

The \( 1 \rightarrow 2 \) implication in Thm. 5 can be read as follows: if it is granted that causal relationships are indicative of underlying deterministic dynamics, and that the qualitative part of Reichenbach’s principle is valid, then, on pain of irrationality, an agent’s assignment \( P(X_1 \ldots X_n) \) must be Markov for the original graph.

The \( 2 \rightarrow 1 \) implication in Thm. 5 like that of Thm. 4 is pertinent for causal inference. It asserts that if one observes a distribution \( P(X_1 \ldots X_n) \), then of the causal models that are compatible with this distribution, the only ones that do not require fine-tuning of the parameters are those involving DAGs relative to which the distribution is Markov.

**V. QUANTUM CAUSAL MODELS**

**A. The proposed definition**

In our treatment of the simple causal scenario where \( A \) is a complete common cause of \( B \) and \( C \) (the DAG of Fig. 5), we focussed on what form is implied for the
quantum channel \(\rho_{BC|A}\). But there has not been any attempt to define a quantity analogous to the classical joint distribution, that is, a quantity analogous to \(P(XYZ)\) in the case of the DAG of Fig. 4, nor indeed other classical Bayesian conditionals such as \(P(X|YZ)\). For results that aim to achieve such analogues, see Ref. 15, 27. See also Ref. 29, however, where it is shown that there are significant blocks towards establishing an analogue of a classical joint distribution when the set of quantum systems includes some that are causal descendants of others.

This work takes a different approach. The interpretation of a quantum causal model will be that each node represents a local region of time and space, with channels such as \(\rho_{BC|A}\) describing the evolution of quantum systems in between these regions. At each node, there is the possibility that an agent is present with the ability to intervene inside that local region. Each node \(A_i\) will be then associated with two Hilbert spaces, one corresponding to the incoming system (before the agent’s intervention) and the dual space, which corresponds to the outgoing system (after the agent’s intervention). A quantum causal model will consist of a specification of all the quantum channels representing the evolution of systems in between nodes, with the operational significance of a network being that it is used to calculate joint probabilities for the agents to obtain the various possible joint outcomes for their interventions. This way of treating quantum systems over time has appeared in various different approaches in the literature, including the multi-time formalism 30–33, the quantum combs formalism 34–36, the process matrices formalism 37–39, and a number of other formalisms besides 14, 40–42.

The discussion of classical causal models in Sec. IV and the results of Sec. III for the special case of \(A\) a complete common cause of \(B\) and \(C\), suggest the following generalization.

**Definition 10.** A quantum causal model specifies a DAG, with nodes \(A_1, \ldots, A_n\), supplemented with the following. For each node \(A_i\), there is associated a finite-dimensional Hilbert space \(\mathcal{H}_i\) (the ‘input’ Hilbert space), and the dual space \(\mathcal{H}_i^*\) (the ‘output’ Hilbert space). For each node \(A_i\), there is associated a quantum channel, described by an operator \(\rho_{A_i|\text{Parents}(i)} \in \mathcal{B}(\mathcal{H}_i \otimes \mathcal{H}_{\text{Parents}(i)}^*)\), where \(\mathcal{H}_{\text{Parents}(i)}^*\) is the tensor product of the output Hilbert spaces associated with the parents of \(A_i\). These channels commute pairwise, i.e., for any \(i, j\), \(|\rho_{A_i|\text{Parents}(i)} \rho_{A_j|\text{Parents}(j)}^*| = 0\). The overall state is given by

\[
\sigma_{1\ldots n} = \prod_{i=1}^{n} \rho_{A_i|\text{Parents}(i)},
\]

where \(\sigma_{1\ldots n}\) is an operator on \(\bigotimes_{i=1}^{n} (\mathcal{H}_i \otimes \mathcal{H}_i^*)\).

Recall from Section III that, given a quantum channel \(\rho_{BC|A}\), it is compatible with \(A\) being the complete common cause of \(B\) and \(C\) if and only if \(\rho_{BC|A} = \rho_{B|A}\rho_{C|A}\), and if this holds, then \(\rho_{B|A}\rho_{C|A} = 0\). The definition of a quantum causal model, in particular, the stipulation that the channels commute pairwise, generalizes this idea.

**B. Bayesian vs. do-conditionals**

In the theory of classical causal models, there is an important distinction between two types of conditionals: Bayesian conditionals and do-conditionals. Roughly, a Bayesian conditional gives the probability distribution for a variable (or collection thereof), given that another variable (or collection thereof) is discovered to have a certain value, as computed using Bayes’ rule: \(P(Y|X) = P(YX)/P(X)\). A do-conditional, on the other hand, gives the probability distribution for a variable (or collection thereof) given that an agent intervenes deliberately and fixes another variable (or collection thereof) to have a certain value. Do-conditionals are written \(P(Y|\text{do } X)\). For a full treatment, see Ref. 8.

Conceptually, a classical channel with input \(X\) and output \(Y\) is better thought of as defined by a do-conditional \(P(Y|\text{do } X)\), rather than the Bayesian conditional \(P(Y|X)\). A quantum channel with input \(A\) and output \(B\), represented by \(\rho_{B|A}\), is best thought of as the analogue of a do-conditional. Hence using a more careful notation, a quantum channel might be written \(\rho_{B|\text{do } A}\) and, given a channel that satisfies the conditions of Thm. 2, we would say that the outputs \(B\) and \(C\) are quantum do-conditionally-independent given \(A\). But this is cumbersome, which is why we have not made use of these distinctions throughout the bulk of this work.

In the definition of a classical causal model above (Def. 4), the DAG is supplemented with Bayesian conditional probabilities of the form \(P(X_i|\text{Parents}(i))\). It is easy to show that when \(P(X_1\ldots X_n)\) is Markov for the graph, then for all \(i\), \(P(X_i|\text{Parents}(i)) = P(X_i|\text{do } \text{Parents}(i))\). Hence an equivalent definition of a classical causal model would be as a DAG, supplemented with a do-conditional \(P(X_i|\text{do } \text{Parents}(i))\) for each \(i\). Our definition of a quantum causal model might be regarded as analogous to a classical causal model in this sense. In particular, we do not in this work attempt any generalization of the classical concept of a Bayesian conditional to the quantum case. Even more exactly, our definition of a quantum causal model is analogous to a generalization of classical causal models in which the random variable at each node is split into two copies, with the possibility of an intervention at the node.

For similar remarks in the context of a related approach, see 39.

**C. Making predictions**

In order to see how a quantum causal model is used to calculate probabilities for the outcomes of agents’ interventions, consider a quantum causal model with nodes
Let the intervention at node $A_i$ have classical outcomes labelled by $k_i$. The intervention is defined by a quantum instrument (that is, by a set of completely-positive trace-non-increasing maps, one for each outcome) which sum to a trace-preserving map. In order to write the probabilities for the outcomes in a simple form, it is useful to define the instrument in such a way that the maps take operators on $\mathcal{H}_i^*$ into operators on $\mathcal{H}_i^*$. Hence, suppose that the outcome $k_i$ corresponds to the map $\mathcal{E}_i^{k_i} : \mathcal{B}(\mathcal{H}_i^*) \rightarrow \mathcal{B}(\mathcal{H}_i^*)$ and let

$$\tau_i^{k_i} = \sum_{lm} \mathcal{E}_i^{k_i}(|l\rangle_A\langle m|) \otimes |l\rangle_A\langle m|.$$ 

The outcome $k_i$ of the agent’s intervention can then be represented by the (positive, basis-independent) operator $\tau_i^{k_i}$. If an agent does not intervene at the node $A_i$, this corresponds to the operator

$$\tau_i^{id} = \sum_{lm} |l\rangle_A\langle m| \otimes |l\rangle_A\langle m|.$$ 

The joint probability for the agents to obtain outcomes $k_1, \ldots, k_n$ is given by

$$P(k_1 \ldots k_n) = \text{Tr}(\sigma_{1 \ldots n}(\tau_1^{k_1} \otimes \cdots \otimes \tau_n^{k_n})).$$

We can also define operations on the state $\sigma_{1 \ldots n}$ corresponding to marginalization over some nodes, much as classical marginal distributions are written $P(X) = \sum_y P(XY)$. For example,

$$\sigma_{1 \ldots (i-1)(i+1) \ldots n} := \sum_{k_i} \text{Tr}_{A_iA_i^*} (\sigma_{1 \ldots n} \tau_i^{k_i}).$$

If there is no intervention at the node $A_i$, then

$$\sigma_{1 \ldots (i-1)(i+1) \ldots n} := \text{Tr}_{A_iA_i^*} (\sigma_{1 \ldots n} \tau_i^{id}).$$

There is in general no reason to expect that the marginal operator $\sigma_{1 \ldots (i-1)(i+1) \ldots n}$ will be compatible with the remaining network. Even in the classical case, the marginal distribution $P(X_1 \ldots X_{i-1} X_{i+1} \ldots X_n)$ obtained by marginalizing over some variable $X_i$ is not in general Markov for the graph obtained by removing the $X_i$ node, along with its incoming and outgoing arrows.

D. Examples

1. Confounding common cause

Consider a quantum causal model with the DAG depicted in Fig. 10. The DAG is supplemented with the quantum channels $\rho_{C|AB}, \rho_{B|A}$, and $\rho_A$, where the latter is simply a quantum state on $\mathcal{H}_A$ (which can also be thought of as a channel from the trivial, or 1-dimensional system into $A$).

The corresponding state is $\sigma = \rho_{C|AB} \rho_{B|A} \rho_A$, where $\sigma$ acts on the Hilbert space $\mathcal{H}_C \otimes \mathcal{H}_B \otimes \mathcal{H}_B \otimes \mathcal{H}_A \otimes \mathcal{H}_A$.

By stipulation, the channels commute pairwise. This is immediate in the case of, say, $\rho_{B|A}$ and $\rho_A$, since these operators are non-trivial on distinct Hilbert spaces. But it is significant in the case of $\rho_{C|BA}$ and $\rho_{B|A}$, both of which act non-trivially on $\mathcal{H}_A^*$. From Thm. [2] this implies that $\mathcal{H}_A^*$ decomposes as $\mathcal{H}_A^* = \bigoplus_i \mathcal{H}_{A|L}^* \otimes \mathcal{H}_{A|R}^*$, with $\rho_{C|BA}$ acting trivially on (say) the right-hand factors and $\rho_{B|A}$ acting trivially on the left-hand factors.

In words, the evolution undergone by the system emerging from $A$ is as follows: a (possibly degenerate) von Neumann measurement is performed and, depending on the outcome, the $A$ system is split into two pieces. One piece evolves to become the input at $B$. The output at $B$ is then recombined with the other piece, and evolves to become the input at $C$.

There is a noteworthy difference here between classical and quantum causal models. In the case of a classical causal model, with graph indicated by Fig. [10] and ABC random variables, any distribution $P(ABC)$ is Markov for the graph. In the quantum case, however, the fact that the output Hilbert space for the $A$ system decomposes as described is a very strong constraint on the kinds of evolution of a quantum system that can be described by this DAG.

It is also instructive to consider classical and quantum causal models with the causal structure shown in Fig. [11]. Consider a classical causal model with this graph, with $A, B, C, \lambda$ random variables, and a joint distribution $P(ABC\lambda)$. Suppose that $P(\lambda = 0) = 1$. Seeing as $\lambda$ does not vary, the $\lambda$ node can be removed from the graph (along with its outgoing arrows) and the marginal distribution $P(ABC)$ is Markov for the portion of the graph that remains. (Of course, it must be, since any distribution $P(ABC)$ is Markov for the graph of Fig. [10]).

In the quantum case, the DAG of Fig. [11] may represent, for example, the evolution of a qubit over three time steps, $A, B$ and $C$, where the qubit interacts with an environment whose initial state is $\rho_A$. Nodes corresponding to the environment at the second and third time steps are omitted. Given that over the course of this evolution, information can flow from the qubit to the environment, and back again, it is necessary to include an arrow from $C$ to $B$. The resulting DAG is identical to that of the classical case, so the results from the classical case apply.

FIG. 10: A causal network with $A$ a common cause for $B$ and $C$ and with $B$ a parent of $C$. 
A to C, as well as from λ to B and λ to C.

A quantum causal model with this DAG defines commuting channels $\rho_{C|BA\lambda}$, $\rho_{B|A\lambda}$, $\rho_A$, $\rho_{\lambda}$. From the fact that $\rho_{C|BA}$ and $\rho_{B|AA}$ commute, we conclude that the Hilbert space $H_A^r \otimes H_A^s$ decomposes as a direct sum over direct products. However, a decomposition of $H_A^r \otimes H_A^s$ as a direct sum over direct products does not imply a decomposition of the Hilbert space $H_A^r$ alone as a direct sum over direct products. Further, suppose that $\rho_{\lambda}$ is the pure state $|0\rangle\langle 0|$. Then, unlike in the classical case of a point distribution on the variable $\lambda$, we cannot remove the quantum $\lambda$ node and expect the resulting marginal channels to be compatible with the causal structure of Fig. 10. This makes intuitive sense if one takes the view (as per Section III E) that a quantum pure state represents maximal but incomplete information. Uncertainty about the $\lambda$ system may underwrite correlations between $B$ and $C$, in such a way that these correlations cannot be understood as arising from channels that would be compatible with the causal structure of Fig. 10.

2. A simple case of Bayesian updating

This section discusses another sense in which the quantum notion of conditional independence of the outputs of a channel given the input mirrors qualitatively an important aspect of the classical case.

Consider a classical causal model with the DAG of Fig. 1 and distribution $P(XYZ)$ such that $P(YZ|X) = P(Y|X)P(Z|X)$. A particular feature of this causal scenario is that if new information is obtained about the variable $Y$, for example an agent learns that the value of $Y$ is $Y = y$, then the process of Bayesian updating can go as follows. First update the distribution over $X$ by applying the rule

$$\tilde{P}(X) := P(X|Y = y) = \frac{P(Y = y|X)P(X)}{P(Y = y)}.$$ 

Then use the new probabilities for $X$ to get an updated distribution for $Z$:

$$P(Z|Y = y) = \sum_x P(Z|X)\tilde{P}(X),$$  \hspace{1cm} (13)

where the sum ranges over the values that $X$ may take. Roughly speaking, the process of Bayesian updating “follows the arrows” of the graph. For this it is crucial that the joint distribution $P(XYZ)$ satisfies $P(YZ|X) = P(Y|X)P(Z|X)$, otherwise the term $P(Z|X)$ in Eq. (13) would have to be replaced with $P(Z|Y, X = y)$.

Consider now a quantum causal model, with the DAG of Fig. 1 and state $\sigma_{ABC} = \rho_{B|A}\rho_{PC|A}\rho_A$. Suppose that an agent at $B$ intervenes, obtaining outcome $k_B$, corresponding to the operator $\tau_{B}^{k_B}$. The agent wishes to calculate the probability that an intervention at $C$ yields outcome $k_C$ corresponding to $\tau_{C}^{k_C}$, conditioned on having obtained the outcome $k_B$, and assuming that there is no intervention at $A$. This can be done as follows. First, update the state at $A$ to get

$$\tilde{\sigma}_A := \sigma_{A|k_B} = \frac{\text{Tr}_{BB'}(\sigma_{AB}\tau_{B}^{k_B})}{\text{Tr}(\sigma_{AB}(\tau_{A}^{id} \otimes \tau_{B}^{k_B}))}.$$ 

Then apply the channel $\rho_{C|A}$ to get an updated state at $C$:

$$\sigma_{C|k_B} = \text{Tr}_{AA'}(\rho_{C|A}\tilde{\sigma}_A\tau_{A}^{id}).$$

Finally, calculate the probability of $k_C$:

$$P(k_C|k_B) = \text{Tr}(\sigma_{C|k_B}\tau_{C}^{k_C}).$$

Again, the process of Bayesian updating “follows the arrows” of the graph. For this to work, it was crucial that the channel $\rho_{BC|A}$ satisfies $\rho_{BC|A} = \rho_{B|A}\rho_{PC|A}$.

VI. RELATION TO PRIOR WORK

We now present a short review of prior works on quantum causal models and describe how our own proposal relates to these.

Preliminary work in this area took to form of explorations of Bell-type inequalities (and whether they admit of quantum violations) for novel causal scenarios [13, 14]. Several researchers recognized that the formalism of classical causal models could provide a unifying framework in which to pose the problem of deriving Bell-type constraints, and that this framework might be extended to address the problem of deriving constraints on the correlations that can be obtained with quantum resources [2, 10, 11, 14]. Note that such constraints are expressed entirely in terms of classical settings and classical outcomes of measurements. Henson, Lal and Pusey [46] and Fritz [47] independently proposed definitions of quantum causal models with the purpose of expressing such constraints. In these approaches, each node of the DAG represent a process (which may have a classical outcome), while each directed edge is associated
with a system that is passed between processes. However, despite the fact that their frameworks incorporate post-classical resources, they do not have sufficient structure to define conditional independences between quantum systems.

Operational reformulations of quantum theory such as Refs. [14, 33] helped to set the stage for the development of quantum causal models. Although they were conceived independently of the framework of classical causal models, they were quite similar to that framework insofar as they made use of DAGs—in the form of circuit diagrams—to depict structural features of a set of processes. When the authors of these formulations turned their attention to relativistic causal structure, the frameworks they devised drew even closer in spirit to those of causal models. Prominent examples include: the causaloid framework of Hardy [34], the multi-time formalism of Aharonov and co-workers [35], the quantum combs framework of Chiribella, D'Ariano and Perinotti [34–36], the causal categories of Coecke and Lal [23, 24], and the process matrix formalism of Oreshkov, Costa and Brukner [37, 38]. A common aim of these approaches is to be able to compute the consequences of an intervention upon a particular quantum system within the circuit, and this is precisely one of the tasks that a quantum analogue of a causal model should be able to handle.

Many of these frameworks represent a quantum system at a given region of space-time by two copies of its Hilbert space, one corresponding to the system that is input into the region and one corresponding to the system that is output from it. In this way, the region becomes a “locus of intervention” for the system. By inserting a particular quantum process into the “slot”, one determines the nature of the intervention. This is the approach taken, for instance, in the multi-time formalism of Ref. [31], the quantum combs of Ref. [34], and the process matrices of Ref. [37]. This representation of interventions has a counterpart in classical causal models, for instance in the work of Refs. [39, 40].

Costa and Shrapnel [32] in particular have sought to explicitly cast this sort of framework as a quantum generalization of a causal model. In their approach, the nodes of the DAG are associated with a quantum system localized in a region (understood as a potential locus of intervention) and the collection of edges from one set of nodes to another represent causal processes.

An approach of this sort is required if one seeks to find intrinsically quantum versions of important theorems of classical causal models. For instance, while Henson, Lal and Pusey derive a generalization of the d-separation theorem of classical causal models, it only infers conditional independence relations from d-separation relations for the classical variables in the graph. An intrinsically quantum version of the d-separation theorem, by contrast, would be one which concerns the causal relations among quantum systems (see, for instance, [54]). If a set of nodes representing quantum systems can be described by a joint or conditional state, then one can seek to determine whether factorization conditions on this state are implied by d-separation relations among the quantum systems on the graph. Similarly, while both the approaches of Henson, Lal and Pusey and of Fritz allow one to derive, from the structure of the DAG, constraints on the joint distribution over classical variables embedded therein, they do not address an intrinsically quantum version of this problem. If a set of nodes representing quantum systems can be described by a joint or conditional state, then one can seek to derive constraints on this state directly from the structure of the DAG.

Our own approach aims at an intrinsically quantum generalization of the notion of a causal model. We therefore associate to each node of the DAG a quantum system localized to a space-time region, and we represent it by a pair of Hilbert spaces, corresponding to the input and output of an intervention upon the system. Consequently, our approach is very similar to that of Costa and Shrapnel. Nonetheless, there is a significant difference in how we represent common causes.

In Costa and Shrapnel's work, any node with multiple outgoing edges is represented as a locus of intervention where the output Hilbert space is a tensor product of Hilbert spaces, one for each outgoing edge. As such, any node acting as a common cause must be associated with a composite quantum system. It cannot, for instance, be associated with a single qubit. By contrast, our approach does not constrain the representation of common causes in this fashion. Any quantum system, including a single qubit, may constitute a complete common cause of a collection of other quantum systems. This extra generality is required since, as our examples have shown, the complete common cause of a set of systems can be a single qubit. Second, and more importantly, our work has shown that for a quantum channel whose input is the complete quantum common causes of its n outputs, it is not the case that the channel must split the input into n components, each of which exerts a causal influence on a different output. This is merely one special case of the most general form that such a channel can take. Third, if the complete common causes consist of multiple nodes in the DAG, then it is only the joint Hilbert space of these that must satisfy the condition of factorizing-in-subspaces, while each individual Hilbert space need not.

These differences are likely to have a significant impact on the form of any intrinsically quantum d-separation theorem.

Finally, we mention a third purpose to which quantum causal models can be put. Theorems about classical causal models often concern the sorts of inferences one can make about one variable given information about another. As an example, if Z is a common effect of X and Y, then learning Z can induce correlations between X and Y. As such, one might expect quantum causal models to also constrain the sorts of inferences one can make among quantum variables. Early work by Leifer and Spekkens [18] had this purpose in mind. The authors noted various scenarios in which their proposal could not
be applied, and subsequent work has narrowed down the scope of possibilities for any such proposal. Our own proposal provides the means of making many of the Bayesian inferences considered in Ref. [18]. The case discussed in Sec. VI D 2 is one such example.

There is also prior work on quantum causal models that takes a significantly different approach to the ones described above and for which the relation to our work is less clear. The work of Tucci [57, 58], which is in fact the earliest attempt at a quantum generalization of a causal model, represents causal dependences by complex transition amplitudes rather than quantum channels.

VII. CONCLUSIONS

The field of classical statistics has benefited greatly from analysis provided by the formalism of causal models [8, 9]. In particular, this formalism allows one to infer facts about the underlying causal structure purely from uncontrolled statistical data, a tool with significant applications in all branches of science, economics and the social sciences. Given that some paradoxical features of classical correlations have found satisfying resolutions when viewed through a causal lens, one might wonder to what extent the same is true of quantum correlations.

Starting with the idea that whatever innovation quantum theory might hold for causal models, the intuition contained in Reichenbach’s principle ought to be preserved, we motivated the problem of finding a quantum version of the principle. This required us to determine what constraint a channel from A to BC must satisfy if A is the complete common cause of B and C. We solved the problem by considering a unitary dilation of the channel and by noting that there is no ambiguity in how to represent the absence of causal influences between certain inputs and certain outputs of a unitary. From this, we derived a notion of quantum conditional independence for the outputs of the channel given its input. This inference from a common-cause structure to quantum conditional independence was then generalized to obtain our quantum version of causal models.

Given a state on a quantum causal model, we described how to construct a marginal state for a subset of nodes. We discussed a number of simple examples of quantum channels and causal structures. A theme of the examples is that when there is a difference between the quantum and classical cases, this can often be understood if one takes the view that a quantum pure state represents maximal but incomplete information about a system, hence may underwrite correlations between other systems in a way that a classical pure state cannot.

There are many directions for future work. In the case of classical causal networks, an important theorem states that d-separation is sound and complete for conditional independence in the joint probability distribution [8]. Here, for arbitrary subsets of nodes S, T and U, subsets S and T are said to be d-separated by U if a certain criterion holds, where this is determined purely by the structure of the DAG. An important question is therefore whether d-separation is sound and complete for some natural property of the state σ on a quantum causal network. It is also desirable to relate properties of a quantum causal network to operational statements involving the outcomes of agents’ interventions: under what circumstances, for example, does it follow that there is an intervention by the agents at nodes in a subset U, such that conditioned on the outcome, the interventions by agents at S and T must be independent? Such a result would have an application to quantum protocols. Imagine, for example, a cryptographic scenario in which agents at S and T desire shared correlations that are not screened off by the information held by agents at U.

In the classical case, there has been a great deal of work on the problem of causal inference [8, 59–61]: given only certain facts about the joint probabilities, e.g., a set of conditional independences, what can be inferred about the underlying causal structure? For an initial approach to quantum causal inference, with a quantum-over-classical advantage in a simple scenario, see [14]. The formalism of quantum causal networks described here is the appropriate framework for inferring facts about underlying, intrinsically quantum, causal structure, given observed facts about the outcomes of interventions by agents.

Recently, there has been much interest in deriving bounds on the correlations achievable in classical causal models [59, 60, 62, 63]. Such bounds constitute Bell-like inequalities for arbitrary causal structures. The main technical challenge in deriving such inequalities is that they—unlike standard Bell inequalities—are nonlinear in character, hence standard techniques for deriving Bell inequalities are not applicable. By adapting these new techniques to the formalism presented here, one could perhaps systematically derive bounds on the quantum correlations achievable in certain quantum causal models thereby providing a general method of deriving Tsirelson-like bounds for arbitrary causal structures.

Finally, it would be interesting to extend the formalism to explore the possibility that certain quantum scenarios are best understood as involving a quantum-coherent combination of different causal structures [57, 58, 64, 65]. It has been argued that the possibility of such indefinite causal structure may be significant for the project of unifying quantum theory with general relativity [54].

Acknowledgments

RWS thanks Elie Wolfe for helpful discussions. This work was supported by EPSRC grants, the EPSRC National Quantum Technology Hub in Networked Quantum Information Technologies, an FQXi Large Grant, University College London, and by the Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is
supervised by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

[1] H. Reichenbach, *The direction of time*, edited by M. Reichenbach (Univ. of California Press, Berkeley - Los Angeles, 1956).
[2] C. J. Wood and R. W. Spekkens, New Journal of Physics 17, 033002 (2015), arXiv:1204.4119 [quant-ph].
[3] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University Press, Cambridge, 1964).
[4] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. 23, 880 (1969).
[5] B. Hensen, H. Bernien, A. Dréau, A. Reiserer, N. Kalb, M. Blok, J. Ruitenberg, R. Vermeulen, R. Schouten, C. Abellán, et al., Nature 526, 682 (2015).
[6] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierbaum, A. Dréau, A. Reiserer, N. Kalb, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, et al., Physical review letters 115, 250402 (2015).
[7] M. Giustina, M. A. Verrsteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, et al., Physical review letters 115, 250401 (2015).
[8] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, Cambridge, 2009).
[9] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*, 2nd ed. (Cambridge University Press, Cambridge, 2001).
[10] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Physical Review Letters 114, 140403 (2015), arXiv:1411.4648 [quant-ph].
[11] R. Chaves, C. Majenz, and D. Gross, Nature Communications 6, 5766 (2015), arXiv:1407.3800 [quant-ph].
[12] M. S. Leifer and D. Poulin, Annals of Physics 323, 1899 (2008), arXiv:0708.1337 [quant-ph].
[13] J. F. Fitzsimons, J. A. Jones, and V. Vedral, Scientific reports 5 (2015).
[14] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, Nature Physics 11, 414 (2015).
[15] E. G. Cavalcanti and R. Lal, Journal of Physics A: Mathematical and Theoretical 47, 424018 (2014), arXiv:1311.6852 [quant-ph].
[16] A. Jamiołkowski, Rep. Math. Phys. 3, 275 (1972).
[17] M. D. Choi, Linear Algebra Appl. 10, 285 (1975).
[18] M. S. Leifer and R. W. Spekkens, Physical Review A 88, 052130 (2013), arXiv:1107.5849 [quant-ph].
[19] B. Schumacher and M. D. Westmoreland, Quantum Information Processing 4, 13 (2005), arXiv:quant-ph/0406223.
[20] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Physical Review A 64, 052309 (2001), arXiv:quant-ph/0102043.
[21] T. Eggerling, D. Schlingemann, and R. F. Werner, EPL (Europhysics Letters) 57, 782 (2002).
[22] P. Hayden, R. Jozsa, D. Petz, and A. Winter, Communications in Mathematical Physics 246, 359 (2004), arXiv:quant-ph/0304007.
[23] B. Coecke and R. Lal, Foundations of Physics 43, 458 (2013).
[24] C. M. Caves, C. A. Fuchs, and R. Schack, Phys. Rev. A 65, 022305 (2002).
[25] C. A. Fuchs, Arxiv preprint quant-ph/0205039 (2002).
[26] R. W. Spekkens, Phys. Rev. A 75, 032110 (2007), eprint arXiv:quant-ph/0401052.
[27] M. S. Leifer, Physical Review A 74, 042310 (2006), arXiv:quant-ph/0606022.
[28] R. W. Spekkens, in *Quantum Theory: Informational Foundations and Foils* (Springer, 2016) pp. 83–135.
[29] D. Horsman, C. Heunen, M. F. Pusey, J. Barrett, and R. W. Spekkens, arXiv:1607.03637 (2016).
[30] Y. Aharonov and L. Vaidman, Time in Quantum Mechanics, 399 (2007).
[31] Y. Aharonov, S. Popescu, J. Tollaksen, and L. Vaidman, Phys. Rev. A 79, 052110 (2009).
[32] Y. Aharonov, S. Popescu, and J. Tollaksen, Quantum Theory: A Two-Time Success Story: Yakir Aharonov Festschrift, 21 (2013).
[33] R. Silva, Y. Guryanova, N. Brunner, N. Linden, A. J. Short, and S. Popescu, Physical Review A 89, 012121 (2014).
[34] G. Chiribella, G. M. D’Ariano, and P. Perinotti, *Causation, Prediction, and Search* (Cambridge University Press, Cambridge, 2001).
[35] R. Chaves, C. Majenz, and D. Gross, Nature Communications 6, 5766 (2015), arXiv:1407.3800 [quant-ph].
[36] M. S. Leifer and D. Poulin, Annals of Physics 323, 1899 (2008), arXiv:0708.1337 [quant-ph].
[37] J. F. Fitzsimons, J. A. Jones, and V. Vedral, Scientific reports 5 (2015).
[38] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, Nature Physics 11, 414 (2015).
[39] E. G. Cavalcanti and R. Lal, Journal of Physics A: Mathematical and Theoretical 47, 424018 (2014), arXiv:1311.6852 [quant-ph].
[40] A. Jamiołkowski, Rep. Math. Phys. 3, 275 (1972).
[41] M. D. Choi, Linear Algebra Appl. 10, 285 (1975).
[42] M. S. Leifer and R. W. Spekkens, Physical Review A 88, 052130 (2013), arXiv:1107.5849 [quant-ph].
[43] B. Schumacher and M. D. Westmoreland, Quantum Information Processing 4, 13 (2005), arXiv:quant-ph/0406223.
[44] D. Beckman, D. Gottesman, M. A. Nielsen, and J. Preskill, Physical Review A 64, 052309 (2001), arXiv:quant-ph/0102043.
[45] T. Eggerling, D. Schlingemann, and R. F. Werner, EPL (Europhysics Letters) 57, 782 (2002).
[46] P. Hayden, R. Jozsa, D. Petz, and A. Winter, Communications in Mathematical Physics 246, 359 (2004), arXiv:quant-ph/0304007.
[47] B. Coecke and R. Lal, Foundations of Physics 43, 458 (2013).
Appendix A: Proof of Theorems 2 and 3

We here provide the proof of Thms. 2 and 3. This amounts to proving that for a channel \( \rho_{BC|A} \), the following four conditions are equivalent:

1. \( \rho_{BC|A} \) admits of a unitary dilation where \( A \) is a complete common cause of \( B \) and \( C \).

2. \( \rho_{BC|A} = \rho_{B|A} \rho_{C|A} \).

3. \( I(B : C|A) = 0 \) where \( I(B : C|A) \) is the quantum conditional mutual information evaluated on the (positive, trace-one) operator \( \hat{\rho}_{BC|A} \).

4. The Hilbert space for the \( A \) system can be decomposed as \( \mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i} \otimes \mathcal{H}_{A_i^R} \) and \( \rho_{BC|A} = \sum_i \left( \rho_{B|A_i} \otimes \rho_{C|A_i^R} \right) \), where for each \( i \), \( \rho_{B|A_i} \) represents a completely positive map \( \mathcal{B}(\mathcal{H}_{A_i}) \to \mathcal{B}(\mathcal{H}_B) \), and \( \rho_{C|A_i^R} \) a completely positive map \( \mathcal{B}(\mathcal{H}_{A_i^R}) \to \mathcal{B}(\mathcal{H}_C) \).

We will show various implications that collectively give Thm. 2.

Proof that (3) \( \iff \) (4).

This follows easily from the results of Ref. 22, where a characterization is given of tripartite quantum states over systems \( A, B, C \) that satisfy \( I(B : C|A) = 0 \).

**Lemma 1** (22 Thm. 6). For any tripartite quantum state \( \rho_{ABC} \), the quantum conditional mutual information \( I(B : C|A) = 0 \) if and only if the Hilbert space of the \( A \) system decomposes as \( \mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i} \otimes \mathcal{H}_{A_i^R} \), such that

\[
\rho_{ABC} = \sum_i p_i \left( \rho_{BA_i} \otimes \rho_{CA_i^R} \right), \quad p_i \geq 0, \quad \sum_i p_i = 1, \tag{A1}
\]

where for each \( i \), \( \rho_{BA_i} \) is a quantum state on \( \mathcal{H}_B \otimes \mathcal{H}_{A_i} \) and \( \rho_{CA_i^R} \) is a quantum state on \( \mathcal{H}_C \otimes \mathcal{H}_{A_i^R} \).
Thm. 2 concerns the channel operator $\rho_{BC|A}$, which satisfies $\text{Tr}_{BC}(\rho_{BC|A}) = I_A$. Applying Lem. 1 to the operator $\hat{\rho}_{BC|A} = (1/d_A)\rho_{BC|A}$ yields the decomposition

$$\hat{\rho}_{BC|A} = \sum_i p_i \left( \hat{\rho}_{B|A_i} \otimes \hat{\rho}_{C|A_i} \right).$$

Using $\text{Tr}_{BC}(\hat{\rho}_{BC|A}) = (1/d_A)I_A$, it follows that for each $i$, the components satisfy $\text{Tr}_B(\hat{\rho}_{B|A_i}) = (1/d_{A_i})I_{(A_i^*)}$, and $\text{Tr}_C(\hat{\rho}_{C|A_i}) = (1/d_{A_i})I_{(A_i^*)}$, with $p_i = (d_{A_i}d_{A_i^*})/d_A$. The result follows.

**Proof that** (1) $\rightarrow$ (4).

Let $\rho_{BFC|A\lambda B \lambda C}^U$ be the Choi-Jamiołkowski operator for the unitary $U$, defined according to the conventions set out in the main text. Let missing indices indicate that a partial trace is taken, as also in the main text. Note that in general $\rho_{BFC|A}^U \neq \rho_{BC|A}$, since the latter is obtained via a particular choice of input states for $\lambda_B$ and $\lambda_C$. The proof proceeds by proving relations between quantum conditional mutual informations evaluated on the renormalized operator $\rho_{BFC|A\lambda B \lambda C}^U = (1/d_{\lambda_B}d_{A}d_{\lambda_C})\rho_{BFC|A\lambda B \lambda C}$, and its partial traces.

First,

$$I(B : FC|\lambda_B A \lambda_C) = 0. \tag{A2}$$

This follows by expanding in terms of von Neumann entropies:

$$I(B : FC|\lambda_B A \lambda_C) = S(\rho_{B|\lambda_B A \lambda C}^U) + S(\rho_{FC|\lambda_B A \lambda C}^U) - S(\rho_{BFC|\lambda_B A \lambda C}^U) - S(\rho_{\lambda_B A \lambda C}^U). \tag{A3}$$

The third term is zero, since the unitarity of $U$ implies that $\rho_{BFC|\lambda_B A \lambda C}^U$ is a pure state. The final term is $\log(d_{\lambda_B}d_{A}d_{\lambda_C})$, since $\rho_{B|\lambda_B A \lambda C}^U = (1/d_{\lambda_B}d_{A}d_{\lambda_C})I_{(\lambda_B A \lambda C)^*}$. Noting also that $\text{Tr}_{\lambda_B A \lambda C}(\rho_{BFC|\lambda_B A \lambda C}^U) = (1/d_{\lambda_B}d_{A}d_{\lambda_C})I_{(\lambda_B A \lambda C)^*}$, and using the fact that the von Neumann entropy of the partial trace of a pure state is equal to the von Neumann entropy of the complementary partial trace, yields that the first two terms equal $\log(d_{F}d_{C})$ and $\log(d_B)$ respectively, hence their sum is equal to $\log(d_{\lambda_B}d_A d_{\lambda_C})$, and Eq. (A2) follows.

Second,

$$I(\lambda_B : \lambda_C|A) = 0. \tag{A4}$$

This follows immediately from $\rho_{B|\lambda_B A \lambda C}^U = (1/d_{\lambda_B}d_A d_{\lambda_C})I_{(\lambda_B A \lambda C)^*}$.

Third,

$$I(B : \lambda_C|\lambda_B A) = 0. \tag{A5}$$

To see this, write

$$I(B : \lambda_C|\lambda_B A) = S(\rho_{B|\lambda_B A}^U) + S(\rho_{\lambda_B A \lambda C}^U) - S(\rho_{B|\lambda_B A \lambda C}^U) - S(\rho_{\lambda_B A}^U). \tag{A6}$$

The second and fourth terms are entropies of maximally mixed states on their respective systems, hence sum to $\log d_{\lambda_C}$. For the first and third terms, it follows from the assumption that there is no causal influence from $\lambda_C$ to $B$ in $U$ that $\rho_{B|\lambda_B A \lambda C}^U = \rho_{B|\lambda_B A} \otimes (1/d_{\lambda_C})I_{(\lambda_C)^*}$. Hence the third term is equal to $S(\rho_{\lambda_B A}^U) + \log(d_{\lambda_C})$, which gives Eq. (A5).

Fourth,

$$I(C : \lambda_B|A \lambda_C) = 0. \tag{A7}$$

This follows from a similar argument as Eq. (A5), using the assumption that there is no influence from $\lambda_B$ to $C$ in $U$.

The aim is now to use Eqs. (A2)–(A7) to show that $\hat{\rho}_{BC|A}$ satisfies $I(B : C|A) = 0$. This follows using a result from Ref. 12, which states that quantum conditional mutual informations on partial traces of a multipartite quantum state satisfy the semi-graphoid axioms familiar from the classical formalism of causal networks 8. The semi-graphoid axioms are as follows:

$$[I(X : Y|Z) = 0] \Rightarrow [I(Y : X|Z) = 0] \tag{A8}$$

$$[I(X : Y W|Z) = 0] \Rightarrow [I(X : Y|Z) = 0] \tag{A9}$$

$$[I(X : Y | W|Z) = 0] \Rightarrow [I(X : Y |ZW) = 0] \tag{A10}$$

$$[I(X : Y|Z) = 0] \land [I(X : W|Y Z) = 0] \Rightarrow [I(X : Y W|Z) = 0] \tag{A11}$$
Applying Eqs. \(A3\) to Eqs. \(A2\) gives

\[
[I(B : FC)_{|\lambda_B A\lambda_C} = 0] \Rightarrow [I(B : C)_{|\lambda_B A\lambda_C} = 0] \quad (A12)
\]

\[
[I(C : \lambda_B A\lambda_C) = 0] \land [I(B : C)_{|\lambda_B A\lambda_C} = 0] \Rightarrow [I(C : B)_{|\lambda_B A\lambda_C} = 0] \quad (A13)
\]

\[
[I(\lambda_B : C_{|A}) = 0] \land [I(\lambda_C : B_{|\lambda_B A}) = 0] \Rightarrow [I(\lambda_C : A_{|\lambda_B B}) = 0] \quad (A14)
\]

\[
[I(B_{\lambda_B} : \lambda_C_{|A}) = 0] \land [I(B_{\lambda_B} : C_{|\lambda_B A}) = 0] \Rightarrow [I(B_{\lambda_B} : C_{|A}) = 0] \quad (A15)
\]

Hence condition (1) of the theorem implies that \(I(B_{\lambda_B} : C_{\lambda_C})_{|A} = 0\), where this quantity is calculated on the trace-one Choi-Jamiołkowski operator representing the dilation unitary \(U\). Using Lem. \(I\) gives

\[
\hat{\rho}^U_{BC|\lambda_B A\lambda_C} = \sum_i p_i \left( \hat{\rho}^U_{B|\lambda_B A_i^L} \otimes \hat{\rho}^U_{C|A_i^R\lambda_C} \right), \quad (A16)
\]

for some appropriate decomposition of \((\mathcal{H}_A)^r\) and probability distribution \(\{p_i\}_i\). The form of the decomposition, and the fact that \(\text{Tr}_{BC}(\rho^U_{BC|\lambda_B A\lambda_C}) = (1/d_{\lambda_B} d_A d_{\lambda_C}) I_{(\lambda_B A\lambda_C)^r}^r\), gives

\[
\rho^U_{BC|\lambda_B A\lambda_C} = \sum_i \left( \rho^U_{B|\lambda_B A_i^L} \otimes \rho^U_{C|A_i^R\lambda_C} \right), \quad (A17)
\]

where for each \(i\), the components satisfy \(\text{Tr}_B(\rho^U_{B|\lambda_B A_i^L}) = I_{(A_i^L)^r}\), and \(\text{Tr}_C(\rho^U_{C|A_i^R \lambda_C}) = I_{(A_i^R)^r}\). The operator \(\rho_{BC|A}\) is obtained by acting with this channel on the input states \(|0\rangle_{\lambda_B}\) for \(\lambda_B\) and \(|0\rangle_{\lambda_C}\) for \(\lambda_C\). This gives

\[
\rho_{BC|A} = \sum_i \left( \rho_{B|A_i^L} \otimes \rho_{C|A_i^R} \right),
\]

where \(\text{Tr}_B(\rho_{B|A_i^L}) = I_{(A_i^L)^r}\) and \(\text{Tr}_C(\rho_{C|A_i^R}) = I_{(A_i^R)^r}\), as required.

**Proof that \((4) \Rightarrow (1)\).**

Let \(\mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i}\), with \(\mathcal{H}_{A_i} = \mathcal{H}_{A_i^L} \otimes \mathcal{H}_{A_i^R}\), and \(\rho_{BC|A} = \sum_i \left( \rho_{B|A_i^L} \otimes \rho_{C|A_i^R} \right)\). Each term \(\rho_{B|A_i^L}\) corresponds to a valid quantum channel, i.e., a CPTP map \(\mathcal{B}(\mathcal{H}_{A_i^L}) \to \mathcal{B}(\mathcal{H}_B)\). Similarly, each term \(\rho_{C|A_i^R}\) corresponds to a CPTP map \(\mathcal{B}(\mathcal{H}_{A_i^R}) \to \mathcal{B}(\mathcal{H}_C)\).

The channel \(\rho_{B|A_i^L}\) can be dilated to a unitary transformation \(V_i\), with ancilla input \(\lambda_B\) in a fixed state \(|0\rangle_{\lambda_B}\), such that \(V_i\) acts on the Hilbert space \(\mathcal{H}_{\lambda_B} \otimes \mathcal{H}_{A_i^L}\). Similarly, \(\rho_{C|A_i^R}\) can be dilated to a unitary transformation \(W_i\), with ancilla \(\lambda_C\) in a fixed state \(|0\rangle_{\lambda_C}\), acting on \(\mathcal{H}_{A_i^R} \otimes \mathcal{H}_{\lambda_C}\). By choosing the dimension of \(\lambda_B\) large enough, we can identify the system \(\lambda_B\) and the state \(|0\rangle_{\lambda_B}\) that are used for each value of \(i\), and similarly \(\lambda_C\).

Let \(V_i'\) be the operator that acts as \(V_i \otimes I_{A_i^L} \otimes I_{\lambda_C}\) on the subspace \(\mathcal{H}_{\lambda_B} \otimes \mathcal{H}_{A_i} \otimes \mathcal{H}_{\lambda_C}\), and as zero on the subspace \(\mathcal{H}_{\lambda_B} \otimes \mathcal{H}_{A_j} \otimes \mathcal{H}_{\lambda_C}\) for \(j \neq i\). Similarly, let \(W_i'\) be the operator that acts as \(I_{\lambda_B} \otimes I_{A_i^L} \otimes W_i\) on the subspace \(\mathcal{H}_{\lambda_B} \otimes \mathcal{H}_{A_i} \otimes \mathcal{H}_{\lambda_C}\), and as zero on the subspace \(\mathcal{H}_{\lambda_B} \otimes \mathcal{H}_{A_j} \otimes \mathcal{H}_{\lambda_C}\) for \(j \neq i\). Let

\[
V = \sum_i V_i', \quad (A18)
\]

\[
W = \sum_i W_i', \quad (A19)
\]

where \(W\) and \(V\) are unitary and \([V, W] = 0\). Note that \(V\) acts as the identity on the \(\lambda_C\) subsystem, and \(W\) acts as the identity on the \(\lambda_B\) subsystem.

The channel represented by \(\rho_{BC|A}\) can be dilated to the unitary transformation \(V W\), with ancillas \(\lambda_B\) and \(\lambda_C\). From the form of \(V\) and \(W\), it follows immediately that there is no causal influence from \(\lambda_C\) to \(B\) in \(V W\). From \([V, W] = 0\) and the form of \(V\) and \(W\), it follows immediately that there is no influence from \(\lambda_B\) to \(C\) in \(V W = V W\).

**Proof that \((2) \Rightarrow (3)\).**
As remarked in the main text, taking the Hermitian conjugate of $\rho_{BC|A} = \rho_{B|A\rho_{C|A}}$ immediately gives $[\rho_{B|A}, \rho_{C|A}] = 0$. Hence

$$\rho_{BC|A} = \rho_{B|APC|A}$$  \hspace{1cm} (A20)
$$\rho_{BC|A} = \exp \left[ \log \rho_{B|A} + \log \rho_{C|A} \right]$$  \hspace{1cm} (A21)
$$\log \rho_{BC|A} = \log \rho_{B|A} + \log \rho_{C|A}$$  \hspace{1cm} (A22)
$$\log(\rho_{BC|A} + \log \rho_{A}) = \log(\rho_{B|A} + \log \rho_{C|A})$$  \hspace{1cm} (A23)
$$\log(d_A^{-1} \rho_{BC|A}) + \log(d_A^{-1} \rho_{A}) = \log(d_A^{-1} \rho_{B|A}) + \log(d_A^{-1} \rho_{C|A})$$  \hspace{1cm} (A24)

The second line follows as $[\rho_{B|A}, \rho_{C|A}] = 0$; the fourth because $\rho_{A} = I_A$, and therefore has the zero matrix as its logarithm; and the final line by adding $2 \log d_A^{-1}$ to both sides. It is proved in Ref. [66] that for any trace-one density operator $\rho_{XYZ}$ then $\log \rho_{XYZ} + \log \rho_{Z} = \log \rho_{XZ} + \log \rho_{YZ}$ is equivalent to the condition $I(X:Y|Z) = 0$.

**Proof that (4) → (2).**

Condition (4) is that $\mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i^t} \otimes \mathcal{H}_{A_i^R}$, with $\rho_{BC|A} = \sum_i \left( \rho_{B_i|A_i^t} \otimes \rho_{C_i|A_i^R} \right)$. It follows that

$$\rho_{B|A} = \sum_i \left( \rho_{B_i|A_i^t} \otimes I_{(A_i^R)^t} \right),$$  \hspace{1cm} (A25)
$$\rho_{C|A} = \sum_i \left( I_{(A_i^R)^t} \otimes \rho_{C_i|A_i^R} \right).$$  \hspace{1cm} (A26)

The product is

$$\rho_{B|A\rho_{C|A}} = \sum_{i,j} \left( \rho_{B_i|A_i^t} \otimes I_{(A_j^R)^t} \right) \left( I_{(A_j^R)^t} \otimes \rho_{C_j|A_j^R} \right).$$  \hspace{1cm} (A27)

The only non-zero terms correspond to $i = j$, hence

$$\rho_{B|A\rho_{C|A}} = \sum_i \rho_{B_i|A_i^t} \otimes \rho_{C_i|A_i^R} = \rho_{BC|A}.$$  \hspace{1cm} (A28)

**Appendix B: Proof of Theorem 4**

**Proof that (3) → (2).**

The proof proceeds via an inductive argument. Consider

$$\rho_{B_1...B_n|A} = \text{Tr}_{B_{n+1}...B_k} \left( \rho_{B_1...B_k|A} \right),$$
with $2 \leq n < k$, and assume that the claim holds for this channel, hence

$$\rho_{B_1...B_n|A} = \rho_{B_1|A} \cdots \rho_{B_n|A},$$  \hspace{1cm} (B1)

with $[\rho_{B_i|A}, \rho_{B_j|A}] = 0$ for $i, j = 1, \ldots, n$. It will be shown that the claim remains true if one fewer system is traced out. To see this, recall that condition (3) gives $I(B_{n+1} : B_{n+1}^{\perp}|A) = 0$, where $I(B_{n+1} : B_{n+1}^{\perp}|A)$ is evaluated on $\rho_{B_1...B_k|A}$. Using Thm. [2] gives

$$\rho_{B_1...B_k|A} = \rho_{B_{n+1}^{\perp}|A} \rho_{B_{n+1}|A},$$

with $[\rho_{B_{n+1}^{\perp}|A}, \rho_{B_{n+1}|A}] = 0$. Tracing out systems $B_{n+2} \cdots B_k$ results in

$$\rho_{B_1...B_{n+1}|A} = \rho_{B_1...B_n|A} \rho_{B_{n+1}|A},$$
with $[\rho_{B_1 \ldots B_n | A}, \rho_{B_{n+1} | A}] = 0$. Since $\rho_{B_1 \ldots B_n | A}$ satisfies Eq. (B1) it follows that

$$\rho_{B_1 \ldots B_{n+1} | A} = \rho_{B_1 | A} \cdots \rho_{B_{n+1} | A}.$$  

For any $i = 1, \ldots, n$, trace out all systems but $B_i, B_{n+1}$ and $A$ to see that $[\rho_{B_i | A}, \rho_{B_{n+1} | A}] = 0$.

Hence if $\rho_{B_1 \ldots B_n | A}$ satisfies the claim, so too does $\rho_{B_1 \ldots B_{n+1} | A}$. As $\rho_{B_1 B_2 | A} = \rho_{B_1 | A} \rho_{B_2 | A}$, with $[\rho_{B_1 | A}, \rho_{B_2 | A}] = 0$ follows from $I(B_1 : \overline{B}_1 | A) = 0$, and tracing out all but systems $B_1, B_2$, and $A$, the proof is complete.

**Proof that** (2) $\Rightarrow$ (3).

This is immediate from Thm. 2 by grouping outputs into $B_i$ and $\overline{B}_i$ for each $i$.

**Proof that** (3) $\Leftrightarrow$ (4).

The proof that (4) $\Rightarrow$ (3) is immediate from Thm. 2 by grouping outputs into $B_i$ and $\overline{B}_i$ for each $i$.

It remains to show that if $I(B_i : \overline{B}_i | A) = 0$ for all $i$, then there exists a decomposition

$$\mathcal{H}_A = \bigoplus_i \left( \bigotimes_{j=1}^k \mathcal{H}_{A_i^j} \right),$$

(B2)

with $\rho_{B_1 \ldots B_k | A} = \sum_i (\rho_{B_1 | A_i^1} \otimes \cdots \otimes \rho_{B_i | A_i^k})$.

Given $I(B_1 : \overline{B}_1 | A) = 0$, Thm. 2 implies that $\mathcal{H}_A$ decomposes as

$$\mathcal{H}_A = \bigoplus_i \mathcal{H}_{A_i^L} \otimes \mathcal{H}_{A_i^R},$$

with $\rho_{B_1 \ldots B_k | A} = \sum_i \rho_{B_1 | A_i^L} \otimes \rho_{B_2 \ldots B_k | A_i^R}$. By assumption, $I(B_2 : \overline{B}_2 | A) = I(B_2 : B_1, B_3, \ldots, B_k | A) = 0$. As the conditional mutual information never increases if systems are discarded, we have $0 = I(B_2 : B_1, B_3, \ldots, B_k | A) \geq I(B_2 : B_3, \ldots, B_k | A)$. Non-negativity of the conditional mutual information then yields $I(B_2 : B_3, \ldots, B_k | A) = 0$.

The above decomposition ensures

$$\hat{\rho}_{B_2 \ldots B_k | A} = \sum_i p_i \left( \frac{I_{A_i^L}}{d_{A_i^L}} \right) \otimes \hat{\rho}_{B_2 \ldots B_k | A_i^R},$$

with $p_i = d_{A_i^L} d_{A_i^R} / d_A$. As the terms in the sum on the RHS have support on orthogonal subspaces,

$$S(\hat{\rho}_{B_2 B_3 \ldots B_k | A}) = H(p) + \sum_i p_i \log d_{A_i^L} + \sum_i p_i S(\hat{\rho}_{B_2 \ldots B_k | A_i^R}),$$

$$S(\hat{\rho}_{B_2 | A}) = H(p) + \sum_i p_i \log d_{A_i^L} + \sum_i p_i S(\hat{\rho}_{B_2 | A_i^R}),$$

$$S(\hat{\rho}_{B_3 \ldots B_k | A}) = H(p) + \sum_i p_i \log d_{A_i^L} + \sum_i p_i S(\hat{\rho}_{B_3 \ldots B_k | A_i^R}),$$

$$S(\hat{\rho}_{| A}) = H(p) + \sum_i p_i \log d_{A_i^L} + \sum_i p_i S(\hat{\rho}_{| A_i^R}).$$

Substituting into

$$I(B_2 : B_3, \ldots, B_k | A) = S(\hat{\rho}_{B_2 | A}) + S(\hat{\rho}_{B_3 \ldots B_k | A}) - S(\hat{\rho}_{B_2 B_3 \ldots B_k | A}) - S(\hat{\rho}_{| A}),$$

the $H(p)$ terms and the $\sum_i p_i \log d_{A_i^L}$ terms cancel, and one is left with

$$I(B_2 : B_3, \ldots, B_k | A) = \sum_i p_i I(B_2 : B_3, \ldots, B_k | A_i^R) = 0.$$

Non-negativity of both the conditional mutual information and the $p_i$ implies

$$I(B_2 : B_3 \ldots B_k | A_i^R) = 0.$$
Hence each $\mathcal{H}_{AR}$ in the above decomposition further decomposes into a direct sum of tensor products. Iterating this procedure results in the required decomposition.

Proof that (1) $\leftrightarrow$ (4).

The proof that (4) $\rightarrow$ (1) is a straightforward extension of the proof in Appendix A that condition (4) $\rightarrow$ (1) in Thm. 2.

To show that (1) $\rightarrow$ (4), first use Def. 4 to show that if, for each $i$, there is no causal influence from $\lambda_i$ to $\overline{B}_i$, it follows that, for each $i$, there is no causal influence from $\overline{\lambda}_i$ to $B_i$. Partitioning the output systems into $B_i$ and $\overline{B}_i$, and the ancilla systems $\lambda_1, \ldots, \lambda_k$ into $\lambda_i$ and $\overline{\lambda}_i$, it follows from Thm. 2 that $I(B_i : \overline{B}_i | A) = 0$. Hence condition (1) $\rightarrow$ (3), and since condition (3) $\rightarrow$ (4), the result follows.