Quantum causal modelling

To cite this article: Fabio Costa and Sally Shrapnel 2016 New J. Phys. 18 063032

View the article online for updates and enhancements.

Related content
- A graph-separation theorem for quantum causal models
  Jacques Pienaar and aslav Brukner

- The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning
  Christopher J Wood and Robert W Spekkens

- Quantum superpositions of 'common-cause' and 'direct-cause' causal structures
  Adrien Feix and aslav Brukner

Recent citations
- Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning
  Eric G. Cavalcanti

- A quantum causal discovery algorithm
  Christina Giarmatzi and Fabio Costa

- Operational Markov Condition for Quantum Processes
  Felix A. Pollock et al
Quantum causal modelling

Fabio Costa $^{1,4}$ and Sally Shrapnel $^{2,3}$

$^1$ Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia
$^2$ School of Historical and Philosophical Inquiry, The University of Queensland, St Lucia, QLD 4072, Australia
$^3$ School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia
$^4$ Author to whom any correspondence should be addressed.
E-mail: f.costa@uq.edu.au and s.shrapnel@uq.edu.au

Keywords: quantum information, causal modelling, foundations of quantum mechanics, quantum causality, quantum causal networks, Markovianity

Abstract

Causal modelling provides a powerful set of tools for identifying causal structure from observed correlations. It is well known that such techniques fail for quantum systems, unless one introduces 'spooky' hidden mechanisms. Whether one can produce a genuinely quantum framework in order to discover causal structure remains an open question. Here we introduce a new framework for quantum causal modelling that allows for the discovery of causal structure. We define quantum analogues for core features of classical causal modelling techniques, including the causal Markov condition and faithfulness. Based on the process matrix formalism, this framework naturally extends to generalised structures with indefinite causal order.

1. Introduction

1.1. Why cause

Causality often appears as an axiom in the formulation of physical theories, typically as the assumption that effects cannot precede their causes. However, what is often left unclear is which general principles should be used to define causal relations and relata.

The recent technical developments of causal modelling provide a consistent mathematical formalism that relates causal influence to the possibility of signalling [1, 2]. This approach provides an explanation for why causation is not 'merely correlation' and has shed light on contentious issues across a variety of scientific fields [3–5].

Causal models rest on essentially two concepts: the existence of autonomous mechanisms, responsible for producing the observed correlations, and the possibility of interventions, which modify part of the mechanism according to some factor external to the model. It is the latter notion that differentiates causal relations from simple correlations. Together, mechanisms and interventions define what is known as the causal Markov condition, a constraint imposed on the structure of a causal model.

Causal models have proved tremendously useful for understanding causal relations across a broad range of disciplines. They provide a powerful mathematical toolbox for efficiently extracting causal information from observed data, figures 1 and 2. They enable scientists to predict, manipulate and explain. The application of such models to physics represents a promising direction for a deeper and unified understanding of causality [6].

Unfortunately, one cannot straightforwardly apply these methods to quantum systems. Causal models presume the existence of objective properties that can be observed and manipulated locally. Such assumptions are incompatible with quantum mechanics, as most prominently demonstrated by Bell’s theorem [7]. Of course, one can apply the classical methodology, but only at the expense of introducing hidden, fine tuned [8] mechanisms, such as action at a distance [9] or retrocausality [10]. Although seemingly conservative, such approaches violate the very pillars of causal reasoning: they posit the existence of mechanisms that cannot be discovered and variables that cannot be controlled.

An alternative approach is to reformulate causal models from the ground up, in a way that makes direct use of the quantum formalism. Indeed, quantum theory can be seen as a theory of interventions [11] and, from the
perspective of quantum information, causal relations are identified with \textit{signalling}, in direct agreement with the intuitions underlying classical models. Additionally, quantum circuits \cite{12} are often interpreted as being representative of causal structure. This suggests that quantum causal models should be rather natural within quantum theory. However, despite much work in this direction \cite{13–22}, a general framework for quantum causal models has not been found. As identified in \cite{21}, a key missing step is the formulation of a quantum version of the causal Markov condition. In classical models, this condition enables one to use observational and interventionist inference to deduce causal structure from data. Arguably, a formalism for quantum causal modelling would require a similar condition.

1.2. This work

Here a complete framework for quantum causal modelling is presented. The framework rests on quantum definitions of mechanisms and spatio-temporally localised interventions. A quantum version of the causal Markov condition is defined, allowing for the possibility of \textit{causal discovery}, namely, the identification of causal structure from empirical data. Classical causal models are recovered as a limiting case of the formalism. Furthermore, the formalism allows for natural quantum extensions of classical concepts, such as faithfulness and the distinction between direct and indirect causes.

In the present framework, causal relations are identified with the possibility of signalling, that is, with the possibility of modifying the statistics of an event through intervention on a different event. The role of interventions was not clearly identified in previous works, which rather attempted to generalise various aspects of ‘causally neutral’ Bayesian networks \cite{13, 14, 16–21}. Without a definition of intervention, it is not clear in what sense such generalisations should be thought of as \textit{causal}.

Although some notion of interventions is present in \cite{15}, in that formalism not all correlations have a causal explanation, as a new form of ‘contemporaneous’ correlation is postulated to account for entanglement. By contrast, in the causal models introduced here all correlations arise from direct, indirect, or common causes.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{Correlations. A typical task in science is to provide a model that can explain the correlations between some observed variables.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure2.png}
\caption{Causal model. A causal model produces correlations in terms of a mechanism. The rain fills the stream that moves the wheel, activating the mechanism that grinds the grains to produce flour.}
\end{figure}
Classical causal models use probability distributions and it is natural to ask whether quantum causal models can be defined by simply replacing such distributions with density matrices. However, a classical joint probability distribution can describe common-cause as well as direct-cause relations, while a joint density matrix cannot describe direct-cause relations between quantum events. Attempts have been made to generalise density matrices in order to solve this problem. In such attempts each quantum event is associated with a single Hilbert space, and it has been shown that such approaches face serious problems. In contrast, the present work identifies a quantum event with a local operation and thus is associated with a pair of Hilbert spaces, describing respectively the input and the output of the operation. This follows the quantum networks and process matrix formalisms. Ultimately, ‘splitting’ the Hilbert space of an event is what allows for a unified representation of arbitrary causal structures. This fact was also recognised in, which presented a special case of the general framework introduced here. Other recent works have focussed on the correlations between classical variables produced in experiments involving quantum systems. These works adopt a device independent perspective, where the observed data contain no information about the functioning of or the physical theory describing the instruments used. Such frameworks cannot distinguish direct from indirect causes, or more generally formulate a quantum version of the causal Markov condition. Such limitations are overcome here by adopting a device dependent quantum formalism, where the description of events contains all relevant information regarding the physical devices that produced them.

The work is organised as follows. In section 2 the framework of classical causal models is briefly reviewed. The quantum formalism is introduced in section 3, where Markov quantum causal models (MQCMs) are defined and it is shown that non-Markovian, causally ordered models can always be reduced to Markovian ones. In section 4, faithful causal models are defined and the basic tools for quantum causal discovery introduced. It is shown in section 5 that classical causal models are recovered as a limiting case of quantum ones. Possible extensions of the formalism to indefinite and cyclic causal structures are briefly discussed in section 6. An introduction to the framework presented here is also considered from a broader philosophical perspective in.

### 2. Classical causal models

The brief review of classical causal models presented here is based on the monograph by Pearl.

#### 2.1. Mechanisms

In classical physics, an event is identified with the value assumed by some property of a system at a given moment in time (denoted \( x, y, \ldots \)). Potential events are described by random variables \((X, Y, \ldots)\), with values distributed according to some probability \( \Pr(X = x, Y = y, \ldots) \). The short-hand \( P(x) \equiv \Pr(X = x) \) is used when there is no risk of confusion.

The building blocks of a causal model are mechanisms that connect events to each other according to classical laws of dynamics, see figure 3. Deterministic laws imply a functional dependence between variables:

\[
Y = f(X).
\]  

Unknown noise affecting the mechanism can be modelled as an additional variable \( \Lambda \) with an associated probability distribution \( P(\lambda) \). The functional relation is then \( Y = f(X, \Lambda) \), which defines a stochastic mechanism that transforms \( X \) to \( Y \) according to the conditional probability.

![Figure 3. Classical mechanism. A classical mechanism is defined by a function \( Y = f(X) \), relating random variables. The variables are depicted as circles and the mechanisms depicted via connecting arrows. External noise is represented by an additional variable \( \Lambda \) with an associated probability distribution. Noise variables are typically not drawn explicitly and the arrows are interpreted as generic stochastic processes, defined by a conditional probability \( P(Y|X) \).](image-url)
where $\delta_{ab}$ is the Kronecker delta.

2.2. Interventions

What distinguishes a causal mechanism from a generic conditional probability is the possibility of changing it via intervention, figure 4. Causal relations are indeed defined by the possibility of intervening on a variable in the model and forcing it to assume some chosen value, thereby overriding the natural process that generates it. If $X$ is a cause of $Y$, then setting $X$ to a particular value will not change the conditional probability $P(y|x)$. On the other hand, setting $Y$ to a specific value requires breaking the process that generates $Y$, thus removing the correlations between $X$ and $Y$.

More generally, an intervention on a variable $Y$ can be any modification of the mechanism that produces $Y$, without necessarily breaking all pre-existing causal influences. This can be modelled by some control variable $I_Y$ that parametrises the mechanism producing $Y$. Intervention variables can be treated as random variables themselves, thus causal mechanisms are defined as conditional probabilities

$$P(y|x) = \sum_{\lambda} P(y|x, \lambda) P(\lambda),$$

$$P(y|x, \lambda) = \delta_{y,f(x,\lambda)},$$

(2)

where $\delta_{ab}$ is the Kronecker delta.

2.3. Causal models

A causal model captures the causal relations between a set of variables in terms of mechanisms and interventions, figure 5. The qualitative structure of cause-effect relations in a model defines a causal structure, which is represented as a directed acyclic graph (DAG).

**Definition 1 (DAG).** A directed graph is a pair $G = \langle V, E \rangle$, where $V = \{ V_0, \ldots, V_n \}$ is the set of vertices and $E \subseteq V \times V$ is a set of ordered pairs of vertices, representing directed edges. A directed path is a sequence of directed edges $\{e_k = (V_i^k, V_j^k) \in E \}_{k=1}^m$, with $V_k^0 = V_{k+1}^1$, for $k = 1, \ldots, m - 1$. A directed cycle is a directed path with $V_m^0 = V_1^1$. A DAG is a directed graph that contains no directed cycles.

---

5 Note the small abuse of notation, i.e. $P(y|x)$ does not represent the marginal of $P(y, i|\lambda)$, as the notation would imply.
A functional causal model is defined by a set of functions that determine the observed variables given their direct causes and possibly some noise. The Markov assumption states that all noise variables are uncorrelated, leading to the following definition:

**Definition 2 (Classical causal model).** A classical causal model for a set of random variables \( \mathcal{X} = \{X_j\}_{j=1}^n \) is defined by

1. A DAG that has the random variables as vertices, \( G = (\mathcal{X}, \mathcal{E}) \);
2. A list of conditional probabilities for each variable given its parents, \( P(x_j|p_{A_j}) \), \( j = 1, \ldots, n \), where the set of parents \( P_{A_j} \) of a vertex \( X_j \) in a DAG is defined as the set of vertices \( X \) with an edge from \( X \) to \( X_j \):

\[
P_{A_j} := \{ X \in \mathcal{X} | (X, X_j) \in \mathcal{E} \}. \tag{6}
\]

The conditional probabilities \( P(x_j|p_{A_j}) \) represent autonomous causal mechanisms that can be modified through interventions. As mentioned above, the intervention variables are often left implicit, i.e. it is understood that \( P(x_j|p_{A_j}) \equiv \Pr(X_j = x_j|P_{A_j} = p_{A_j}, I_j = \text{idle}) \).

A causal model generates a probability distribution

\[
P(x_1, \ldots, x_n) = \prod_{j=1}^n P(x_j|p_{A_j}) \tag{7}
\]

for the observed random variables. Decomposition (7) is called the Markov condition and can be used to test the compatibility of given data with a causal structure, even in the absence of interventions. However, without additional assumptions, the Markov condition does not identify a unique DAG and the causal structure remains underdetermined\(^6\). Explicitly, the causal information of a model is encoded in the conditional probability

\[
P(x_1, \ldots, x_n|I_1, \ldots, I_n) = \prod_{j=1}^n P(x_j|p_{A_j}, I_j), \tag{8}
\]

which reduces to condition (7) when \( I_j = \text{idle} \) for all \( j \). Equation (8) can be called causal Markov condition, although often this is defined as condition (7), leaving intervention variables and causal interpretation implicit. Equation (8) can also be used to define causal models in the language of influence diagrams [28].

If a set of variables \( Y_1, \ldots, Y_r \) in a causal model is not observed, the marginal probability

\[
P(x_1, \ldots, x_n) = \sum_{y_1, \ldots, y_r} P(x_1, \ldots, x_n, y_1, \ldots, y_r) \]

does not necessarily satisfy the Markov condition, and the unobserved variables are called latent, see figure 6. Given a probability that does not fulfill the Markov condition, it is always possible to extend the set of variables to include latent variables that restore the condition. Implicit in such a move is the assumption that latent variables can be accessed at least in principle and their causal role verified through interventions. In the context of quantum correlations, such an extension leads to hidden

---

\(^6\) In fact, even under additional assumptions such as faithfulness and causal sufficiency, the causal structure will often remain underdetermined. In such situations, interventions are required to uncover the correct structure.
variable models. However, a causal interpretation of such models implies a deviation from quantum mechanics: intervening on hidden variables would allow signalling at a distance, in contradiction with quantum predictions. Models in which the hidden variables cannot be accessed, even in principle, are a logical possibility, but they do not have a causal interpretation in the interventionist sense that underpins classical causal modelling.

3. Quantum causal models

The process matrix formalism reviewed below was introduced in [26] as an extension of causally ordered quantum networks [25, 29, 30]. Only finite-dimensional systems are considered here, a generalisation to infinite dimensions was studied in [31].

3.1. Quantum events

In contrast to classical systems, quantum systems do not possess objective properties that can be assigned values prior to and independently of measurement. Thus, the relata of causal relations are not classical random variables—a genuinely quantum notion of event is needed. A quantum event can be identified with all available information about a system localised in space and time. Typical quantum events are 'the spin was found aligned with the z axis' and 'the spin was rotated by π around the z axis'. A combination of the type 'the spin was found aligned with the z axis and then reprepared aligned with the x axis' can be considered as an event too. In general, a quantum event is associated with an operation, which can be deterministic (as a fixed transformation) or non-deterministic (as one associated with the outcome of a measurement).

Formally, such a quantum operation is represented by a completely positive (CP) map $\mathcal{M} : A_I \rightarrow A_O$, where input and output spaces are the spaces of linear operators over input and output Hilbert spaces, $A_I \equiv \mathcal{L}(\mathcal{H}_I)$, $A_O \equiv \mathcal{L}(\mathcal{H}_O)$, respectively (here identified with the corresponding matrix spaces). Complete positivity means that, for arbitrary dimensions of an ancillary system $A'$, the map $\mathcal{I}_{A'} \otimes \mathcal{M}$ transforms positive operators into positive operators, where $\mathcal{I}_{A'}$ is the identity map on $A'$. Input and output spaces can have different dimensions $d_{A_I}$, $d_{A_O}$, as ancillas can be added or discarded.

Using the Choi–Jamiołkowski isomorphism [32, 33], a CP map can be represented as a matrix$^7$

$$M^{A_IA_O} = \sum_{j,l} |j\rangle \langle j|^{A_I} \otimes \mathcal{M}(|j\rangle \langle l|)A_O^\dagger,$$

(9)

$$\mathcal{M}(\rho)^{A_O} = [\text{tr}_{A_I}(\rho^{A_I} \otimes 1^{A_O} \cdot M^{A_IA_O})]^T,$$

(10)

where $|j\rangle^{A_I}$ is an orthonormal basis in $\mathcal{H}_I$ and $^T$ denotes transposition in that basis.$^8$ According to Choi’s theorem, $\mathcal{M}$ is CP if and only if $M^{A_IA_O} \geq 0$. Henceforth CP maps will be identified with their Choi–Jamiołkowski representation unless otherwise stated.

An example of a quantum event is a projective measurement that yields as an outcome some pure state $|\psi\rangle$ and resets the systems in the same state. As a CP map, this is represented as

$$M^{A_IA_O}_\psi = |\psi\rangle^{A_I} \otimes (|\psi\rangle A_O^\dagger)^T,$$

(11)

$^7$ This definition aligns with the convention in [26]. Different choices, amounting to a transposition or partial transposition, are found elsewhere.

$^8$ The convention used here is that the superscript denotes the space in which a matrix is defined—hence $M^T$ is a short-hand for $M \in X$ and $J^Z$ for $J \subset X$, if $J$ is a set of matrices)—but not the matrix itself. $T^{A\overline{B}} = M^A \otimes M^\overline{B}$ means that $T$ is the tensor product of two equal matrices $M$. Identity matrices are sometimes omitted: $M^A \otimes I^B \equiv M^A$, etc.
where the notation \(|\psi\rangle := |\psi\rangle\langle\psi|\) is used \[^{[34]}\]. Another example is a transformation of the system occurring with unit probability, for example as defined by a unitary matrix \(U\). Introducing the notation \[^{[35,36]}\] \(\langle j|U|j\rangle\), \(\langle U\rangle := \sum_j |j\rangle U |j\rangle\), \(\langle U\rangle| = \langle U\rangle|U\rangle\),

\[
|\langle U\rangle| := \sum_j |j\rangle U |j\rangle, \quad \langle U\rangle := \langle U\rangle|U\rangle, \quad \langle U\rangle|U\rangle = \langle U\rangle|U\rangle,
\]

(12)

the local event corresponding to the unitary transformation \(U\) is given by

\[
M_{\alpha A_0}^{\alpha A_0} = \langle U\rangle| U\rangle = \langle U\rangle|U\rangle,
\]

(13)

where * is the complex conjugation in the chosen basis. A more general deterministic transformation is a CP and trace-preserving (CPTP) map, represented as

\[
M_{\alpha A_0}^{\alpha A_0} \geq 0, \quad tr_{A_0}M_{\alpha A_0}^{\alpha A_0} = 1_A.
\]

(14)

A CPTP map is also called a quantum channel.

The space of potential events is identified with the set of CP maps between an input \((A_1)\) and an output \((A_2)\) space, which is isomorphic to (the cone of positive matrices in) the space \(A_1 \otimes A_2\). Input and output spaces can be respectively identified with the past and the future space-like boundaries of the space–time region where the event takes place, figure 7, as in Oeckl’s ‘general boundary’ formalism \[^{[37]}\]. The space of potential events, which plays a role analogous to a classical random variable, is called a local laboratory.

While the name implies some anthropocentric element, the term ‘local laboratory’ is taken to represent an observer-independent spatio-temporal region where quantum events can take place. In direct analogy to the classical case \[^{[38]}\], one can argue whether it is possible to entirely remove the concept of observer, though we shall not enter into such discussions here.

### 3.2. Mechanisms

As in the classical case, a causal model is built on the notion of physical mechanisms that are responsible for mediating causal influences between events. A quantum mechanism maps the output space of a local laboratory, say \(A_0\), to the input space of another one, say \(B_0\). The analogue of a deterministic mechanism is a unitary\(^9\) map \(U\), which transforms states as \(\rho \mapsto U\rho U^\dagger\). External noise can be described by an interaction with an environment which is then traced out, leading to the more general definition of a mechanism as a CPTP map.

In the present approach, CPTP maps representing events in local laboratories are distinguished from CPTP maps representing connecting mechanisms. This distinction is reflected by a different representation of the two: a CPTP map \(T\) corresponding to a connecting mechanism is described by the transpose of representation (9):

\[
T^{A_0B_0} = \sum_{j,l} |j\rangle \langle l| A_0 \otimes T(|j\rangle \langle l|) A_0,
\]

(15)

\[
T(\rho)^{B_0} = tr_{A_0}[\rho^{A_0} \cdot T^{A_0B_0}],
\]

(16)

\[
T \geq 0, \quad tr_{B_0}T^{A_0B_0} = 1_A,
\]

(17)

where \(T^{X^Y}\) denotes partial transposition on subsystem \(X\).

As an easy example, a unitary transformation from the output space of \(A\) to the input space of \(B\) is represented by the projector \([U\rangle|_ {A_0B_0}\), the transpose of (13). A mechanism that connects a space of dimension 1 to the input of a local laboratory is given by a density matrix \(\rho^{B_0} \geq 0, \quad tr\rho = 1\), which corresponds to the situation where state \(\rho\) is sent to the input space \(A_0\).

\[^{9}\] In fact, the direct analogue of a deterministic mechanism is an isometry, for which \(V^\dagger V = 1\) but it is not necessarily true that \(VV^\dagger = 1\). Unitaries represent reversible deterministic mechanisms.
3.3. Interventions

Recall, the possibility of intervention is required to give causal models empirical meaning. Quantum interventions can be formalised using the notion of instrument \[39\], which represents the collection of all possible events that can be observed given a specific choice of probing the system. An intervention is thus defined as a choice of instrument. Given a local laboratory \(A_I \otimes A_O\), an instrument is formally defined as a set of CP maps that sum up to a CPTP map:

\[
\mathcal{J} = \{ M_{k}^{A_O A_I}, M_{k}^{A_I A_O} \geq 0, \ \text{tr}_{A_O} \sum_{k=1}^{m} M_{k}^{A_I A_O} = 1 \}
\]

(The trace-preserving condition for the sum guarantees that probabilities sum up to 1.)

A typical example of an instrument is the measurement of the incoming system according to a positive operator valued measure (POVM) \(\{ E_{x}^{\mu_{x}}, E_{x}^{n_{x}} \geq 0, \sum_{x=1}^{m} E_{x}^{n_{x}} = 1 \}\), followed by the preparation of a state \(\rho\). Using representation (9), this is given by

\[
\mathcal{J} = \{ E_{x}^{n_{x}} \otimes (\rho^{A_O})^{\mu_{x}} \}_{x=1}^{m}.
\]

By keeping the POVM fixed and choosing different states, one breaks the flow of information through the local laboratory and obtains the equivalent of classical ‘do’ interventions.

As discussed in section 5 below, classical ‘idle’ interventions correspond to quantum projective measurements in a fixed basis. These, however, are not interpreted as passive observations in the quantum formalism. Indeed, it is one of the essential features of quantum mechanics that measurements necessarily disturb the system.

3.4. Process matrices and causal relations

The general situation of interest is described by a set of local laboratories \(L = \{ L_j \otimes O_j \}_{j=1}^{n}\), interpreted as representing \(n\) disjoint space–time regions, each bounded by two space-like surfaces. In an individual run of an experiment, instruments \(\mathcal{J}_{j}^{l_{j}}, \ldots, \mathcal{J}_{n}^{l_{n}}\) are implemented in these local laboratories and the corresponding events recorded. The events are described by CP maps \(M_{l_{j}, \ldots, l_{n}}\).

By assuming the local validity of quantum mechanics, with no further assumption about causal relations between local laboratories, it is possible to prove that the probability for such events to occur is given by the generalised Born rule \[26\]

\[
P(M_{l_{j}, \ldots, l_{n}}|\mathcal{J}_{j}^{l_{j}}, \ldots, \mathcal{J}_{n}^{l_{n}}) = \text{tr}[M_{l_{j}, \ldots, l_{n}} \otimes W^{l_{1}, \ldots, l_{n}}],
\]

where \(W^{l_{1}, \ldots, l_{n}} \geq 0\) is called the process matrix, figure 8, and represents the information about the outside world available in the local laboratories\[10\]. In particular, the process matrix encodes all the information about the causal relations between laboratories.

As in the classical case, causal relations are defined by signalling. By definition, a laboratory \(L_k\) can signal to a laboratory \(L_k\) if there exists a set of instruments \(\{ \mathcal{J}_{j}^{l_{j}} \}_{j=1}^{n}\) such that

\[10\text{To be more precise, the conditional probability is defined as equation (20) when all the maps belong to the corresponding instruments, } M_{j} \in \mathcal{J}_{j}, \text{ for } j = 1, \ldots, n, \text{ and it is 0 otherwise.}
for some instrument $\mathcal{J}_h$, where the marginal probability is defined as

$$P(M^h_k|\mathcal{J}^{i_1}_{j_1}, \ldots, \mathcal{J}^{i_n}_{j_n}) = \sum_{M \in \mathcal{S}} \text{tr}[M^{i_1} \otimes \ldots \otimes M^{i_n} \cdot W^{j_1 \ldots j_n}].$$

(22)

Signalling between sets of laboratories is similarly defined. By definition, a local laboratory $A$ represents a cause for a distinct laboratory $B$ if $A$ can signal to $B$.

In classical causal models there is an important distinction between direct and indirect causes. Intuitively, a direct cause always influences its effect, no matter what else is changed in the model. This is formalised as follows:

**Definition 3 (Direct cause).** Given a set of local laboratories $\mathcal{L} = \{L_i = I_j \otimes O_i, \eta = 1\}$ and a process matrix $\mathcal{W}^{L_1 \ldots L_m}$, a laboratory $L_i$ represents a direct cause for a laboratory $L_k$ if, for any possible set of instruments $\{\mathcal{J}_j\}_{j=1}^n$, there exist instruments $\mathcal{J}_k$ and $\mathcal{J}_h$ such that

$$P(M^{L_k}_k|\mathcal{J}^{i_1}_{j_1}, \ldots, \mathcal{J}^{i_n}_{j_n}) = P(M^{L_k}_k|\mathcal{J}^{i_1}_{j_1}, \ldots, \mathcal{J}^{i_n}_{j_n}, \mathcal{J}^{h_1}_{j_1}, \ldots, \mathcal{J}^{h_n}_{j_n}).$$

(23)

### 3.5.1. Single laboratory

Consider first the case of a single laboratory $A$, with only non-trivial input space $A_i$. In this case, process matrices reduce to $W^A \equiv \rho^A \geq 0$, CP maps reduce to POVM elements $E^A \geq 0$, and the generalised Born rule (20) reduces to the ordinary Born rule: $P(E^A) = \text{tr}E^\rho$. This describes the situation where laboratory $A$ receives a quantum state $\rho$ from the outside environment and performs a measurement on it. The environment, responsible for preparing state $\rho$, describes some part of the world that is not under experimental control. Of course it is always possible to consider a different scenario in which the state preparation is controlled. This is formalised by introducing a second laboratory $B$, with only output space $B_0$, in which the state is prepared. Now $B$ has the possibility to choose from different instruments, i.e. prepare different states $\rho_i$. Recall that, as local events, such state preparations are represented as $(\rho_i^B)^T$. The process matrix now describes how the two laboratories are connected: $\mathcal{W}^A \rho^A \mathcal{W}^B \rho^B$ if $B$ is connected to $A$ via the identity channel, where notation (12) is used,

$$[I]^{R_0A_i} = \sum_{i,k} |i\rangle \langle k| \rho^i_{0k} \otimes |i\rangle \langle k|^{A_i}.$$  

(24)

The generalised Born rule (20) now reads

$$P(E^A)[\rho^{R_0B}^A] = \text{tr}[E^A \otimes (\rho^{R_0B}^A) \cdot [I]^{R_0A_i}] = \text{tr}E\rho.$$  

(25)

This reduces to the previous single-laboratory case if $B$ prepares $\rho_i \equiv \rho$. As the probability for any $E \neq 1$ depends non-trivially on $\rho$, $B$ represents a cause for $A$ in this example.

### 3.5.2. Common cause

The next example consists of two laboratories with non-trivial input and output spaces, $A = A_1 \otimes O_0$ and $B = B_1 \otimes O_0$. A process matrix of the form $W^{AB} = \rho^{AB} \otimes I^{A_0B_0}$ describes the situation where the two laboratories have no causal influence over each other (figure 9).

The generalised Born rule reduces to

$$P(M^A, N^B) = \text{tr}(E^A \otimes F^B \cdot \rho^{AB}),$$  

(26)

where $E^A := \text{tr}_{A_0} M^{A_0A_0}$ and $F^B := \text{tr}_{B_0} N^{B_0B_0}$. Thus, the above process matrix describes a bipartite state shared by the two laboratories.

As before, one can allow for control over the state preparation by introducing a third laboratory $C$ with only non-trivial output space $C_0$. As $C$ can prepare bipartite states, its output space must decompose as a tensor product $C_0 = C_0^1 \otimes C_0^2$. The subsystems $C_0^1$ and $C_0^2$ are isomorphic to $A_1$ and $B_1$ respectively. The process matrix that links $C_0^1$ to $A_1$ and $C_0^2$ to $B_1$, through identity channels, is...
given by \( W_{ABC}^{W} = [[|>]_{AB}^{G} \otimes [|[>]_{C}^{G} \otimes \mathcal{I}_{A_0}^{B_0} \otimes \mathcal{I}_{C_0}^{B_0} \otimes \mathcal{I}_{C}^{B_0}, \mathcal{I}_{C}^{B_0} \otimes \mathcal{I}_{C}^{B_0}} \). It is easy to see that, when the instrument \( \mathcal{J}^{C} \equiv \left( \rho^{G} \otimes \mathcal{I}_{C}^{B_0} \otimes \mathcal{I}_{C}^{B_0} \right)^{T} \) is chosen, the generalised Born rule reduces to expression (26).

As a particular case, state \( \rho \) can be entangled and \( A, B \) can perform measurements that violate Bell inequalities. The common cause for such correlations is simply associated with laboratory \( C \), i.e. with the possibility of preparing different states.

3.5.3. Direct and indirect cause

Still in the scenario with two laboratories \( A = A_1 \otimes A_0, B = B_1 \otimes B_0 \), consider now the process matrix \( W_{AB}^{W} = \rho^{A} \otimes [[|>]_{U}^{G} \otimes \mathcal{I}_{A_0}^{B_0} \otimes \mathcal{I}_{C_0}^{B_0} \otimes \mathcal{I}_{C}^{B_0} \right) = \mathcal{U}_{10}, \) where \( \mathcal{U} \) is a unitary matrix. In this case the generalised Born rule (20) reduces to

\[
\text{tr}(M_{A}^{A} \otimes N_{B}^{B} \cdot W_{AB}^{W}) = \text{tr}[F_{U_{10}} \mathcal{M}(\rho) U_{10}^{\dagger}],
\]

where \( F \) is defined by \( F_{10} = \text{tr}_{A_1} N_{B_1}^{B_0} \) and \( \mathcal{M} \) is related to \( M \) through isomorphism (10). Equation (27) describes a state \( \rho \) which is first measured in laboratory \( A \) and, after evolving via unitary \( \mathcal{U} \), is measured in laboratory \( B \). For arbitrary \( U \) and \( \rho, A \) can signal to \( B \) through an appropriate choice of instruments (for example preparing different pure states). Thus \( A \) represents a cause for \( B \).

As in the previous examples, the connecting mechanism can be included as an event in the model by introducing a third laboratory \( C = C_1 \otimes C_0, \) with the new process matrix \( W_{ABC}^{W} = \rho^{A} \otimes [[|>]_{U}^{G} \otimes [|[>]_{C}^{G} \otimes \mathcal{I}_{A_0}^{B_0} \otimes \mathcal{I}_{C_0}^{B_0} \otimes \mathcal{I}_{C}^{B_0} \right) \).

The unitary evolution \( \mathcal{U} \) is now represented by the single-element instrument \( [[|>]_{U}^{G} \otimes [|[>]_{C}^{G} \otimes \mathcal{I}_{A_0}^{B_0} \otimes \mathcal{I}_{C_0}^{B_0} \) (see equation (13)). It is now possible to consider interventions in the additional laboratory \( C \). For example, \( C \) can implement an instrument that breaks the flow of information from \( A \) to \( B \), e.g. the maximally noisy channel \( F / d_{c_{0}} \). In this case, no choice of instrument at \( A \) can affect the probability for any POVM element in \( B \). Thus, \( A \) is a direct cause of \( B \) given the process \( W_{AB}^{W} \) but it is an indirect cause given the process \( W_{ABC}^{W} \). As in the classical case, whether a causal relation is direct or indirect is relative to the set of variables included in the model.

3.6. Markov quantum causal models

As for classical causal models, it is useful to depict causal relations graphically. Local laboratories are represented as nodes and the causal mechanisms connecting them as arrows.

Multiple arrows departing a single node represent different physical systems. Thus the output space of a laboratory factorises as a tensor product, with a tensor factor associated to each outgoing arrow, see figure 10. Formally, a source space \( S_{S} \) is associated to the origin of each directed edge \( e \), such that each output space factorises as \( A_{O} = \otimes_{e \in OUT_{S}} s_{e} \), where \( OUT_{S} \) is the set of edges departing from \( A \).

As with classical models, arrows should represent a generic possibility of causal influence (and not just the undisturbed transfer of a system, as in the representation of quantum circuits). Thus, a graph with arrows that connect \( A \) to \( C \) and \( B \) to \( C \) represents quantum systems exiting from \( A \) and \( B \) and, possibly after interaction, entering \( C \). In other words, the graph represents a generic quantum channel \( T^{(A_0 \otimes B_0)_G} \) from the output of laboratories \( A \) and \( B \) to the input of laboratory \( C \), figure 11 (or, more generally, from a subsystem of \( A_{O} \otimes B_{O} \) to \( C_{O} \) if \( A \) and \( B \) have outgoing edges that do not end on \( C \).

When there are no outgoing edges from a laboratory, the output space can be understood as connecting to the trivial space (of dimension one). As the only CPTP map from a space \( D_{O} \) to the trivial space is \( T^{D_{O}} = \mathcal{I}_{D_{O}} \), a process matrix must be equal to the identity on the output space of all such laboratories.
A quantum causal model can be defined as a set of unitaries that connect (a subsystem of) the output spaces of the parents of a laboratory, plus some noisy environment, to the input space of the laboratory (possibly after discarding part of the system).

Recall for classical causal models, the causal Markov condition is guaranteed by the assumption that all environmental noise remains uncorrelated. Also known as the independent noise assumption, this feature of causal modelling is retained in the quantum case. Thus a quantum causal model is a list of quantum channels, the structure of which forms a DAG, figure 12.

Definition 4 (MQCM). Given a set of local laboratories \( \mathcal{L} = \{ L_j = I_j \otimes O_j \}_{j=1}^n \), a Markov quantum causal model is a pair \( (G, W) \), where

1. \( G = (\mathcal{L}, \mathcal{E}) \) is a DAG that has the local laboratories as vertices;
2. to each edge \( e \in \mathcal{E} \) is associated a space \( S_e \) such that \( O_j = \bigotimes_{e \in \text{OUT}_j} S_e, j = 1, \ldots, n \), where
   \[
   \text{OUT}_j := \{ e \in \mathcal{E} | e = (L_j, L_k) \}
   \] (28)
   is the set of edges departing from the vertex \( L_j \);
3. \( W \) is a process matrix of the form
   \[
   W_{L_k \rightarrow L_j} = \bigotimes_{j=1}^n T_j^{PS_L_j} \otimes I_{O_k},
   \] (29)
   where \( O_D := \bigotimes_{k \in D} O_k \) is the output space of the laboratories with no outgoing edges, \( D := \{ k | \text{OUT}_k = \emptyset \} \); \( PS_j := \bigotimes_{e \in \text{IN}_j} S_e \) is the parent space associated with laboratory \( L_j \), with
   \[
   \text{IN}_j := \{ e \in \mathcal{E} | e = (L_k, L_j) \}
   \] (30)
   the set of incoming edges to \( L_j \) and
   \[
   T_j^{PS_L_j} \geq 0, \quad \text{tr}_j T_j^{PS_L_j} = 1^{PS_L_j}, \quad j = 1, \ldots, n.
   \] (31)

The matrices \( T_j \) in equation (31) define the quantum channels that connect local laboratories to each other and replace the conditional probabilities that define causal mechanisms in a classical model. A process matrix with the structure (29) is said to factorise over the DAG \( G \). As this factorisation is a direct analog of the causal Markov condition (8), it is natural to name (29) the quantum causal Markov condition.
It is useful to relate Markov causal models to Markovian time evolution, a concept familiar to most physicists [40]. Consider a DAG $L_1 \rightarrow L_2 \rightarrow \ldots \rightarrow L_n$, with $d_{ij} = d_{Oi}$ (the generalisation to different dimensions is straightforward). The quantum causal Markov condition (29) reduces to
\[
(W)_{ij} = \text{tr}[E_{ij} T_{jk}(\rho_i)],
\]
where $\circ$ denotes composition of maps. Equation (34) defines a discrete-time Markovian evolution, while equation (33) shows that the $k$th laboratory receives the result of such an evolution applied to the initial state $\rho_i$, provided that no intermediate operation is performed. Channels of the form (34) are also called divisible [41].

### 3.7 Latent laboratories and non-Markovian models

A quantum process can involve unobserved events. Such events may correspond to naturally occurring evolution in the environment surrounding the observed events, or to non-selective measurements, with unknown outcomes, performed by adversary agents. A local laboratory in which the events are not observed will be called latent, in direct analogy to latent variables in classical models.

**Definition 5.** A Markovian causal explanation for a process matrix $W_{L_1 \ldots L_n}$ is an MQCM $(G, \tilde{W})$, where $G$ is a DAG with vertices containing $L_0, \ldots, L_m$, and possibly latent laboratories $\tilde{L} = \{\tilde{L}_0, \ldots, \tilde{L}_m\}$, such that
\[
W = \text{tr}_\tilde{L}[C_{L_1} \otimes \ldots \otimes C_{L_m} \cdot \tilde{W}^{\tilde{L}_1, \ldots, \tilde{L}_m}],
\]
for some CPTP maps $C_{L_1}, \ldots, C_{L_m}$, where $\text{tr}_\tilde{L}$ denotes the partial trace over all the laboratories in $\tilde{L}$.

In equation (35), $W$ is called a reduced process matrix [42, 43] and provides a full description of the physical situation for the observed laboratories, once the CPTP maps in the latent laboratories are fixed. The observed laboratories can now be linked by non-Markovian evolution, possibly with initial system-environment correlations: all the information regarding initial state and evolution of the environment is encoded in the latent laboratories. A formalism closely related to the one described here has in fact been developed for the study of non-Markovian dynamics [44, 45].

A process matrix is called causally ordered if it is possible to define a relation of partial order among the laboratories, and signalling from a laboratory $L_1$ to another $L_2$ is possible only if $L_1$ precedes $L_2$ according to the
assigned partial order. An important result from the theory of quantum networks [25, 29] is that every causally
ordered process matrix (formally equivalent to a quantum strategy or a quantum comb) can be realised by
combining a sequence of quantum channels with memory. This implies that every causally ordered process
matrix has a Markovian causal explanation. In other words, causally ordered, non-Markovian models can
always be reduced to Markovian ones through the introduction of latent laboratories.

An MQCM can be further extended by including connecting mechanisms as locally observed events. There
is a general procedure to re-write a connecting mechanism as a local event. Consider a local laboratory
\( A = A_I \otimes A_O \), with parent space \( \mathcal{P}_i \), and a process matrix that factorises as \( W = \mathcal{P}_{\mathcal{P}_i} \otimes \mathcal{W} \),
where \( \mathcal{P}_{\mathcal{P}_i} \) is the mechanism to be included as an event. The extended process matrix is
\[ \tilde{W} = [\mathcal{P}_i] \mathcal{P}_i \otimes [\mathcal{P}_i] \mathcal{P}_i \otimes \mathcal{W} \], with the additional local laboratory \( B = B_I \otimes B_O, B_I \cong \mathcal{P}_i \text{ and } B_O \cong A_I. \]
The original process matrix is recovered as the reduced process matrix
\[ W = \text{tr}_B([\mathcal{P}_i] \mathcal{P}_i \otimes \mathcal{W}). \]

With sufficient resources, it is in principle possible to take control over all the quantum channels that define
an MQCM. In this scenario, all non-trivial evolution takes place in local laboratories and can be in principle
controlled, while the process matrix only describes connections between laboratories. In the resulting causal
model, each edge \( e \) is associated both with a source space \( S_e \) and a target space \( T_e \), such that \( S_e \cong T_e \),
\( O_j = \otimes_{e \in \mathcal{E}_O} S_e \), and \( I_j = \otimes_{e \in \mathcal{E}_I} T_e \) for every \( j \). The process matrix is
\[ W_{I^n,\ldots,I^n} = \otimes_{e \in \mathcal{E}} [\mathcal{P}_i] \mathcal{P}_i \mathcal{R}_e. \]
This process represents the set of ‘wires’ obtained from a quantum circuit after removing all the gates.

Thus it is possible to give a causal interpretation to quantum circuits and related structures [46–48] by
allowing the possibility of replacing each gate with an arbitrary operation.

4. Causal discovery

The success of classical causal models is largely due to the existence of efficient tools for discovering causal
structures. Apart from practical utility, causal discovery has foundational significance, as it gives empirical
meaning to causal structures. This section shows that it is always possible to uniquely determine the causal
structure of a MQCM from experimental observations. The assumptions required are similar to those imposed
on classical causal models. The question of whether this can be done efficiently will not be addressed here.

4.1. Process-matrix tomography

A process matrix can always be reconstructed using informationally complete instruments. This is directly
analogous to informationally complete measurements [49–51]. For this purpose, note that a process matrix
\( W_{I^n,\ldots,I^n} \) can be formally regarded as a density matrix with different normalisation:
\[ \rho_{I^n,\ldots,I^n} = \frac{1}{d_O} W_{I^n,\ldots,I^n}, \tag{37} \]
where \( d_O = \prod_{j=1}^n d_{O_j} \) is the product of the dimensions of all output spaces. The matrix (37) is normalised as a
quantum state, \( \text{tr} \rho = 1 \), because process matrices are normalised as \( \text{tr} W = d_O \) [26, 42]. It is always possible to
find an informationally complete POVM with product elements \( E_{I^n,\ldots,I^n}^{I^n} \otimes \cdots \otimes E_{I^n,\ldots,I^n}^{I^n}, \sum_{x_n} E_{I^n,\ldots,I^n}^{I^n} = I^I \), such that \( \rho \) is a function of the probability distribution
\[ P(x_1,\ldots,x_n) = \text{tr}[E_{I^n,\ldots,I^n}^{I^n} \otimes \cdots \otimes E_{I^n,\ldots,I^n}^{I^n} \cdot \rho_{I^n,\ldots,I^n}]. \tag{38} \]
This means that \( W \) is also a function of the same distribution. Define now \( M_{x_j} = E_{x_j}/d_{O_j} \) for \( j = 1,\ldots,n \). As the
sets \( J_j = \{ M_{x_j} \} \) are properly normalised instruments, the statistics (38) can be obtained by implementing such
instruments on the original process matrix:
\[ P(x_1,\ldots,x_n) = \text{tr}[M_{x_1}^{I^n} \otimes \cdots \otimes M_{x_n}^{I^n} \cdot W_{I^n,\ldots,I^n}]. \tag{39} \]
Thus, it is possible to reconstruct \( W \) by measuring the instruments \( J_1,\ldots,J_n \). An example of this procedure is the
causal tomography considered in [22].

Alternatively, one can use the fact that a process matrix defines a CPTP map from all outputs, \( \otimes_{j=1}^n O_j \), to all
inputs, \( \otimes_{j=1}^n I_j \), and such a CPTP map can be reconstructed using conventional process tomography [52].
Informationally complete testers [25] also allow reconstructing a process matrix, although in general they require
a priori knowledge of the causal order between laboratories. In the context of causal discovery, such a knowledge
is not available and thus arbitrary testers cannot be used.
Given a process with a Markovian causal explanation, it is in principle possible to gain control over all connecting mechanisms and, through tomography, reconstruct the process matrix of ‘wires’ (36). This matrix contains full information about the causal structure. However, this may well be unfeasible in practice; the methods discussed below can provide information about causal structure in the absence of full control.

4.2. Discovery using Hilbert–Schmidt (HS) basis

Consider a space of linear operators over a Hilbert space with finite dimension $d_X$, $X \equiv \mathcal{L}(H^X)$. A HS basis is a set of self-adjoint matrices $\{ \sigma^X_{\mu} \}_{\mu=0}^{d^X-1}$ with $\sigma_0 = 1$, $\text{tr}\sigma_\mu\sigma_\nu = d_X \delta_{\mu\nu}$. This implies $\text{tr}\sigma_j = 0$ for $j \geq 1$. Local operations and process matrices can be expanded in HS bases. For example, a generic CP map for a local laboratory $A = A_I \otimes A_O$ decomposes as

$$M^{A_IA_O} = \sum_{\mu,\nu=0}^{d^I-1} \nu_{\mu\nu} \sigma^A_{\mu} \otimes \sigma^A_{\nu}, \quad \nu_{\mu\nu} \in \mathbb{R}. \quad (40)$$

Following [26], the following terminology applies.

- The term proportional to identity, $\nu_{00} I$, is called of type 1.
- Terms equal to the identity on all subsystems except subsystem $X$ are called of type $X$. For example, the terms $\nu_{0j} 1^h \otimes \sigma^X_j$, for $j > 0$, are of type $A_O$.
- Terms equal to the identity on all subsystems except $X_1 \otimes \cdots \otimes X_k$ are called of type $X_1 \cdots X_k$. For example, the terms $\nu_{0j} \sigma^X_1 \otimes \sigma^X_j$, for $j \geq 1$, are of type $A_I A_O$.
- The sum of terms of different types $X_1, \ldots, X_k$ is called of type $\left( \sum_{j=1}^k X_j \right)$. For example, a generic CP map (40) has terms of type $(1 + A_I + A_O + A_I A_O)$.

Some simple algebra can be applied to the above notation, such as $1 \cdot X = X$ and $(\sum X_i) \cdot (\sum Y_i) = (\sum_{i,j} X_i Y_j)$ for $X_i = Y_i$.

The HS decomposition can be used to characterise correlations and causal relations. For example, given a bipartite process matrix $W^{A_B}$, terms of type $A_O B_I$ allow signalling from $A$ to $B$. Terms of type $A_I B_O$, on the other hand, are responsible for common cause correlations when POVM elements $E^A \otimes F^B$ are measured.

For a quantum channel $T^{A_O B_I}$, the trace-preserving condition, equation (17), implies that terms of type $A_O$ vanish. Consequently, a process matrix that factorises over a DAG $G$, equation (29), has to satisfy a set of constraints. It is thus possible to test the compatibility of a process matrix with a DAG $G$ by looking for the terms excluded by $G$: if an excluded term is found, the process matrix is not compatible with $G$. Note that, although the constraints are formulated with reference to a HS basis, they are in fact basis independent, as any local change of HS bases does not change the type of terms contained in a process matrix. Such conditions can also be reformulated as basis-independent linear constraints, following the methods introduced in [42], but this technique will not be discussed here.

4.3. Faithfulness

A general process matrix might factorise over more than one DAG. Consider the example, figure 13(a),

$$W^{A_O B_I} = \rho^A_B \otimes \rho^B_I \otimes 1^{A_O B_I}, \quad (41)$$

with $\text{tr}\rho_A = \text{tr}\rho_B = 1$. This can be factorised in the following ways (with tensor product symbols understood):

$$W = \rho^A_X T^{A_O B_I} 1^{A_O B_I} = 1^{A_O B_I} \rho^B_I, \quad (42)$$

$$W = \rho^B_I T^{B_O A_I} 1^{A_O B_I} = 1^{B_O A_I} \rho^A_X, \quad (43)$$

$$W = T^A_A T^B_B 1^{A_O B_I} = \rho^A_A, \quad T^B_B = \rho^B_B, \quad (44)$$

which correspond to the DAGs in figures 13(b)–(d), respectively. However, it is clear that only the third DAG provides an appropriate representation of the causal relations in $W$, as it describes two uncorrelated laboratories. The DAGs in figures 13(c) and (d), on the other hand, require a specially selected mechanism that is unable to carry causal influence. This leads to the following definition (analogous to the definition for classical causal models [1]):

**Definition 6 (Faithfulness).** An MQCM $\langle G, W \rangle$ is faithful if $W$ contains non-vanishing HS terms of all types compatible with the factorisation (29). Such terms are said to be *implied* by $G$.

11 Pauli matrices are an example for $d = 2$. 

---

The terms implied by a DAG can be readily identified. To each laboratory $L_j$ is associated a quantum channel connecting the parent space $P_Sj$ to the input space $I_j$. Such a channel is given by a matrix that satisfies equation (31), $\text{tr}_j T_{P_Sj}^{S_Ij} = I^{PS}$. This condition implies that $T_{P_Sj}^{S_Ij}$ cannot contain terms of type $S_e$, for $e \in \text{IN}_j$, or any product thereof. The remaining terms are of type $1 + \prod_{e \in \text{IN}_j} (1 + S_e)I_j$. Thus, a DAG $G = (\mathcal{L}, \mathcal{E})$, $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^n$, implies all and only terms of type

$$\prod_{j=1}^n [1 + \prod_{e \in \text{IN}_j} (1 + S_e)I_j].$$

In particular, every edge $e$ ending on a laboratory $L_j$ implies terms of type $S_eI_j$. Recall, for an edge $e$ connecting a laboratory $L_k$ to a laboratory $L_j$, the source space $S_e$ is a subsystem of the output space $O_k$. At the price of a slight abuse of terminology, we will refer to terms of type $S_eI_j$, with $e \in \text{OUT}_k \cap \text{IN}_j$, simply as terms of type $O_k I_j$. We can thus say that a term $O_k I_j$ is implied if and only if $L_k$ is a parent of $L_j$ in the DAG.

If $\langle G, W \rangle$ is a faithful MQCM, $W$ is said to be faithful to $G$. A non-faithful causal model is said to be fine tuned. A process matrix $W$ is fine tuned if it only factorises for fine tuned models, while it is faithful if it is faithful to some DAG.

For faithful causal models, the relations between laboratories in a DAG directly correspond to causal relations. In appendix A, the following theorem is proved.

**Theorem 1.** Given a faithful causal model $\langle G, W \rangle$ for a set of laboratories $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^n$, a laboratory $L_k$ is a direct cause for a laboratory $L_h$ if and only if $L_k$ is a parent of $L_h$ in $G$.

The ’only if’ part of the theorem is also true for non-faithful models. This property can be phrased as children are able to screen off the causal influence of their parents to the rest of the world. This means that there exist instruments for the child laboratories that break the causal connections from their common parent to any other laboratory.

### 4.4. Discovery of faithful causal structures

The task of causal discovery is to find a causal structure, i.e. a DAG $G$, which can explain a set of empirical data. Faithful quantum causal models have two important properties in this respect: (i) the causal structure of a faithful MQCM can be discovered unambiguously; (ii) almost all causal models are faithful, in an appropriate measure-theoretic sense. These points are formalised in the following theorems.

**Theorem 2.** If two MQCMs $\langle G, W \rangle$, $\langle G', W \rangle$ for the same process matrix $W$ are both faithful, then $G = G'$.

**Proof.** As $G$ and $G'$ have the same set of vertices, the only thing to prove is that they have the same set of edges. This is equivalent to proving that each laboratory has the same set of parents for $G$ and $G'$. The latter is a direct consequence of the fact that $W$ contains a term of type $O_k I_j$ if and only if $L_k$ is a parent of $L_j$ in equation (45).

Thus, the HS terms contained in a faithful process matrix uniquely identify a causal structure. The HS terms can be fully determined if in each laboratory an informationally complete instrument is available. If only a subset
of the HS terms is measured, the observed data can be compatible with more than one faithful causal model. This is analogous to the classical case: if arbitrary interventions are not available, it is not possible to fully characterise causal structure.

The next theorem shows that, given a sufficiently regular prior knowledge on the causal structure of a model, there is vanishing probability that the model is fine-tuned (see [53] for the classical analogue).

**Theorem 3.** Given a set of laboratories \( \mathcal{L} = \{ L_j = I_j \otimes O_{j} \}_{j=1}^{n} \), a DAG \( G = (\mathcal{L}, \mathcal{E}) \) and a non-singular measure on the space of process matrices that factorise over \( G \), the set of fine-tuned process matrices have measure zero.

**Proof.** The set of process matrices that factorise over \( G \) can be parametrised by \( m \) real parameters, where \( m \) is the number of HS terms implied by \( G \) (excluding the term of type 1, which has a constant coefficient). A non-singular measure on this set is a measure on \( \mathbb{R}^m \) that is non-singular with respect to the Lebesgue measure. The set of fine-tuned process matrices is defined by having exactly vanishing coefficients for some of the implied HS terms, which is a zero-measure set for every non-singular measure. \( \square \)

Note that there is only a finite number of DAGs with the same vertex set. Thus, a non-singular measure over the space of causal models for a given set of laboratories decomposes as a finite sum of non-singular measures for the individual DAGs. This implies the following corollary:

**Corollary 4.** Given a non-singular measure on the space of MQCMs for a given set of laboratories \( \mathcal{L} = \{ L_j = I_j \otimes O_{j} \}_{j=1}^{n} \), the set fine-tuned causal models has measure zero.

The above results indicate that, given a set of possible causal explanations, a faithful one should be preferred because, unless additional information is available, a vanishing probability is assigned to fine-tuned models.

In the presence of latent laboratories, faithfulness might not be sufficient to single out a unique causal model. In this case, further assumptions would be needed to decide among a set of faithful causal explanations.

### 5. Classical limit

It is important to note that classical causal models can be recovered from quantum ones in an appropriate limit. To ensure we are recovering classical models that can be considered causal, it is necessary to show that these models support identification via classical interventions. Recall, this is the hallmark of the classical causal modelling methodology. As it will be shown in this section, to recover classical causal models it is sufficient that all available local operations can be described as classical interventions. The particular case of classical models conditioned on ‘idle’ interventions, or equivalently ‘purely observational’ Bayesian networks, is recovered by further restricting local operations to projective measurements in a fixed basis\(^{12}\). It is crucial that all classical causal structures can be recovered in this way. This was not possible, for example, in the framework of [20], where only a subset of DAGs was recovered in the classical limit.

Quantum systems effectively behave as classical ones in the limit where it is only possible to access and manipulate states in a fixed basis that factorises over separated systems (called pointer basis). Whether this condition is enforced by decoherence [54], collapse models [35], ‘fuzzy measurements’ [56], or in other ways, will not be discussed here. Rather, it will be shown that, provided the transition to classicality takes place, MQCMs reduce to classical ones.

For a local laboratory \( L_j = I_j \otimes O_{j} \) in an MQCM, with output space factorised according to the outgoing edges, \( O_{j} = \bigotimes_{e \in \text{OUT}_{j}} S_{e} \), consider CP maps of the form

\[
M_{x, i}^{(j)} = \sum_{z_i} P(\tilde{o}_j, x_j, z_i) x_j^{i} \otimes [\tilde{o}_j]^i,
\]

where \( x_j \) label the possible measurement outcomes, \( i_j \) label the possible choices of instruments, \( \tilde{o}_j := \{ s_e \}_{e \in \text{OUT}_{j}} \) and \( [\tilde{o}_j]^i := \bigotimes_{e \in \text{OUT}_{j}} [s_e]^i \). For a fixed basis, CP maps of this form can be interpreted as follows: a classical system enters the laboratory in some state labelled by \( z_i \), a value \( x_j \) is recorded as a measurement outcome and, for each outgoing edge \( e \), a system is sent out of the laboratory in the state \( s_e \). The parameter \( i_j \) determines the conditional probability for \( x_j \) and \( \tilde{o}_j \) given \( z_i \).

In this interpretation, \( x_j \) are the observed variables and \( i_j \) are intervention variables. \( z_i \) and \( \tilde{o}_j \), on the other hand, are latent variables, as they are not observed directly. In order to recover a causal model for the variables \( x_j \),

\(^{12}\) Indeed, classical non-invasive measurements correspond, in the quantum formalism, to projective measurements in a fixed basis: as repeating such measurements confirms previously obtained outcomes, they can be interpreted as revealing pre-existing properties of the system without disturbing it.
without such additional latent variables, it is necessary to further restrict the possible local operations. In particular, the variables $\tilde{y}_i$, responsible for transmitting causal influence out of the laboratory, should be influenced directly only by the observed variables $x_i$ (and only indirectly by $i$ and $z_j$):

$$P(\tilde{y}_i, x_j | z_j, i_j) = P(\tilde{y}_i | x_j) P(x_j | z_j, i_j).$$  \hspace{1cm} (47)

Given the above conditions, the statistics generated by a MQCM are equivalent to those generated by a classical causal model with the same causal structure. The proof of the following theorem is in appendix B.

**Theorem 5.** Given an MQCM $(G, W)$ for a set of laboratories $\mathcal{L} = \{L_j = I_j \otimes O_j\}_{j=1}^n$, the conditional probability

$$P(x_1, \ldots, x_n | i_1, \ldots, i_n) = \text{tr} \left[ \bigotimes_{j=1}^n M_{x_j | i_j}^{L_j} \cdot W^{L_{1-n}} \right]$$  \hspace{1cm} (48)

obtained from CP maps of the form (46) with probabilities of the form (47), satisfies the causal Markov condition (8) for a DAG $\mathcal{G}_C$ isomorphic to $G$. $G_C$ has random variables $X_1, \ldots, X_n$ as vertices, where $X_j$ takes as values the labels $x_j$ of the local measurement outcomes.

It is also possible to show that each classical causal model can be reproduced as the classical limit of an MQCM with the same causal structure, i.e. with an isomorphic DAG. Replace each random variable $X_j$ with a local laboratory $L_j = I_j \otimes O_j$, where $I_j$ has a basis element $|x_j\rangle$ for each value of $X_j$ and $O_j$ is the tensor product of a copy of $I_j$ for each outgoing edge. The process matrix is then

$$W = \bigotimes_{j=1}^n T^{PS}_{j},$$

$$T^{PS}_{j} = \sum_{z_j, p_{z_j}} P(z_j | p_{z_j}) |p_{z_j}\rangle |z_j\rangle^j \otimes |x_j\rangle^j,$$  \hspace{1cm} (49)

with coefficients $P(z_j | p_{z_j}) = \Pr(X_j = z_j | P_{A_j} = p_{z_j}, I_j = \text{idle})$ given by the classical model. The ‘observed’ probability distribution (where no interventions are made, from the perspective of the classical model) is obtained when all instruments are restricted to be projective measurements in the pointer basis:

$$M_{x_j}^L = |x_j\rangle^j \bigotimes_{e \in \text{OUT}_j} |e\rangle^j.$$  \hspace{1cm} (50)

Interventions in the classical model correspond to more general diagonal operations of the form (46). In particular, ‘do’ interventions are realised by ignoring the input and preparing the chosen value on the output space: $M_{\text{do}(x_j)}^L = |x_j\rangle^j \bigotimes_{e \in \text{OUT}_j} |e\rangle^j$. The split of classical nodes into input and output reproduces the single node intervention graphs, studied in the context of classical causal models [57].

The result above not only shows that classical causal models can be subsumed by quantum ones, but also provides a direct way to apply classical tools to quantum models. Indeed, classical models are recovered under restrictions on the performed operations. Such restrictions can arise because of uncontrolled factors, such as environmental decoherence, but it is also possible to simply choose to use only instruments compatible with a classical description. Thus, all properties of classical causal models directly extend to quantum ones, conditioned on the choice of classical instruments. For example, in classical causal models, if $X_j$ is the only common cause of $X_1$ and $X_2$, it follows from the classical Markov condition that $X_1$ and $X_2$ are conditionally independent given $X_3$: $P(X_1, X_2 | X_3) = P(X_1 | X_3) P(X_2 | X_3)$, where it is understood that such a property holds conditioned on the ‘idle’ intervention. Such a property still holds for quantum causal models, conditioned on the choice of appropriate instruments: if $I_3$ is a quantum common cause of $L_1$ and $L_2$, then

$$P(M_1^{L_1}, M_2^{L_2} | M_3^{I_3}, J_1^{L_1}, J_2^{L_2}) = P(M_1^{L_1}, M_3^{I_3}, J_1^{L_1}, J_2^{L_2}) P(M_2^{L_2} | M_3^{I_3}, J_1^{L_1}, J_2^{L_2})$$

for all instruments $J_1^{L_1}, \ldots$ satisfying conditions (46) and (47).

6. Beyond definite and acyclic causal structures

The results of the previous section show that the causal structure of a Markov process, represented by a DAG, remains unaltered in the quantum-to-classical transition. Thus, the quantum causal models discussed above are ‘quantum’ only by virtue of the physics describing the systems that convey causal influence. The causal structure itself is not different to that of classical models.

The formalism, however, can be naturally extended to more general structures. Probabilistic mixtures of causal structures are a natural extension. Owing to the linearity of the generalised Born rule, equation (20), these are represented as
\[ W = \sum_G p(G) W^{(G)}, \quad \sum_G P(G) = 1, \quad P(G) \geq 0, \]  

(51)

where the sum is over DAGs \( G \) and \( W^{(G)} \) is a process matrix that factorises over \( G \).

A process of the form (51) is still compatible with a definite, albeit unknown, causal structure, which can be understood as an objective property that exists independently of the operations performed. The natural next step is to look for causal structures that might be, in some sense, indefinite. That space-time itself might have quantum properties in a theory of quantum gravity motivates this idea [58, 59].

An example is the ‘quantisation’ of the mixture (51). Given a set of laboratories \( \mathcal{L} = \{ L_j = I_j \otimes O_j \}_{j=1}^n \), introduce two additional laboratories, \( C \equiv C_0 \) (with trivial input) and \( D \equiv D_1 \) (with trivial output). Let \( W^{(G)} = [w^{(G)}] \) be a rank-one projector which factorises over the DAG \( G \) (this is the case if all mechanisms are unitary matrices or pure-state preparations). A quantum-controlled causal structure is then given by the process matrix

\[ W = [w], \quad |w| = \sum_G |G|^{C_O} \otimes |G|^{I_H} \otimes |w^{(G)}|, \]  

(52)

where the basis vectors \( |G| \) range over the set of controlled DAGs. If \( C \) prepares a state \( |G|^{C_O} \), the laboratories in \( \mathcal{L} \) find themselves in the causal relations dictated by \( G \) (and the control system is transferred undisturbed to \( D \)). However, if \( C \) prepares an arbitrary superposition \( \sum_G \psi_G |G|^{C_O} \), and \( D \) also measures in a superposition basis, one can expect causal relations that are incompatible with any DAG [13]. Indeed, an example of quantum-controlled causal structure is the quantum switch [60, 61], where the above sum ranges over DAGs of type \( L_{r(1)} \rightarrow L_{r(2)} \rightarrow \ldots \rightarrow L_{r(n)} \), with \( r \) a permutation of \( n \) elements. This type of resource can be shown to be incompatible with any definite causal order [42, 43] and thus in particular with any DAG. Additionally, the quantum switch can provide an advantage for several tasks [62–64] and an experimental proof-of-principle was recently demonstrated [65], showing the potential practical relevance of quantum causal structures.

The process (52) can be interpreted as the superposition of different amplitudes, each corresponding to a directed, acyclic causal structure. It is an interesting question whether more general causal models can also be understood in a similar way. An indication that this might not be the case is the fact that processes of the form (52) cannot violate causal inequalities [42, 43] whereas more general processes allowed by the formalism can [26, 66]. Intriguingly, this is also true for classical systems: there exist causal models, locally compatible with classical physics, that are incompatible with any causal order or mixture of causal orders [67, 68].

The possibility of closed time-like curves (CTCs) in general relativity [69] motivates considering models that contain directed cycles. The process matrix formalism provides a natural framework for studying CTCs as cyclic causal structures, as the possibility of interventions, and thus the causal interpretation of the model, does not rely on the acyclicity property. It is not clear, however, whether indefinite and acyclic causal structures are in fact separate concepts. Furthermore, it is unclear whether and how core notions such as the causal Markov condition and faithfulness should be generalised beyond definite, acyclic causal structures.

7. Conclusions

This work has foundational implications insofar as it shows that quantum mechanics has a causal interpretation in a similar manner to classical mechanics. Cause-effect relations are identified with correlations between controlled and observed events, and a causal structure is a set of transformations that define a DAG.

The findings presented here do not reproduce the wealth of results from the literature on classical causal models. However, the surprising similarity between the frameworks suggests that further techniques from classical causal modelling may be applicable to the quantum case. Quantising causal models ought to be a rich and promising field of research.

Causal discovery plays a prominent role in classical machine learning and ought to play a similarly pivotal role in the emerging field of quantum machine learning [70]. The mathematical formalism for quantum causal discovery introduced here can provide a foundation for further development in this direction.

Acknowledgments

We are grateful to Peter Evans, Gerard Milburn, and Magdalena Zych for discussions and feedback. This work was supported in part by the Templeton World Charity Foundation (TWCF 0064/AB38).

[13] This should not be expected, however, when the superposition only involves DAGs with the same partial order between laboratories, as this results in a causally ordered process matrix, which has a Markovian causal explanation, see section 3.7.
Appendix A. Proof of theorem 1

If \( L_k \) is a parent of \( L_h \), then, by virtue of faithfulness, \( W \) contains terms of type \( O_{ik} \). Given an arbitrary set of CPTP maps \( C_j \), \( j \neq k, h \), their tensor product contains the term of type 1, which implies that the reduced process

\[
\mathcal{W}^{L_k L_h} = \text{tr}_{L_h} = \text{tr}_{L_k} [ \bigotimes_{j \neq k, h} C_j^{L_j} \cdot W ]
\]

also contains a term of type \( O_k L_h \). As such a term can be used for signalling, \( L_k \) has causal influence over \( L_h \) for any instruments in the remaining laboratories, thus \( L_k \) is a direct cause for \( L_h \). Conversely, assume that \( L_k \) is not a parent of \( L_h \). Then, every HS term non-trivial on \( O_k \) must also be non-trivial on an input \( I_j \) such that \( L_k \) is a parent of \( L_j \) (and thus \( j \neq h \)). Let \( \sigma_{O_k L_h} \) be the corresponding element of the HS basis (where \( X \) is any additional subsystem on which \( \sigma \) is non-trivial); then, for the CPTP map \( C_j = I_j^L / d_O \), and recalling that HS basis elements have null partial trace for any subsystem in which they are not trivial,

\[
\text{tr}_{L_h} C_j^{L_j} \sigma_{O_k L_h} X \equiv \text{tr}_{L_j} \sigma_{O_k L_h} X / d_O = 0,
\]

which means that the reduced process matrix obtained when \( L_j \) performs the maximally noisy channel does not contain the HS term \( O_l X \). If the maximally noisy channel is performed in every laboratory that has \( L_k \) as a parent, then the resulting reduced model does not contain any term of the form \( O_k X \), which means that no signalling from \( L_k \) is possible in the reduced model. Thus, \( L_k \) is not a direct cause of any laboratory of which it is not a parent.

Appendix B. Proof of theorem 5

The probabilities (48) are equivalently generated by the diagonal process matrix

\[
[\mathcal{V}] = \sum_{z_1} \cdots \sum_{z_n} \text{tr}([\mathcal{V}] \cdot W)[\mathcal{V}],
\]

\[
[\mathcal{V}] = [z_j]_{\mathcal{I}_j} \cdot [\delta_j]_{\mathcal{I}_0}.
\]

(B1)

Using the quantum causal Markov condition (29), the above matrix can be written as a product of CPTP maps diagonal in the pointer basis:

\[
\mathcal{W} = \bigotimes_{j=1}^n P_{I_j}^{PS_j},
\]

\[
T_j^{PS_j} = \sum_{z_p} P(z_j | p_j) [p_j]_{PS_j} \otimes [z_j]_{I_j},
\]

\[
P(z_j | p_j) = \text{tr}([p_j]_{PS_j} \otimes [z_j]_{I_j} \cdot T_j^{PS_j}),
\]

(B2)

with \( p_j = \{ s_j \}_{j \in \mathbb{N}} \).

For diagonal operations, the generalised Born rule reduces to the composition of probabilities:

\[
P(x | i) = \text{tr}[\bigotimes_{j} C_j^{I_j} \cdot W]
\]

\[
= \text{tr}[\bigotimes_{j} C_j^{I_j} \cdot \mathcal{W}]
\]

\[
= \sum_{i_{o}} P_i(o, x | z, i) P_W(z | o),
\]

(B3)

where \( o := [\delta_j]_{j=1} \equiv \{ p_j \}_{j=1} \), \( x := \{ x_j \}_{j=1} \), \( i := \{ i_j \}_{j=1} \), \( z := \{ z_j \}_{j=1} \), and

\[
P_i(o, x | z, i) := \prod_{j=1}^n P(\delta_j, x_j | z_j, i_j),
\]

\[
P_W(z | o) := \prod_{j=1}^n P(z_j | p_j)
\]

(B4)

are the conditional probabilities corresponding to the coefficients of local operations (46) and mechanisms (B2), respectively. Combining the expressions above, and using condition (47), gives
\[ P(x_0, \ldots, x_n | i_0, \ldots, i_n) = \sum_{\omega} \prod_{j=1}^{n} P(\omega_j | x_j, i_j) P(x_j | p_{j,\omega}) \]

\[ = \sum_{\omega} \prod_{j=1}^{n} P(\omega_j | x_j) P(x_j | p_{j,\omega}) \]

\[ = \sum_{\omega} \prod_{j=1}^{n} P(\{x_k \in \text{OUT}_j | x_j\}) P(x_j | \{x_k \in \text{IN}_j\}, i_j) \]

\[ = \prod_{j=1}^{n} P(x_j | p_{j,\omega}) \]

where \( p_{j,\omega} \equiv \{ x_j \mid \text{IN}_j \subseteq E \} \) is the set of outcomes associated with the parents of node \( j \) with respect to the original graph \( G \).

References

[26] Oreshkov O, Costa F and Brukner Č 2012 Quantum correlations with no causal order Nat. Commun. 3 1092
[33] Choi M-D 1975 Positive semidefinite biquadratic forms Linear Algebra Appl. 12 95–100
[34] Cohen S M 2011 When a quantum measurement can be implemented locally, and when it cannot Phys. Rev. A 84 052322
[56] Peres A 1992 Emergence of local realism in fuzzy observations of correlated quantum systems Found. Phys. 22 819–28
[57] Richardson T S and Robins J M 2013 Single world intervention graphs (SWIGs): a unification of the counterfactual and graphical approaches to causality CSSS Working Paper 128