Causal Models for a Quantum World

by

Katja Stephanie Ried

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2016

Author's declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Statement of contributions

The following is a list of all material in this thesis of which I am not the sole author.

Sections 2.3 and 2.4 summarize and extend the foundations on which the framework of quantum causal models is built, as developed by Leifer and Spekkens in [1].

The material in section 4.2 is based on a collaboration with M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens and K. J. Resch, and a version of it was published in [2]. The concept of observational schemes (section 3.2) was developed as part of this collaboration. The experiment was realized by M. Agnew and L. Vermeyden under the supervision of K. J. Resch, and Figs. 4.5 and 4.6 were prepared by them as part of the publication. I contributed to the design of the experiment and performed the data analysis for Fig. 4.6. The procedure of causal tomography introduced in section 2.5.3 and the proofs presented in appendices A.1 and A.4 were developed by me as part of this project and also included in [2].

Chapter 5 arose from a collaboration with J.-P. MacLean, R. W. Spekkens and K. J. Resch, the results of which are being prepared for publication at the time of writing. The idea of a quantum Berkson effect was conceived as part of this collaboration, but developed predominantly by myself. I designed the indicators introduced in section 5.3. The experiment described in section 5.4 was realized by J.-P. MacLean under the supervision of K. J. Resch; my contributions were to the design and subsequent analysis. Figs. 5.3, 5.4 and 5.5 were prepared by J.-P. MacLean.

Abstract

Quantum mechanics has achieved unparalleled success as an operational theory, describing a wide range of experiments to remarkable accuracy. However, the physical foundations on which it rests remain as puzzling as they were a century ago, and a concise statement of the physical principles that underlie quantum mechanics is still outstanding. One promising approach holds that these principles should be formulated in terms of *information*, since many of the counter-intuitive effects in quantum theory concern questions such as what one can know about a quantum system and how the information encoded therein can be processed and distributed between parties. Another challenging feature of quantum theory is its incompatibility with the other cornerstone of modern physics, general relativity. In order to reconcile the two, one must identify and retain only the essential concepts and principles of each theory, and in the case of general relativity, *causality* has been identified as such a concept. This raises the question of how our classical understanding of causality must change when quantum theory is taken into account. In order to address these questions about information, knowledge and causality, I turn to the framework of *causal models*. In classical statistics, causal models explain the relations among a set of variables in terms of causal influences, which makes them a powerful tool for structuring our knowledge about complex systems and developing strategies for interacting with them. More importantly, the framework provides the conceptual underpinnings and mathematical methods for addressing questions about causation, information and knowledge in a rigorous manner. In this thesis, I develop a version of the classical causal models framework that is compatible with quantum theory and explore its physical implications.

A first step is to define quantum versions of fundamental elements such as variables, conditionals and belief propagation rules. This allows one to consider the question of what one can come to know about a quantum variable from the point of view of causal modelling. The conditionals relating quantum variables are found to have a richer structure than their classical counterparts, which can be exploited for the task of discerning causal relations given limited data – a central problem in classical causal modelling. The mathematical properties of quantum conditionals also establish a correspondence between various classes of two-party correlations, such as bound and distillable entanglement, and the types of causal structures that can give rise to them, which may become a useful tool for entanglement theory and quantum information processing. In the context of open quantum systems dynamics, quantum causal models provide a clear physical explanation of not completely positive maps and, more broadly, of non-Markovian quantum dynamics. Finally, I consider the possibility of non-classical effects in the way that different causal mechanisms are combined – that is, non-classical causal *structures*. For the simple case of two causally ordered variables, I propose indicators that witness different classes of combinations of causal mechanisms, including a non-classical mixture, and describe an experiment realizing examples of the different classes.

As these results illustrate, the framework of quantum causal models provides both greater conceptual clarity and a comprehensive mathematical formalism for studying a diverse set of problems, ranging from foundational questions to applications in open systems dynamics and the exploration of non-classical causal structures, which are likely to feature in a future theory of quantum gravity. The material presented in this thesis is intended as a foundation and inspiration for further applications of quantum causal models.

Acknowledgements

I wish to thank my supervisor, Rob Spekkens, for supporting me, advocating for me and, most importantly, providing an inspiring example of how to do solid, interesting science. Credit also goes to my collaborators and colleagues, with whom I have shared many interactions that were both educational and enjoyable. I am grateful to the University of Waterloo and Perimeter Institute and their sponsors for the financial support I have received, and to the administrative staff at both institutions for keeping everything running smoothly.

To my friends and family: thank you for your support and companionship. I concede, after having been presented with proof by explicit construction, that the classical world can also be pretty fun.

Contents

List of Figures

List of Tables

Notation, conventions and abbreviations

Classical variables

We use upper-case letters, preferentially X, Y, Z, to represent classical random variables and the corresponding lower-case letters to denote the values of that variable.

h_X cardinality of a classical variable X.
$\mu(X) = \frac{1}{d_X}$ uniform distribution.
$\delta(x, y) = 1$ if $x = y$, zero otherwise (Kronecker delta distribution for discrete variables).

Hilbert spaces

We use upper-case letters, preferentially A, B, etc. to represent quantum variables. In the case of a product of Hilbert spaces, subscripts indicate which element of an expression pertains to which Hilbert space, as in $|\psi\rangle_A \otimes |\phi\rangle_B$. The ordering of the factor spaces is not fixed, so that $|\psi\rangle_A \otimes |\phi\rangle_B$ and $|\phi\rangle_B \otimes |\psi\rangle_A$, for example, represent the same state.

$\mathcal{H}_A$ Hilbert space of a variable A.
$\mathcal{H}_d$ Hilbert space of dimension d.
h_A Hilbert space dimension of a quantum variable A.
$\mathrm{i} \equiv \sqrt{-1}$, the imaginary unit.

Generic maps and operators

ρ_A may denote a quantum state on A (definition 11) or a generic operator on $\mathcal{H}_A$.

$\mathcal{L}(\mathcal{H}_A)$ space of linear operators on $\mathcal{H}_A$.
$\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ linear map between spaces of operators.
χ_{BA} operator that is Choi isomorphic to the map $\mathcal{E}_{B|A}$, see eq. (2.15).
$\tau_{B|A}$ operator that is Jamiołkowski isomorphic to the map $\mathcal{E}_{B|A}$, see eq. (2.16).
Π generic projector, $\Pi^2 = \Pi$.

Special maps, operators and states

$\mathbb{I}$ identity matrix.

$\mathcal{I}$ identity channel; $\mathcal{I}(\rho) = \rho \ \forall \rho \in \mathcal{L}(\mathcal{H})$.
$T_A(\rho_A) = T(\rho_A) = \rho_A^T$ transposition (basis-dependent).
$T_A(\rho_{AB}) = \rho_{AB}^{T_A}$ partial transposition with respect to A.
σ_i with $i \in \{1, 2, 3\}$ Pauli operators on a qubit, see eq. (2.63).
$\sigma_0 \equiv \mathbb{I}_{2\times 2}$ identity operator on a qubit, see eq. (2.63).
$|\pm_i\rangle$ with $i \in \{1, 2, 3\}$ Pauli eigenstates associated with eigenvalues ± 1.
$\{|0\rangle, |1\rangle\} = |\pm_3\rangle$ alternate notation.
$|\pm\rangle = |\pm_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$ for brevity.
$|\psi_\perp\rangle$ denotes a vector orthogonal to $|\psi\rangle$
$|\Phi^+\rangle \equiv \frac{1}{\sqrt{d}} \sum_j |jj\rangle$ for an orthonormal basis $\{|j\rangle\}$, see eq. (2.14).

Abbreviations

CP completely positive; coCP completely co-positive (definition 14).

POVM positive-operator valued measure (definition 20).
PPT positive (rigorously: positive-semidefinite) under partial transposition.
NPT not positive under partial transposition.
CC common-cause.
CE cause-effect.

Nomenclature

(in order of appearance)

Belief propagation (classical): eq. (2.4), p. 7.
Bayesian inversion (classical): proposition 4, p. 7.
Causal model (classical), causal structure: definition 6, p. 8.
Path (common-cause, cause-effect): definition 7, p. 9.
Causal parents, children, ancestors and descendants: definition 8, p. 9.
Quantum variable: definition 10, p. 12.
Belief propagation (quantum): eq. (2.13), p. 13.
Choi operator and Jamiołkowski operator: definitions 12 and 13, p. 13.
Quantum conditional: definition 16, p. 15.
Cause-effect and common-cause quantum conditionals: propositions 18 (p. 18) and 25 (p. 22).
Quantum instrument: definition 21, p. 20.
Pre- and post-intervention variables; splitting: proposition 26, p. 23.
Causal model (quantum): proposition 29, p. 28.
Causal map: definition 30, p. 29.
Causal tomography: definition 32, p. 30.
Informational symmetry and observational schemes: definition 33, p. 41.
Quantum observational scheme: example 35, p. 42.
Partial tomography: definition 37, p. 44.
Inference ellipsoid: definition 42, p. 55.
Classification of quantum conditionals:
 left- and right-handed: definition 43 (p. 56),
 extremal: definition 44 (p. 56),
 undecidable: definition 46 (p. 57).
Markov chain: definition 55, p. 66.
Classical common-cause and cause-effect relations: definition 58, p. 71.
Physical mixture of CC and CE: definition 62, p. 75.
Intrinsically quantum physical mixture of CC and CE: definition 65, p. 80.

1 Introduction

"Why?" is, simply put, the driving question of science. Humans have a deeply ingrained tendency to attempt to make sense of the complex world they live in by asking why it is the way it is. The first formal discussion of the problem dates back at least to the classical Greek philosophers. Aristotle, for example, wrote about the topic in his *Physics* and *Metaphysics*, where he divides the possible answers to the question of "why?" into classes such as efficient causes, which are external agents who bring about an observed change, and final causes, or purpose. The early concept of causation often carried strong connotations of responsibility, and indeed the discussion of causation, agency and free will remains relevant to questions in moral philosophy to this day.

Science, however, began to shift away from causal explanations. Galileo's *Discorsi*, for example, advises us to focus primarily on *describing* observed phenomena in the form of equations, before attempting to explain *why* they occur. The attitude that causal explanations are secondary was furthered by philosophers like David Hume, who essentially reduced causation to a mere impression that arises from "constant conjunction" (that is, the fact that two events are always observed together) [3]. In a similar vein, the mathematician Karl Pearson made seminal contributions to the discipline of statistics, but at the same time argued that causation should be abandoned as "a fetish amidst the inscrutable arcana of modern science" [4]. Only in the last decades has causality resurfaced as both a tool and a topic for rigorous scientific study. The benefits are clear: firstly, formulating a description of a complex system in terms of causal mechanisms relating its constituent parts allows one to predict how the system will respond when parts of it are manipulated or replaced, and extrapolate the effects of changing circumstances. Moreover, by using the predictive power of a causal model, one can *control* the system's behaviour, choosing appropriate actions and policies to bring about a desired outcome. Neither of these feats can be achieved using just a table of statistical correlations. These features have made causal reasoning highly successful in a diverse set of other fields, ranging from econometrics and social policy to epidemiology and computer science. This prompted the development of a thorough mathematical and conceptual framework of causal models on which we can now base ourselves [5, 6].

We begin by contrasting some proposed definitions of causality that are relevant to the present work; then we present the motivation for studying quantum theory from the point of view of causal models. Section 1.3 presents an outline of the remainder of this thesis.

1.1 Causality (disambiguation)

Over its long history, the term "causality" has acquired many definitions and meanings across the various disciplines that use it. This section reviews and comments on a few examples that are relevant to the present work.

The interventionist definition of causation is both intuitive and probably the most useful for practical applications: X has a causal effect on Y if, when altering X but leaving all else unchanged, one also sees a change in the value of Y. This immediately provides a canonical way of testing causal influences experimentally, namely by intervening on the putative cause and tracking whether this changes the putative effect.

In practice, this method may sometimes be unavailable: it may be impractical, unethical or outright impossible to manipulate the putative cause. Examples include medical trials, studies of galaxy formation and the question of whether humanity played a role in global warming. However, even in these cases, one can still generally *define* causation in terms of (possibly hypothetical) interventions – one may merely have to resort to more subtle techniques for detecting it.

More fundamentally, we note that this notion of causation hinges on how one defines and realizes interventions. For example, one must ensure that an agent's choice of how to alter X is not influenced by some unobserved factor that also influences Y, since in that case, seeing changes in Y that

correlate with the changes of X would not imply a causal influence of X on Y. But this condition on what constitutes an intervention in turn invokes a notion of X being "(causally) influenced by other factors", making the definition circular. This issue and others like it are the subject of an ongoing debate in philosophy, an introduction to which can be found in [7]. However, for practical purposes such as causal modelling, one can generally leave these questions aside and simply assume that there exist agents with sufficient free will to support an interventionist notion of causation.

Deterministic causation, which is sometimes referred to as "causality" or a "cause-and-effect relation", is the idea that a given configuration of the causes must always entail a particular effect; in other words, that the effect is completely determined by its causes. This meaning of the term causality may conflict with quantum theory, since the results of measurements on a quantum system are conventionally thought not to be predetermined by any cause, but intrinsically probabilistic [8]. (On the other hand, if quantum mechanics does generate outcomes that have no causal ancestors, this could support a much stronger interventionist notion of causation, since the possibility of an unobserved common cause is ruled out by hypothesis.) However, the question of whether one can give a deterministic account of quantum physics – and whether this is necessary or desirable, on philosophical grounds – lies outside the scope of this thesis. We will focus on a concept of causation that applies equally to probabilistic theories.

In theories about space-time and gravity, causality normally refers to the principle that constrains physical objects and information to travel inside the light-cone and thereby limits which events can influence which others. Whether or not a given event actually affects a second event is generally of little interest. (Saying that an event influences another of course presumes an underlying notion of causation, but, for the purpose of discussions about relativity, it is normally sufficient to use the intuitive, interventionist understanding of the term.)

Bell's "local causality" is defined in [9] as follows, with reference to the space-time structure depicted in Fig 1.1: "Full specification of what happens in Z makes events in X irrelevant for predictions about Y in a locally causal theory". This statement embodies the relativistic constraints on causal influences, for example by demanding that X does not have a direct causal influence on Y. For this reason, the violation of Bell inequalities is often cited as the textbook example of the tension between relativity and quantum theory. However, in order to obtain Bell's condition from relativistic causality, one must also assume that causal influences must be *mediated* by localized degrees of freedom along a continuous path through space-time. This ensures that there cannot be a common cause of X and Y in their distant shared past which influences Y directly, without being mediated by Z. This idea will appear again in the development of quantum causal models. Finally, we note that the definition proposed by Bell refers to a "full specification" of what happens in Z and considers what one can infer conditionally on that information. Both these steps are problematic if Z is a quantum system, as we will discuss in section 2.2.4.

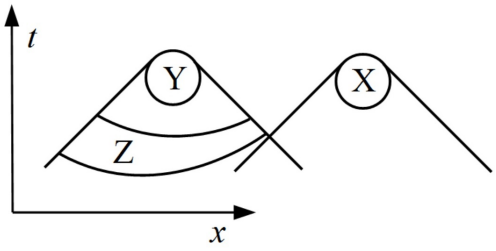

Figure 1.1: Bell's notion of local causality for variables living in a background space-time: X is independent of Y, conditional on Z. The axes representing time (t) and one spatial dimension (x) are scaled such that light rays travel at 45° angles with them. The resulting cones shown in the figure are the past light-cones of the regions Y and X, which consist of all space-time points that can send light rays or massive particles to the respective regions.

The framework of classical causal models implicitly assumes the interventionist notion of causation; indeed, the defining purpose of the formalism is essentially to predict the effects of the interventions of free agents. The practical limitations on *detecting* causal influences using the interventionist definition can be circumvented using various techniques designed to reveal information about the causal relations without resorting to interventions.

1.2 The promise of quantum causal models

The motivation for integrating the existing framework of classical causal models with quantum theory is twofold: firstly, causal models promise to provide a better understanding of the physical principles on which quantum theory is founded. Beyond that, developing a theory of quantum causality is a natural step on the way towards a futures theory of quantum gravity.

Although the basic mathematical machinery of quantum mechanics was established almost a century ago, the search for the physical principles underlying this branch of physics is still ongoing. By identifying such principles, we hope to remedy the dissatisfying situation that, as Richard Feynman put it, "nobody understands quantum mechanics" [10]. More pragmatically speaking, identifying the central concepts and principles of a theory is an important prerequisite for successfully generalizing it, while also providing an inspiration for new ways to harness its unique effects for practical purposes. In the case of quantum theory, it has been suggested – and borne out by a number of results – that the physical principles should be formulated in terms of information, governing how information can be transmitted and processed and what one can know and infer about the world [11, 12, 13, 14, 15, 16, 17]. Causal models are a natural framework for addressing questions about information processing, knowledge and inference. The process of designing a version of causal models that can accommodate quantum systems instead of classical variables promises a new perspective – one that is firmly grounded in operational notions – on the principles that govern knowledge and information in a quantum world and how they differ from their classical counterparts.

Conversely, generalizing the formalism of causal models to a new type of variable that behaves in a fundamentally different way, namely quantum variables, may provide new insights and inspiration for the discipline of causal modelling. For example, the results presented in section 4.1 reveal a new paradigm for causal discovery, and the challenge of witnessing an intrinsically quantum combination of causal relations has motivated us to propose a classification of the combinations of causal relations that can hold between *classical* variables, including several criteria for detecting the different classes (sections 5.1 and 5.3).

In addition to gaining a better understanding of quantum theory by itself, one of the biggest challenges of modern theoretical physics is to reconcile quantum theory with relativity to develop a unified theory of quantum physics and gravity. The current theory of relativity and quantum mechanics are incompatible in profound ways: for example, special and general relativity are deterministic, whereas quantum mechanics is probabilistic, and the role of time in the two theories is fundamentally different. The clash is epitomized by Bell inequality violations: the predictions of quantum mechanics are incompatible with local causality, which is in turn justified by relativistic constraints on causal influences. In order to reconcile the two theories, we will almost certainly need to give up familiar elements of both, including, quite possibly, the idea of space-time as a fundamental concept. Mathematically, it is known that one can reconstruct almost the entire metric of a space-time manifold (up to a conformal scale factor) given only the underlying structure of light-cones [18, 19, 20, 21]. This has inspired several compelling approaches to quantum gravity which consider causal structure to be the fundamental element from which space-time emerges, such as [22, 23, 24, 25, 26]. On these grounds, quantum causal models could become a cornerstone of a future theory of quantum gravity.

1.3 Outline

The remainder of this thesis is organized as follows:

Chapter 2 begins by introducing the mathematical and conceptual framework of classical causal models. We then lay the mathematical groundwork for quantum causal models, proposing definitions of variables, conditionals and a belief propagation rule that share the relevant properties of their classical counterparts. In simple scenarios, such as two variables related only via a shared common cause, we show that this formalism reproduces the results of conventional quantum mechanics. In order to describe more general causal relations, we introduce the additional requirement of *splitting* quantum variables, which is necessary in order to accommodate a full description of what one learns about a quantum variable by probing it with a general instrument. With this provision, the causal relations between quantum variables can be represented by a suitable operator. We discuss some necessary mathematical properties of such operators and a scheme for characterizing them experimentally.

Chapter 3 considers what one can come to know about a quantum variable by probing it in different ways. By studying this question, we hope to gain a new perspective on the fundamental rules governing information encoded in quantum systems, which underlie many of the puzzling behaviours typically exhibited by quantum systems. Specifically, we classify probing schemes for quantum variables according to their usefulness for causal discovery, which is the task of inferring the causal structure from observed correlations and arguably one of the central applications of classical causal models. We consider probing schemes that resemble passive observation of a classical variable in this sense; we term these "observational schemes". We analyse weak measurements, but find that they do not meet the requirement. However, we show that a rank-one projective measurement constitutes an observational scheme, if one had no prior information about the variable. We derive an algebraic characterization of what one can learn by this scheme and consider how one's powers of causal discovery change under small deviations from informational symmetry.

Chapter 4 proceeds to study the *conditionals* that relate two quantum variables; specifically the implications of how they differ from classical conditional probability distributions. For example, classical conditionals always take the same mathematical form, regardless of how the variables are causally related, whereas quantum conditionals must be positive-semidefinite if they describe a relation via a common cause (CC), but positive under partial transposition (PPT) if they describe a cause-effect (CE) relation. In the case of qubits, this distinction translates to a series of geometric criteria in terms of the Bloch sphere representation that allow one to determine the underlying causal structure. In the context of causal discovery, it powers a new, uniquely quantum tool for distinguishing causal relations given only data obtained by the aforementioned observational scheme. Section 4.2 reports on a quantum optics experiment that realizes an unknown probabilistic mixture of CC and CE relations and uses the theoretical result to estimate the mixing parameter from observational data.

An equivalent statement of the distinction between CC and CE relations is that maps representing inferences between variables that are related by a common cause need not be completely positive. This suggests that the framework of quantum causal models could bring mathematical and conceptual clarity to the discussion about non-completely positive maps in the context of open quantum systems, as well as the related phenomenon of non-Markovian quantum dynamics. In a similar vein, the fact that conditionals describing a cause-effect relation are positive under partial transposition provides an operational interpretation of PPT and NPT states in terms of the causal structures that can give rise to them, which may provide a new perspective on problems in quantum information and entanglement theory. We explore these connections in sections 4.3 and 4.4.

Chapter 5 takes up the possibility that not only the conditionals describing individual causal mechanisms, but also the ways in which they are combined to form a causal structure is in some sense non-classical. In an effort to make this statement precise, we consider how one can combine causal mechanisms that realize common-cause and cause-effect relations between two causally

ordered variables. We begin by proposing a distinction between probabilistic and non-probabilistic (physical) mixtures, which applies even in the classical case, and note that one can discern the two using Berkson's paradox. This statistical effect allows one to induce correlations between two variables, D and E, by conditioning on their common effect, B, and the strength of the induced correlations reflects how the mechanisms by which D and E each influence B were combined. We introduce a quantum version of the Berkson effect and note that it defines a natural third category in our classification, namely intrinsically quantum physical mixtures of common-cause and cause-effect relations. Having established these definitions, we propose several indicators that allow one to draw conclusions about the causal structure given limited experimental data. Finally, we report on a table-top experiment that realizes an intrinsically quantum physical mixture of common-cause and cause-effect relations between two photons, as witnessed by the quantum Berkson effect.

Chapter 6 draws conclusions from the findings of the previous chapters and points out promising directions for further investigation.

2 A mathematical framework for quantum causal models

This chapter lays the mathematical groundwork on which the remainder of the thesis is built. We begin by stating some basic properties of classical probability theory and introducing the formalism of classical causal models. We then propose quantum versions of the basic elements which reproduce the properties that are relevant for our purposes. The following sections illustrate that these generalizations behave as intended in simple scenarios involving just two variables, and subsequently extend them to more general cases.

2.1 Classical foundations

2.1.1 Probability theory

Let us begin by formalizing some familiar definitions and properties. The subsequent proposals of quantum analogues will build on these definitions.

Definition 1. Let X be a classical random variable that can take values denoted x. A *probability distribution* over X, denoted $P(X)$, is a function that assigns to each possible value x a non-negative real number $P(X = x) \geq 0$ such that the sum over all possible values is one, $\sum_x P(X = x) = 1$.

We use $P(X)$ to denote the function over the set of values $\{x\}$, whereas $P(X = x)$ denotes the value that this function takes for a certain argument. We will use the shorthand

$$\sum_X P(X) \equiv \sum_x P(X = x), \tag{2.1}$$

so that the normalization condition in definition 1, for example, can be written $\sum_X P(X) = 1$. Operations on probability distributions are to be performed element by element; for example, the product of two distributions $P(X)P(Y)$, is a function over ordered pairs of values, $\{(x, y)\}$, and the value of that function for an argument (x, y) is given by $P(X = x)P(Y = y)$.

If one wishes to describe two classical random variables, one can combine their sets of possible values by Cartesian product and analogously define a joint probability distribution, denoted $P(X, Y)$, over the product space. Given a joint distribution over several variables, one can derive *marginal* distributions over subsets of variables by summing over the possible values of the variables to be eliminated:

$$\begin{cases} P(X) = \sum_Y P(X, Y) \equiv \sum_y P(X, Y = y) \\ P(Y) = \sum_X P(X, Y) \equiv \sum_x P(X = x, Y). \end{cases} \tag{2.2}$$

Relating the joint distribution and the marginals, we define the following property:

Definition 2. Two classical variables X and Y are *statistically independent* or *uncorrelated* if their joint distribution is simply the product of its marginals,

$$P(X, Y) = P(X)P(Y). \tag{2.3}$$

The central question in the context of causal models, however, is this: if one finds X to take the value x, what can one infer about a second variable Y? The answer is encoded in the conditional probability distribution:

Definition 3. Let X and Y be two classical random variables that can take values denoted x and y, respectively. A *conditional probability distribution* over Y conditional on X, denoted $P(Y|X)$, is a set of probability distributions over Y indexed by x, which we denote $\{P(Y|X = x)\}$. That is, for each x, the distribution $P(Y|X = x)$ assigns to each possible value y a non-negative real number, $P(Y = y|X = x) \geq 0$, such that the sum over all possible values of y is one, $\sum_Y P(Y = y|X = x) = 1$.

Now, given a probability distribution $P(X)$ over a variable X and a conditional $P(Y|X)$ that encodes what one can infer about Y given X, then what one can deduce about Y is given by the probability distribution

$$P(Y) = \sum_X P(Y|X) P(X). \tag{2.4}$$

This is the rule of *belief propagation*, also known as Bayesian updating.

One can define a joint distribution over X and Y that combines information about how Y depends on X, as specified by a conditional $P(Y|X)$, and a given marginal distribution $P(X)$: the appropriate form is

$$P(X,Y) = P(Y|X) P(X). \tag{2.5}$$

It is easy to verify that this is in fact a joint probability distribution, and when marginalizing over Y one recovers

$$\sum_Y P(X,Y) = \left[\sum_Y P(Y|X)\right] P(X) = P(X). \tag{2.6}$$

If one marginalizes over X, expression (2.5) reproduces the belief propagation rule. Conversely, given a joint distribution $P(X,Y)$, one may wish to isolate a conditional encoding what one can infer about Y given certain information about X. Inverting expression (2.5) and substituting $P(X) = \sum_Y P(X,Y)$, one obtains

$$P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{P(X,Y)}{\sum_Y P(X,Y)}. \tag{2.7}$$

For those x for which $P(X = x) = 0$, the conditional probability obtained in this manner is not defined. However, considering that these x occur with probability zero, the fact that we do not know what we should infer about Y if we were to observe them is not problematic.

Bayesian inversion. We will see that causal models explicitly encode how each variable depends on its causal parents (see definitions 6 and 8 below), in the form of conditional probability distributions over each variable given its parents. However, one may wish to transform this information so as to permit inferences about any subset of variables given any other subset of variables. Bayesian inversion is the basic tool for making such transformations.

Proposition 4. *Given two classical random variables X, Y distributed according to the marginal $P(X)$ and the conditional $P(Y|X)$, what one can infer about X given Y is described by the conditional probability distribution*

$$P(X|Y) = \frac{P(Y|X) P(X)}{\sum_Y P(Y|X) P(X)}. \tag{2.8}$$

Proof. This follows by switching the roles of X and Y in expression (2.7) and combining it with expression (2.5). □

Note that the Bayesian inverse, $P(X|Y)$, depends not only on $P(Y|X)$, but also on the marginal $P(X)$. The marginals $P(X)$ and $P(Y) = \sum_X P(Y|X) P(X)$ are said to encode our *prior* knowledge. When one acquires new information about Y, updating one's beliefs to some $P'(Y)$ – and under the central assumption that the causal mechanism relating X and Y remains unchanged – then what can one infer about X is given by $P'(X) = \sum_Y P(X|Y) P'(Y)$. The updated distributions $P'(Y)$ and $P'(X)$ are termed *posterior*. We note that the terms prior and posterior, as introduced in this context, refer not to the *physical* time in the system under study, but rather to *epistemological* time: they distinguish what an agent believed before gaining new information about Y from what she believes after gaining that new information.

The following example illustrates how the conditional $P(X|Y)$ combines information about the causal mechanism, encoded in the conditional $P(Y|X)$, with the prior $P(X)$.

Example 5. Let X and Y be two classical variables that range over the values $x \in \{\pm 1, \pm 2\}$ and $y \in \{\pm 1\}$, respectively, and let

$$P(Y|X) = \delta\left(Y, \frac{X}{|X|}\right) = \begin{cases} Y = +1 & \text{if } X \in \{+1, +2\} \\ Y = -1 & \text{if } X \in \{-1, -2\} \end{cases}. \tag{2.9}$$

Let us assume at first that we have no prior information about X, which implies the uniform distribution, $P(X) = \mu(X) = \frac{1}{4}$ for all x. In this case $P(X,Y) = \frac{1}{4}\delta\left(Y, \frac{X}{|X|}\right)$, so $P(Y) = \frac{1}{2}$ for both values of Y, and

$$P(X|Y) = \frac{1}{2}\delta\left(Y, \frac{X}{|X|}\right). \tag{2.10}$$

That is, upon finding $Y = +1$, one would retrodict $X = +1$ or $X = +2$ with equal probability. Now suppose that we do have non-trivial prior knowledge about X, namely that $\tilde{P}(X = \pm 1) = \frac{1}{6}$, whereas $\tilde{P}(X = \pm 2) = \frac{1}{3}$. Now $\tilde{P}(X,Y) = \frac{1}{6}\delta\left(Y, \frac{X}{|X|}\right)[1 + \delta(|X|, 2)]$, which still makes $P(Y) = \frac{1}{2}$, but

$$\tilde{P}(X|Y) = \frac{1}{3}\delta\left(Y, \frac{X}{|X|}\right)[1 + \delta(|X|, 2)]. \tag{2.11}$$

Upon finding $Y = +1$, one must now conclude that $X = +1$ occurred with probability $\frac{1}{3}$, whereas $X = +2$ has probability $\frac{2}{3}$. The prior knowledge (that $|X| = 2$ is twice as likely as $|X| = 1$) is combined with the inference about the sign of X in order to determine the overall $\tilde{P}(X|Y)$.

2.1.2 The framework of classical causal models

Based on definitions 7.1.1 and 7.1.6 in [5], we will adopt the following definition:

Definition 6. A *causal model* over a set of classical random variables $\{X_i\}$ consists of (1) the *causal structure*, which is a directed acyclic graph whose nodes represent the X_i and whose edges ("arrows") represent direct causal influences, and (2) the *conditionals*: conditional probability distributions that encode how each X_i depends on those that directly influence it, according to the graph.

Elaborating on this definition, let us begin by specifying what constitutes a variable for the purpose of classical causal modelling. The variables in a causal model are classical random variables, which describe degrees of freedom in a physical system. We do not demand that every independent degree of freedom be represented by a distinct variable in the model; for example, it may simplify the discussion to summarize the position and momentum of a particle in a single variable in the model. However, considering that classical causal models describe systems which live in a well-defined space-time and obey the constraint of relativistic causality, it is generally advantageous to use distinct variables to represent the degrees of freedom associated with distinct localized regions of space-time, since this simplifies the description of the pattern of causal influences among the variables.

More importantly, we note that not every degree of freedom of the underlying physical system will be represented by a variable in the causal model. The purpose of a causal model is to summarize the pattern of influences between a selected set of degrees of freedom, which are relevant to a particular application, in a convenient way, and this implies coarse-graining over all other degrees of freedom. In the literature, the variables that are represented in the causal model are termed the observed variables.

The causal structure mentioned in definition 6 encodes a qualitative description of which of these variables directly influence which others. We stress the requirement of *direct* influences, in the sense that they are not mediated by other *observed* variables. (In general, any causal influence is mediated by a sequence of unobserved variables, such as the field degrees of freedom at every point along the trajectory of a photon.)

In order to discuss the inferences that one can make based on a given causal model, it is useful to introduce the concept of a path through the causal structure, which is a sequence of nodes (variables) such that each one allows one to make an inference about the next. An example is given in Fig. 2.1. Note that it is not necessary that every variable in the sequence causally *influence* the next.

Definition 7. Let X and Y be two variables that are part of a causal structure $\mathcal{G}$, and consider an ordered sequence $\{Z_i\}$ of nodes on the graph such that (a) every consecutive pair (Z_i, Z_{i+1}) is connected by an edge in the graph, and (b) the first element is X and the last element is Y. Such a sequence defines a *path* connecting X and Y.
If all edges along the path are directed away from X and towards Y, it is termed a *cause-effect* (CE) *path*, with X being the cause and Y, the effect. If there exists a node Z_j that is a cause of both X and Y (that is, such that there is a cause-effect path from Z_j to X and a cause-effect path from Z_j to Y), then the union of the two paths, which connects X and Y, is termed *common-cause* (CC) and Z_j is a common cause of X and Y.

Based on this, let us furthermore define the following terms:

Definition 8. For two nodes X and Y in a causal structure, if there exists a cause-effect path from X to Y, with X being the cause, then X is said to be a (causal) ancestor of Y and Y is said to be a (causal) descendant of X. If there is a single edge directly from X to Y, then X is a (causal) parent of Y and Y is a (causal) child of X.

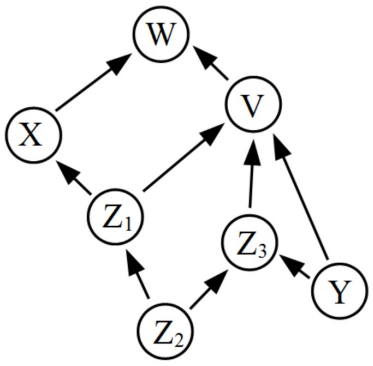

Figure 2.1: Example of a causal structure illustrating causal paths and relationships. The ordered sequence $\{X, Z_1, Z_2, Z_3, Y\}$ defines a path from X to Y. One can see that Z_1 and W are connected by a cause-effect path, with Z_1 being the cause and W, the effect. It follows that Z_1 is a parent of X and an ancestor of W, whereas W is a child of X and a descendant of Z_1. On the other hand, X and Z_3 are connected by a common-cause path, with Z_2 being a common cause of X and Z_3.

The uses of the causal models framework can be broadly grouped into two categories. On the one hand lies causal discovery, also known as causal inference: the problem of finding a suitable causal model for describing a many-variable system given statistical data about its constituent variables, such as a joint probability distribution. There are several tests and criteria for determining objectively which causal model best explains the data. Notably, a number of them do not require interventions. We will return to this topic in chapter 4.1. For the purpose of the discussion section 2.2.4, let us explicitly mention one necessary condition (based on definition 1.2.2 and the parental Markov condition, theorem 1.2.7, in [5]):

Proposition 9. *A causal structure $\mathcal{G}$ and a probability distribution P over the variables represented in $\mathcal{G}$ are compatible only if, in the probability distribution P, every variable is independent of all its non-descendants (in $\mathcal{G}$) conditional on its parents.*

The second major application of the causal models framework is, once the causal structure and the parameters are known, to make various inferences: mathematically speaking, one can derive conditional probability distributions relating any two subsets of variables, regardless of how they are causally related. The simplest inferences are based on cause-effect relations, for example the fact that rain causes the grass to become wet. Based on this causal statement, one can predict that, if it has rained, then the grass must be wet. Conversely, upon observing the grass to be wet, one can *retrodict* that it has likely rained. Inferences can also be made along common-cause paths. For example, if rain causes both the grass and the driveway to become wet, then finding the driveway to be wet allows one to infer that the grass will also likely be wet. The framework can also address much more complex and subtle questions, for instance about the implications of policies that prescribe interventions under certain conditions or complicated counterfactual claims. In the present work, however, we will focus on relatively straightforward inferences, and refer the reader to Pearl [5] for an overview of the alternatives.

Knowing the causal model allows one to construct conditional probability distributions $P(Y|X)$ for any two variables, which encode what one can infer about Y if one knows that X took a certain value. The conditionals specified as part of the causal model, which describe inferences about each variable given its causal parents, are a subset of these. They are special, however, because if X is a parent of Y, then $P(Y|X)$ also encodes how Y is *causally* affected if one *forces* X to take a certain value. That is, they represent not just statistical inferences, but causal influences. If one has the power to intervene on the individual variables, then these conditionals are easy to determine. We will argue in section 2.4.1 that, in the quantum case, conditionals to this type must have particular mathematical properties not demanded of other conditionals.

2.2 Features to be retained in quantum causal models

This section highlights some principles and features of classical causal models that we would like to retain in the quantum version:

2.2.1 Interventions by free agents

We will retain the assumption, made in classical causal models, that there exist experimenters with free will who can choose, for example, how to prepare or measure a quantum system. While philosophy and classical statistics have provided many interesting insights on how one can meaningfully define and detect causation without relying on this power, we will leave that problem aside and focus on generalizing the mathematical framework of classical causal models, which describes causal influences that can *in principle* be verified by interventions.

2.2.2 Autonomous causal mechanisms

A crucial feature of causal models, though it is often not stated explicitly, is that each conditional – that is, the mechanism by which each variable is (probabilistically) determined from its parents – is autonomous. This means that, if one alters one part of the system, for instance suspending the mechanism that would normally determine a variable X and setting it manually instead, then all other parts – in particular the mechanisms by which X in turn affects its descendants – remain unaffected, and our knowledge about them remains valid. In chapter 5, we explore some circumstances that can prevent one from describing the overall causal relations among a set of quantum variables as a collection of simple, autonomous mechanisms.

However, as a conservative first step towards fully quantum causal models, we will for now restrict ourselves to the case wherein the causal structure encoding how the variables influence each other is a directed acyclic graph and that it is supplemented by autonomous conditionals that specify how each variable depends on its parents, just as in the classical case. The assumption of a causal structure that can be represented as a directed acyclic graph is appealing because of the strong expectation

that causal influences must be mediated by physical systems, whose trajectories through space and time can be tracked experimentally and which are often strongly localized. A prime example is a Bell experiment, where we expect causal mechanisms to be mediated by the particles (such as photons) that are sent out to two parties and subsequently measured. If we dropped the assumption that causal influences must be mediated by localized carriers, then we would have no reason to propose causal structures wherein only certain pairs of variables are connected by causal influences, but should instead allow every variable to depend directly on all other variables that lie in its past light-cone. Such a model would not be particularly informative.

More importantly, we stress that models with a conventional causal structure can, in fact, capture interesting effects in quantum causality. For instance, a non-local state on two quantum variables, which is the resource that powers Bell inequality violations, can be explained by a simple common-cause structure – if one allows the common cause to be a quantum variable. (If the common cause is classical, by contrast, it can generate at most classical correlations, that is, a separable state.) It is therefore a worthwhile first step to simply replace classical random variables with a quantum analogue. The possibility of more exotic causal structures will be taken up in chapter 5.

2.2.3 Reichenbach's principle

The principle, stated originally by Reichenbach [27], essentially demands that every correlation be explained by a causal relation. Without this principle, causal models become meaningless: if we were to allow correlations without demanding some causal connection, then any pattern of correlations in a multivariate system could be "explained" by a model that contains no causal links at all.

Let us state this principle in more detail: two variables X and Y can only exhibit correlations if they are connected by a path[1] in the causal structure that supports a chain of inferences from X to Y. This can be achieved in two[2] ways: either by a cause-effect path or by a common-cause path (definition 7). However, the causal structure is only a qualitative description. The existence of a path (either cause-effect or common-cause) connecting X and Y is therefore necessary for explaining correlations between them, but it is not sufficient: one must furthermore demand that the conditionals associated with each edge along the path satisfy certain conditions. In classical causal models, one can establish the following sufficient condition: a causal model is said to explain correlations between X and Y – either by a common cause Z or by a cause-effect path through Z – if conditioning on Z makes X and Y statistically independent. At this point, the problem of formulating a quantum version of Reichenbach's principle ties into a larger issue, which we discuss next.

2.2.4 Screening

In classical causal models, one can observe the following phenomenon: For a causal chain $X \to Z \to Y$, if one already knows Z, then learning Y provides no new information about X. The same holds for the fork $X \leftarrow Z \to Y$: if one did not know Z, then learning Y would allow one to retrodict Z, which in turn has new implications for X. But if Z is already known, then Y provides no new information about X. In both cases, we say that knowing a variable Z can *screen off* inferences between X and Y. Formally, if one conditions on Z, then X and Y are statistically independent, $P(X, Y|Z) = P(X|Z)\,P(Y|Z)$. The same effect can arise in more complex causal structures, and a

[1] There may be more than one such path in a given causal structure, which may interfere with each other in non-trivial ways, but there must exist at least one in order to explain correlations between X and Y.

[2] We consider the case wherein X and Y are correlated *unconditionally* on any other variables. If one allows conditioning on a third variable, Z, then there are other types of paths between X and Y that support a chain of inferences. The simplest example is if Z is a common effect of X and Y, that is, $X \to Z \leftarrow Y$: if one knows both how X and Y influence Z and one also knows the value that Z took, then learning X allows one to make an inference about Y.

general necessary and sufficient condition for when this so-called d-separation occurs is given in [5]. However, for the purpose of this thesis, we will focus on the two simple scenarios discussed above.

Conditional independence plays an important role throughout the framework of causal models: in the context of Reichenbach's principle, a common-cause or cause-effect relation that is mediated by Z is deemed to explain observed correlations between X and Y only if conditioning on Z makes X and Y statistically independent. By the parental Markov condition (proposition 9), a probability distribution is only compatible with a given causal structure if every variable is independent of all its non-descendants conditional on its parents. In the context of causal discovery, a number of algorithms use observed conditionals independences to rule out potential causal structures. These facts provide a strong motivation for retaining some form of screening in the framework of quantum causal models.

The problem lies in generalizing the formal statement in terms of conditional independence – more specifically, in the act of conditioning on a quantum variable. If Z is a classical variable, one can unambiguously divide statistical data, such as results from many runs of an experiment, into subsets according to the values of Z in each run. Each subset defines one element of the conditional distribution $P(X, Y|Z = z)$. In the case of quantum variables, by contrast, there is no single, preferred set of mutually exclusive states on which one can condition in this manner. In order to explore the ways in which one *can* meaningfully condition on a quantum variable, we now turn to a discussion of quantum causal models.

2.3 Basic elements of quantum causal models

This section proposes quantum generalizations of basic elements of the causal models framework, such as conditionals and the belief propagation rule.

2.3.1 Quantum variables

Definition 10. A *quantum variable* A is a set of degrees of freedom which is associated with a Hilbert space $\mathcal{H}_A$. In a setting where there exists a background space-time, it is convenient to define each quantum variable as a subset of the degrees of freedom contained in a particular localized region of space-time.

This definition is designed to reproduce the relevant properties of classical random variables: they represent a coarse-grained set of degrees of freedom, which are grouped into variables in a way that (hopefully) puts the causal relations between them in a simple form. In particular, if one studies physical systems that are conventionally described as spatially localized systems that evolve in time, such as a register in a quantum circuit, then we will describe them as a series of distinct quantum variables indexed by time, rather than a single variable that evolves in time. This convention is common in classical causal modelling, where it is used to ensure that one's beliefs about a given variable are static, rather than evolving under some dynamics.

The quantum analogue of a classical probability distribution is straightforward:

Definition 11. Let A be a quantum variable associated with a Hilbert space $\mathcal{H}_A$. A *quantum state* on A, denoted ρ_A, is a positive-semidefinite operator on $\mathcal{H}_A$, $\rho_A \geq 0$, with unit trace: $Tr_A \rho_A = 1$.

We demand that the operator ρ_A be positive-semidefinite and normalized in order to ensure that the statistics generated by measurements, in accordance with the usual rules of quantum mechanics (specifically the Born rule), constitute a valid probability distribution according to definition 1.

Since there is a Hilbert space associated with each quantum variable, it is natural that the Hilbert space associated with a set of variables is obtained by taking the tensor product of the individual Hilbert spaces; a generalization of the Cartesian product by which we combine the sets of possible values of classical random variables. Using the tensor product ensures that we recover the conventional mathematical structure of the state space in the case where several quantum variables

are related by a single common cause, that is, where they are prepared in a multipartite quantum state.

Joint and conditional probability distributions are therefore generalized to linear operators on the tensor product of Hilbert spaces. We will see that these operators cannot always have the properties one demands of multipartite quantum states, such as positivity (see section 2.4.1), and if there are more than two variables involved, for example, then the simple construction we are developing in this section may not allow for any operator that reproduces the appropriate marginals on all subsets of variables (see section VII.B of [1]). The modifications required to overcome these difficulties are developed in section 2.5.

Regardless of the exact properties of the appropriate operators on several variables, one can establish a prescription for reducing them to the corresponding marginal operators on a subset of the variables, which reflects our beliefs about that subset if we ignore the remaining variables. The natural choice is the partial trace, which is well established in conventional quantum mechanics:

$$\rho_A = Tr_B \rho_{AB}. \tag{2.12}$$

This finally leads to the following: if probability distributions over several variables are generalized to operators on the tensor product Hilbert space and marginalization is achieved by the partial trace, then we expect the belief propagation rule for two quantum variables to take the form

$$\rho_B = Tr_A \left[\tau_{B|A} \left(\mathbb{I}_B \otimes \rho_A \right) \right], \tag{2.13}$$

for some operator $\tau_{B|A} \in \mathcal{L}\left(\mathcal{H}_B \otimes \mathcal{H}_A\right)$ with suitable properties, which we will term a *quantum conditional*. A formal definition will be derived in the following section and stated at its end; definition 16.

2.3.2 The Choi-Jamiołkowski isomorphism and quantum conditionals

A rule for quantum belief propagation, which maps states on a quantum variable A to states on a second quantum variable, B, is most naturally described as a map $\mathcal{E}_{B|A} : \mathcal{L}\left(\mathcal{H}_A\right) \to \mathcal{L}\left(\mathcal{H}_B\right)$. In order to put it in the desired form of the belief propagation rule, eq. (2.13), one must establish a correspondence between linear operators on a product Hilbert space, $\tau_{B|A} \in \mathcal{L}\left(\mathcal{H}_B \otimes \mathcal{H}_A\right)$, and linear maps of the form $\mathcal{E}_{B|A} : \mathcal{L}\left(\mathcal{H}_A\right) \to \mathcal{L}\left(\mathcal{H}_B\right)$. This section introduces two dualities between these spaces, and, based on them, proposes a definition of quantum conditionals.

Let us begin by defining two operators associated with a given map $\mathcal{E}_{B|A}$:

Definition 12. Let $\mathcal{E}_{B|A} : \mathcal{L}\left(\mathcal{H}_A\right) \to \mathcal{L}\left(\mathcal{H}_B\right)$ be a positivity-preserving linear map between linear operators on quantum variables A and B, and let $\mathcal{I} : \mathcal{L}\left(\mathcal{H}_A\right) \to \mathcal{L}\left(\mathcal{H}_A\right)$ denote the identity map on operators on A. Let $\{|\phi_j\rangle\}$, with $j = 1, ...h_A$ ranging up to the dimension h_A of $\mathcal{H}_A$, be a basis of the Hilbert space and let

$$\left|\Phi^+\right\rangle \equiv \frac{1}{\sqrt{h_A}} \sum_{j=1}^{h_A} |\phi_j\rangle \otimes |\phi_j\rangle \tag{2.14}$$

be a symmetric, maximally entangled state between A and an ancilla A' with the same Hilbert space dimension. The operator

$$\chi_{BA} \equiv \left(\mathcal{E}_{B|A'} \otimes \mathcal{I}_A\right) \left[\left|\Phi^+\right\rangle \left\langle\Phi^+\right|_{A'A}\right], \tag{2.15}$$

obtained by applying the map to the ancilla A', is termed the *Choi operator* – or *Choi state*, if it is positive-semidefinite and trace-one – that corresponds to the map $\mathcal{E}_{B|A}$ [28].

Note that the Choi operator associated with a given map $\mathcal{E}_{B|A}$ depends on the basis $\{|\phi_a\rangle\}$ chosen to define the symmetric state $|\Phi^+\rangle$. However, in most uses of the Choi operator, the choice of basis is ultimately irrelevant; and unless required, we will not specify it. A basis-independent alternative to the Choi operator can be obtained as follows:

Definition 13. Let $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a positivity-preserving linear map between linear operators on quantum variables A and B, and let $\mathcal{I} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_A)$ denote the identity map on operators on A. Let h_A denote the dimension of the Hilbert space $\mathcal{H}_A$, let $\mathcal{H}_{A'}$ be a Hilbert space, associated with an ancilla A', of the same dimension, and let $\{|\phi_j\rangle\}$ with $j = 1, ...h_A$ be a basis of $\mathcal{H}_A$ or $\mathcal{H}_{A'}$. The operator

$$\tau_{B|A} \equiv \left(\mathcal{E}_{B|A'} \otimes \mathcal{I}_A\right) \left[\sum_{j,k=1}^{h_A} |\phi_j\rangle \langle\phi_k|_{A'} \otimes |\phi_k\rangle \langle\phi_j|_A\right], \tag{2.16}$$

obtained by applying the map to the ancilla A', is termed the *Jamiołkowski operator* that corresponds to the map $\mathcal{E}_{B|A}$ [29].

The two operators are related by a simple partial transposition on A, denoted T_A, and renormalization:

$$\tau_{B|A} \equiv h_A T_A \chi_{BA}. \tag{2.17}$$

We note that the operator in parentheses in (2.16) is not generally positive-semidefinite, and therefore $\tau_{B|A}$ is not generally a valid quantum state. However, $\tau_{B|A}$ has the merit of being independent of the choice of basis: one can verify that applying unitary rotations that take $\{|\phi_j\rangle\}$ to some other basis $\{|\bar{\phi}_j\rangle\}$ on both A and A' leaves $\tau_{B|A}$ invariant.

The relevance of these two representations of maps $\mathcal{E}_{B|A}$ as operators on $\mathcal{H}_B \otimes \mathcal{H}_A$ is established by two closely related but distinct isomorphisms, named after Choi and Jamiołkowski, respectively. Before we state it, let us establish the following definition for future reference:

Definition 14. A positivity-preserving linear map $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is termed *completely positive* (CP) if and only if its extension with the identity channel $\mathcal{I}$ on an ancilla A' of arbitrary dimension, that is, the map $\mathcal{E}_{B|A} \otimes \mathcal{I}_{A'}$, is positivity-preserving. A positivity-preserving map $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ is termed *completely co-positive* (coCP) if and only if its composition with partial transposition, $\mathcal{E}_{B|A} \circ T_A$, is completely positive.

Theorem 15. *(Choi and Jamiołkowski isomorphisms [28, 29]) Let $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ be a positivity-preserving linear map between linear operators on quantum variables A and B. The representation of $\mathcal{E}_{B|A}$ as an operator $\chi_{BA} \in \mathcal{H}_B \otimes \mathcal{H}_A$ defined by eq. (2.15) constitutes an isomorphism between maps $\mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$ and operators on $\mathcal{H}_B \otimes \mathcal{H}_A$, termed the Choi isomorphism. The representation of $\mathcal{E}_{B|A}$ as an operator $\tau_{B|A} \in \mathcal{H}_B \otimes \mathcal{H}_A$ defined by eq. (2.16) constitutes an isomorphism of the same type, termed the Jamiołkowski isomorphism.*

Moreover, the mathematical properties of the map and the operators are related by the following necessary and sufficient conditions:

1. *$\mathcal{E}_{B|A}$ is trace-preserving $\Leftrightarrow Tr_B \chi_{BA} = \frac{1}{d_A}\mathbb{I}_A \Leftrightarrow Tr_B \tau_{B|A} = \mathbb{I}_A$*
2. *$\mathcal{E}_{B|A}$ is completely positive $\Leftrightarrow \chi_{BA} \geq 0 \Leftrightarrow T_A \tau_{B|A} \geq 0$*
3. *$\mathcal{E}_{B|A}$ is completely co-positive $\Leftrightarrow T_A \chi_{BA} \geq 0 \Leftrightarrow \tau_{B|A} \geq 0$.*

Proof. Proof of the isomorphism and the necessary and sufficient condition for complete positivity is given in the original references. The necessary condition for $\mathcal{E}_{B|A}$ to be trace-preserving follows directly from the definitions of χ_{BA} and $\tau_{B|A}$; the sufficient condition follows from (2.18) and (2.20) below. □

Although the two forms, χ_{AB} and $\tau_{B|A}$, differ only by a partial transpose and the choice of normalization, their physical interpretations and mathematical uses are different. Assuming that $\mathcal{E}_{B|A}$ represents a quantum channel realized by some black box, the corresponding Choi operator can be obtained by starting with the maximally entangled state $|\Phi^+\rangle$ on A and A' and applying the black box to A', as illustrated in Fig. 2.2. (This is precisely the setup used in entanglement-assisted process tomography, which characterizes an unknown quantum process by performing quantum

state tomography on the associated Choi operator χ_{BA} [30, 31].) In this case, wherein the map $\mathcal{E}_{BA}$ is completely positive and trace-preserving, it follows that the operator χ_{BA} must be positive-semidefinite and trace-one, hence being a valid quantum state. This is why χ_{BA} termed the "Choi state".

In order to discuss the physical interpretation of the Jamiołkowski operator, let us consider how one can describe the effect of the map on a generic input $\rho_A \in \mathcal{L}(\mathcal{H}_A)$ in terms of the representations introduced above: in terms of the Choi state, we can write

$$\mathcal{E}_{B|A}(\rho_A) = h_A \mathrm{Tr}_A\left[(T_A \chi_{BA})\, \mathbb{I}_B \otimes \rho_A\right] \quad (2.18)$$
$$= h_A \mathrm{Tr}_A\left[\chi_{BA} \cdot (\mathbb{I}_B \otimes T_A \rho_A)\right] \quad (2.19)$$

where one must include (partial) transposition on A, denoted T_A, that is applied either to the input ρ_A or to the Choi state itself. (The identity operator on B, which is formally required in order for us to multiply τ_{BA} and ρ_A, is often omitted for brevity.) The Jamiołkowski operator, on the other hand, is the natural choice[3] for writing the effect of the map as

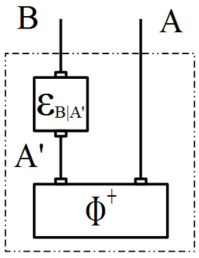

Figure 2.2: Operational interpretation of the Choi state χ_{AB} which is isomorphic to the map $\mathcal{E}_{B|A}$: the circuit inside the dashed box prepares χ_{BA}.

$$\mathcal{E}_{B|A}(\rho_A) = \mathrm{Tr}_A\left[\tau_{B|A} \mathbb{I}_B \otimes \rho_A\right], \quad (2.20)$$

which bears a close resemblance to belief propagation using a classical conditional probability distribution, eq. (2.4),

$$P(B) = \sum_a P(B|A)\, P(A). \quad (2.21)$$

Furthermore, the map is trace-preserving if and only if the Jamiołkowski operator satisfies $Tr_B \tau_{B|A} = \mathbb{I}_A$, in close analogy with the classical normalization requirement for conditional distributions, $\sum_B P(B|A) = 1\ \forall a$. This motivates us to introduce the following terminology:

Definition 16. A *quantum conditional* is an operator $\tau_{B|A}$ that is Jamiołkowski-isomorphic to a trace-preserving, positivity-preserving linear map $\mathcal{E}_{B|A}$. The fact that $\mathcal{E}_{B|A}$ is trace-preserving implies that $Tr_B \tau_{B|A} = \mathbb{I}_A$. The fact that $\mathcal{E}_{B|A}$ is positivity-preserving, but not necessarily completely positive, implies that $\tau_{B|A}$ is Hermitian, but not necessarily positive-semidefinite.

The definition of a quantum conditional proposed here was chosen because of its desirable properties for the purpose of representing inferences between quantum variables, following previous work in this direction, in particular [32, 1] and references therein. If one pursues a different purpose, then a different generalization of classical conditional probability distributions may be more appropriate. For example, Cerf and Adami have proposed a "conditional amplitude operator" that is suitable for studying quantum conditional entropy [33, 34].

2.3.3 Manipulating quantum conditionals

In order to perform belief propagation, Bayesian inversion, and other manipulations, one must often multiply and divide conditionals and marginals. While these operations can be defined in a relatively straightforward manner in classical probability theory, where they simply act element by element, similar operations on operators pertaining to quantum variables are more subtle. Different quantum

[3] Eq. (2.20) follows from eq. (2.16), which defines $\tau_{B|A}$.

generalizations of Bayesian inversion and retrodiction have been proposed, for example in [35, 12]. We will follow the proposal by Leifer and Spekkens [1].

We begin by noting that we may wish to combine operators that act on different Hilbert spaces, such as ρ_{AB} and ρ_{BC}. For these cases, we introduce the convention that each operator is extended as required by taking the tensor product with the identity operator on the complementary Hilbert space. In the example just mentioned, the extensions are $\rho_{AB} \otimes \mathbb{I}_C$ and $\mathbb{I}_A \otimes \rho_{BC}$. For brevity, the extra factors of $\mathbb{I}$ are often left implicit.

A natural choice of a multiplication operation on operators is provided by the matrix product. We will see in the following sections that a symmetrized version of the matrix product, termed the *star product* [1], can be used as a suitable generalization of multiplication in manipulating quantum conditionals. Given two linear operators τ_X and τ_Y, whose Hilbert spaces $\mathcal{H}_X$ and $\mathcal{H}_Y$ can be extended to $\mathcal{H}_X \otimes \mathcal{H}_{\bar{X}} = \mathcal{H}_Y \otimes \mathcal{H}_{\bar{Y}}$, let

$$\tau_X \star \tau_Y \equiv (\tau_Y \otimes \mathbb{I}_{\bar{Y}})^{\frac{1}{2}} (\tau_X \otimes \mathbb{I}_{\bar{X}}) (\tau_Y \otimes \mathbb{I}_{\bar{Y}})^{\frac{1}{2}}. \tag{2.22}$$

(The square root of the operators appears because, instead of multiplying with τ_Y either from the left or from the right, we wish to distribute the factor symmetrically, with $\tau_Y^{\frac{1}{2}}$ on either side.) Note that this product is neither associative nor commutative. Its extension to more than two operators is therefore not unique, which makes it complicated to formalize belief propagation involving more than two quantum variables. However, the chosen form does ensure other convenient properties:

Proposition 17. *If τ_X is a Hermitian operator and τ_Y is positive-semidefinite, then $\tau_X \star \tau_Y$ is also Hermitian. Moreover, if both τ_X and τ_Y are positive-semidefinite, then $\tau_X \star \tau_Y$ is also positive-semidefinite.*

Proof. One can see the first implication by noting that, if τ_Y is positive-semidefinite, then $\tau_Y^{\frac{1}{2}}$ is well-defined and also Hermitian, and therefore $\left[\tau_Y^{\frac{1}{2}} \tau_X \tau_Y^{\frac{1}{2}}\right]^\dagger = \left(\tau_Y^{\frac{1}{2}}\right)^\dagger (\tau_X)^\dagger \left(\tau_Y^{\frac{1}{2}}\right)^\dagger = \tau_Y^{\frac{1}{2}} \tau_X \tau_Y^{\frac{1}{2}}$, hence $\tau_X \star \tau_Y$ is also Hermitian. For the second part, note that the hypotheses imply that $\langle\phi| \tau_X \otimes \mathbb{I}_Y |\phi\rangle \geq 0$ for any vector $|\phi\rangle \in \mathcal{H}_{XY}$, and that $\left(\mathbb{I}_X \otimes \tau_Y^{\frac{1}{2}}\right) |\psi\rangle$ is such a vector for all $|\psi\rangle \in \mathcal{H}_{XY}$. It follows that $\langle\psi| \tau_X \star \tau_Y |\psi\rangle = \langle\phi| \tau_X \otimes \mathbb{I}_Y |\phi\rangle \geq 0$ for any $\psi \in \mathcal{H}_{XY}$, which implies positive-semidefiniteness. □

Division by an operator ρ_A can be defined analogously as taking the star-product with its pseudo-inverse, ρ_A^{-1}, that is, the inverse on the support of A.

These operations allow us to combine a conditional $\tau_{B|A}$ and a marginal quantum state ρ_A to form an operator that is analogous to a classical joint probability distribution in the sense that it encodes both the correlations between the two variables specified by the conditional $\tau_{B|A}$ and the marginal ρ_A:

$$\tau_{B|A} \star \rho_A = \rho_A^{\frac{1}{2}} \tau_{B|A} \rho_A^{\frac{1}{2}}. \tag{2.23}$$

Note that, since $Tr_B \tau_{B|A} = \mathbb{I}_A$ and the marginal on A is normalized to unit trace, we have $Tr\left(\tau_{B|A} \star \rho_A\right) = 1$.

On the other hand, given a joint operator ρ_{AB} whose marginal we denote $\rho_B = Tr_A \rho_{AB}$, one can multiply by the pseudo-inverse ρ_B^{-1} to obtain

$$\rho_{AB} \star \rho_B^{-1} = \rho_B^{-\frac{1}{2}} \rho_{AB} \rho_B^{-\frac{1}{2}} = \tau_{A|B}, \tag{2.24}$$

which satisfies definition 16 (since $Tr_A \tau_{A|B} = \rho_B^{-\frac{1}{2}} \rho_B \rho_B^{-\frac{1}{2}} = \mathbb{I}_B$ by design) and therefore constitutes a quantum conditional. If ρ_B does not have full rank, the conditional $\tau_{A|B}$ is only well-defined on the support of ρ_B (tensored with $\mathcal{H}_A$).

Combining the two previous steps produces a quantum analogue of Bayesian inversion, introduced in proposition 4: given a conditional $\tau_{B|A}$ and a prior quantum state ρ_A, and denoting $\rho_B = Tr_A \tau_{B|A} \rho_A$, one can obtain the conditional $\tau_{A|B}$ by taking

$$\tau_{A|B} = \left(\rho_A^{\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}}\right) \tau_{B|A} \left(\rho_A^{\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}}\right). \tag{2.25}$$

Note that the prior ρ_B is positive-semidefinite because the original $\tau_{B|A}$ represents a positivity-preserving map. It follows by the same reasoning as above that this $\tau_{A|B}$ is also a quantum conditional. As in the classical case, $\tau_{A|B}$ describes how one should update one's beliefs about A upon acquiring new information about B: the posterior ρ'_A can be obtained from the posterior ρ'_B by the belief propagation rule $\rho'_A = Tr_B \left[\tau_{A|B} \rho'_B\right]$. For future reference, we note that positivity of the conditional is preserved under Bayesian inversion: if $\tau_{B|A} \geq 0$, then $\tau_{A|B} \geq 0$.

2.3.4 The classical limit

As in conventional quantum mechanics, quantum variables can become effectively classical, in the sense that they can be described completely using classical probability theory. This limit is reached when all operators on the associated Hilbert space that appear[4] in the problem – states, conditionals, observables and so forth – are diagonal in a single basis[5], sometimes called the preferred basis. Many of the effects we consider typically quantum, such as coherence terms and the existence of non-commuting operators, vanish. Density matrices on a h-dimensional Hilbert space are reduced to vectors of h non-negative numbers that add to one – that is, to probability distributions over a set of h classical values.

Variables that are effectively classical in this sense can still be combined with non-trivial quantum variables by the usual tensor product structure and one can define conditionals relating the two types. Some examples are given in section 2.4.2 below. We will use the labels X and Y preferentially for classical variables, whereas A and B denote variables that exhibit non-trivial quantum behaviour. In analogy with classical conditional probability distributions such as $P(Y|X)$, which is defined as a set of probability distributions $P(Y|X = x)$, we introduce the notation

$$\tau_{A|X=x} \equiv \langle x| \tau_{A|X} |x\rangle \tag{2.26}$$

for the elements of a quantum-classical conditional $\tau_{A|X}$ associated with distinct values of the classical variable X, and similarly for other operators over both classical and quantum variables.

2.4 Inferences between two variables connected by a single causal path

This section illustrates how the formalism of quantum conditionals described in the previous section applies to simple scenarios involving two variables. Standard quantum mechanics provides representations of causal influences between quantum variables, but the treatment of other inferences is cumbersome at best. By contrast, the quantum conditionals formalism allows us to express belief propagation in a unified form regardless of the causal relation (at least in the simple scenarios discussed below) and therefore treat different causal relations in a more even-handed manner. The scenarios below were first analysed using quantum conditionals in [1], which also presents a number of additional applications of the formalism. More general scenarios, which require a physical and mathematical framework that goes beyond what we have introduced so far, are discussed in the next sections.

[4]Note that it is possible for a quantum variable to behave in an effectively classical way due to constraints on how an experimenter can interact with it, rather than intrinsic properties of the system. For example, one can effectively reduce a qubit to a classical bit by demanding that it may only be prepared in eigenstates of the Pauli observable σ_z, that one may only measure the observable σ_z on it, and that any coupling to other systems must be diagonal in the σ_z eigenbasis as well.

[5]The transpose, where required, must also be taken in that basis to avoid ambiguities.

In this section and much of the rest of the thesis, we are concerned in particular with two simple causal structures that can hold between two causally ordered variables: either they are only related as cause and effect (CE) or only via a common cause (CC). In either case, the nodes in the graph representing the causal structure are connected only by a single path[6].

2.4.1 Inferences between cause and effect

If A and B are related as cause and effect, respectively, which is conventionally described by a completely positive, trace-preserving map $\mathcal{E}_{B|A}$, then a rule of inference from states prepared on A to states subsequently found on B can be written in a form analogous to classical Bayesian inference,

$$\rho_B = \mathcal{E}_{B|A}(\rho_A) = Tr_A\left(\tau^{ce}_{B|A}\rho_A\right), \tag{2.27}$$

where $\tau_{B|A}$ is simply the Jamiołkowski operator isomorphic to $\mathcal{E}_{B|A}$, defined in eq. (2.16):

$$\tau^{ce}_{B|A} \equiv d_A T_A\left(\mathcal{E}_{B|A'} \otimes \mathcal{I}_A\right)\left[\left|\Phi^+\right\rangle\left\langle\Phi^+\right|_{A'A}\right]. \tag{2.28}$$

Let us highlight a mathematical property that will play a large role in later chapters:

Proposition 18. *If two quantum variables A and B are related as cause and effect by a map $\mathcal{E}^{ce}_{B|A}$, then $\mathcal{E}^{ce}_{B|A}$ must be trace-preserving and completely positive. Equivalently, the corresponding Jamiołkowski operator $\tau^{ce}_{B|A}$ is a quantum conditional and furthermore positive under partial transposition (PPT), i.e. $T_A\tau_{B|A} \geq 0$. Conversely, every PPT conditional can be realized as a cause and effect relation.*

Formally, this follows from theorem 15. The physical reason for demanding complete positivity is that a quantum channel $\mathcal{E}_{B|A}$, which represents a causal influence of a variable on its causal child, leaves the experimenter free to prepare any input they want, including inputs that are entangled with an ancilla A'. We demand that the output of the larger channel defined by $\mathcal{E}_{B|A} \otimes \mathcal{I}_{A'}$ (the tensor product with the identity channel on the ancilla) be a positive-semidefinite operator in all cases, so as to ensure that it predicts non-negative probabilities for the outcomes of all measurements one could perform on it. This means that the map $\mathcal{E}_{B|A}$ must be completely positive. Cause-effect conditionals are Jamiołkowski isomorphic to such maps. By the same physical arguments, they must therefore be PPT.

Using the star product, one can derive a quantum generalization of the joint probability distribution over cause and effect: a Hermitian operator that encodes the same correlations between A and B as when they are related by the conditional $\tau_{B|A}$, but also has the appropriate marginals on both B and A. Letting ρ_A denote one's prior beliefs about the input variable A, the desired operator takes the form $\rho_A^{\frac{1}{2}}\tau^{ce}_{B|A}\rho_A^{\frac{1}{2}}$. We do not term this a joint *state* because the operator need not be positive-semidefinite, although it must be PPT, which follows from the fact that the conditional $\tau^{ce}_{B|A}$ is PPT.

As a next step, we will derive from this object a conditional that encode *retrodictions*: inferences about a cause given its effect. Using Bayesian inversion, we have the following:

Proposition 19. *Let A and B be two quantum variables that are related as cause and effect by a conditional $\tau_{B|A}$ and let ρ_A denote the prior on A. In this case, the prior state on B is $\rho_B = Tr_A\left[\rho_{B|A}\rho_A\right]$. Then inferences from the effect B to the cause A are described by the operator*

$$\tau_{A|B} = \left(\rho_A^{\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}}\right)\tau_{B|A}\left(\rho_A^{\frac{1}{2}} \otimes \rho_B^{-\frac{1}{2}}\right), \tag{2.29}$$

[6]If one is interested only in the type of causal relation between A and B, then, following the discussion at the beginning of section 2.1.2, one can coarse-grain over all other variables, in particular merging them where appropriate. For example, if A and B have two common causes, G_1 and G_2, one can combine these into a single variable G and conclude that A and B are connected by a single (common-cause) path. In order to streamline the discussion, we will consider these coarse-grained scenarios, which support only one common-cause path and one cause-effect path between two causally ordered variables.

which is a quantum conditional and has a positive partial transpose.

The fact that a conditional describing inferences about a cause given its effect is also positive under partial transposition follows from proposition 18 and the fact that Bayesian inversion preserves not only positivity, but, by extension, also positivity under partial transposition.

2.4.2 Classical-quantum cause-effect conditionals: preparations and measurements

Taking the classical limit of cause-effect conditionals gives rise to three possibilities: classical-to-classical, classical-to-quantum and quantum-to-classical conditionals. We will provide physical interpretations for forward-in-time cause-effect conditionals of these forms.

Classical-to-classical. A classical-to-classical conditional in the quantum formalism takes the form

$$\tau_{Y|X} = \sum_{yx} \alpha_{yx} \left|y\right\rangle \left\langle y\right| \otimes \left|x\right\rangle \left\langle x\right|, \tag{2.30}$$

and the defining properties of cause-effect quantum conditionals (as stated in definition 16 and proposition 18) imply that

$$\begin{cases} T_X \tau_{Y|X} \geq 0 & \Leftrightarrow \alpha_{yx} \geq 0 \\ Tr_Y \tau_{Y|X} = \mathbb{I}_X & \Leftrightarrow \sum_y \alpha_{yx} = 1 \ \forall x. \end{cases} \tag{2.31}$$

These are precisely the defining properties of a conditional probability distribution; that is, the diagonal elements of $\tau_{Y|X}$ define a distribution $P\left(Y|X\right)$ over the classical variables.

Preparation of an ensemble of states. A classical-to-quantum conditional can be written as

$$\tau_{B|X} = \sum_{x} \rho_B^x \otimes \left|x\right\rangle \left\langle x\right|, \tag{2.32}$$

subject to the requirements

$$\begin{cases} T_X \tau_{B|X} \geq 0 & \Leftrightarrow \rho_B^x \geq 0 \\ Tr_B \tau_{B|X} = \mathbb{I}_X & \Leftrightarrow Tr_B \rho_B^x = 1 \ \forall x. \end{cases} \tag{2.33}$$

(There are no explicit coefficients α_x; instead, these are absorbed into the operators ρ_B^x.) These properties characterize physically valid quantum states. A classical-to-quantum cause-effect conditional therefore defines an ensemble of states $\{\rho_B^x\}$, indexed by the classical variable X. It can be interpreted as a preparation procedure that realizes one of a set of possible states depending on a setting X.

Destructive measurement. A quantum-to-classical conditional takes the form

$$\tau_{Y|A} = \sum_{y} \left|y\right\rangle \left\langle y\right| \otimes \tau_A^y, \tag{2.34}$$

where

$$\begin{cases} T_A \tau_{Y|A} \geq 0 & \Leftrightarrow \tau_A^y \geq 0 \\ Tr_Y \tau_{Y|A} = \mathbb{I}_A & \Leftrightarrow \sum_y \tau_A^y = \mathbb{I}_A. \end{cases} \tag{2.35}$$

(Note that $T_A \tau_A^y \geq 0$ if and only if $\tau_A^y \geq 0$.) It can be interpreted as a measurement on a quantum variable A which destroys the quantum system and yields different values of a classical variable Y as outcomes. Indeed, the set of operators $\{\tau_A^y\}_y$ fits the definition of how a general destructive measurement is represented in the conventional framework of quantum mechanics:

Definition 20. A *positive-operator valued measure* (POVM) is an indexed set of positive-semidefinite operators $\{\tau_A^y\}_y$ that sum to the identity operator, $\sum_y \tau_A^y = \mathbb{I}_A$. The individual operators in the set are termed *POVM elements*.

The probability of finding an outcome y if A was prepared in a state ρ_A can be obtained by the belief propagation rule: we find $P(Y = y|\rho_A) = Tr_A[\tau_A^y \rho_A]$, recovering the Born rule. To gain some intuition for POVMs, consider how they can represent a measurement of an observable M, which is often introduced as the first example of measurements on quantum systems: in this case, the possible values y of the outcome variable Y are the eigenvalues of M , and the POVM elements τ_A^y are projectors onto the associated eigensubspaces.

Non-destructive measurement. Finally, let us introduce a slightly more general cause-effect conditional relating classical and quantum variables, which we will refer to frequently in the coming chapters. This conditional takes two inputs, the classical variable X and the quantum variable A, and generates two outputs, the classical variable Y and the quantum variable B. It takes the form

$$\tau_{YB|XA} = \sum_{yx} |y\rangle\langle y| \otimes |x\rangle\langle x| \otimes \tau_{B|A}^{xy}, \tag{2.36}$$

subject to the constraints

$$\begin{cases} T_{XA}\tau_{YB|XA} \geq 0 & \Leftrightarrow T_A \tau_{B|A}^{xy} \geq 0 \\ Tr_{YB}\tau_{YB|XA} = \mathbb{I}_{XA} & \Leftrightarrow Tr_B\left(\sum_y \tau_{B|A}^{xy}\right) = \mathbb{I}_A \forall x. \end{cases} \tag{2.37}$$

The operators $\tau_{B|A}^{xy}$ are PPT, hence isomorphic to completely positive maps $\mathcal{E}_{B|A}^{xy}$, but one only obtains a trace-preserving map by summing over all y. For each x, this defines a quantum instrument:

Definition 21. A *quantum instrument* is a set of completely positive maps $\mathcal{E}_{B|A}^y : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, indexed by y, such that $\sum_y \mathcal{E}_{B|A}^y$ is trace-preserving. It can also be represented by the set of Jamiołkowski-isomorphic operators $\left\{\tau_{B|A}^y\right\}$, which must be PPT and satisfy $Tr_B\left(\sum_y \tau_{B|A}^y\right) = \mathbb{I}_A$.

An indexed set of quantum instruments, such as we have here, represents the most general kind of measurement: a classical setting X allows one to choose which instrument is realized, then the instrument takes as input a quantum variable A and produces a classical outcome Y, but also returns a quantum variable B, which is obtained from A by some transformation $\mathcal{E}_{B|A}^{xy}$ that may depend on both the chosen setting and on the outcome generated by the measurement. The individual $\mathcal{E}_{B|A}^{xy}$ associated with each outcome y need not be trace-preserving for the following reason: the trace of the resulting operator on B, by design, gives the probability of finding the outcome y, given the setting x and an input state ρ_A:

$$Tr_B\left[\mathcal{E}_{B|A}^{xy}(\rho_A)\right] = Tr_{BA}\left[\tau_{Y=y,B|X=x,A}\rho_A\right] = P(Y = y|X = x, \rho_A). \tag{2.38}$$

The probability of any individual outcome y is generally less than unity (depending on ρ_A); only the sum over possible outcomes recovers probability 1.

Example 22. Trivial measurement. If we do not interact with system A or its evolution to B at all, we gain no information about either – in other words, both the set of possible settings $\{x\}$ and the set of possible outcomes $\{y\}$ are singular, containing only a single element each, which we denote x_0 and y_0, respectively. The operator $\tau_{B|A}^{x_0y_0}$ is therefore by itself a quantum conditional, since it satisfies $Tr_B\tau_{B|A}^{x_0y_0} = Tr_B\left(\sum_y \tau_{B|A}^{xy}\right) = \mathbb{I}_A$. In particular, if we assume that B has the same dimension as A

and that the free evolution is modelled by the identity channel, definition 2.16 implies

$$\tau_{B|A}^{x_0 y_0} = \sum_{j,k=1}^{h_A} |\phi_j\rangle\langle\phi_k|_B \otimes |\phi_k\rangle\langle\phi_j|_A . \tag{2.39}$$

We note that this instrument preserves all coherence between A and B.

Example 23. Measure and reprepare. Consider an instrument from A to B that is realized by performing a measurement on A, with setting X and outcome Y, followed by a preparation of the variable B, which is determined only by the classical variables X and Y. The first step can be represented by an operator of the form

$$\tau_{Y|XA} = \sum_{yx} |y\rangle\langle y| \otimes |x\rangle\langle x| \otimes \tau_A^{xy} \tag{2.40}$$

and a set of POVMs: each setting x defines one POVM $\{\tau_A^{xy}\}_y$, whose elements are positive-semidefinite operators indexed by y that add up to the identity operator:

$$\begin{cases} \tau_A^{xy} \geq 0 \\ \sum_y \tau_A^{xy} = \mathbb{I}_A \forall x. \end{cases} \tag{2.41}$$

The second step can be modelled as a preparation on B controlled by a two-variable setting, YX, which is represented by an operator of the form

$$\tau_{B|YX} = \sum_{yx} \rho_B^{xy} \otimes |y\rangle\langle y| \otimes |x\rangle\langle x| \tag{2.42}$$

and an ensemble $\{\rho_B^{xy}\}_{xy}$ of positive-semidefinite, trace-one operators indexed by Y and X. One can combine the two steps just as one would classical conditional probability distributions, using the fact that X and Y are classical, so that the ordering of operators does not matter, and the fact that the preparation of B is independent of A, $\tau_{B|YXA} = \tau_{B|YX}$, to write

$$\tau_{YB|XA} = \tau_{B|YX}\tau_{Y|XA} = \sum_{yx} |y\rangle\langle y| \otimes \rho_B^{xy} \otimes \tau_A^{xy} \otimes |x\rangle\langle x| . \tag{2.43}$$

This implies that the operators associated with the overall measure-and-reprepare instrument take the product form,

$$\tau_{B|A}^{xy} = \rho_B^{xy} \otimes \tau_A^{xy}. \tag{2.44}$$

Physically, this implies that the transformation of A into B is not coherence-preserving.

As one can see by these examples, quantum instruments can transform the systems on which they act in very general ways, and it is sometimes convenient to assume particular, simple forms for these transformations. One form in particular will be used repeatedly in the coming chapters, hence we introduce it here. It is cast in the language of conventional quantum mechanics, wherein an instrument in considered to provide both a specification of a measurement on the original state of a quantum system and an update rule for predicting the state of the system after the measurement. It is often assumed that the measurement is realized in such a way that the update rule takes the following form:

Proposition 24. *(von Neumann-Lüders update rule) If a quantum system, initially in a state ρ, is subjected to a projective measurement and yields an outcome associated with the projector Π_y, the then state of the system after the measurement is*

$$\rho' = \frac{1}{Tr\left[\rho\Pi_y\right]}\Pi_y\rho\Pi_y. \tag{2.45}$$

(If the projector is rank-one, this reduces to $\rho' = \Pi$.)

The above are merely examples, illustrating how common elements of the conventional formalism can be represented as quantum conditionals. Similar methods can be used to obtain the classical limit of other operators, for instance the common-cause conditionals introduced below.

2.4.3 Inferences along a common-cause path

The relation between two quantum variables that are connected by a common cause is conventionally represented by a joint state ρ_{AB}. One can acquire information about A by performing a POVM measurement $\{E_A^y\}_y$, whose elements are indexed by the classical measurement outcome Y. Upon finding one value of Y, one can update one's beliefs about B to the normalized state

$$\rho_B^y = \frac{1}{P\left(Y=y\right)} Tr_A\left(\rho_{AB} E_A^y\right). \tag{2.46}$$

This manner of updating one's beliefs has been studied under the name Einstein-Podolsky-Rosen (EPR) steering [36, 37].

We wish to recast this rule in a form that resembles the general belief propagation rule more closely, specifically as

$$\rho_{B|Y=y} = Tr_A\left(\tau_{B|A}\tau_{A|Y=y}\right), \tag{2.47}$$

thereby putting this scenario on equal footing with inference along CE paths and allowing us to read off the appropriate definition of a quantum conditional that describes a common-cause relation. To this end, note that a conditional $\tau_{B|A}$ encodes a map from *what one knows* about A to what one can infer about B. That is, the input should be a state of belief on A, just as the input into a CE conditional is a state ρ_A. The POVM $\{E_A^y\}$, however, is represented in the inference framework as a conditional of the form $\tau_{Y|A}$. We therefore use Bayesian inversion to find $\tau_{A|Y} = \sum_y \tau_{A|Y=y} \left|y\right\rangle \left\langle y\right|$, which encodes what one can retrodict about A upon finding a measurement outcome y. We find

$$\tau_{A|Y=y} = \frac{1}{P\left(Y=y\right)} \rho_A^{\frac{1}{2}} E_A^y \rho_A^{\frac{1}{2}}, \tag{2.48}$$

which predictably depends on our prior knowledge ρ_A. Comparing with eq. (2.47), we identify the appropriate operator for belief propagation as

$$\tau_{B|A} = \rho_A^{-\frac{1}{2}} \rho_{AB} \rho_A^{-\frac{1}{2}}, \tag{2.49}$$

as one would have expected using Bayesian inversion[7]. Note that, if the marginal is maximally mixed, $\rho_A = \frac{1}{h_A}\mathbb{I}_A$, then the CC conditional is simply proportional to the joint state ρ_{AB}.

Proposition 25. *If two quantum variables A and B are related by a common cause that prepares the bipartite density operator ρ_{AB} with marginal $\rho_A = Tr_B \rho_{AB}$, then the Jamiołkowski operator describing inferences from A to B is $\tau_{B|A}^{cc} = \rho_A^{-\frac{1}{2}} \rho_{AB} \rho_A^{-\frac{1}{2}}$, which is a quantum conditional and furthermore positive-semidefinite. The corresponding map $\mathcal{E}_{B|A}^{cc}$ is therefore trace-preserving and completely co-positive. Conversely, every positive-semidefinite conditional can be realized in a common-cause structure.*

Proof. Let us first verify that the operator $\tau_{B|A}^{cc}$ defined by a CC relation is in fact a quantum conditional, that is, that the map $\mathcal{E}_{B|A}^{cc}$ is trace-preserving. This is ensured by construction: since

[7] The work on EPR steering provides an interesting perspective on the conditional $\tau_{B|A} = \rho_A^{-\frac{1}{2}} \rho_{AB} \rho_A^{-\frac{1}{2}}$: In the study of steering, the analysis is easier in the cases wherein the marginal ρ_A is maximally mixed. If this is not the case for the state ρ_{AB} that one is originally given, one can generate a modified state that does have the desired property, $\tilde{\rho}_{AB} = \frac{1}{d_A} \rho_A^{-1/2} \rho_{AB} \rho_A^{-1/2}$, which is of course proportional to the quantum conditional derived here. In the context of steering, $\tilde{\rho}_{AB}$ is obtained by *local filtering*: a quantum instrument followed by post-selection on one particular outcome.

$\tau^{cc}_{B|A}$ is the star product of ρ_{AB} with ρ_A^{-1}, it follows that $Tr_B \tau^{cc}_{B|A} = \mathbb{I}_A$. Similarly, it is easy to see that $\tau^{cc}_{B|A}$ inherits positivity from the state ρ_{AB}. In the special case wherein the marginal $\rho_A = Tr_B \rho_{AB}$ is the maximally mixed state, the two differ only by a normalization factor: $\tau^{cc}_{B|A} = d_A \rho_{AB}$. Finally, the properties of the corresponding map follow by theorem 15. □

We note that $\mathcal{E}^{cc}_{B|A}$ need not be completely positive. A quantum channel, which corresponds to a cause-effect conditional, must be CP in order to ensure consistency (rule out negative probabilities) because an experimenter can prepare any input for the channel, in particular an entangled state with an ancilla. The map $\mathcal{E}^{cc}_{B|A}$, on the other hand, represents only an inference, and the input into this particular inference map are retrodictive states of belief about A, $\tau_{A|Y=y}$. One can construct scenarios wherein one can assign a bipartite operator $\tau_{AA'}$ that represents a joint retrodictive state of belief about A and an ancilla A', which is then input into the inference map $\mathcal{E}^{cc}_{B|A}$, yielding $\tau_{BA'} = \left(\mathcal{E}^{cc}_{B|A} \otimes \mathcal{I}_{A'}\right)(\tau_{AA'})$. However, the operator $\tau_{BA'}$ is not a conventional joint quantum state, since it does not represent a preparation on two systems, and it is not required to have the same mathematical properties. Specifically, it need not be positive-semidefinite, but (as one can show), it must have a positive partial transpose. For this reason, we do not demand complete positivity of common-cause inference maps, but rather complete co-positivity.

2.5 General causal relations

While we have so far considered only cases wherein two variables, A and B, are related either purely as cause and effect or purely by a common cause, in general the mechanisms realizing these two possibilities can act simultaneously. This scenario serves as a toy model for exploring general causal relations between any number of quantum variables. In these cases, we will argue that one must split a quantum variable into two versions of itself in order to meaningfully discuss what one knows about the variable and its causal relations. Based on this, we will introduce an operator that encodes the general causal relations among a set of quantum variables and propose a causal model to explain them.

In our two-variable toy model, we will continue to assume that the variables are causally ordered, with A preceding B, so that the cause-effect relation has A as the cause of B. The general causal structure in this case is depicted in Fig. 2.3a. A concrete realization of this general causal relation is the circuit depicted in Fig. 2.3b: if the gate $\mathcal{E}_{B|AE}$ takes the form $\mathcal{E}_{B|A} \otimes Tr_E$, transforming the input A into B, then it realizes a CE relation; if $\mathcal{E}_{B|AE} = \mathcal{E}_{B|E} \otimes Tr_A$, then A and B are related via E and therefore via the common cause, and in general, one finds a combination of both scenarios.

2.5.1 Splitting quantum variables

We will argue the following:

Proposition 26. *In order to describe the general causal relation between a quantum variable A and other surrounding variables, one must replace A with two distinct variables, denoted C and D, that are associated with distinct Hilbert spaces, $\mathcal{H}_C$ and $\mathcal{H}_D$, each of the same dimension as $\mathcal{H}_A$. The causal structure is modified as follows: all causal influences acting on A, originating from its causal parents, are made to act on C, whereas all causal influences of A on its causal children are made to originate from D. We refer to C and D as* pre-intervention *and* post-intervention *versions of A, respectively.*

This splitting and the attendant changes to a generic causal structure are illustrated in Fig. 2.4. We note that, among the variables in the original causal structure, C has no causal children and D has no causal parents. This allows an agent to learn about the variable A by measuring C (which

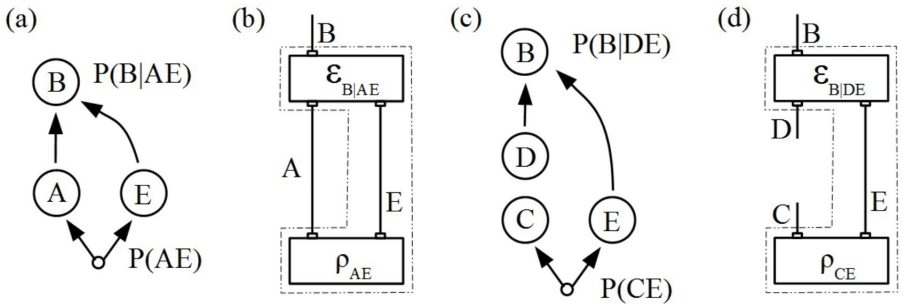

Figure 2.3: (a,b) The general causal relation between two causally ordered variables, A and B, is a combination of a CC relation, which is mediated by an ancillary variable E, and a CE relation. It can be depicted as an abstract causal structure and implemented as a circuit fragment (dashed box). (c) A complete description of the causal relation requires one to distinguish two versions of A: the pre-intervention version of A, denoted C, which is purely CC-related to B, and the post-intervention version of A, denoted D, which is purely CE-related (see proposition 26). (d) The general causal structure relating A and B can be realized with just two circuit elements: the preparation of a bipartite state on CE and a gate from DE to B. The input-output functionality of such a circuit is represented by the *causal map*, $\mathcal{E}_{CB|D}$ (see definition 30).

summarizes the influences of all parents of A) and preparing[8] D (thereby probing its effect on all children of A) *without* overriding any of the causal mechanisms that relate the original A to the surrounding variables.

The necessity of associating not one, but two Hilbert spaces with a quantum variable has manifested itself in several contexts, such as the process matrix framework [38] and the analysis of open quantum systems dynamics [39]. However, the framework of causal models provides a much clearer physical motivation for this mathematical trick, which is discussed in the remainder of this section. The arguments are stated for the case of just two variables, but one can see that a similar reasoning holds for any causal structure surrounding A. In the two-variable case, after splitting, the post-intervention version of A, D, is related to B purely as cause and effect, whereas the pre-intervention version of A, C, has a purely common-cause relation to B. In this manner, we return to the simple case wherein any pair of variables is related only by a single causal path. The causal structure and corresponding quantum circuit are shown in Fig 2.3c,d.

Linearity of inference rules. Let us begin by noting that, given a general causal relation between A and B (for example a combination of CC and CE mechanisms, as shown in Fig. 2.3a), there exists no linear map from a single version of $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$ that describes inferences from A to B. The reason is that there are generally *two* paths by which learning about A allows one to update one's beliefs about B. On the one hand, A has a direct causal influence on B, via the CE path, so that learning about A allows one to make an inference about B. On the other hand, learning about A allows one to make an inference about E, which in turn allows one to make an inference about B. If A and B were classical variables, this would not cause any further complications: one could still define a conditional probability distribution, since each value of A would entail a well-defined probability distribution over B, $P(B|A=a)$. However, if A and B are quantum variables, the existence of two paths along which one can make inferences causes an unusual effect. We illustrate this with an example:

[8] In general, the instrument chosen by the agent may not be of the measure-and-reprepare type, but allow a non-trivial causal influence of C on D. However, the important point is that this instrument is chosen by the agent and that it is not part of the pre-existing causal relations between A and the surrounding variables. For this reason, we do not include a causal influence of C on D in depictions of the modified causal structure, such as Fig. 2.4.

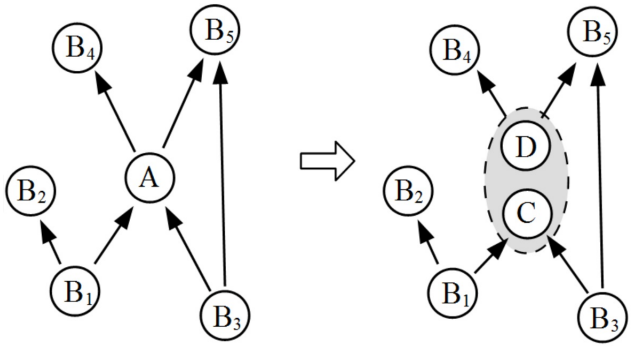

Figure 2.4: Splitting a quantum variable A into the pre- and post-intervention versions, C and D: all incoming causal influences from causal parents of A are transferred to C, whereas all outgoing causal influences acting on causal children of A now originate from D.

Example 27. In the circuit shown in Fig. 2.3d, suppose that C, D and E are qubits and let $\rho_{CE} = |\Phi^+\rangle\langle\Phi^+|$, where $|\Phi^+\rangle \equiv \frac{1}{\sqrt{2}}(|0\rangle \otimes |0\rangle + |1\rangle \otimes |1\rangle)$ and $\{|0\rangle, |1\rangle\}$ is the eigenbasis of the Pauli operator σ_z. Suppose that we acquire information about A using a projective measurement that gives an outcome $Y = 0$ if the state was $\Pi_0 \equiv |\psi_0\rangle\langle\psi_0|$, and $Y = 1$ for $\Pi_1 = |\psi_1\rangle\langle\psi_1|$, for some orthogonal pair $\{|\psi_0\rangle, |\psi_1\rangle\}$. Consider first what we can retrodict about C upon leaning Y: this is encoded in a classical-to-quantum conditional, which takes the form $\tau_{C|Y} = \sum_y \rho_C^y \otimes |y\rangle\langle y|$, where the ρ_C^y are quantum states. Considering that the prior $\rho_C = \frac{1}{2}\mathbb{I}_C$ is maximally mixed, our retrodictions about C are simply pure states, $\rho_C^y = \Pi_y$. In order to derive an inference about E, we construct

$$\tau_{E|C} = \rho_C^{-\frac{1}{2}} \rho_{CE} \rho_C^{-\frac{1}{2}} = \sum_{j,k=0}^{1} |jj\rangle\langle kk| . \tag{2.50}$$

Now we use $\tau_{E|Y} = Tr_C \tau_{E|C} \tau_{C|Y} = \sum_y \left[Tr_C \tau_{E|C} \rho_C^y\right] \otimes |y\rangle\langle y|$ to make inferences about E, finding

$$\rho_E^y \equiv Tr_C\left[\tau_{E|C}\rho_C^y\right] = \sum_{j,k=0}^{1} |j\rangle\langle k|_E \cdot \langle k|\rho_C^y|j\rangle = \Pi_y^T. \tag{2.51}$$

The projector Π_y^T is the transpose of ρ_C^y relative to the basis $\{|0\rangle, |1\rangle\}$ on which ρ_{CE} was defined.

Let us now consider what we can infer about D. To this end, suppose that the measurement obeys the von Neumann-Lüders update rule (proposition 24), so that the state after the measurement is simply the one corresponding to the outcome: $\rho_D^y = \Pi_y$. For the purpose of this pedagogical example, we furthermore take $\mathcal{H}_B$ to be isomorphic to the product $\mathcal{H}_D \otimes \mathcal{H}_E$ and take the the map from DE to B to be the identity map, $\mathcal{I}$. In this case, we find that the state on B that we infer if the measurement on A produced outcome y is

$$\rho_B^y = \mathcal{I}_{B|DE}\left[(\Pi_y)_D \otimes \left(\Pi_y^T\right)_E\right] = \Pi_y \otimes \Pi_y^T, \tag{2.52}$$

which contains two copies (up to transposition) of the state Π_y.

The circuit in this example realizes a map that essentially takes a single generic pure state to two copies of that state (up to transposition of one copy): if the measurement on A yields the outcome associated with the projector Π_y, then the inferred state on B is $\rho_B^y = \Pi_y \otimes \Pi_y^T$, for any Π_y. This is highly unconventional for quantum mechanics, but it does not actually constitute a violation of the no-cloning principle. To see this, recall that the no-cloning theorem is only violated if we construct

a circuit that takes a quantum system in an unknown state as *input* and produces two copies of that state as output. In example 27, by contrast, an agent interacting with the circuit does not have the power to input a generic state ρ_A of his choice, but can at most choose on which basis she wishes to measure A and then passively record which states are found. Equivalently, one may note that conventional quantum channels, which must obey consistency constraints such as linearity and the no-cloning theorem, always describe causal influences. By contrast, the map from A to B that is defined by the procedure in example 27 describes only an inference, and for this reason it is not subject to the same consistency conditions.

Although non-linearity does not violate any fundamental principles of physics, it is generally an undesirable property in an inference map. The reason is that convex combinations play a prominent role in probability theory by representing ignorance. For example, if one acquires information about A by a measurement that distinguishes among a set of orthogonal states $\{|\psi_i\rangle\}_i$ and finds either $|\psi_0\rangle$ or $|\psi_1\rangle$, but one is not sure which of the two was obtained, then one would expect the corresponding state on B to be given by a suitable convex combination of what one would infer upon finding $|\psi_0\rangle$ and what one would infer upon finding $|\psi_1\rangle$.

Independent information. A second, stronger reason for not attempting to define an inference map from $\mathcal{L}(\mathcal{H}_A)$ to $\mathcal{L}(\mathcal{H}_B)$ is that, in the general case, what one learns about the variable A by interacting with it in some manner cannot be represented by a linear operator on a single copy of $\mathcal{H}_A$. To see this, suppose that one intervenes on A with an instrument $\left\{\mathcal{E}^y_{D|C}\right\}_y$ and learns that the outcome variable Y took some value y, and let ρ_C be a quantum state encoding any prior information one may have had about the system. Based on this information, one can make a prediction about the post-intervention state, which is

$$\rho^y_D = \frac{1}{Tr\left[\mathcal{E}^y_{D|C}(\rho_C)\right]}\mathcal{E}^y_{D|C}(\rho_C), \tag{2.53}$$

and a retrodiction about the pre-intervention state,

$$\rho^y_C = \frac{1}{Tr\left[\mathcal{E}^y_{D|C}(\rho_C)\right]}\rho_C^{\frac{1}{2}}\left(Tr_D\tau^y_{D|C}\right)\rho_C^{\frac{1}{2}}, \tag{2.54}$$

where $\tau^y_{D|C}$ denotes the Jamiołkowski representation of $\mathcal{E}^y_{D|C}$. Except under special circumstances[9], the two do not coincide. In this case, therefore, one can only describe what one learns from the instrument with two operators, one on each version of A. For a general instrument, what one learns about C and D is represented by a general operator on $\mathcal{H}_C \otimes \mathcal{H}_D$, which need not take the form of a product, $\rho^y_C \otimes \rho^y_D$. (For example, in the case of the trivial measurement (example 22), for instance, we observed that the element $\tau^{x_0y_0}_{D|C}$ does not factorize.)

Characterizing the causal relation to other variables. Distinguishing between C and D is not only a matter of representing a complete description of the variable A, but also key to fully characterizing its causal relation to other variables. In the case where A is related to B by a combination of cause-effect and common-cause mechanisms, distinguishing C and D is the easiest way to discern these two causal mechanisms. This is well known in classical statistics. The paradigmatic example is a drug trial, with the variable A encoding whether a patient took the drug and the variable B representing their recovery. The goal is to verify that A causally influences B and to rule out the possibility that the observed correlations between A and B are due to an unobserved common cause. The most straightforward way to verify a causal influence is by an intervention:

[9]This possibility is further explored in section 3.2.

one assigns each patient either the drug or a placebo in a way that guarantees its independence of any potential common causes – that is, at random. The assigned treatment is the post-intervention version of A, denoted D. The relation between D and B encodes the cause-effect component of the relation between A and B. At the same time, one could track patients' intent to take the drug, which is the pre-intervention version of A, denoted C. The relation of C to B encodes the common-cause component of the relation between A and B. Splitting A into C and D therefore allows one to give a full description of the causal relation between A and B.

When splitting is not necessary. The above arguments become moot in cases wherein only one of the two pieces of information about A is actually relevant for making inferences; in other words, wherein A has either only parents or only children among the remaining observed variables. (Indeed, the definitions of C and D in proposition 26 are designed to take advantage of this simplification: the fact that C has no children among the other variables in the model ensures that only retrodictions about C are relevant, and conversely for D. If C or D had both parents and children, then they would again need to be split into two versions, leading to an infinite regress.)

In the example where A is related to a single, causally posterior B, if the relation is common cause, then only ρ_C is relevant, whereas if the relation is purely cause-effect, then only ρ_D is relevant. In either case, inferences from A to B do take the form of a linear map $\mathcal{E}_{B|A} : \mathcal{L}(H_A) \to \mathcal{L}(\mathcal{H}_B)$. This scenario and its consequences are explored in section 4.1. Similarly, since we assume that A and B are causally ordered, with A preceding B, we note that B has only causal parents among the remaining variables. For this reason, the post-intervention version of B is irrelevant for our purposes, and it is sufficient to consider a single copy of the variable. Without the promise of causal ordering, we would have to split B as well.

Note that the above arguments implicitly assume that one wants to make some statement involving beliefs about A: what one learns from measurements or preparations, what one can infer about B given certain information about A, or how to distinguish the different ways in which knowledge of A implies knowledge of B. In order to make such statements operationally meaningful, one must probe A by some instrument, and this is what forces us to split the variable. If one wants to make statements that do not involve beliefs about A, for instance if A is only an intermediary variable along a cause-effect path between two other quantum variables, B and F, and one seeks $\tau_{B|F}$, then there is no need to probe A. Formally, one can represent the fact that one does not probe A by inserting the trivial instrument (that is, the identity channel from example 22) between the pre- and post-intervention versions and tracing over the associated Hilbert spaces.

2.5.2 General joint and conditional operators and quantum causal models

Splitting a quantum variable A ensures that both of the resulting new variables are accessible to agents who seek to acquire information about them, in the following sense: Since C has no children among the variables in the model, an agent can apply a measurement on C without overriding the existing causal relations in the model, and, based on the outcome, make retrodictions about C. Similarly, D has no parents in the model, so that an agent can prepare D without modifying the existing causal relations and make predictions about D based on the preparation setting. Once all the variables $\{A_i\}$ in a causal model have been split into $\{C_i\}$ and $\{D_i\}$, the fact that all the new variables are accessible to an agent – either by prediction or by retrodiction – ensures that one can define joint and conditional operators on more than two variables in an operationally meaningful manner. We will argue the following:

Proposition 28. *Consider a set of quantum variables $\{A_i\}$ and let $\mathcal{C} \equiv \{C_i\}$ and $\mathcal{D} \equiv \{D_i\}$ denote the sets of pre- and post-intervention versions of the $\{A_i\}$, as defined in proposition 26. Then the causal relations among the $\{A_i\}$ can be represented as a quantum conditional $\tau_{\mathcal{C}|\mathcal{D}} \in \mathcal{L}(\mathcal{H}_\mathcal{C} \otimes \mathcal{H}_\mathcal{D})$ that is positive under partial transposition on the post-intervention variables, $T_\mathcal{D}\tau_{\mathcal{C}|\mathcal{D}} \geq 0$.*

To see why this holds, let us consider the necessary mathematical properties of joint and conditional operators on subsets of $\{C_i, D_i\}$. For this discussion, it is useful to recall that joints and conditionals are related by the star product, as detailed in section 2.3.3, and that switching between the two forms leaves properties of positivity and positivity under partial transposition unchanged.

One can distinguish three classes of operators: those that act only on pre-intervention variables (i.e. a subset of $\mathcal{C}$), those that act only on the post-intervention variables ($\mathcal{D}$) and those that combine both. An operator over a set of pre-intervention variables can be interpreted as a multipartite preparation, which can be probed by any measurement of suitable dimension. It is therefore subject to the same consistency constraints that apply to single-variable quantum states (definition 11) and two-variable common-cause joint operators (section 2.4.3): that is, it must be positive-semidefinite. Similarly, an operator over a set of post-intervention variables can be interpreted as part of a measurement (specifically an unnormalized POVM element, see definition 20), being subject to the same constraints, which dictate that it must also be positive-semidefinite.

Now consider an operator that acts on both pre- and post-intervention versions of the variables (for notational simplicity, suppose that it acts on all of $\mathcal{C}$ and $\mathcal{D}$). If we cast it in the form of a conditional, conditioning on $\mathcal{D}$, then it can be interpreted as a quantum channel[10]: given the inputs $\mathcal{D}$, which an agent is free to prepare, the conditional specifies the outputs $\mathcal{C}$, which the agent can then measure. As we have argued before, such a scenario must be represented by a completely positive map. If we wish to maintain the form of the quantum belief propagation rule, eq. (2.13), this means that the corresponding conditional must be positive when taking the partial transpose with respect to the inputs of the channel, i.e. the post-intervention variables $\mathcal{D}$.

Given the conditional $\tau_{\mathcal{C}|D}$, one can derive various inferential conditionals and joint states over different subsets of variables using the mathematical tools developed earlier. Among other implications, this resolves a limitation of the star product, pointed out in section 2.3.3: that the mathematical prescription (2.22) does not necessarily yield a Hermitian operator, or may not even be uniquely defined, in more general cases such as the three-variable chain. In retrospect, one may argue that there is no reason why there should exist a generalization of a joint state, that is, an operator that encodes joint beliefs, in cases such as the three-variable chain. Operationally, learning about the three variables requires one to probe all of them, in particular the middle one, which has both a parent and a child among the observed variables and therefore must be split in order to represent what one has learned. After the splitting, all variables are once again only connected by single causal paths, in which case the star product causes no problems.

Quantum causal models. Having established how to generalize joint and conditional probability distributions over a set of quantum variables in an operationally meaningful manner, let us finally propose the following:

Proposition 29. *Consider a set of quantum variables $\{A_i\}$ and suppose that the causal influences between them, as determined by interventions, can be represented by a directed acyclic graph. Then one can describe their relations by a causal model consisting of (1) the* causal structure*, which is a directed acyclic graph whose nodes represent the A_i and whose edges ("arrows") represent causal influences, and (2) the* conditionals*: quantum cause-effect conditionals that encode how each A_i depends on its causal parents, which are specified by the graph.*

In principle, it is possible that a set of quantum variables exhibit patterns of influences that cannot be represented as a directed acyclic graph (see [38], for an example). In these cases, our remarks about splitting still hold, but instead of autonomous causal mechanisms (represented by

[10] In general, if the events described by the causal model unfold over some period of time $[t_1, t_N]$, then different C_i may become available for measurement at different times and different D_i may be required to be prepared at different times. This is unusual for a quantum channel with inputs $\{D_i\}$ and outputs $\{C_i\}$. However, it is always possible to prepare all D_i together before t_1 and measure all C_i together after t_N, thereby recovering the more familiar setting of a quantum channel.

distinct conditionals), one may have to allow an overall map from the $\{D_i\}$ to the $\{C_i\}$ which cannot be decomposed into individual maps for each C_i (at least in the simple manner that is familiar from classical causal modelling). It is an open question how the central ideas of causal modelling can be applied to such scenarios, but the results presented in chapter 5 of this thesis may provide a foundation for considering this question.

2.5.3 General causal relations between two causally ordered variables

This final section focuses on special cases of the previous results for the scenario of only two causally ordered quantum variables, which will be used heavily throughout the remainder of this thesis as a pedagogical example to illustrates more general effects.

The causal map is an instance of the general conditional introduced in proposition 28 (recalling that it is not necessary to split B):

Definition 30. Let A and B be two quantum variables, with A causally prior to B and let C and D denote the two versions of A introduced in proposition 26. A *causal map* relating A and B is a completely positive and trace-preserving map $\mathcal{E}_{CB|D} : \mathcal{L}(\mathcal{H}_D) \rightarrow \mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_B)$ such that its marginal factorizes as

$$Tr_B \mathcal{E}_{CB|D} = \rho_C \otimes Tr_D, \tag{2.55}$$

for some ρ_C.

In order to provide some intuition for the physical meaning of causal maps, let us consider two simple examples: If the causal relation between A and B is purely cause-effect, then the causal map reduces to

$$\mathcal{E}^{ce}_{CB|D} = \rho_C \otimes \mathcal{E}_{B|D}, \tag{2.56}$$

that is, a quantum channel from D to B combined with the preparation of a fixed state on C. If the causal relation between A and B is purely common-cause, the causal map prepares a bipartite state on CB, regardless of the input provided at D:

$$\mathcal{E}^{cc}_{CB|D} = \rho_{CB} \otimes Tr_D. \tag{2.57}$$

Using this definition, we can state the following:

Theorem 31. *Let A and B be two quantum variables, with A causally prior to B and let C and D denote the two versions of A introduced in proposition 26. Then the causal relation between A and B can be represented by a causal map $\mathcal{E}_{CB|D}$.*

Proof. In order to see that the general relation between two causally ordered quantum variables can be represented as a causal map, consider a circuit fragment that implements this relation, which is illustrated in Fig 2.3d: the circuit prepares the pre-intervention version of A, denoted by C, possibly in a joint state with some ancillary degrees of freedom, which we denote E. The circuit then receives the post-intervention version of A, denoted by D, and can use both D and E to generate its final output, B. The functionality of this circuit, which specifies the causal relation between A and B, can be represented – like any other quantum circuit – by a linear map from the input, D, to the outputs, C and B, which we denote $\mathcal{E}_{CB|D} : \mathcal{L}(\mathcal{H}_D) \rightarrow \mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_B)$. This map must be completely positive in order to ensure that, if one prepares D in an entangled state with an ancilla D' and inputs D into the circuit while applying the identity channel to D', then the resulting operator on D', C and B is positive-semidefinite.

Unlike the maps representing conventional quantum channels, the causal map is also subject to a second constraint: because the output C, which is the pre-intervention version of A, is causally prior to the input D, it is impossible for C to be causally influenced by D. This implies that the marginal $Tr_B \mathcal{E}_{CB|D}$ must factorize, and therefore $\mathcal{E}_{CB|D}$ satisfies the definition of a causal map. □

Conversely, any causal map can be realized by a circuit of the form Fig. 2.3d, by choosing the appropriate circuit elements ρ_{CE} and $\mathcal{E}_{B|DE}$. One can see this explicitly as follows: let d_A denote the Hilbert space dimension of C and D, and let E be a quantum variable of the same Hilbert space dimension. Given the causal map $\mathcal{E}_{CB|D}$, denote $\rho_C \equiv Tr_B\left[\mathcal{E}_{CB|D}\left(\frac{1}{d_A}\mathbb{I}_D\right)\right]$ and take

$$\rho_{CE} = \rho_C^{\frac{1}{2}}\left(\left|\Phi^+\right\rangle\left\langle\Phi^+\right|_{CE}\right)\rho_C^{\frac{1}{2}}. \tag{2.58}$$

The second component of the circuit, the gate $\mathcal{E}_{B|DE}$, can be specified by its Jamiołkowski operator, $\tau_{B|DE}$, which we choose as follows. Let $\tau_{CB|D}$ be the Jamiołkowski operator representing the causal map and consider the operator

$$\tau_{B|CD} \equiv T_C\left(\rho_C^{-\frac{1}{2}} \otimes \mathbb{I}_{BD}\tau_{CB|D}\rho_C^{-\frac{1}{2}} \otimes \mathbb{I}_{BD}\right). \tag{2.59}$$

Note that $Tr_B\tau_{B|CD} = \mathbb{I}_C \otimes \mathbb{I}_D$ (because of the defining property (2.55) of causal maps), hence $\tau_{B|CD}$ is a quantum conditional, justifying the notation we chose. The partial transpose appearing in its definition ensures that the conditional $\tau_{B|CD}$ is PPT with respect to both of the variables on which we are conditioning, $T_{CD}\tau_{B|CD} \geq 0$. This is the property we expect to see in $\tau_{B|DE}$, which is a cause-effect conditional. Finally, noting that $\tau_{B|CD}$ is an operator on $\mathcal{H}_B \otimes \mathcal{H}_C \otimes \mathcal{H}_D$ but that formally $\mathcal{H}_C = \mathcal{H}_E$, consider an operator of the same form acting on $\mathcal{H}_B \otimes \mathcal{H}_E \otimes \mathcal{H}_D$: take $\tau_{B|DE} = \tau_{B|CD}$, where we identify the Hilbert spaces of C and E. Using this conditional together with ρ_{CE} from above, it is easy to verify that we recover the conditional that is Jamiołkowski-isomorphic to the original causal map, $\mathcal{E}_{CB|D}$:

$$\tau_{CB|D} = Tr_E\left[\tau_{B|DE}\rho_{CE}\right]. \tag{2.60}$$

Mathematical objects that are closely related to the causal maps and more general conditionals defined here have appeared before in the context of alternative formulations of quantum theory and have been studied by a number of authors, for example under the name of quantum combs [40], operator tensors [41], process matrices [38] and the $\mathcal{M}$-matrix [39]. However, the interpretation and implications in terms of causal models and inference have not received much consideration except in [1], on which the formalism used in this thesis builds.

Causal tomography is a method for characterizing a causal map:

Definition 32. The term *causal tomography* refers to an experimental procedure and subsequent analysis that yields a complete characterization of an unknown causal map.

This section details how to obtain an experimental characterization of an unknown causal map that is provided in the form of a black-box quantum circuit. The analysis builds on the fact that the space of operators on a Hilbert space, $\mathcal{L}(\mathcal{H})$, is a vector space itself. It is equipped with an inner product, the Hilbert-Schmidt inner product $\langle\rho,\tau\rangle \equiv Tr[\rho\tau]\ \forall\rho,\tau \in \mathcal{L}(\mathcal{H})$, and admits an orthonormal basis.

In essence, one can use informationally complete sets of preparations on the input (D) and measurements on the outputs (C and B) to evaluate the inner product of the unknown Jamiołkowski operator $\tau_{CB|D}$ with the elements of bases of the operator space $\mathcal{L}(\mathcal{H}_C)\otimes\mathcal{L}(\mathcal{H}_B)\otimes\mathcal{L}(\mathcal{H}_D)$, then use the respective dual bases to reconstruct $\tau_{CB|D}$ explicitly. Let the set of operators $\{P^s\}$, indexed by s, be a basis of the operator space $\mathcal{L}(\mathcal{H}_C)$, and similarly $\{Q^u\}$ for $\mathcal{L}(\mathcal{H}_B)$ and $\{R^t\}$ for $\mathcal{L}(\mathcal{H}_D)$. Let $\{\bar{P}^s\}$ denote the basis of $\mathcal{L}(\mathcal{H}_C)$ that is dual to $\{P^s\}$, that is, such that $Tr\left[P^s\bar{P}^{s'}\right] = \delta(s,s')$, and similarly let $\{\bar{Q}^u\}$ denote the dual of $\{Q^u\}$ and $\{\bar{R}^t\}$ the dual of $\{R^t\}$. In terms of these operators, one can define the coefficients

$$\mathcal{C}_{stu} = Tr\left[\tau_{CB|D}P_C^s \otimes Q_B^u \otimes R_D^t\right], \tag{2.61}$$

and explicitly reconstruct

$$\tau_{CB|D} = \sum_{stu} \mathcal{C}_{stu} \bar{P}^s_C \otimes \bar{Q}^u_B \otimes \bar{R}^t_D. \tag{2.62}$$

In the following, we illustrate the abstract paradigm with a concrete example for qubits. In this case, the extended set of Pauli operators,

$$\sigma^0 \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \sigma^1 \equiv \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 \equiv \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \quad \sigma^3 \equiv \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{2.63}$$

forms an orthonormal, self-dual[11] basis of the space of linear operators. Therefore, if we evaluate the inner products

$$\mathcal{C}_{stu} \equiv Tr\left[\tau_{CB|D} \sigma^s_C \otimes \sigma^u_B \otimes \sigma^t_D\right] \tag{2.64}$$

for all combinations of $s, t, u \in \{0, 1, 2, 3\}$, then we simply have

$$\tau_{CB|D} = \frac{1}{8} \sum_{stu=0}^{3} \mathcal{C}_{stu} \sigma^s_C \otimes \sigma^u_B \otimes \sigma^t_D. \tag{2.65}$$

Note that the Pauli operators with $j \neq 0$ are traceless, and therefore not valid quantum states that can be prepared by themselves, nor are they positive-semidefinite, which implies that they cannot be elements of a POVM. However, one can use pure-state preparations of individual eigenstates of the non-trivial Pauli operators and projective measurements that distinguish these states. To this end, let $s', t', u' \in \{1, 2, 3\}$ denote the restriction of s, t, u to the indices corresponding to non-trivial Pauli observables and let the projector $\Pi^{s'c}$ denote the eigenstate of $\sigma^{s'}$ with eigenvalue[12] c, and similarly $\Pi^{u'b}$ and $\Pi^{t'd}$, with $c, b, d \in \{\pm 1\}$. This allows us to evaluate

$$\mathrm{Tr}_{CBD}\left[\tau_{CB|D}(\Pi^{s'c}_C \otimes \Pi^{u'b}_B \otimes \Pi^{t'd}_D)\right] = P\left(cb|ds't'u'\right) \tag{2.66}$$

by finding the probability of obtaining the eigenvalues c and b in measurements of the observables $\sigma^{s'}_C \otimes \sigma^{u'}_B$, assuming that the input on D was an eigenstate of $\sigma^{t'}$ with eigenvalue d. The experiment is shown schematically in Fig. 2.5. In order to obtain the inner product with the full Pauli operators, we note that multiplying each projector $\Pi^{s'c}$ by the associated eigenvalue c and adding them gives $\sum_c c\Pi^{s'c} = \sigma^{s'}$ for $s' \neq 0$. It follows that we can obtain the correlators for $s', t', u' \in \{1, 2, 3\}$ by taking

$$\sum_{c,b,d=\pm 1} cbd\ P\left(cb|ds't'u'\right) = \mathrm{Tr}\left[\tau_{CB|D} \sigma^{s'}_C \otimes \sigma^{u'}_B \otimes \sigma^{t'}_D\right] = \mathcal{C}_{s't'u'}. \tag{2.67}$$

We now consider how to obtain the $\mathcal{C}_{stu}$ when one or more indices are zero. Note that if we add the projectors without multiplying by the respective eigenvalues, we obtain $\sum_c \Pi^{s'c} = \sigma^0$ for any choice of $s' \in \{1, 2, 3\}$. It follows that the marginals of $P\left(cb|ds't'u'\right)$ obtained by tracing out one of the eigenvalues are independent of the choice of the corresponding observable:

$$\begin{cases} \sum_d P\left(cb|ds't'u'\right) = 2P\left(cb|s'u'\right) & \forall t' \\ \sum_b P\left(cb|ds't'u'\right) = P\left(c|ds't'\right) & \forall u' \\ \sum_c P\left(cb|ds't'u'\right) = P\left(b|dt'u'\right) & \forall s' \end{cases} \quad \text{and} \quad \begin{cases} \sum_{bd} P\left(cb|ds't'u'\right) = 2P\left(c|s'\right) & \forall u't' \\ \sum_{cb} P\left(cb|ds't'u'\right) = 1 & \forall s'u't' \\ \sum_{cd} P\left(cb|ds't'u'\right) = 2P\left(b|u'\right) & \forall s't'. \end{cases} \tag{2.68}$$

[11] Up to normalization, since $Tr\left[\sigma^j \sigma^j\right] = 2$.

[12] Note that here lower-case c denotes the value of a classical variable, namely an eigenvalue, that is obtained by a measurement on a quantum variables denoted by capital C, and similarly the eigenvalue d indexes preparations of the quantum variable D. In most cases we will use different upper-case letters for a quantum and a classical variable, reserving a lower-case version of the latter for the *value* that the classical variable takes. However, in the present section (and several others that build on it), there is little need to distinguish between the classical variable and the value it takes, which allows us to use the same letter to label the associated quantum variable. We hope that, by labelling the eigenvalues in such an evocative manner, it will become easier to keep track of which quantum system they pertain to.

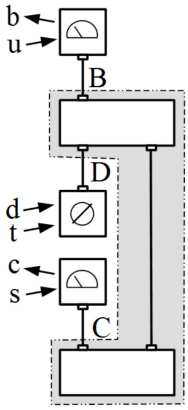

Figure 2.5: Causal tomography provides a complete characterization of the causal relation between two causally-ordered quantum variables, A and B. It requires tomographically complete sets of measurements on C and B, indexed by the settings (choice of observable) s and u and producing outcomes (eigenvalues) c and b, respectively, and a complete set of preparations on D, indexed by the choice of observable, t, and eigenvalue, d.

Once one has computed these marginals from the observed distributions, one can directly read off

$$\begin{cases} \mathcal{C}_{s'0u'} = 2\sum_{cb} cbP\left(cb|s'u'\right) \\ \mathcal{C}_{s't'0} = \sum_{cd} cdP\left(c|ds't'\right) \\ \mathcal{C}_{0t'u'} = \sum_{bd} bdP\left(b|dt'u'\right) \end{cases} \quad \text{and} \quad \begin{cases} \mathcal{C}_{s'00} = 2\sum_{c} cP\left(c|s'\right) \\ \mathcal{C}_{0t'0} = \sum_{d} d = 0 \\ \mathcal{C}_{00u'} = 2\sum_{b} bP\left(b|u'\right) \end{cases} \tag{2.69}$$

Note that, since C precedes D in time, it cannot depend on D. Hence $P\left(c|ds't'\right) = P\left(c|s'\right)$ for all d and t', and consequently $\mathcal{C}_{s'0t'} = 0$. By a similar reasoning, one can also deduce that the final coefficient is $\mathcal{C}_{000} = \sum_d 1 = 2$. This completes the set of required inner products and allows one to reconstruct $\tau_{CB|D}$ explicitly using eq. (2.65).

We note that this algebraic reconstruction is mainly intended to prove that experimental data of the form $\mathcal{C}_{stu}$ contains sufficient information for a reconstruction of $\tau_{CB|D}$, and for this reason it is derived under the assumption of an ideal experiment. In a more realistic setting, the measured values of the $\mathcal{C}_{stu}$ are generally subject to noise, and consequently a naïve linear inversion as described here would likely yield unphysical results (such as a causal map $\mathcal{E}_{CB|D}$ that is not completely positive or violates the condition $Tr_B\mathcal{E}_{CB|D} = \rho_C \otimes Tr_D$). Under these circumstances, one can use a more robust least-squares fit to find the causal map that best fits the data. This technique is detailed in section 5.4, which describes an experiment involving causal tomography.

In a similar vein, note that the choice of using projective measurements on C and B and preparations of pure states on D only served to simplify the mathematical derivation. In principle, one can use generic POVM measurement on B and instruments on A (i.e. maps from C to D) or other experiments, such as joint measurements on C and B, to probe $\tau_{CB|D}$. The only requirement for ensuring a complete characterization of the causal map is that the chosen operators form a spanning set of $\mathcal{L}\left(\mathcal{H}_C \otimes \mathcal{H}_B \otimes \mathcal{H}_D\right)$.

Finally, we note that, in the limiting cases of purely cause-effect or purely common-cause relations, causal tomography reduces to conventional tomography of channel and bipartite states, respectively. In these cases, the causal map factorizes as shown in eqs. (2.56) and (2.57), respectively. Appendix A.1 details how, in these two scenarios, causal tomography yields a characterization of the channel $\mathcal{E}_{B|D}$ and the state ρ_{CB}, respectively.

3 Probing quantum variables

A central application of causal models, and in fact a core task of science in general, is to make predictions or retrodictions in a network of related variables. To this end, one must first acquire information about the subset of variables on which one wishes to condition; that is, one must probe them using some interaction. In a classical world, there is little subtlety to this process: it is generally possible to simply observe the value of a variable. Moreover, this value constitutes a complete description of the variable, in the sense that any previous or subsequent observation (until its state is changed by some transformation) must yield the same information: the variable is found to take the same value. Classical measurements do not, by default, disturb the variable, in the sense that the variable both before and after the interaction has the same value, which is merely revealed by observation.

A quantum variable, by contrast, is not completely described by a single real value, because there exists a set of observables, each of which produces one value if evaluated, yet which cannot be evaluated jointly. Measuring one observable can alter the probability distribution over outcomes in a subsequent measurement of an incompatible observable. This operational fact is often loosely described by saying that the act of measuring somehow disturbs the quantum system, although the hypothesis that a quantum system even has underlying states (in the sense of specifications of the outcomes of all possible measurements) leads to a host of other issues.

This section analyses different ways of probing quantum variables and investigates how they differ from the straightforward passive observation of classical variables, from the point of view of causal models. Specifically, we are interested in the implications of different probing schemes for a paradigmatic problem in causal discovery: a variable A is related to a later variable B either as cause and effect or via a common cause, and the task is to determine which relation is realized. In the case of classical variables, passive observation of A is insufficient to answer this question, because it does not reveal independent information about the pre- and post-intervention versions of A. By this standard, under what conditions does the act of probing a quantum variable become like classical passive observation?

A likely candidate for a quantum probing scheme that resembles classical passive observation are weak measurements, which we analyse in section 3.1. However, a discussion of what makes a probing scheme resemble passive observation for the purpose of causal discovery (section 3.2) reveals that weak measurements do not satisfy this condition. Instead, we provide a different example of a quantum probing scheme that does have the desired property. Implications of this probing scheme are discussed in the remainder of this chapter. Chapter 4 proceeds by exploring a typically quantum effect in causal models that persist despite the probing scheme being "effectively classical", in the sense introduced here.

Characterizing the (causal) relation of a quantum variable A to other variables. Before we compare different ways of probing a quantum variable A, let us first give a formal description of how one gains information about A and what constitutes a complete characterization of its connections to a larger causal network.

We will continue to draw on the example of a simple causal model wherein A is related only to a single, later variable, B, by the causal structure and circuit depicted in Fig. 2.3. (The generalization to other causal structures is outlined at the end of this section.) As discussed in sections 2.5.1, in this scenario one must split A into C and D, as defined in proposition 26. The relation between A and B is then described by a causal map $\mathcal{E}_{CB|D}$, or equivalently the quantum conditional $\tau_{CB|D} \in \mathcal{L}\left(\mathcal{H}_C \otimes \mathcal{H}_B \otimes \mathcal{H}_D\right)$. This is essentially the object we wish to characterize.

Following section 2.5.3, this can be achieved by using a set of measurements on B and a set of instruments from C to D to probe the causal relations. As per definition 21, each instrument used to probe A is represented by a family of maps from C to D, $\left\{\mathcal{E}^{sm}_{D|C}\right\}_m$, where we introduce the

notation S and M for the classical variables encoding the setting and outcome, respectively. The instrument is equivalently specified by the corresponding Jamiołkowski operators $\left\{\tau^{sm}_{D|C}\right\}$, which have the properties

$$T_C \tau^{sm}_{D|C} \geq 0 \quad Tr_D \left(\sum_m \tau^{sm}_{D|C} \right) = \mathbb{I}_C \ \forall m. \tag{3.1}$$

Similarly, B can be probed by a POVM, whose outcomes we will denote by Y, represented by a set of positive operators $\{\tau^y_B\}$. The statistics generated by these experiments can be written in terms of the Jamiołkowski operators as

$$P\left(ym|s\right) = Tr\left[\left(\tau^{sm}_{D|C} \otimes \tau^y_B\right) \tau_{CB|D}\right], \tag{3.2}$$

which allows us to evaluate the Hilbert-Schmidt inner product of the unknown $\tau_{CB|D}$ with a family of operators indexed by S, M and Y. Assuming that the $\{\tau^y_B\}$ span $\mathcal{L}\left(\mathcal{H}_B\right)$ and that the $\left\{\tau^{sm}_{D|C}\right\}$ span $\mathcal{L}\left(\mathcal{H}_C \otimes \mathcal{H}_D\right)$, this allows us to reconstruct $\tau_{CB|D}$.

However, in the present chapter, we focus only on what one learns about the two versions of A depending on the available instruments $\left\{\tau^{sm}_{D|C}\right\}$. We will assume implicitly that we have access to a suitable set $\{\tau^y_B\}$ of POVM elements on B and consider only the problem of characterizing the resulting effective operators on CD,

$$\tilde{\tau}^y_{C|D} \equiv Tr_B\left(\tau^y_B \tau_{CB|D}\right). \tag{3.3}$$

The notation $\tilde{\tau}^y_{C|D}$ is chosen so as to highlight the fact that these operators are only *components* of the full causal map: they are elements of the conditional that one obtains by combining the conditional $\tau_{Y|B} = \sum_y |y\rangle\langle y| \otimes \tau^y_B$ (which represents a measurement on B) with $\tau_{CB|D}$, which represents the causal map:

$$\tau_{YC|D} = Tr_B\left(\tau_{Y|B}\tau_{CB|D}\right) = \sum_y |y\rangle\langle y| \otimes \tilde{\tau}^y_{C|D}. \tag{3.4}$$

We stress that, although the notations $\tau^{sm}_{D|C}$ and $\tilde{\tau}^y_{C|D}$ are somewhat similar, they represent different physical objects, which are contrasted in Fig. 3.1. The operators $\left\{\tau^{sm}_{D|C}\right\}$ represent instruments that an agent uses to probe the variable A; in other words, an additional circuit element that the agent plugs into a given circuit. The $\left\{\tau^{sm}_{D|C}\right\}$ take C as an input and return D as an output and encode a causal influence of C on D. The operators $\tilde{\tau}^y_{C|D}$, on the other hand, summarize the relations between C and D that are realized by the causal map $\mathcal{E}_{CB|D}$ under post-selection on the POVM elements τ^y_B. They correspond to maps with input D and output C. Considering that D lies in the causal future of C, we note that these maps encode only inferences, not causal influences.

If we can characterize every $\tilde{\tau}^y_{C|D}$ for a set of τ^y_B that span $\mathcal{L}\left(\mathcal{H}_B\right)$, then one can reconstruct the full causal map. To this end, note that the experimental statistics can be rewritten as the inner product of the component $\tilde{\tau}^y_{C|D}$ of the causal map with the instrument used to probe A (in the Jamiołkowski representation), $\left\{\tau^{sm}_{D|C}\right\}_m$:

$$P\left(ym|s\right) = Tr\left[\tau^{sm}_{D|C}\tilde{\tau}^y_{C|D}\right]. \tag{3.5}$$

We will study how different types of instruments impact one's ability to characterize a generic component $\tilde{\tau}^y_{C|D}$.

In general, the variable of interest, A, may be causally connected to more than just one other variable. In this case, rather than just measuring B, one may have to range over preparations and

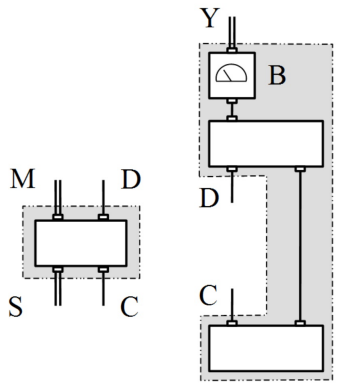

Figure 3.1: Operators on C and D can represent two different types of objects, which we illustrate with the circuit elements implementing each type: (left) An instrument $\left\{\tau^{sm}_{D|C}\right\}_m$ (specified by a setting S and yielding an outcome M) can be plugged into the principal circuit in order to probe it. It encodes a causal influence of C on D. (right) Given a causal map specified by a Jamiołkowski operator $\tau_{CB|D}$, if one post-selects on finding a POVM element τ^y_B in a measurement on B, one induces a relation between C and D that supports inferences encoded in the (unnormalized) operator $\tilde{\tau}^y_{C|D} = Tr_B\left(\tau^y_B \tau_{CB|D}\right)$. This chapter considers what one can learn about $\tilde{\tau}^y_{C|D}$ using different types of instruments $\left\{\tau^m_{D|C}\right\}_m$.

measurements on a number of other variables in order to characterize the causal relations in the network completely. However, the object of interest for the present discussion continues to be the generic component $\tilde{\tau}^y_{C|D} \in \mathcal{L}\left(\mathcal{H}_C \otimes \mathcal{H}_D\right)$ that one obtains by tracing the conditional describing the global causal relations with generic preparations and measurements on all variables aside from A.

3.1 Weak measurements

The fact that one cannot measure a quantum system without affecting its subsequent behaviour in some way is especially troublesome if one's goal is to characterize the dynamics of a system over a period of time. The reason is that measuring the system at any one time will significantly alter its subsequent states, so that the results of a second measurement at a later time may not necessarily reflect the natural dynamics of the system itself, but rather the side-effects of the earlier probing. This conundrum led to the proposal of weak measurements, which minimize the disturbance due to any single measurement[13] by *weakly* coupling the principal system that one wishes to probe, A, to an ancilla, N, and subsequently measuring the ancilla. This section explores what one learns about A, that is, about the two versions of A, C and D, in such a scenario.

For concreteness, we consider the following standard way of realizing a weak measurement, illustrated in Fig. 3.2. The choice of measurement (setting) is encoded in the classical variable S, and the outcome is a classical variable M, whereas the ancilla to which the principal system is coupled is a quantum system N. The dimension of the ancilla, h_N, is taken to be the same as that of the principal system, h_A. Let $\{|m\rangle\}_{1,\ldots h_A}$ denote some fixed basis of $\mathcal{H}_N$ and take $|\psi_0\rangle \equiv \sqrt{\frac{1}{h_A}} \sum_m |m\rangle$ to be the initial state of N. Next we specify the coupling between the system and the ancilla. The setting S specifies a set of orthogonal rank-one projectors Π^{sa} on A that the measurement aims to distinguish, which correspond to an orthonormal basis $\{|a_s\rangle\}_{a=1,\ldots h_A}$ by $\Pi^{sa} = |a_s\rangle\langle a_s|$. The

[13] However, there may still be cumulative effects due to a series of measurements, or, as we will see, due to post-selection.

coupling between A and the ancilla N takes the form

$$U^s_{NA} = \sum_a \bar{U}^{m=a}_N \otimes \Pi^{sa}_A, \tag{3.6}$$

associating a unitary $\bar{U}^m_N$ on the ancilla to each projector Π^{sa} on A. For each m, the transformation $\bar{U}^m_N$ rotates the initial state of the ancilla, $|\psi_0\rangle$, slightly towards $|m\rangle$, so that the subsequent measurement in the $|m\rangle$ basis becomes biased towards that outcome. We formalize this as follows: let ϵ denote the strength of the coupling and define

$$f(\epsilon) = \sqrt{\frac{1}{h_A} - \frac{h_A - 1}{h_A^2}\epsilon^2} - \frac{\epsilon}{h_A} \quad \Leftrightarrow \quad h_A f^2(\epsilon) + 2f(\epsilon)\epsilon + \epsilon^2 = 1, \tag{3.7}$$

which ensures that the state vector $|\phi^m_0\rangle \equiv f(\epsilon)\sqrt{h_A}\,|\psi_0\rangle + \epsilon\,|m\rangle$ is normalized (noting that $\langle m|\psi_0\rangle \neq 0$). Let $\{|\psi_j\rangle\}$ and $\{|\phi^m_j\rangle\}$, indexed by $j = 0, ...h_A - 1$, be two orthonormal bases of $\mathcal{H}_N$ that include the $|\psi_0\rangle$ and $|\phi^m_0\rangle$ defined above. In terms of these parameters, we take

$$\bar{U}^m_N \equiv \sum_{j=0}^{h_A-1} |\phi^m_j\rangle\langle\psi_j| = \left[f(\epsilon)\sqrt{h_A}\,|\psi_0\rangle + \epsilon\,|m\rangle\right]\langle\psi_0| + ... \tag{3.8}$$

After the coupling, the quantum system N is subjected to a projective measurement that distinguishes the basis elements $\{|m\rangle\}$, reducing it to the classical variable M, whose value m reflects which state was found in the measurement.

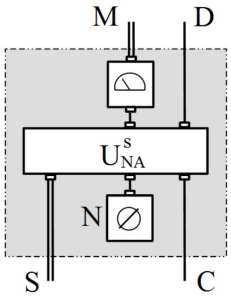

Figure 3.2: Standard realization of a weak measurement. The weak measurement apparatus (dashed box) takes as input a quantum variable C and the classical setting S, and outputs the quantum variable D and the classical outcome M. Inside the box, an ancilla N is initialized in some fixed state $|\psi\rangle$ and subsequently coupled to the principal system, A. The coupling is weak, in the sense that it is ϵ-close to being a product of identity operators on A and N. The setting S controls the choice of unitary, and, by extension, which measurement is realized on A. After the coupling, the ancilla N is measured on a preferred basis, yielding the classical outcome M.

Description in conventional quantum mechanics. We wish to derive concise expressions for what we can deduce about C and D from such a measurement. To this end, let us first analyse the situation in the conventional framework, whose elements may be more intuitive to the reader. Let $P(m|s)$ denote the probability of obtaining an outcome m given a setting s. For each pair s, m, one can assign a quantum state to D, which we denote ρ^{sm}_D. Both $P(m|s)$ and ρ^{sm}_D can be obtained from a single expression that gives the *unnormalized* state on D after finding a particular outcome

m in terms of the initial state ρ_C of the principal system:

$$\begin{aligned} P(m|s)\,\rho_D^{sm} &= Tr_N\left[|m\rangle\langle m|_N U^s\left(|\psi_0\rangle\langle\psi_0|_N \otimes \rho_C\right) U^{s\dagger}\right] \\ &= \langle m|_N U^s|\psi_0\rangle_N \rho_C \langle\psi_0|_N U^{s\dagger}|m\rangle_N. \\ &\equiv K^{sm}\rho_C K^{sm\dagger} \end{aligned} \tag{3.9}$$

Let us introduce a shorthand for the operators

$$K^{sm} \equiv \langle m|_N U^s|\psi_0\rangle_N = f(\epsilon)\sum_a \Pi^{sa} + \epsilon\Pi^{sm} = f(\epsilon)\,\mathbb{I} + \epsilon\Pi^{sm} \tag{3.10}$$

associated with a choice of basis s and an outcome m in the weak measurement. In this notation,

$$P(m|s)\,\rho_D^{sm} = K^{sm}\rho_C K^{sm\dagger}, \tag{3.11}$$

from which we can isolate the normalized state on D given m,

$$\rho_D^{sm} = \frac{1}{P(m|s)} K^{sm}\rho_C K^{sm\dagger} \tag{3.12}$$

and the probability of obtaining m,

$$P(m|s) = Tr\left[K^{sm}\rho_C K^{sm\dagger}\right] = Tr\left[\rho_C K^{sm\dagger}K^{sm}\right]. \tag{3.13}$$

This last expression indicates that, if we ignore the state of the principal system after the measurement, D, then the weak measurement specified by s effectively realizes a POVM on C. Its elements, which we denote τ_C^{sm}, are given by

$$\tau_C^{sm} = K^{sm\dagger}K^{sm}, \tag{3.14}$$

indexed by the outcomes $m = 1, ...h_A$. If we discard instead the outcome m, we are effectively implementing a noisy quantum channel from C to D, which we represent by the map $\mathcal{E}^s_{D|C}$. Its output, given a generic input ρ_C, is given by

$$\mathcal{E}^s_{D|C}(\rho_C) = \sum_m P(m|s)\,\rho_D^{sm} = \sum_m K^{sm}\rho_C K^{sm\dagger}. \tag{3.15}$$

This form shows that the $\{K^{sm}\}_{m=1,...h_A}$ are Kraus operators representing the channel $\mathcal{E}^s_{D|C}$.

Description in the framework of quantum conditionals. Although the results obtained in the previous paragraph are useful descriptions of the measurement, the formal treatment of versions C and D is rather different, and in particular it is not clear what we can retrodict about C. For this reason, we will now recast the analysis in the framework of quantum conditionals, which will allow us to derive not only predictions about D, but also retrodictions about C.

The weak measurement, which describes two quantum variables related as cause and effect along with associated classical variables, can be represented by an overall conditional

$$\tau_{MD|SC} = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \tau^{sm}_{D|C}, \tag{3.16}$$

where, for each setting s, the $\left\{\tau^{sm}_{D|C}\right\}_m$ define a quantum instrument, as per definition 21. In terms of the Kraus operators introduced above, one can show that the Jamiołkowski operators $\tau^{sm}_{D|C}$ take the form

$$\tau^{sm}_{D|C} = \sum_{i,j=1}^{h_A} K^{sm}|i\rangle\langle j|_D\, K^{sm\dagger} \otimes |j\rangle\langle i|_C\,, \tag{3.17}$$

where $|i\rangle$ and $|j\rangle$ denote elements of an arbitrary[14] basis of $\mathcal{H}_A$ (recalling that $\mathcal{H}_C$ and $\mathcal{H}_D$ are both isomorphic to $\mathcal{H}_A$). In other words, each $\tau^{sm}_{D|C}$ is simply isomorphic to the map

$$\mathcal{E}^{sm}_{D|C}(\rho_C) \equiv K^{sm}\rho_C K^{sm\dagger}. \tag{3.18}$$

From the overall $\tau_{MD|SC}$, one can derive what one learns about C and D given the measurement outcome. Given an initial state ρ_C, belief propagation yields a conditional of the form

$$\tau_{MD|S} = Tr_C\left[\tau_{MD|SC}\rho_C\right] = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \left[K^{sm}\rho_C K^{sm\dagger}\right]_D. \tag{3.19}$$

Tracing over the resulting quantum variable D reduces this to

$$\tau_{M|S} = Tr_D\tau_{MD|S} = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes Tr\left[K^{sm}\rho_C K^{sm\dagger}\right], \tag{3.20}$$

which simply encodes the classical probability distribution

$$P(m|s) = Tr\left[K^{sm}\rho_C K^{sm\dagger}\right]. \tag{3.21}$$

On the other hand, conditioning on M gives

$$\tau_{D|MS} = \sum_{ms} \frac{1}{P(m|s)} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \left[K^{sm}\rho_C K^{sm\dagger}\right]_D = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \rho^{sm}_D, \tag{3.22}$$

which represents the preparation of an ensemble of (normalized) quantum states,

$$\rho^{sm}_D = \frac{1}{P(m|s)}\left[K^{sm}\rho_C K^{sm\dagger}\right]_D. \tag{3.23}$$

With regards to D, the quantum conditionals framework reproduces the results of the conventional analysis.

In order to derive a retrodiction about C, let us instead trace out D from $\tau_{MD|SC}$, which yields a conditional of the form

$$\tau_{M|SC} = Tr_D\tau_{MD|SC} = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \tau^{sm}_C, \tag{3.24}$$

which describes a POVM on C whose elements are

$$\tau^{sm}_C = Tr_D\tau^{sm}_{D|C} = K^{sm\dagger}K^{sm}. \tag{3.25}$$

In order to derive inferences about C, we use Bayesian inversion: given a prior ρ_C,

$$\tau_{C|SM} = \frac{1}{P(M|S)}\rho_C^{\frac{1}{2}}\tau_{M|SC}\rho_C^{\frac{1}{2}} = \sum_{ms} \rho^{sm}_C \otimes |s\rangle\langle s| \otimes |m\rangle\langle m|, \tag{3.26}$$

hence our retrodictions about C take the form

$$\rho^{sm}_C = \frac{1}{P(m|s)}\rho_C^{\frac{1}{2}}K^{sm\dagger}K^{sm}\rho_C^{\frac{1}{2}}. \tag{3.27}$$

[14]Recall from the discussion of eq. (2.16) that the above expression remains unchanged if one changes the choice of basis.

What one can learn by a weak measurement. Substituting (3.10) into (3.21), (3.23) and (3.27), one can obtain expressions for ρ_D^{sm}, ρ_C^{sm} and $P(m|s)$ that are exact for all ϵ. In the limit of weak measurements ($\epsilon \ll 1$), the expressions for what one can infer simplify to[15]

$$\begin{cases} \rho_D^{sm} = \frac{1}{P(m|s)} \left[f^2(\epsilon) \rho_C + \epsilon f(\epsilon) (\Pi^{sm} \rho_C + \rho_C \Pi^{sm}) + \mathcal{O}(\epsilon^2) \right] \\ \rho_C^{sm} = \frac{1}{P(m|s)} \left[f^2(\epsilon) \rho_C + 2\epsilon f(\epsilon) \rho_C^{\frac{1}{2}} \Pi^{sm} \rho_C^{\frac{1}{2}} + \mathcal{O}(\epsilon^2) \right] \\ P(m|s) = f^2(\epsilon) + 2\epsilon f(\epsilon) Tr[\Pi^{sm} \rho_C] + \mathcal{O}(\epsilon^2). \end{cases} \tag{3.28}$$

One can see that the weak measurement does indeed provide information about C, by noting that the retrodictive state ρ_C^{sm} contains a non-trivial dependence on the measurement outcome m. The post-intervention state ρ_D^{sm} also depends on the indices specifying the measurement, s and m; this is conventionally interpreted as a disturbance of the system due to the measurement. Note that ρ_D^{sm} depends on s and m at the same order in the weak parameter ϵ as ρ_C^{sm}. This implies that, regardless of how weak one makes the measurement, one cannot effectively suppress the disturbance, because it is always of the same order as the achieved information gain. However, if one's goal is to characterize the causal relations of A to both its parents and its children, then one must vary over operators on both $\mathcal{H}_C$ and $\mathcal{H}_D$. The sm-dependence of ρ_D^{sm} is therefore no longer a disturbance to be minimized, but rather an indispensable resource.

Indeed, the form of ρ_D^{sm} derived above allows one to range over a spanning set of operators on $\mathcal{H}_D$. To see this, recall that the projectors $\{\Pi^{sm}\}$ for each s are defined by an orthonormal basis of $\mathcal{H}_A$, hence, if one ranges over suitable combinations of s and m, they span $\mathcal{L}(\mathcal{H}_A)$. On the other hand, D, being the post-intervention version of A, has by proposition 26 the same Hilbert space dimension as A, therefore $\mathcal{L}(\mathcal{H}_D)$ is isomorphic to $\mathcal{L}(\mathcal{H}_A)$. Finally, since ρ_D^{sm} is linear in Π^{sm}, it follows that the $\{\rho_D^{sm}\}$ form a spanning set of the space of operators on D, and by similar arguments the $\{\rho_C^{sm}\}$ span $\mathcal{L}(\mathcal{H}_C)$.

In order to obtain a *complete* characterization of how A relates to other variables, one must range over spanning sets of operators on $\mathcal{H}_C$ and $\mathcal{H}_D$ *independently*. However, in the weak measurement scenario, the information gained about C and D is not independent, but conditioned on the same two indices, s and m. Consequently, what one learns about C and D from a weak measurement is not generally a product of the ρ_C^{sm} and ρ_D^{sm} derived above. Instead, one can use Bayesian inversion on eq. (3.16) to find

$$\tau_{CD|SM} = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \rho_{CD}^{sm}, \tag{3.29}$$

where ρ_{CD}^{sm} are joint states of belief about C and D conditional on s and m, given by

$$\rho_{DC}^{sm} = \rho_C^{\frac{1}{2}} \tau_{D|C}^{sm} \rho_C^{\frac{1}{2}}. \tag{3.30}$$

Substituting expression (3.17) for $\tau_{D|C}^{sm}$, one can verify that the marginals of ρ_{DC}^{sm} reproduce the expressions for ρ_D^{sm} and ρ_C^{sm} derived above, (3.23) and (3.27).

However, for the purpose of characterizing the causal relation between variable A and other quantum variables, the Jamiołkowski operator $\tau_{D|C}^{sm}$ is actually more convenient, because, as detailed at the beginning of this chapter, its inner product with the component $\tilde{\tau}_{C|D}^{y}$ can be read off directly from the experimental statistics,

$$Tr\left[\tau_{D|C}^{sm} \tilde{\tau}_{C|D}^{y}\right] = P(ym|s). \tag{3.31}$$

[15] The normalization of the probability distribution $P(M|S)$ can be verified for the general expression, $P(m|s) = f^2(\epsilon) + 2\epsilon f(\epsilon) Tr[\Pi^{sm}\rho_C] + \epsilon^2 Tr[\Pi^{sm}\rho_C]$. Summing over m gives $\sum_m 1 = d_A$ and $\sum_m Tr[\Pi^{sm}\rho_C] = Tr\rho_C = 1$, hence $\sum_m P(m|s) = df^2(\epsilon) + 2\epsilon f(\epsilon) + \epsilon^2$. Referring back to eq. (3.7), one can see that $f(\epsilon)$ was defined so as to ensure that this is unity.

The operator $\tau_{D|C}^{sm}$ is given by expression (3.17); substituting (3.10) for the Kraus operators and subsequently taking the weak limit, $\epsilon \ll 1$, we have

$$\begin{aligned}\tau_{D|C}^{sm} &= f^2(\epsilon)\left(\sum_{ij}|i\rangle\langle j|_D \otimes |j\rangle\langle i|_C\right) \\ &+\epsilon f(\epsilon)\left(\sum_{j}|j_s\rangle\langle m_s|_D \otimes |m_s\rangle\langle j_s|_C\right) \\ &+\epsilon f(\epsilon)\left(\sum_{j}|m_s\rangle\langle j_s|_D \otimes |j_s\rangle\langle m_s|_C\right) + \mathcal{O}\left(\epsilon^2\right),\end{aligned} \tag{3.32}$$

where $|j_s\rangle$ and $|m_s\rangle$ denote elements of the orthonormal basis on A specified by the setting s, so that $|m_s\rangle\langle m_s| = \Pi^{sm}$ and consequently $\Pi^{sm}|j_s\rangle = \delta_{j,m}|j_s\rangle$. (Note that the first term in the above expression is written in terms of a basis that is independent of s, since in fact this term takes the same form in terms of any orthonormal basis of $\mathcal{H}_A$.)

The first term of $\tau_{D|C}^{sm}$, of order zero in ϵ, is the conditional representing the identity channel from C to D, that is, the trivial measurement. When used in expression (3.31), it yields the statistics one would expect to see if one did not probe A at all,

$$Tr\left[\left(\sum_{ij}|i\rangle\langle j|_D \otimes |j\rangle\langle i|_C\right)\tilde{\tau}_{C|D}^{y}\right] = P_0(ym|s) = P_0(y)\ \forall m, s. \tag{3.33}$$

The information extracted by a weak measurement is based on the first-order terms, which contribute to the statistics $P(ym|s)$ an expression proportional to the inner product

$$Tr\left[\left(\sum_{j}|j_s\rangle\langle m_s|_D \otimes |m_s\rangle\langle j_s|_C\right)\tilde{\tau}_{C|D}^{y}\right] + Tr\left[\left(\sum_{j}|m_s\rangle\langle j_s|_D \otimes |j_s\rangle\langle m_s|_C\right)\tilde{\tau}_{C|D}^{y}\right]. \tag{3.34}$$

In order to extract these terms from the overall $P(ym|s)$, one adds up experimental data obtained for different m, which yields

$$\begin{aligned}\textstyle\sum_m P(ym|s) &= \left(f^2(\epsilon) + 2\epsilon f\right) Tr\left[\left(\sum_{ij}|i\rangle\langle j|_D \otimes |j\rangle\langle i|_C\right)\tilde{\tau}_{C|D}^{y}\right] + \mathcal{O}\left(\epsilon^2\right) \\ &= \left(f^2(\epsilon) + 2\epsilon f\right) P_0(ym|s)\end{aligned} \tag{3.35}$$

and subtracts it in order to isolate the terms of interest,

$$\begin{aligned}P(ym|s) - \tfrac{f^2(\epsilon)}{f^2(\epsilon)+2\epsilon f(\epsilon)}\textstyle\sum_{m'} P(ym'|s) &= \epsilon f(\epsilon)\, Tr\left[\left(\sum_{j}|j_s\rangle\langle m_s|_D \otimes |m_s\rangle\langle j_s|_C\right)\tilde{\tau}_{C|D}^{y}\right] \\ &+\epsilon f(\epsilon)\, Tr\left[\left(\sum_{j}|m_s\rangle\langle j_s|_D \otimes |j_s\rangle\langle m_s|_C\right)\tilde{\tau}_{C|D}^{y}\right] + \mathcal{O}\left(\epsilon^2\right).\end{aligned} \tag{3.36}$$

The physical interpretation of these isolated terms is not straightforward – indeed, the operator in parentheses in each term, taken by itself, need not even be Hermitian. However, we note that both the complete expression for $\tau_{D|C}^{sm}$ and the terms that one isolates in the weak measurement

paradigm are symmetric under an exchange of $C \leftrightarrow D$. Consequently, inner products of the form $Tr\left[\tau^{sm}_{D|C}\tilde{\tau}^{y}_{C|D}\right]$ do not allow one to evaluate any components of $\tilde{\tau}^{y}_{C|D}$ that are anti-symmetric under this exchange, since the inner product with these components must vanish. It is therefore impossible to obtain a complete characterization of $\tilde{\tau}^{y}_{C|D}$, and, by extension, of the causal relations between A and the surrounding variables.

In summary, we stress that weak measurements on a quantum system do not provide a faithful quantum generalization of passive observation in the sense of causing no disturbance at all. One can minimize the disturbance in the post-intervention state ρ^{sm}_{D} by reducing the coupling strength ϵ, but this simultaneously reduces the information gain in ρ^{sm}_{C} by the same factor. However, for the purpose of characterizing the causal relations between A and other variables – both parents and children –, then the fact that D depends on s and m is in fact a necessary resource. A weak measurement scheme provides some sm-dependence in ρ^{sm}_{D}, but not enough to allow for a complete characterization of the component $\tilde{\tau}^{y}_{C|D}$, which encodes the relation between A, its parents and children. One can characterize at most the part that is symmetric under the exchange $C \leftrightarrow D$, and an exact specification of the components that can be characterized is not straightforward, given the inconvenient form of the instruments $\left\{\tau^{sm}_{D|C}\right\}_{m}$ that one can implement.

3.2 Observational schemes for classical and quantum variables

We have noted in the previous section that measurements which disturb the system are actually advantageous for characterizing causal relations, because they provide some amount of independence between the information one obtains about C and D. In the context of causal models, the question of whether or not a probing scheme disturbs the variable under study is therefore secondary to the question of how useful the probing scheme is for the purpose of characterizing the causal relations. This section proposes a classification of probing schemes based on this second criterion. We will argue that passive observation is but one example of a class we term *observational* schemes (see definition 33 below) , and show that other schemes in this class do admit of a quantum version.

In order to allow for a complete characterization of the causal relations of a variable, a probing scheme must provide some degree of independence between the information acquired about the pre- and post-intervention versions of the variable. We argued this point in section 2.5.1, illustrating it with the example of a drug trial in classical statistics, wherein taking a new drug (variable A) is causally related to the treatment outcome (B) by a combination of an unknown common cause and a cause-effect relation. In this scenario, one can characterize the causal relation between treatment and recovery – and, in particular, answer the question of whether it is cause-effect or common-cause (or some combination of both) – by assigning each patient either the drug or a placebo perfectly at random. This ensures that the assigned treatment (variable D, the post-intervention version of A) is not influenced in any way by the patients' intent to treat (C, the pre-intervention version), and consequently by any potential common causes. However, perfect randomization is not necessary: any amount of statistical independence between C and D can be leveraged to characterize how each of them relates to B, and therefore the causal relation between A and B. It follows that, for the purpose of causal discovery, the following property is relevant:

Definition 33. A scheme for probing a variable A is termed *informationally symmetric* or *observational* if and only if the retrodictions it yields about the pre-intervention version of A are the same as the predictions about the post-intervention version of A, for any outcome the probing may produce. Denoting the two versions of A by C and D, respectively, and the setting and outcome of the instrument by the classical variables S and M, the condition takes the form

$$\tau_{C|SM} = \tau_{D|SM}. \tag{3.37}$$

One motivation for studying observational probing schemes for quantum variables is their significance for causal discovery, which we take up in more detail in section 4.1. However, we also note

that observational schemes imply that it is effectively not necessary to split the variable under study. In the classical case, variables are split simply because, depending on how one probes the variable, the interaction may change its value. In the case of quantum variables, on the other hand, the need for splitting is a manifestation of several counter-intuitive features of quantum theory, such as the fact that one cannot simply assign values to quantum variables.By restricting ourselves to observational probing schemes on quantum variables, we can effectively suspend this particular feature of quantum causal models and instead explore other, unrelated quantum phenomena.

Classically, passive observation is one example of an observational scheme, but it is not the only way to ensure informational symmetry. We illustrate this with an explicit classical example of a probing scheme that disturbs the variables under observation, yet exhibits informational symmetry.

Example 34. Consider a classical system that has two properties, described by the variables X and Y. Suppose that the probing scheme is such that measuring the value of one disturbs the system in the sense that it randomizes the other, so that one's posterior beliefs about the latter variable are best represented by the uniform distribution, denoted μ. Let X_C, Y_C denote the versions of the variables before the measurement, whereas X_D, Y_D denote the versions after the measurement; let the setting $S = \{\mathbb{X}, \mathbb{Y}\}$ encode which variable one chooses to measure and let M encode the outcome. Now, if one measures the variable X and finds a value x_0, one's prediction of the post-intervention state is

$$P\left(X_D, Y_D | S = \mathbb{X}, M = x_0\right) = \delta\left(X_D - x_0\right) \mu\left(Y_D\right). \tag{3.38}$$

The retrodiction about the pre-intervention state, on the other hand, depends on one's prior information, $P\left(Y_C\right)$. Using standard Bayesian inversion, one finds

$$P\left(X_C, Y_C | S = \mathbb{X}, M = x_0\right) = \delta\left(X_C - x_0\right) P\left(Y_C\right), \tag{3.39}$$

which is in general different from the predictions about X_D and Y_D. However, in the particular case wherein one has no prior information, that is, if $P\left(Y_C\right) = \mu\left(Y_C\right)$, then

$$P\left(X_C, Y_C | S = \mathbb{X}, M = x_0\right) = P\left(X_D, Y_D | S = \mathbb{X}, M = x_0\right) \tag{3.40}$$

(and, by a similar reasoning, the same holds for all other values of S and M), hence the scheme satisfies the informational symmetry condition.

Unlike passive observation, the above example of an observational probing scheme admits a close quantum analogue:

Example 35. Let A be a quantum variable which is probed with a non-destructive projective measurement that distinguishes between the rank-one projectors $\{\Pi^{sm}\}_m$ and obeys the von Neumann-Lüders update rule (proposition 24). In terms of the pre- and post-intervention versions of A, denoted C and D, and the classical setting S and outcome M, each s defines an of instruments $\left\{\mathcal{E}^{sm}_{D|C}\right\}_m$ with $\mathcal{E}^{sm}_{D|C}\left(\cdot\right) = \Pi^{sm}\left(\cdot\right)\Pi^{sm}$. Equivalently, the measurement can be described by the conditional $\tau_{MD|SC} = \sum_{ms} |m\rangle\langle m| \otimes \tau^{sm}_{D|C} \otimes |s\rangle\langle s|$, with the individual operators $\tau^{sm}_{D|C}$ isomorphic to the maps $\mathcal{E}^{sm}_{C|D}$. Given a prior ρ_C, the unnormalized post-intervention state associated with values s and m is $P\left(m|s\right)\rho^{sm}_D = \mathcal{E}^{sm}_{D|C}\left(\rho_C\right)$, hence the state on D conditional on the setting s and outcome m is

$$\rho^{sm}_D = \frac{1}{P\left(m|s\right)} \Pi^{sm} \rho_C \Pi^{sm} = \Pi^{sm}. \tag{3.41}$$

In order to derive a retrodiction about C, note that $\tau_{M|SC} = Tr_D \tau_{MD|SC} = \sum_{ms} |m\rangle\langle m| \otimes |s\rangle\langle s| \otimes \Pi^{sm}_C$, form which one obtains by Bayesian inversion $\tau_{C|SM} = \sum_{ms} \rho^{sm}_C \otimes |m\rangle\langle m| \otimes |s\rangle\langle s|$, with

$$\rho^{sm}_C = \frac{1}{P\left(m|s\right)} \rho_C^{\frac{1}{2}} \Pi^{sm} \rho_C^{\frac{1}{2}}. \tag{3.42}$$

In general, the retrodictive state depends on one's prior information about the system, ρ_C, but if one has no prior information, that is, if $\rho_C = \frac{1}{h_A}\mathbb{I}_C$, then

$$\rho_C^{sm} = \Pi^{sm}, \tag{3.43}$$

so that

$$\rho_C^{sm} = \rho_D^{sm} \tag{3.44}$$

for all s and m, or equivalently $\tau_{C|SM} = \tau_{D|SM}$, and therefore the scheme satisfies the informational symmetry condition.

Having established that there exists an observational probing scheme for quantum variables, let us now explore the consequences of informational symmetry in the context of quantum causal models. To this end, we introduce a slightly modified account of how one probes a quantum variable A: instead of an instrument described by a conditional

$$\tau_{MD|SC} = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \tau_{D|C}^{sm}, \tag{3.45}$$

whose elements $\tau_{D|C}^{sm}$ represent causal influences of C on D, we represent what we can infer about C and D given the setting S and the outcome M by a conditional

$$\tau_{CD|SM} = \sum_{ms} |m\rangle\langle m|_M \otimes |s\rangle\langle s|_S \otimes \rho_{CD}^{sm}, \tag{3.46}$$

so as to put C and D on equal footing. The elements ρ_{CD}^{sm} are bipartite quantum states. Accordingly, the functionality of the external circuit that connects A to other variables should be represented as a conditional $\tau_{B|CD}$ that takes states of knowledge about D and C to states of knowledge about B. This operator can be obtained from the Jamiołkowski representation of the causal map by $\tau_{B|CD} = \rho_C^{-\frac{1}{2}}\tau_{CB|D}\rho_C^{-\frac{1}{2}}$. In terms of these operators, the statistics once again take the form of an inner product,

$$P(y|ms) = Tr\left[(\rho_{DC}^{sm} \otimes \tau_B^y)\,\tau_{B|CD}\right]. \tag{3.47}$$

Analogously to the earlier definition of the component $\tilde{\tau}_{C|D}^y \equiv Tr_B\left[\tau_B^y \tau_{CB|D}\right]$, we will introduce a shorthand for the operator on CD that is probed by the $\{\rho_{CD}^{sm}\}$: let

$$\tilde{\tau}_{CD}^y \equiv Tr_B\left[\tau_B^y \tau_{B|CD}\right] = \rho_C^{-\frac{1}{2}}\tilde{\tau}_{C|D}^y\rho_C^{-\frac{1}{2}}. \tag{3.48}$$

Recall that the $\left\{\tilde{\tau}_{C|D}^y\right\}$ are components of the conditional $\tau_{CB|D}$, which represents the causal map, in the sense that characterizing the $\left\{\tilde{\tau}_{C|D}^y\right\}$ induced by a set $\{\tau_B^y\}$ that span $\mathcal{L}(\mathcal{H}_B)$ is sufficient to characterize the entire causal map. In the same way, the $\{\tilde{\tau}_{CD}^y\}$ defined here are components of the operator $\tau_{B|CD}$. They can be obtained by combining $\tau_{B|CD}$ with the conditional $\tau_{Y|B}$ that represents the measurement on B, giving

$$\tau_{Y|CD} = Tr_B\left(\tau_{Y|B}\tau_{B|CD}\right) = \sum_y |y\rangle\langle y| \otimes \tilde{\tau}_{CD}^y. \tag{3.49}$$

In terms of these operators, we have

$$P(y|ms) = Tr\left[\rho_{DC}^{sm}\tilde{\tau}_{CD}^y\right], \tag{3.50}$$

and ranging over a spanning set of operators ρ_{CD}^{sm} allows one to characterize the component $\tilde{\tau}_{CD}^y$ of $\tau_{B|CD}$.

3.3 Partial tomography

This section provides an explicit, algebraic characterization of what one can learn about a qubit A and its causal relations to other variables by the probing scheme introduced in example 35, which is an example of an observational scheme.

Theorem 36. *Let A be a quantum variable, whose pre- and post-intervention versions we denote C and D, and suppose that A is probed by the following observational scheme: a projective measurement obeying the von Neumann-Lüders update rule, under the assumption that the prior ρ_C is maximally mixed (see example 35). By this scheme, one can only access a subspace $\mathcal{A} \subset \mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_D)$. If A is a qubit, then the accessible subspace $\mathcal{A}$ is generated by the nine basis elements G^{+k} ($k = 1, ...6$) and $G^{7,8,9}$ from definition 38 below.*

Proof. We prove this by explicit construction in the remainder of this section. □

Definition 37. Any probing scheme on A that completely characterizes the subspace $\mathcal{A}$ is termed *partial tomography*.

Let us now prove the claim of theorem 36. If one uses the observational scheme from example 35, then the information one acquires about the pre- and post-intervention versions of A by finding an outcome m in a measurement with setting s takes the form

$$\rho_{CD}^{sm} = \Pi_C^{sm} \otimes \Pi_D^{sm}. \tag{3.51}$$

We will introduce a self-dual[16] basis $\left\{G^k\right\}_k$ of the space of two-qubit operators such that the elements G^{+k} ($k = 1, ...6$) and $G^{7,8,9}$ can be obtained by linear combinations of the available ρ_{CD}^{sm}, while the remaining elements, G^{-k} ($k = 1, ...6$) and G^0, cannot. Consequently, the statistics

$$P(y|ms) = Tr\left[\rho_{DC}^{sm} \tilde{\tau}_{CD}^{y}\right] \tag{3.52}$$

can reveal the inner product of $\tilde{\tau}_{CD}^{y}$ with the first subset of basis elements, but not the second. It follows that the accessible subspace $\mathcal{A}$ is generated by the first subset of G^k.

Let us now construct the basis $\left\{G^k\right\}$. One conventional choice of basis for the space of two-qubit operators is formed by products of the extended set of Pauli operators defined in eq. (2.63), but this is not ideal for the purpose of the present discussion. Instead, we will motivate the following choice:

Definition 38. In terms of the Pauli operators σ_j ($j = 1, 2, 3$) and the identity operator, σ_0, let

$$\begin{cases} G^{\pm j} \equiv \sigma_0 \otimes \sigma_j \pm \sigma_j \otimes \sigma_0 & [j = 1, 2, 3] \\ G^{\pm(i+j+1)} \equiv \sigma_i \otimes \sigma_j \pm \sigma_j \otimes \sigma_i & [(i, j) = (1, 2), (1, 3), (2, 3)] \end{cases} \tag{3.53}$$

$$\begin{cases} G^0 \equiv \sigma_0 \otimes \sigma_0 - \sum_{j \neq 0} \sigma_j \otimes \sigma_j \\ G^7 \equiv \sigma_0 \otimes \sigma_0 + \frac{1}{3} \sum_{j \neq 0} \sigma_j \otimes \sigma_j \\ G^8 \equiv \sigma_1 \otimes \sigma_1 + \sigma_2 \otimes \sigma_2 - 2\sigma_3 \otimes \sigma_3 \\ G^9 \equiv \sigma_1 \otimes \sigma_1 - \sigma_2 \otimes \sigma_2. \end{cases} \tag{3.54}$$

The reasons for choosing this basis are as follows. Firstly, note that ρ_{CD}^{sm} in expression (3.51) is manifestly symmetric under the exchange $C \leftrightarrow D$, hence no linear combination of projectors of this form can generate an operator that is anti-symmetric. For this reason, we introduce

$$\begin{cases} G^{\pm j} \equiv \sigma_0 \otimes \sigma_j \pm \sigma_j \otimes \sigma_0 & [j = 1, 2, 3] \\ G^{\pm(i+j+1)} \equiv \sigma_i \otimes \sigma_j \pm \sigma_j \otimes \sigma_i & [(i, j) = (1, 2), (1, 3), (2, 3)] \end{cases} \tag{3.55}$$

[16] The basis in definition 38 is self-dual up to normalization factors, which we do not include in order to simplify the notation.

noting that the six anti-symmetric operators G^{-k} (with $k = 1, ...6$) lie, by construction, outside $\mathcal{A}$. The six symmetric operators G^{+k}, on the other hand, can be obtained by linear combinations of products $\Pi_C^{sm} \otimes \Pi_D^{sm}$: for example,

$$G^{+1} = \frac{1}{2}(\sigma_0 + \sigma_1) \otimes (\sigma_0 + \sigma_1) - \frac{1}{2}(\sigma_0 - \sigma_1) \otimes (\sigma_0 - \sigma_1), \tag{3.56}$$

where each term is the product of projectors onto an eigenstate of σ_1. It follows that G^{+k} with $k = 1, ...6$ are part of a basis of $\mathcal{A}$.

The operators $G^{\pm k}$ with $k = 1, ...6$ span 12 dimension of the space of two-qubit operators, out of 16. The remaining four dimensions are spanned by the products of Paulis $\{\sigma_j \otimes \sigma_j\}$ with $j = 0, 1, 2, 3$. However, this is again an inconvenient choice of basis if our goal is to characterize the subspace spanned by products of the form (3.51), for the following reason: since the measurements in the quantum observational scheme are sharp, the projectors Π^{sm} are rank-one, viz. pure states. In the Pauli basis, such a projector takes the form $2\Pi^{sm} = \sigma_0 + \vec{v} \cdot \vec{\sigma}$, with a Bloch vector $\vec{v}$ that has unit norm, $\vec{v} \cdot \vec{v} = 1$. The product of two identical projectors Π^{sm} therefore has an expansion of the form

$$4\Pi_C^{sm} \otimes \Pi_D^{sm} = \sigma_0 \otimes \sigma_0 + \sum_i v_i^2 \sigma_i \otimes \sigma_i + \left[\sum_{i \neq j} v_i v_j \sigma_i \otimes \sigma_j\right], \tag{3.57}$$

which implies that the inner product

$$Tr\left[(\Pi_C^{sm} \otimes \Pi_D^{sm})\left(\sigma_0^C \otimes \sigma_0^D - \sum_{j \neq 0} \sigma_j^C \otimes \sigma_j^D\right)\right] = 1 - \vec{v} \cdot \vec{v} = 0 \tag{3.58}$$

is zero. In other words, the accessible subspace $\mathcal{A}$, which is spanned by the products $\{\Pi_C^{sm} \otimes \Pi_D^{sm}\}$, does not include the operator $\sigma_0^C \otimes \sigma_0^D - \sum_{j \neq 0} \sigma_j^C \otimes \sigma_j^D$. We take this operator to be G^0. Any other combination of the $\{\sigma_j \otimes \sigma_j\}$ that is orthogonal to G^0, however, can be generated by suitable combinations of the $\{\Pi_C^{sm} \otimes \Pi_D^{sm}\}$: by adding the products of projectors onto eigenvectors of σ_j, one obtains

$$(\sigma_0 + \sigma_j) \otimes (\sigma_0 + \sigma_j) + (\sigma_0 - \sigma_j) \otimes (\sigma_0 - \sigma_j) = 2(\sigma_0 \otimes \sigma_0 + \sigma_j \otimes \sigma_j) \tag{3.59}$$

for $j = 1, 2, 3$. The combinations G^k with $k = 7, 8, 9$ therefore lie in the accessible subspace. This completes the construction of the basis.

3.3.1 Complete characterization of purely CE or CC relations between two qubits

Theorem 39. *If two qubits are related either purely as cause and effect or purely by a common cause, so that their relation can be represented by a one-to-one qubit inference map,*

$$\rho_B = \mathcal{E}_{B|A}(\rho_A) = Tr_A\left[\tau_{B|A}\rho_A\right], \tag{3.60}$$

and they are probed by an observational scheme, then the data thus obtained are sufficient to completely characterize $\tau_{B|A}$.

Proof. Without loss of generality, we can label the variables such that A is causally prior to B. This implies that it is not necessary to split B, since only the pre-intervention version of B can be related to A, and consequently the restriction to observational probing on B is irrelevant. Splitting A into the pre- and post-intervention versions, C and D, the causal relation between A and B is then completely described by the causal map $\mathcal{E}_{CB|D}$ (definition 30).

Depending on the causal structure, the Jamiołkowski representation of the causal map takes one of two forms:

$$\tau_{CB|D} = \begin{cases} \rho_{CB} \otimes \mathbb{I}_D & \text{with } Tr_B \rho_{CB} = \frac{1}{2}\mathbb{I}_C, \text{ if purely CC} \\ \frac{1}{2}\mathbb{I}_C \otimes \tau_{B|D} & \text{with } Tr_B \tau_{B|D} = \mathbb{I}_D, \text{ if purely CE.} \end{cases} \tag{3.61}$$

Note that the marginal on C must be maximally mixed in either case, because, as we saw in example 35, this is required in order to ensure informational symmetry. Consequently the Jamiołkowski operator that we can access takes the form

$$\tau_{B|CD} = \rho_C^{-\frac{1}{2}} \tau_{CB|D} \rho_C^{-\frac{1}{2}} = \begin{cases} \tau_{B|C} \otimes \mathbb{I}_D & \text{with } \tau_{B|C} = \rho_C^{-\frac{1}{2}} \rho_{CB} \rho_C^{-\frac{1}{2}} = \rho_{CB}, \text{ if purely CC} \\ \mathbb{I}_C \otimes \tau_{B|D} & \text{, if purely CE.} \end{cases} \tag{3.62}$$

Our goal is to characterize the inference map $\tau_{B|A}$, which is given respectively by $\tau_{B|C}$ or $\tau_{B|D}$.

Since we are restricted to an observational scheme on A but not on B, we can evaluate the inner products of $\tau_{B|CD}$ with the accessible G^k on CD and generic Pauli observables on B. In a purely CC scenario, we find

$$Tr\left[\tau_{B|CD} G^k_{CD} \otimes \sigma^l_B\right] = \begin{cases} 2Tr\left[\tau_{B|C}\left(\sigma^k_C \otimes \sigma^l_B\right)\right] & \text{for } k = +1, 2, 3 \\ 0 & \text{for } k = +4, 5, 6 \\ 2Tr\left[\tau_{B|C}\left(\sigma^0_C \otimes \sigma^l_B\right)\right] & \text{for } k = 7 \\ 0 & \text{for } k = 8, 9, \end{cases} \tag{3.63}$$

where all terms in the G^k that contain a non-trivial Pauli observable ($j = 1, 2, 3$) on D vanish, because $Tr\left[\mathbb{I}_D \sigma^j_D\right] = 2\delta(j, 0)$. Since we can range over all σ^l ($l = 0, 1, 2, 3$) on B, this provides a complete characterization of $\rho_{CB} = \tau_{B|C}$. Similarly, in the purely CE scenario, any terms with non-trivial Paulis on C vanish, leaving

$$Tr\left[\tau_{B|CD} G^k_{CD} \otimes \sigma^l_B\right] = \begin{cases} 2Tr\left[\tau_{B|D}\left(\sigma^l_B \otimes \sigma^k_D\right)\right] & \text{for } k = +1, 2, 3 \\ 0 & \text{for } k = +4, 5, 6 \\ 2Tr\left[\tau_{B|D}\left(\sigma^l_B \otimes \sigma^0_D\right)\right] & \text{for } k = 7 \\ 0 & \text{for } k = 8, 9, \end{cases} \tag{3.64}$$

which also provides a complete characterization of $\tau_{B|D}$. In either case, one learns the inference map $\tau_{B|A}$. □

3.4 Implications of small deviations from the informational symmetry condition

This section details how the previous results are modified as one deviates from the informational symmetry condition, taking as an example the case wherein the measurement on A is still a perfect rank-one projective measurement that obeys the von Neumann-Lüders update rule, but allowing the prior ρ_C to deviate slightly from the maximally mixed state.

One could instead violate informational symmetry by fixing ρ_C to be maximally mixed, but allowing a generic POVM measurement. However, this opens to door to generic state update rules, including, for example, an instrument that performs a measurement, followed by one of a set of possible unitary rotations, and reports both the outcome of the measurement and which unitary was realized. With appropriate choices of the measurements and unitaries, such a scheme essentially allows one to reprepare D independently of C. In the present chapter, however, we are interested in restrictions to what one can learn about C and D. For this reason, we focus on the effects of a biased prior ρ_C in conjunction with perfectly sharp, von Neumann-Lüders-type measurements.

In this case, one still cannot obtain a complete characterization of the causal map, since the two pieces of information that one obtains about the pre- and post-intervention versions of A, respectively, are *different*[17], but not *independent*, as we will show below. However, one does recover other physical effects that arise from a violation of informational symmetry (even if C and D do not become statistically independent): for example, if Alice has a cause-effect relation to Bob and the pre-intervention state of Alice's variable is not maximally mixed, then she can signal him by her choice of measurement. In classical causal modelling, signalling is the paradigmatic example of a phenomenon that heralds a CE relation. By considering two variables whose causal relation is either CE or CC, we derive a closely analogous phenomenon that is similarly characteristic of CC relations.

3.4.1 No full causal tomography

In general, performing a rank-one projective measurement with the Lüders update rule on a state that is not maximally mixed reveals information about C and D that is different, but not independent: in terms of the projector $\Pi^{sm} \in \mathcal{L}(\mathcal{H}_A)$ associated with a setting s and outcome m, what one learns about C and D takes the form

$$\rho_{CD}^{sm} = \rho_C^{sm} \otimes \rho_D^{sm} = \rho_C^{\frac{1}{2}} \Pi_C^{sm} \rho_C^{\frac{1}{2}} \otimes \Pi_D^{sm}. \tag{3.65}$$

By ranging over different s and m, one has access to a set of projectors $\{\Pi^{sm}\}$ that span $\mathcal{L}(\mathcal{H}_A)$. However, given the particular form of eq. (3.65), this is not sufficient to make the corresponding ρ_{CD}^{sm} span the space of bipartite operators $\mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_D)$ (recalling that $\mathcal{H}_C$ and $\mathcal{H}_D$ are isomorphic to $\mathcal{H}_A$). If the variables in question are qubits, we can express the constraints on the accessible ρ_{CD}^{sm} explicitly, as we illustrate in the following, noting that they are closely related to those that characterize partial tomography.

If the prior ρ_C of a single qubit is not maximally mixed, it is biased towards one state, which we take to be the $+1$ eigenstate of σ^3 without loss of generality. Let us write

$$\rho_C = \frac{1}{2}\sigma^0 + \frac{\epsilon}{2}\sigma^3, \tag{3.66}$$

where $\epsilon \ll 1$ parametrizes the deviation from informational symmetry. In this case, the following linearly independent combinations of Pauli products can be obtained by combining the available ρ_{CD}^{sm}:

$$\left\{ \sigma_C^1 \otimes \sigma_D^2 + \sigma_C^2 \otimes \sigma_D^1 \right. \tag{3.67}$$

$$\begin{cases} \sigma_C^0 \otimes \sigma_D^i + \frac{1}{\sqrt{1-\epsilon^2}} \sigma_C^i \otimes \sigma_D^0 - \frac{\epsilon}{\sqrt{1-\epsilon^2}} \sigma_C^i \otimes \sigma_D^3 \\ \sigma_C^3 \otimes \sigma_D^i + \frac{1}{\sqrt{1-\epsilon^2}} \sigma_C^i \otimes \sigma_D^3 - \frac{\epsilon}{\sqrt{1-\epsilon^2}} \sigma_C^i \otimes \sigma_D^0 \end{cases} \quad \text{with } i = 1, 2 \tag{3.68}$$

$$\begin{cases} \sigma_C^i \otimes \sigma_D^i + \frac{\epsilon}{\sqrt{1-\epsilon^2}} \sigma_C^0 \otimes \sigma_D^0 + \frac{\epsilon}{\sqrt{1-\epsilon^2}} \sigma_C^3 \otimes \sigma_D^0 & \text{with } i = 1, 2 \\ \sigma_C^3 \otimes \sigma_D^3 + \sigma_C^0 \otimes \sigma_D^0 \\ \sigma_C^3 \otimes \sigma_D^0 + \sigma_C^0 \otimes \sigma_D^3 \end{cases} \tag{3.69}$$

Note that this subspace has only nine dimensions, as opposed to the $4 \times 4 = 16$ dimensions that would become accessible if we could intervene and reprepare D independently of C. In the limit of

[17] As an aside, we note that the difference between ρ_C^{sm} and ρ_D^{sm} in this scenario – that is, the violation of informational symmetry – follows from the non-commutativity of generic operators on quantum systems: for a general (not necessarily rank-one) projective measurement obeying the Lüders rule, we have $\rho_C^{sm} = \rho_C^{\frac{1}{2}} \Pi^{sm} \rho_C^{\frac{1}{2}}$, while $\rho_D^{sm} = \Pi^{sm} \rho_C \Pi^{sm}$. This consequence of the algebraic nature of quantum mechanics had likely not been noted before, and attempting to interpret it physically may lead to interesting insights.

the bias $\epsilon \to 0$, this reduces to a spanning set of the accessible subspace from partial tomography, introduced in the previous section:

$$\left\{\sigma_C^1 \otimes \sigma_D^2 + \sigma_C^2 \otimes \sigma_D^1 = G^{+4} \right. \tag{3.70}$$

$$\begin{cases} \sigma_C^0 \otimes \sigma_D^i + \sigma_C^i \otimes \sigma_D^0 = G^{+1,2} \\ \sigma_C^3 \otimes \sigma_D^i + \sigma_C^i \otimes \sigma_D^3 = G^{+5,6} \end{cases} \quad \text{with } i = 1, 2 \tag{3.71}$$

$$\begin{cases} \sigma_C^i \otimes \sigma_D^i + \sigma_C^0 \otimes \sigma_D^0 = G^7 + \frac{1}{2}\left(G^8 \pm G^9\right) & \text{with } i = 1, 2 \\ \sigma_C^3 \otimes \sigma_D^3 + \sigma_C^0 \otimes \sigma_D^0 = G^7 - \frac{1}{3}G^8 \\ \sigma_C^3 \otimes \sigma_D^0 + \sigma_C^0 \otimes \sigma_D^3 = G^{+3} \end{cases} \tag{3.72}$$

By relaxing the assumption of a uniform prior ρ_C, we gain access to a different set of components of the operator $\tilde{\tau}_{CD}^y$ on $\mathcal{H}_C \otimes \mathcal{H}_D$ than those which are accessible under the informational symmetry assumption, but we are nevertheless constrained to a nine-dimensional subspace of $\mathcal{L}\left(\mathcal{H}_C \otimes \mathcal{H}_D\right)$.

3.4.2 Physical implications of breaking informational symmetry

One important implication of informational symmetry is that, even though there exists a cause-effect path from A to B, Alice cannot signal Bob by the choice of her measurement (that is to say, by setting the value of S) alone. (If Alice is allowed to also fix the value of the outcome M, using post-selection, then she can effectively prepare a pure state of her choosing, Π^{sm}, to send to Bob, which makes signalling trivial.) Indeed, the setting S doesn't affect the post-intervention variable D, which in turn influences B. One can see this by noting first that knowledge of the setting S by itself, i.e. if one ignores the measurement outcome M, does not allow any inference about pre-intervention variable, C, because C is causally prior to S but not an ancestor of S. By informational symmetry, this implies that knowledge of S alone allows no inferences about D either. Mathematically, the retrodictive states on C conditional on different outcomes m,

$$\rho_C^{sm} = \frac{1}{P\left(m|s\right)}\left[\rho_C^{\frac{1}{2}} \Pi^{sm} \rho_C^{\frac{1}{2}}\right], \tag{3.73}$$

weighted by the probabilities of obtaining each outcome,

$$P\left(m|s\right) = Tr\left[\Pi^{sm}\rho_C\right], \tag{3.74}$$

give an average that is independent of the choice of setting:

$$\sum_m P\left(m|s\right)\rho_C^{sm} = \sum_m \rho_C^{\frac{1}{2}} \Pi^{sm} \rho_C^{\frac{1}{2}} = \rho_C^{\frac{1}{2}} \mathbb{I} \rho_C^{\frac{1}{2}} = \rho_C \ \ \forall s. \tag{3.75}$$

If informational symmetry holds, then the weighted average of the states on D, which we denote simply ρ_D, is also independent of the measurement setting:

$$\rho_D = \sum_m P\left(m|s\right)\rho_D^{sm} = \sum_m P\left(m|s\right)\rho_C^{sm} = \rho_C \ \ \forall s. \tag{3.76}$$

Now suppose that informational symmetry is broken because the prior state ρ_C is not maximally mixed. In this case, one must instead explicitly use the expression of the states on D conditional on sm,

$$\rho_D^{sm} = \Pi^{sm}, \tag{3.77}$$

which implies that the weighted average state on D takes the form

$$\rho_D = \sum_m Tr\left[\Pi^{sm}\rho_C\right]\Pi^{sm}; \tag{3.78}$$

that is, the projection of ρ_C onto the eigenbasis $\{\Pi^{sm}\}_m$ specified by the setting s. Notably, if ρ_C is not maximally mixed, then ρ_D depends on s, hence Alice can change the post-intervention state by her choice of setting alone (without using any post-selection to alter the probabilities with which different m appear in the mixture). If some later variable B is causally influenced by D, realizing a CE relation between A and B, then Alice can signal Bob by her choice of measurement setting, in the sense that the average of states on B,

$$\rho_B \equiv \sum_m P(m|s)\, \rho_B^{sm}, \tag{3.79}$$

may depend on s. For example, if the channel from D to B is the identity channel, then

$$\rho_B = \rho_D = \sum_m Tr\left[\Pi^{sm}\rho_C\right]\Pi^{sm}. \tag{3.80}$$

One can use this phenomenon to witness a CE relation, assuming one has access to a not-maximally mixed prior ρ_C.

A violation of informational symmetry also enables a CC analogue of this phenomenon. Just as the *weighted* average of retrodictive states on C is independent of s, the average *with equal weights* of the predictive states on D, denoted $\bar{\rho}_D$, is

$$\bar{\rho}_D \equiv \sum_m \frac{1}{h_A}\rho_D^{sm} = \sum_m \frac{1}{h_A}\Pi^{sm} = \frac{1}{h_A}\mathbb{I} \;\; \forall s \tag{3.81}$$

(where $h_A = h_C = h_D$ denotes the dimensionalities of the respective Hilbert spaces). The projective measurement acts as a probabilistic preparation of D, which, for a given setting s, generates one out of an ensemble of states $\{\Pi^{sm}\}_m$ with probabilities $P(m|s)$. If we use post-selection to form a subset of runs within which each m occurs with equal probability, then the state prepared on D on average over this subset is $\frac{1}{h_A}\mathbb{I}$. If informational symmetry holds, then the evenly weighted average of the states on C is also the maximally mixed state:

$$\bar{\rho}_C \equiv \sum_m \frac{1}{h_A}\rho_C^{sm} = \sum_m \frac{1}{h_A}\rho_D^{sm} = \frac{1}{h_A}\mathbb{I}\;\forall s. \tag{3.82}$$

However, if $\rho_C \neq \frac{1}{h_A}\mathbb{I}$, breaking informational symmetry, then

$$\bar{\rho}_C = \sum_m \frac{1}{h_A}\frac{1}{P(m|s)}\left[\rho_C^{\frac{1}{2}}\Pi^{sm}\rho_C^{\frac{1}{2}}\right], \tag{3.83}$$

which generally depends on s.

If A and B are related by a common cause, we can identify an intermediary (quantum) variable E that shares a common cause with A (more specifically, with the pre-intervention version of A, C) and causally influences B. Given the retrodictive conditional $\tau_{C|SM} = \sum_{sm}\rho_C^{sm}\otimes|s\rangle\langle s|\otimes|m\rangle\langle m|$ (whose elements ρ_C^{sm} are given by eq. (3.73)), which exhibits an analogue of signalling, one can construct a conditional that encodes inferences about E, using $\tau_{E|SM} = Tr_C\left[\tau_{E|C}\rho_{C|SM}\right]$. From there, one can make inferences about B, and those will inherit the S-dependence of the equally weighted average,

$$\bar{\rho}_B \equiv \sum_m \frac{1}{h_A}\rho_B^{sm}. \tag{3.84}$$

In summary, we have established that, if the variable A is probed using an informationally symmetric scheme, then neither the average $\bar{\rho}_B$, eq. (3.84), nor the weighted average ρ_B, eq. (3.79), of the conditionals states produced on B can depend on Alice's choice of setting, s, regardless of whether A and B are related as cause and effect or by a common cause. However, suppose that

informational symmetry is broken by virtue of a bias in the prior ρ_C. In that case, and only if the causal relation is CE, the setting s may affect the weighted average ρ_B, thereby allowing Alice to signal to Bob. The contrapositive is that only if the causal relation is CC may Alice's choice of setting be reflected in the equally weighted average $\bar{\rho}_B$. This second effect has not been explored, to our knowledge, possibly because one must perform post-selection in order to generate the equally weighted average. However, it is formally analogous to signalling in the sense that both effects bear witness to the respective underlying causal structure.

3.4.3 Scaling with the magnitude of the bias in ρ_C

Let us consider how these two effects scale with the magnitude of the bias in ρ_C. This is particularly relevant for experiments like the one described in section 4.2, which aim to discern the causal structure despite the restriction to an observational scheme: since one cannot guarantee that the conditions for informational symmetry are met exactly in an experimental implementation, one must bound how large an effect a certain deviation from informational symmetry can have.

In the case of CE signalling, it is easy to see that the effect is linear in ρ_C: since

$$\rho_D = \sum_m Tr\left[\Pi^{sm}\rho_C\right]\Pi^{sm}, \tag{3.85}$$

a small deviation in ρ_C from the maximally mixed state implies a proportionally small dependence of $\sum_m P\left(m|s\right)\rho_D^{sm}$ on the setting s. The scaling of the s-dependence in the analogous CC effect,

$$\bar{\rho}_C = \sum_m \frac{1}{h_A}\frac{1}{P\left(m|s\right)}\left[\rho_C^{\frac{1}{2}}\Pi^{sm}\rho_C^{\frac{1}{2}}\right], \tag{3.86}$$

is less straightforward to derive, since the expression contains not only two separate factors of $\rho_C^{\frac{1}{2}}$, but also the probability $P\left(m|s\right)$, which in turn depends on ρ_C as well. In this case, linear scaling with the bias in ρ_C may only hold in the limit of small bias, when one can neglect possible higher-order terms.

3.4.4 Visual representation for qubits

If the variables under study are qubits, we can represent both signalling and its CC analogue graphically, as shown in Fig. 3.3. This is based on the fact that any trace-one operator ρ, when decomposed on the basis formed by the Pauli operators (including $\sigma_0 = \mathbb{I}$), corresponds to a three-component real vector $\vec{v}$, with

$$\rho = \frac{1}{2}\left[\sigma_0 + \sum_j v_j\sigma_j\right] \quad \Leftrightarrow \quad v_j = Tr\left[\rho\sigma_j\right]. \tag{3.87}$$

For pure states, i.e. rank-one operators, $\vec{v}\cdot\vec{v} = 1$.

For simplicity, we will choose the $\hat{z}$-axis of the Bloch sphere such that the Bloch vector representing the prior ρ_C takes the form $\vec{q} = (0, 0, \epsilon)$, where ϵ measures the strength of the bias. We write the Bloch vector representing the projector $\Pi^{s,m=+1}$ associated with a setting s and outcome $m = +1$ as $\vec{p}^{s}$, and note that the complementary measurement outcome, $m = -1$, corresponds to the opposite Bloch vector, $-\vec{p}^{s}$. In general, Π^{sm} has a Bloch vector $m\vec{p}^{s}$. Let $\vec{c}^{sm}$ and $\vec{d}^{sm}$ represent the states ρ_C^{sm} and ρ_D^{sm}, respectively, while we reserve $\vec{\mathbf{d}}$ for the Bloch vector of ρ_D, the weighted average of states on D that exhibits signalling, and $\vec{\mathbf{c}}$ for $\bar{\rho}_C$, the equally weighted average of states on C that exhibits the analogous s-dependence.

Since the post-intervention state is determined simply by the von Neumann-Lüders rule (24), independently of the prior ρ_C, we have

$$\rho_D^{sm} = \Pi^{sm} \quad \Leftrightarrow \quad \vec{d}^{sm} = m\vec{p}^{s}. \tag{3.88}$$

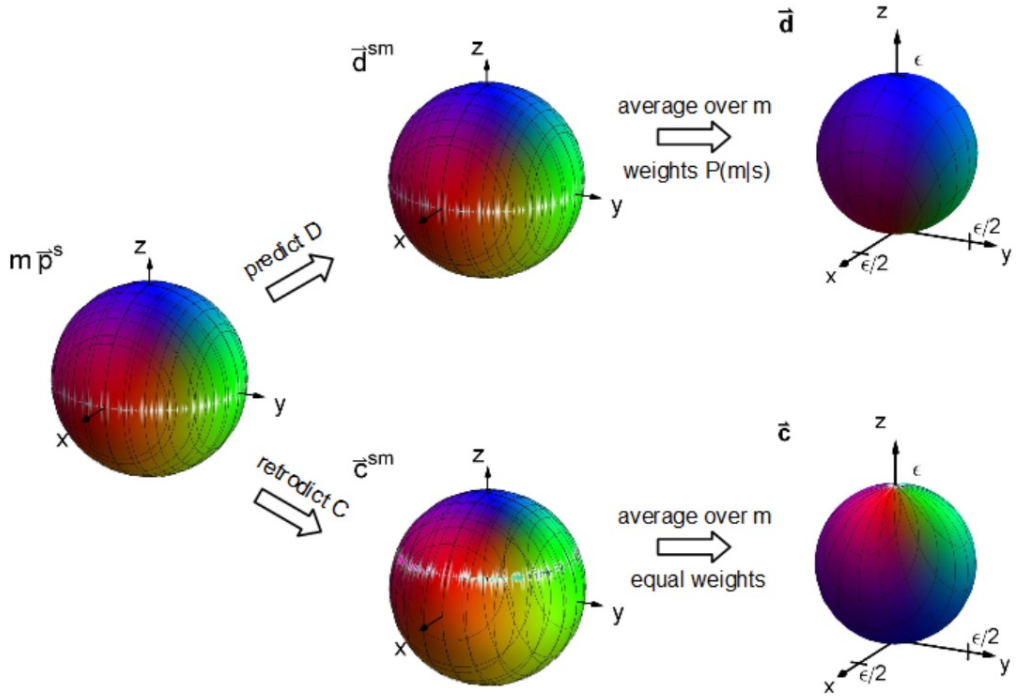

Figure 3.3: Inferences and signalling based on observational probing of a qubit; represented graphically as transformations of the Bloch sphere. The Bloch vectors $m\vec{p}^s$ ($m \in \{\pm 1\}$) represents the states found in a projective measurement, while $\vec{c}^{sm}$ and $\vec{d}^{sm}$ represent the inferences about the pre- and post-intervention versions of the variable, conditional on finding $m\vec{p}^s$, respectively. Note that the post-intervention state is simply the state found in the measurement, $\vec{d}^{sm} = m\vec{p}^s$, whereas the retrodictive pre-intervention state $\vec{c}^{sm}$ also depends on prior information, which causes a distortion of the Bloch sphere, as detailed in eq. (3.90). The figures are based on prior information represented by the Bloch vector $\vec{q} = (0, 0, \epsilon)$.
The Bloch vector $\vec{\mathbf{d}}$ represents the average of the post-intervention states $\vec{d}^{sm}$ associated with opposite outcomes, weighted by the probabilities with which these outcomes are found: it is what one obtains by fixing the basis of the measurement, but ignoring the outcome. The fact that $\vec{\mathbf{d}}$ ranges over a finite region of the Bloch sphere as we vary $\vec{p}^s$ shows that this state contains some information about the choice of basis. If a later variable B is causally influenced by D, one can send signals to B by choosing different measurement bases. The magnitude of this effect scales with our prior information about the system: the radius of the sphere in the last panel is ϵ.
The Bloch vector $\vec{\mathbf{c}}$ represents the equally weighted average of the retrodicted states $\vec{c}^{sm}$ associated with opposite outcomes; it is generated by post-selection. The fact that $\vec{\mathbf{c}}$ ranges over a finite region of the Bloch sphere as we vary $\vec{p}^s$ shows that this state also contains information about the choice of basis. If a later variable B shares a common cause with C, one can alter the state at B by choosing different measurement bases. This effect, too, scales with the parameter quantifying our prior information, ϵ.

The retrodictive state on C does not take such a simple form, since it incorporates both information about from the measurement ($\vec{s}$) and prior knowledge ($\vec{p}$),

$$\rho_C^{sm} = \frac{1}{P(m|s)} \left[\rho_C^{\frac{1}{2}} \Pi^{sm} \rho_C^{\frac{1}{2}} \right]. \tag{3.89}$$

The Bloch vector $\vec{c}^{sm}$ is given by different rules for the two components that are orthogonal and

collinear to $\vec{q}$, respectively. Since we chose the axes such that $\vec{q} = (0, 0, \epsilon)$, they simplify to (see appendix A.2)

$$c_3^{sm} = \frac{1}{1 + m\epsilon p_3^s} \left(m p_3^s + \epsilon \right) \quad c_{1,2} = \frac{\sqrt{1-\epsilon^2}}{1 + m\epsilon p_3^s} m p_{1,2}^s. \tag{3.90}$$

The probability of obtaining an outcome associated with a Bloch vector $\vec{p}^s$, given a prior specified by $\vec{q}$, is

$$P\left(m|s\right) = Tr\left[\Pi^{sm} \rho_C\right] = \frac{1}{2}\left(1 + m\epsilon p_3^s\right). \tag{3.91}$$

Now consider the weighted average over states on D that exhibits signalling, ρ_D, from eq. (3.78). We have

$$\vec{\mathbf{d}} = \sum_{m \in \{\pm 1\}} P\left(m|s\right) \vec{d}^{sm} = \frac{1}{2}\left(1 + \epsilon p_3^s\right) \vec{p}^s + \frac{1}{2}\left(1 - \epsilon p_3^s\right)\left(-\vec{p}^s\right) = \left(\epsilon p_3^s\right) \vec{p}^s. \tag{3.92}$$

Note that the average over m, $\vec{\mathbf{d}}$, depends only on the axis onto which we project, $\vec{p}^s$, but not on its orientation, i.e. whether we use $\vec{p}^s$ or $-\vec{p}^s$. Also note that $\vec{\mathbf{d}} = \vec{0}$ when $\vec{p}^s \perp \vec{q}$. In the CC scenario, we take the equally weighted mixture of states on C, $\bar{\rho}_C$ defined in (3.83):

$$\begin{cases} \mathbf{c}_3 = \frac{1}{2}\left[\frac{\epsilon + p_3^s}{1 + \epsilon p_3^s} + \frac{\epsilon - p_3^s}{1 - \epsilon p_3^s}\right] = \frac{1}{1 - \left(\epsilon p_3^s\right)^2}\left[-\left(\epsilon p_3^s\right) p_3^s + \epsilon\right] \\ \mathbf{c}_{1,2} = \frac{\sqrt{1-\epsilon^2}}{2}\left[\frac{p_{1,2}^s}{1 + \epsilon p_3^s} - \frac{p_{1,2}^s}{1 - \epsilon p_3^s}\right] = \frac{\sqrt{1-\epsilon^2}}{1 - \left(\epsilon p_3^s\right)^2}\left[\left(\epsilon p_3^s\right) p_{1,2}^s\right]. \end{cases} \tag{3.93}$$

These expressions, too, are independent of whether we use $\vec{p}^s$ or $-\vec{p}^s$. Furthermore, note that $\mathbf{c}_{1,2} = \vec{0}$ if $\vec{p}^s \perp \vec{q}$ or $\vec{p}^s \parallel \vec{q}$, and $\mathbf{c}_3 = 0$ if $\pm\vec{p}^s$ is the unit vector parallel to $\vec{q}$, whereas if $\vec{p}^s \perp \vec{q}$, then $\mathbf{c}_3 = \epsilon$.

Fig. 3.3 shows that, despite the distortion due to the biased prior, the weighted averages $\vec{\mathbf{d}}$ and $\vec{\mathbf{c}}$ range over spherical regions inside the Bloch sphere. The radius of these spheres quantifies the strength of the signalling effects, since it determines how well B can distinguish between different measurement bases A could have chosen. One can see that both $\vec{\mathbf{d}}$ and $\vec{\mathbf{c}}$ range over spheres of radius ϵ, showing that, at least for qubits, both effects scale linearly with the deviation from the informational symmetry condition.

4 Form and function of quantum inference maps

An appealing feature of classical causal models is that all inferences based on them can be represented by the same class of mathematical object: conditional probability distributions such as $P(B|A)$, whose only defining property is the normalization, $\sum_B P(B|A=a) = 1 \ \forall a$. This form of the conditionals holds regardless of the causal structure relating A and B – in other words, the two constituents of the causal model, the structure and the conditionals, are independent. In the quantum limit, this separation breaks down: depending on the causal relation between A and B, the quantum inference map $\mathcal{E}_{B|A}$ must have different mathematical properties, as discussed in section 2.4. This chapter explores the implications of this fact, ranging from applications as a tool for causal discovery and the analysis of open systems dynamics to more fundamental insights about the physical significance of PPT operators.

4.1 The quantum advantage for causal discovery

The central question of most causal modelling problems – and, indeed, much of science – can be cast as some variation of the following: what can one deduce about the causal relations between a set of variables, given certain observed correlations? Ideally, one can probe the causal relations with interventions: changing one variable, while leaving all else equal, to test how this affects other variables. However, interventions may be limited by various factors: ethical considerations in medical trials, practical limitations, or even fundamental physical impossibility, as is the case in astronomy. In those cases, one must resort to a more sophisticated analysis of the available statistics, looking for features of non-interventionist statistics that reflect properties of the underlying causal structure. One example is conditional independence, as suggested in [42]: if X is a common cause of Y and Z, then conditioning on X makes the other two statistically independent, i.e. $P(YZ|X) = P(Y|X)\,P(Z|X)$, whereas if Y and Z have some other common cause and X is merely a descendant of one of them, one will find no such independence in the statistics. These methods have been successfully applied to a range of problems involving several variables; but the seemingly simple problem of just two variables defies them, since their statistics exhibit little structure that could provide clues about the causal relation. Other methods have been proposed that determine the best-fitting causal explanation by demanding the smallest Kolmogorov complexity, for example, but these may require additional assumptions about the particular scenario under study [43]. This section introduces a new criterion for discerning causal relations, which is fundamentally different from existing techniques in that it exploits the strictly richer structure of correlations that can arise between quantum variables.

We illustrate this criterion using a task at which most conventional causal discovery methods fail: given two quantum variables, A and B, we aim to distinguish whether they are related as cause and effect (CE) or by an unobserved common cause (CC), but under the constraint that one is only allowed to probe A by an observational scheme. Our scheme does not distinguish whether A causes B or B causes A; hence, to avoid ambiguity, we assume that the two are causally ordered, with A referring to the early variable. (There is therefore no need to impose a restriction on how one can probe B, since the pre-intervention version of B can be correlated with A.)

In a classical world, observational data about two variables takes the form of a joint distribution $P(AB)$. By the rules of classical statistics, any such probability distribution can also be decomposed into valid conditionals and marginals of the form

$$P(A) = \sum_B P(AB) \quad \text{and} \quad P(B|A) = \frac{P(AB)}{P(A)}. \tag{4.1}$$

These distributions parametrize a causal model in which A causally influences B, whereas the joint distribution $P(AB)$ parametrizes a model wherein A and B share a common cause. Either explanation is equally valid and compatible with the available data.

If A and B are quantum variables, on the other hand, then an observational scheme – at the operational level – yields a map from pure states on A (namely the projectors realized in the measurement) to the generic quantum states subsequently found on B. If the causal relation is either purely CC or purely CE, or a convex combination of the two, then this map must be linear, positivity-preserving and trace-preserving, and therefore admit a representation as a quantum conditional[18]: a Hermitian operator $\tau_{B|A}$ such that $Tr_B \tau_{B|A} = \mathbb{I}_A$, which relates the projectors found on A to the corresponding states on B by

$$\rho_B = Tr_A \left[\tau_{B|A} \left(\mathbb{I}_B \otimes \Pi_A \right) \right]. \tag{4.2}$$

By ranging over tomographically complete sets of projectors on A and measurements on B, one can characterize $\tau_{B|A}$. If the causal structure is CC, then $\tau_{B|A}$ is a common-cause conditional, as introduced in section 2.4.3: in terms of the joint state ρ_{AB} prepared by the common cause and its marginal, $\rho_A = Tr_B \rho_{AB}$, we have

$$\tau^{cc}_{B|A} = d_A \left(\rho_A^{-1/2} \otimes \mathbb{I}_B \right) \rho_{AB} \left(\rho_A^{-1/2} \otimes \mathbb{I}_B \right). \tag{4.3}$$

By proposition 25, a CC relation implies that $\tau_{B|A}$ is a positive-semidefinite operator, since it is closely related to the bipartite state ρ_{AB}. If the relation is CE, on the other hand, then it is most naturally described by a completely positive and trace-preserving map $\mathcal{E}_{B|A}$, which, under the Jamiołkowski isomorphism, corresponds to a conditional $\tau_{B|A}$ that is positive under partial transposition (see section 2.4.1 and in particular proposition 18). Taking the contrapositive, the mathematical properties of $\tau_{B|A}$ allow one to draw the following conclusions about the underlying causal structure:

Theorem 40. *Consider a linear, positivity-preserving and trace-preserving map $\mathcal{E}_{B|A}$ between two quantum variables, A and B, and its Jamiołkowski representation, $\tau_{B|A}$. If $\tau_{B|A}$ is not a positive-semidefinite operator, then A and B cannot be related by a purely CC structure. If $\tau_{B|A}$ is not positive under partial transposition (i.e. $\tau_{B|A}$ is NPT), this rules out a purely CE relation. (A map that fails to be linear, positivity-preserving and/or trace-preserving cannot be explained by either causal structure.)*

Proof. This follows directly from the above discussion, which draws on propositions 18 and 25. □

Note that this constitutes a uniquely quantum criterion for distinguishing the two causal relations, which becomes ineffective in the classical limit. Indeed, when there exist preferred bases of the Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ on which all operators involved in the problem are diagonal, then the eigenvalues of $\tau_{B|A}$ are unchanged under partial transposition (since one could perform the partial transposition in the preferred basis, which leaves $\tau_{B|A}$ unaltered). In this limit, both CC and CE structures give rise to $\tau_{B|A}$ that are both positive-semidefinite and PPT, and therefore provide no clues about the underlying causal structure.

4.1.1 Geometric and visual representation for qubits

If A and B are qubits, the effect of a quantum channel $\mathcal{E}_{B|A}$ is commonly represented by plotting the image of the Bloch sphere, that is, the set of resulting states $\rho_B = \mathcal{E}_{B|A}(\rho_A)$ as the inputs ρ_A range over the entire surface of the Bloch sphere. (The ρ_B arising from mixed inputs ρ_A, which lie in the interior of the Bloch sphere, follow by linearity. For this reason, the following discussion focuses only on the surfaces of regions of interest.) If $\mathcal{E}_{B|A}$ applies a unitary transformation, the Bloch sphere is simply rotated; if the channel is noisy, it generally shrinks the Bloch sphere towards the origin, i.e. the maximally mixed state. Verstraete [44] introduces an analogous representation for steering,

[18] If the causal relation is not a mere probabilistic mixture, then generally the inference map from A to B, or even the part of it that is accessible under partial tomography, need not take the simple form $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, as discussed in section 2.5.1.

which can easily be extended to represent inferences along a CC connection, as discussed in section 2.4.3. Indeed, the same representation applies to all one-to-one-qubit maps $\mathcal{E}_{B|A}$ that are linear, positivity-preserving and trace-preserving, with inference maps that arise from either purely CE or purely CC relations being particular examples.

The geometric and resulting visual representation is based on the following mathematical fact:

Lemma 41. *Consider a single-qubit conditional $\tau_{B|A} \in \mathcal{L}(\mathcal{H}_2 \otimes \mathcal{H}_2)$, which represents a trace-preserving one-to-one-qubit inference map, and let the positive-semidefinite, trace-one operators ρ_A and $\rho_B = Tr_A\left[\tau_{B|A}\rho_A\right]$ be its input and output, respectively. The conditional can be decomposed on the basis formed by Pauli operators as*

$$\frac{1}{4}Tr\left(\tau_{B|A}\sigma_B^i \otimes \sigma_A^j\right) = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{c} & T_{3\times 3} \end{pmatrix}_{ij}, \tag{4.4}$$

and the states ρ_A and ρ_B can be represented by the three-component Bloch vectors $\vec{a}$ and $\vec{b}$, respectively:

$$\begin{cases} a_j = \frac{1}{2}Tr\left(\rho_A\sigma_A^j\right) & j = 1,2,3 \\ b_i = \frac{1}{2}Tr\left(\rho_B\sigma_B^i\right) & i = 1,2,3. \end{cases} \tag{4.5}$$

Then the Bloch vectors are related directly by

$$\vec{b} = \vec{c} + T\vec{a}. \tag{4.6}$$

In geometrical terms, the image of the set of all pure states on A forms an ellipsoid in the Bloch sphere of B whose centre lies at $\vec{c}$ and whose axes (direction and scaling) are specified by the eigenvectors and the square roots of the eigenvalues of TT^T.

Proof. This is an extension of an unnumbered claim in [45]. Note that $\mathrm{Tr}\left(\tau_{B|A}\sigma_B^0 \otimes \sigma_A^j\right) = 0$ for $j \neq 0$ because the quantum conditional $\tau_{B|A}$ is trace-preserving, i.e. $Tr_B\tau_{B|A} = \sigma_A^0$. The transformation rule for Bloch vectors, $\vec{b} = \vec{c} + T\vec{a}$, follows from the corresponding rule for operators, $\rho_B = Tr_A\left[\tau_{B|A}\rho_A\right]$, using the fact that the Pauli observables form an orthonormal basis of operator space. The centre of the resulting ellipsoid can be read off directly from the resulting expression; its axes follow using standard tools of analytic geometry. □

If we perform partial tomography, given the promise that the causal relation is either purely CC or purely CE, as described in section 3.3.1, then the inner products that we can determine give precisely the non-zero elements of the Pauli basis representation of $\tau_{B|A}$, denoted $\vec{c}$ and T. Let us introduce the following terminology:

Definition 42. The term *inference ellipsoid* refers to the geometric representation of a single-qubit conditional introduced in lemma 41. The 3×3 real matrix T in the representation of $\tau_{B|A}$ on the Pauli basis, eq. (4.4), is termed the *correlation matrix*, since it arises from correlations between the Pauli observables $\sigma_{1,2,3}$ on the variables A and B.

Note that the geometry of the ellipsoid, which is specified by $\vec{c}$ and TT^T, does not encode a complete description of the inference map: one must also specify which pure-state input on A implies which state ρ_B on the ellipsoid. In order to include this information in a visual representation of $\mathcal{E}_{B|A}$, we assign a unique colour to each pure input: as a function of the Bloch vector $\vec{a}$, the colour is specified by the RGB components

$$\begin{cases} R = \frac{1}{2}(1 + a_1 - a_2 - a_3) \\ G = \frac{1}{2}(1 - a_1 + a_2 - a_3) \\ B = \frac{1}{2}(1 - a_1 - a_2 + a_3). \end{cases} \tag{4.7}$$

This marks the $+1$ eigenstates of σ_1 ($\vec{a} = (1,0,0)$) as red, σ_2 ($\vec{a} = (0,1,0)$) as green and σ_3 ($\vec{a} = (0,0,1)$) as blue, while their opposites, with eigenvalues -1, are coloured cyan (anti-red), magenta (anti-green) and yellow (anti-blue), respectively. Some examples of colour-coded ellipsoids are illustrated in Fig. 4.1. This scheme allows one to visualize, for example, the effect of different unitary channels: although all unitary transformations map the set of all pure states on A to the full Bloch sphere of B, different colour distributions reflect how the Bloch sphere is rotated differently by various unitaries. More importantly, this colour-coding introduces an easy visual distinction between CC and CE conditionals. To this end, we introduce the following nomenclature:

Definition 43. An inference ellipsoid is termed *right-handed* if the transformation T of the Bloch sphere that it represents is proper, i.e. does not involve a reflection: $\det T > 0$. If $\det T < 0$, the ellipsoid is termed *left-handed* . If $\det T = 0$, then the ellipsoid is degenerate (disk, line or point) and does not have well-defined handedness.

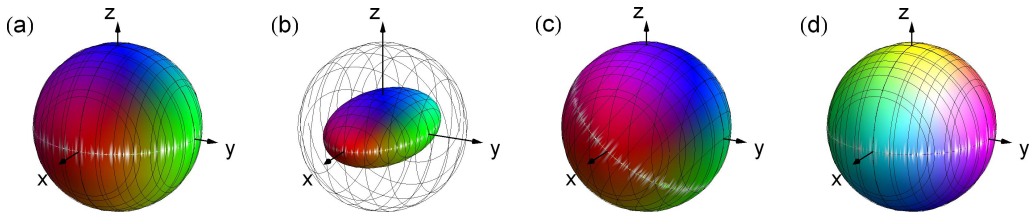

Figure 4.1: Visual representation of single-qubit conditionals by their effect on the Bloch sphere: (a) the identity channel, (b) a noisy channel, (c) a unitary rotation, (d) steering via a bipartite state (CC relation), specifically the singlet state. The geometry of the inference ellipsoid distinguishes extremal conditionals (definition 44) from noisy ones, whereas the differences between various unitary rotations and between extremal CE and extremal CC conditionals are only captured by the colouring scheme.

The handedness of an ellipsoid is easily identifiable in its coloured representation: Consider the images of the unit vectors $\{\hat{x}, \hat{y}, \hat{z}\}$, identified by red, green and blue, respectively. In the original Bloch sphere, they are mutually orthogonal and form a right-handed triad, in the sense that $(\hat{x} \times \hat{y}) \cdot \hat{z} > 0$. A proper transformation ($\det T > 0$) can distort their magnitudes, but only by non-negative factors, and the angles between them, but only within the range $(0, \pi)$. Thus, their images still form a right-handed triad. On the other hand, the triad of vectors RGB becomes left-handed if and only if there is a reflection involved, that is, under improper rotations of the Bloch sphere.

4.1.2 Causal relations reflected in the geometric and visual representation

In order to establish what the transformation parameters $(\vec{c}, T)$ and their visual representation as a coloured ellipsoid reveal about the causal structure, let us consider at first only the implications of the geometry of the ellipsoid, in two extreme cases. The first arises for the following set of causal relations:

Definition 44. A CE relation between two qubits is termed *extremal* if and only if it realized a unitary transformation. A CC relation between two qubits is termed extremal if and only if it realizes a pure, entangled state[19].

[19] A necessary and sufficient condition is that the conditional $\tau_{B|A}$ be pure and maximally entangled. However, for any pure two-qubit state ρ_{BA} that has non-zero entanglement (i.e. which is not a product state), the multiplication with $\rho_A^{-\frac{1}{2}}$ in $\tau_{B|A} = \rho_A^{-\frac{1}{2}} \rho_{BA} \rho_A^{-\frac{1}{2}}$ ensures that the resulting conditional is maximally entangled (see section 2.3.3).

Theorem 45. *Given a single-qubit conditional $\tau_{B|A}$, the inference ellipsoid representing $\tau_{B|A}$ is the unit sphere if and only if the only causal explanation of $\tau_{B|A}$ is either an extremal CE relation or an extremal CC relation.*

Proof. See appendix A.3. □

On the other hand, it is useful to have a characterization of the following set of inference maps:

Definition 46. Given a quantum conditional $\tau_{B|A}$, it is termed *undecidable* if it can be explained both by a purely CE relation and by a purely CC relation. If the properties of $\tau_{B|A}$ rules out at least one of the explanations, on the other hand, it is termed *decidable*.

Note that this definition of a decidable conditional does not actually guarantee that it can be explained by either CC or CE relations in general.

Theorem 47. *A single-qubit conditional $\tau_{B|A}$ is undecidable if and only if the inference ellipsoid fits inside a tetrahedron which is in turn circumscribed by the Bloch sphere.*

Proof. Note that $\tau_{B|A}$ is compatible with both CC and CE if and only if it is both positive-semidefinite and PPT, which in the case of qubits is a necessary and sufficient condition for separability. Furthermore, appendix D of [45] establishes that $\tau_{B|A}$ is separable if and only if the ellipsoid fits inside a nested tetrahedron. (The theorem in the reference refers to two-qubit joint states in the context of steering, but following section 2.4.3, $\tau_{B|A}$ can be considered a special case thereof.) □

The geometric condition is illustrated in Fig. 4.3c. The remaining cases of single-qubit inference maps, which are decidable but not extremal, require more refined criteria for determining what causal structure can account for them. We begin by focusing on a subset that admits a simpler analysis; the general case is discussed afterwards.

Unital single-qubit conditionals. A single-qubit map is termed unital if and only if it leaves the centre of the Bloch sphere, viz. the maximally mixed state, invariant. By extension, a Jamiołkowski operator $\tau_{B|A}$ is termed unital if and only if it is isomorphic to a unital map, which is equivalent to the mathematical condition $\mathrm{Tr}_A\left(\tau_{B|A}\mathbb{I}_A\right) = \mathbb{I}_B$. We begin by establishing the following:

Lemma 48. *A unital single-qubit conditional $\tau_{B|A}$ can be represented on the Pauli basis as*

$$\frac{1}{4}Tr\left(\tau_{B|A}\sigma_B^i \otimes \sigma_A^j\right) = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & T_{3\times 3} \end{pmatrix}_{ij}, \tag{4.8}$$

with the eigenvalues of the correlation matrix T forming a three-vector $\vec{t}$. Then $\tau_{B|A}$ admits a CC explanation if and only if $\vec{t}$ lies within the tetrahedron with vertices

$$\{(1,1,-1),(1,-1,1),(-1,1,1),(-1,-1,-1)\}, \tag{4.9}$$

and a CE explanation if and only if it lies within the tetrahedron

$$\{(1,1,1),(1,-1,-1),(-1,1,-1),(-1,-1,1)\}. \tag{4.10}$$

Proof. See appendix A.3. □

The two regions in $\vec{t}$-space are illustrated in Fig. 4.2. The tetrahedra introduced here are an alternative representation of the sets of positive-semidefinite respectively PPT operators on two qubits, formulated in terms of their Pauli coefficients. (See e.g. section 10.2 of Bengtsson and Zyczkowski [46].) This rather technical criterion gives rise to a simple condition on the visual representation of the inference map:

Theorem 49. *Let $\tau_{B|A}$ be a unital single-qubit conditional and assume that $\tau_{B|A}$ is decidable. Then $\tau_{B|A}$ admits a CE explanation if and only if the colouring of the inference ellipsoid is right-handed, and a CC explanation if and only if the colouring of the inference ellipsoid is left-handed.*

Proof. If $\tau_{B|A}$ is decidable, its representation in the space of eigenvalues $\vec{t}$ lies outside the intersection of the two tetrahedra in lemma 48. The remaining conditionals are positive-semidefinite (i.e. compatible with a CC relation) if and only if they lie in one of four smaller tetrahedra, each of them contained in an octant such that $t_1 t_2 t_3 = \det T < 0$. Conversely, they are negative if and only if they lie in one of four smaller tetrahedra with $\det T > 0$. Definition 43 relates this to the handedness of the steering ellipsoid. □

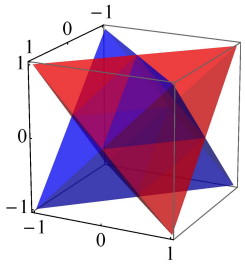

Figure 4.2: Classification of unital one-to-one qubit conditionals in terms of $\vec{t}$, the eigenvalues of the correlation matrix T: those admitting a CC (CE) explanation form a tetrahedron with vertices $\{(1,1,-1),(1,-1,1),(-1,1,1),(-1,-1,-1)\}$ (red) (respectively $\{(1,1,1),(1,-1,-1),(-1,1,-1),(-1,-1,1)\}$, blue). Their intersection, which represents undecidable maps, is an octahedron. Conditionals that admit only a CC (CE) explanation lie in octants with $\det T < 0$ ($\det T > 0$). (Note that all regions of interest are invariant under reorderings of the vector $\vec{t}$, hence there is no need to label the axes.)

Example 50. The conditional

$$\tau_{B|A} = \begin{pmatrix} \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{pmatrix} \tag{4.11}$$

is proportional to the Werner state $\rho_{BA} = \frac{1}{3}|\Phi^+\rangle\langle\Phi^+| + \frac{2}{3}\frac{\mathbb{I}}{4}$. It reduces the image of the Bloch sphere to radius $\frac{1}{3}$, so that it fits inside a tetrahedron, as illustrated in Fig. 4.3c, which implies undecidability (see theorem 47). Indeed, the conditional is separable, i.e. both positive-semidefinite and PPT, and therefore it can be explained by either causal relations: a CC relation that prepares the Werner state, or a CE relation that rotates the qubit by π about the y axis of the Bloch sphere with probability $\frac{1}{3}$ and applies complete dephasing in the σ_3-eigenbasis (i.e. a non-destructive, projective measurement of σ_3) with probability $\frac{2}{3}$. In both cases, one obtains a left-handed inference ellipsoid with radius $\frac{1}{3}$.

Non-unital single-qubit conditionals. Now consider a single-qubit conditional $\tau_{B|A}$ that is not necessarily unital. Again, one might ask under what conditions such conditionals are positive-semidefinite (compatible with CC), PPT (compatible with CE) or both (undecidable). In order to address this question, let us first establish the following useful representation:

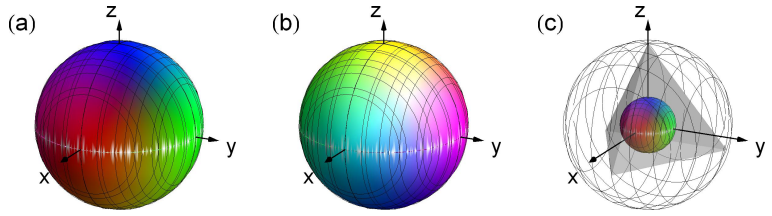

Figure 4.3: Geometric properties of the inference ellipsoid reflect the underlying causal structure: (a) The full sphere with a right-handed colour distribution is generated by unitary rotations (CE), specifically the identity channel in the case shown here. (b) The full sphere with left-handed colouring is generated by maximally entangled states (CC), specifically the singlet state. (c) An ellipsoid that fits inside a nested tetrahedron, regardless of handedness, can be explained by either causal relation. The figure shown here represents the Werner state discussed in example 50.

Lemma 51. *Given a generic single-qubit conditional $\tau_{B|A}$, whose Pauli basis representation is*

$$\frac{1}{4} Tr\left[\tau_{B|A}\sigma_B^i \otimes \sigma_A^j\right] = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{c} & T' \end{pmatrix}_{ij}, \tag{4.12}$$

there exist rotations of the Bloch spheres of A and B, viz. a product of local unitary operations on $\mathcal{H}_A$ and $\mathcal{H}_B$, that put the Pauli basis decomposition in the form

$$\frac{1}{4} Tr\left[(V_B \otimes U_A)\,\tau_{B|A}\,(V_B \otimes U_A)^\dagger\,\sigma_B^i \otimes \sigma_A^j\right] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_1 & t_1 & 0 & 0 \\ c_2 & 0 & t_2 & 0 \\ c_3 & 0 & 0 & t_3 \end{pmatrix}_{ij}. \tag{4.13}$$

The diagonal elements $\vec{t} = (t_1, t_2, t_3)$ are the singular values of the correlation matrix T' in the Pauli representation of the original $\tau_{B|A}$.

Proof. See appendix A.3.

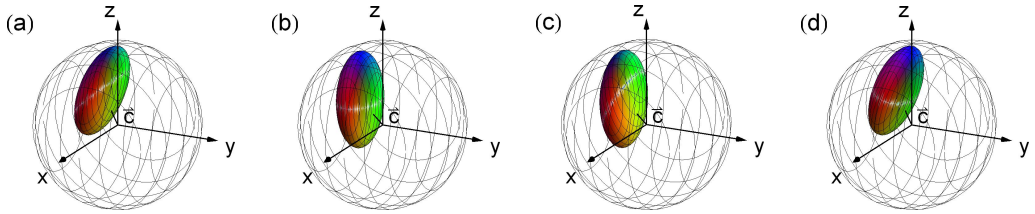

Figure 4.4: Inference ellipsoids (a) of a generic single-qubit conditional $\tau_{B|A}$ and (b) combined with suitable unitaries to put it in the standard form, introduced in lemma 51. The lengths of the axes t_1, t_2, t_3 and the magnitude of the offset from the origin, $|\vec{c}|$, remain fixed, but the orientations generally change. This is effected by the unitary V_B, which acts on the *output* of the inference map: it rotates the ellipsoid so that its axes align with the axes of B's Bloch sphere. The unitary on the input, U_A, on the other hand, changes which inputs (represented by certain colours) are mapped to which points on the ellipsoid. It is chosen such that the axes of the *input* Bloch sphere, which are represented visually by the colours red, green and blue, align with the axes of the ellipsoid. Their separate effects are illustrated in panels (c) and (d), respectively.

□

The inference ellipsoids representing $\tau_{B|A}$ and the suitably rotated $(V_B \otimes U_A)\,\tau_{B|A}\,(V_B \otimes U_A)^\dagger$ are illustrated in Fig. 4.4. Note that the position $\vec{c}$ of the centre of the ellipsoid representing $(V_B \otimes U_A)\,\tau_{B|A}\,(V_B \otimes U_A)^\dagger$ may be different from $\vec{c}$, which now denotes the position of the centre of the ellipsoid representing $\tau_{B|A}$ without the unitary transformations. It is furthermore important to note the following:

Lemma 52. *The set of positive-semidefinite bipartite operators is closed under products of unitaries, as is the set of PPT operators.*

Proof. For the first claim, note that unitary transformations, including products of unitaries on two subsystems, constitute positivity-preserving maps. Then the second claim follows by noting that $T_A\left(U_A\tau_{B|A}U_A^\dagger\right) = U_A^* T_A\left(\tau_{B|A}\right)U_A^T$, and the complex conjugate U_A^* is unitary if and only if U_A is unitary as well: if $\tau_{B|A}$ is PPT, so that $T_A\left(\tau_{B|A}\right) \geq 0$, then so is its image under any unitary U_A^*,

$$U_A^* T_A\left(\tau_{B|A}\right)U_A^T = T_A\left(U_A\tau_{B|A}U_A^\dagger\right) \geq 0, \tag{4.14}$$

and consequently $U_A\tau_{B|A}U_A^\dagger$ is also PPT, for any unitary U_A. □

Since positivity and PPT are unchanged under products of local unitaries, it is sufficient to seek a criterion in terms of the parameters $\vec{c}$ and $\vec{t}$ of the simplified form of $\tau_{B|A}$ that was introduced in lemma 51. Furthermore, the set of all positive-semidefinite operators is related to the set of all PPT by a simple partial transposition, therefore we will only directly characterize the region of $\vec{c}\times\vec{t}$-space that corresponds to PPT operators and from there deduce a characterization of the region that corresponds to positive-semidefinite operators. Ruskai et al. [47] derive necessary and sufficient conditions for a conditional to be PPT:

Lemma 53. *(Corollary 2 in [47], paraphrased) Consider a single-qubit conditional $\tau_{B|A}$ and suitable local unitaries $V_B \otimes U_A$ that put its Pauli representation in the form*

$$\frac{1}{4}Tr\left[(V_B \otimes U_A)\,\tau_{B|A}\,(V_B \otimes U_A)^\dagger\,\sigma_B^i \otimes \sigma_A^j\right] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_1 & t_1 & 0 & 0 \\ c_2 & 0 & t_2 & 0 \\ c_3 & 0 & 0 & t_3 \end{pmatrix}_{ij}. \tag{4.15}$$

The conditional $\tau_{B|A}$ is PPT if and only if
(a)$|c_3| + |t_3| = 1$, $c_1 = c_2 = 0$ and $t_1 = \pm t_2$; or
(b) $|c_3| + |t_3| < 1$, and all of the following hold, where $s = \pm 1$ denotes the sign of t_3:

$$\begin{cases} (t_1+t_2)^2 \leq (1+t_3)^2 - c_3^2 - \left(c_1^2+c_2^2\right)\left(\frac{1+t_3-s|c_3|}{1-t_3-s|c_3|}\right), \\ (t_1-t_2)^2 \leq (1-t_3)^2 - c_3^2 - \left(c_1^2+c_2^2\right)\left(\frac{1-t_3+s|c_3|}{1+t_3+s|c_3|}\right), \end{cases} \tag{4.16}$$

$$\left[1 - \left(t_1^2+t_2^2+t_3^2\right) - \left(c_1^2+c_2^2+c_3^2\right)\right]^2 \geq 4\left[t_1^2\left(c_1^2+t_2^2\right) + t_2^2\left(c_2^2+t_3^2\right) + t_3^2\left(c_3^2+t_1^2\right) - 2t_1t_2t_3\right]. \tag{4.17}$$

(Note that the parameters $\vec{c}$ and $\vec{t}$ must satisfy $|c_i| + |t_i| \leq 1$ for $i = 1,2,3$, since otherwise the image of $|\pm_i\rangle$, viz. the Bloch vector $\vec{a} = \hat{i}$, would lie outside the Bloch sphere and therefore not be a positive-semidefinite operator.)

This characterization of the region of $\vec{c}\times\vec{t}$-space that represents all PPT operators is rather cumbersome. Consider instead the set of $\vec{t}$ that, for a given $\vec{c}$, generate PPT conditionals. If $\vec{c}$ is zero (i.e. in the limit of unital conditionals), this is the blue tetrahedron shown in Fig. 4.2. The effect of non-trivial $\vec{c}$ is essentially to round the corners of the tetrahedron. (The case of $c_1 = c_2 = 0$, $c_3 > 0$

is discussed and represented graphically in [47], and numerical simulations show a similar form for generic $\vec{c}$.)

Let us now characterize the region in $\vec{c} \times \vec{t}$-space that corresponds to the set of all positive-semidefinite operators. This set is obtained from the set of all PPT operators by partial transposition. The effect of partial transposition on the parameters $\vec{c}$ and $\vec{t}$ that describe an individual $\tau_{B|A}$ depends on the basis chosen to perform the transposition: for example, transposition on B in the eigenbasis of σ^3 leaves σ_B^1 and σ_B^3 unchanged, but inverts the sign of σ_B^2, so that $(c_1, c_2, c_3) \to (c_1, -c_2, c_3)$ and $(t_1, t_2, t_3) \to (t_1, -t_2, t_3)$ (reflection about the $1,3$-plane), whereas transposition on B in the eigenbasis of σ^2 inverts the sign of σ_B^3, so that $(c_1, c_2, c_3) \to (c_1, c_2, -c_3)$ and $(t_1, t_2, t_3) \to (t_1, t_2, -t_3)$ (reflection about the $1,2$-plane). However, the images of a given $\tau_{B|A}$ under partial transposition in different bases are related by unitaries on the system on which the transposition acts, and according to lemma 52, the set of positive-semidefinite operators and the set of PPT operators are both closed under these operations. Therefore the image under partial transposition of the region of $\vec{c} \times \vec{t}$-space that corresponds to the set of *all* PPT operators is independent of the basis chosen to realize the transposition. For simplicity, we chose partial transposition on B in the eigenbasis of σ^3, so that the effect of partial transposition is a reflection about the $1,3$-plane, inverting the signs of c_2 and t_2. Note furthermore that the sign of c_2 is irrelevant in the conditions of lemma 53. The region of $\vec{t}$ that generate positive-semidefinite conditionals for a given $\vec{c}$ is therefore the reflection about the $1,3$-plane of the region that generates PPT conditionals for the same $\vec{c}$.

This symmetry allows us to generalize theorem 49, dropping the assumption of unitality:

Theorem 54. *Let $\tau_{B|A}$ be any single-qubit conditional and assume that $\tau_{B|A}$ is decidable, i.e. the inference ellipsoid does not fit inside a nested tetrahedron. Then $\tau_{B|A}$ admits a CE explanation if and only if the colouring of the inference ellipsoid is right-handed, and a CC explanation if and only if the ellipsoid is left-handed.*

Proof. Proof is provided in appendix A.3. □

This almost completes the classification of all single-qubit conditionals in terms of the causal structures that can account for them: if the inference ellipsoid fits inside a nested tetrahedron, then, by theorem 47, it can be explained both by a purely CC and by a purely CE relation. If the inference ellipsoid does not fit inside a nested tetrahedron but has well-defined handedness, then, by theorem 54, it is compatible either with a purely CC relation or with a purely CE relation. What remains to be classified, then, is the set of inference ellipsoids that violate the nested tetrahedron condition but do not have well-defined handedness: degenerate ellipsoids which are reduced to disks of sufficiently large radius. (If the ellipsoid is reduced to a line or a single point, then it automatically satisfies the nested tetrahedron condition.) By theorem 54, these conditionals can be explained neither by purely CC relation nor by purely CE relations alone. However, they can be explained by a convex mixture of the two[20].

Note that this no longer holds when one goes beyond the case of linear one-to-one qubit maps: a full causal map $\mathcal{E}_{CB|D}$, for example, which tracks distinct pre- and post-intervention versions of A, cannot generally be explained as a probabilistic mixture of purely CC and purely CE relations, even if A and B are only qubits. This possibility is further explored in chapter 5.

[20] To see this, consider the representation of the two-dimensional degenerate ellipsoid in terms of the parameters $\vec{c}$ and $\vec{t}$, and note that one component of $\vec{t}$, say t_1, is zero. This ellipsoid can be obtained from a convex combination of two other inference ellipsoids with $\vec{c}' = \vec{c}'' = \vec{c}$ and $t'_{2,3} = t''_{2,3} = t_{2,3}$, but with $t'_1 = -t''_1 \neq 0$. If one were to choose $|t'_1|$ too large, these may not be valid inference ellipsoids (they would exceed the unit sphere), but because the original ellipsoid parametrized by $\vec{c}, \vec{t}$ with $t_1 = 0$ is contained inside the unit sphere, there exists some permissible range of non-zero $|t'_1|$ that ensures the same. The inference ellipsoids parametrized by $\vec{c}', \vec{t}'$ and $\vec{c}'', \vec{t}''$ also violate the nested tetrahedron condition, but have well-defined handedness: one left-handed, one right-handed. It follows that one part of the convex decomposition admits a purely CC explanation, while the other admits a purely CE explanation.

4.2 Experiment

This section describes a linear optics experiment[21] that puts the results of the previous section, specifically theorem 40, into practice. Going beyond a simple decision problem between purely CC and purely CE relations, the task here is quantitative: given a promise that the relation between two qubits is a probabilistic mixture of an unknown quantum channel (CE) and an unknown bipartite state (CC), we aim to estimate the probability with which each scenario is realized, using only observational data. Theorem 39 shows that observational data provides a complete characterization of the linear inference map $\tau_{B|A}$, assuming such a map exists, which is the case if we are promised a probabilistic mixture of CC and CE relations. If one is furthermore promised a mixture of *extremal* CE and CC relations, i.e. a unitary channel and a maximally entangled state between two qubits, this problem admits a unique solution: observational data is sufficient to completely characterize the unitary, the entangled state and their relative weights. This is shown by an explicitly geometric construction in appendix A.4.

The probabilistic mixture of CC and CE relations between two qubits is implemented by a version of the circuit introduced in Fig. 2.3d. Using the notation C and D to distinguish the pre- and post-intervention versions of qubit A, we can specify the circuit elements as follows: two qubits are prepared in the maximally entangled state defined in eq. (2.14),

$$\rho_{CE} = \left|\Phi^+\right\rangle \left\langle\Phi^+\right|, \tag{4.18}$$

and C is measured, after which it is relabelled D. Qubit D is then recombined with E in a two-qubit gate to obtain B and F: letting $\mathcal{I}_{B|D}$ denote the identity channel that transforms D into B (and similarly for other pairs), let

$$\mathcal{E}_{BF|DE} = (1-q)\,\mathcal{I}_{B|D} \otimes \mathcal{I}_{F|E} + q\mathcal{I}_{B|E} \otimes \mathcal{I}_{F|D}. \tag{4.19}$$

This is followed by a partial trace over the qubit F, effectively defining the gate $\mathcal{E}_{B|DE} \equiv Tr_F \mathcal{E}_{BF|DE}$, whose output is B. The probabilistic swap gate $\mathcal{E}_{BF|DE}$ is the central element: with probability $1-q$, it routes the input D to B, implementing a CE relation between A and B, whereas with probability q, it routes E to B, realizing a CC relation between A and B.

The experimental setup is shown in Fig. 4.5. The qubits are encoded in the polarization degree of freedom of photons. The initial preparation generates pairs in the polarization-entangled state $|\Phi^+\rangle$, achieved with 98.5% fidelity, by parametric down-conversion. The gate $\mathcal{E}_{BF|DE}$ is implemented using a displaced Sagnac interferometer (chosen for the passive stability it provides), which realizes either a two-photon identity gate or a swap depending on the phase difference between the clockwise and counter-clockwise paths. The choice is controlled by a liquid crystal retarder (LCR) which, when voltage is applied, introduces a phase shift between the paths. (As shown in Fig. 4.5, the LCR itself extends across both paths, but one can ensure that it affects the two paths differently by inserting wave-plates, which can address each path separately, at various points in the interferometer.) The LCR is switched on or off at random at $.2s$ intervals, with a given probability q^{exp} of being switched on for each interval. On average, this realizes a probabilistic mixture of $(1-q^{exp})$ identity and q^{exp} swap. At the end, the two photons that exit the gate, B and F, are detected in coincidence: only when both detectors click within a window of 3 ns do we conclude that the initial preparation did in fact produce an entangled pair, and count the run.

The polarization of photons is measured using a polarizing beam-splitter (PBS) in conjunction with a single-photon detector at the end of the circuit: if the detector clicks, then the photon passed the PBS and therefore was polarized horizontally at that point. Conversely, if the photon was vertically polarized, it is sent to a dump and the final detector will not click. By itself, a PBS

[21] The experiment was designed, realized and analysed in collaboration with M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens and K. J. Resch and a description of it was published in [2] (see statement of contributions in the front matter).

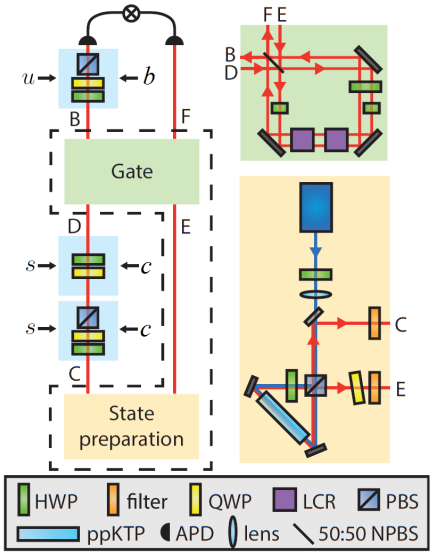

Figure 4.5: Experimental setup implementing a probabilistic mixture of CC and CE relations between two qubits encoded in the polarization of photons. Coloured boxes distinguish functional components of the setup that correspond to the different elements of the abstract circuit (Fig. 2.3d): preparation of an entangled state (yellow), measurement of photon A (blue), probabilistic swap gate (green) and measurement of photon B (yellow). Notation for optical components: half-wave plate (HWP), quarter-wave plate (QWP), liquid-crystal retarder (LCR), polarizing beam splitter (PBS), non-polarizing beam splitter (NPBS), periodically poled KTP crystal (ppKTP), avalanche photo diode (APD).

selectively transmits one of two eigenstates of σ_z. By combining a PBS with half- and quarter-wave plates, one can select other polarization states in a similar manner. This paradigm is different from that of an idealized measurement, which gives different non-null outcomes for both polarizations. However, assuming that photons are produced at a constant rate, on average, one can compensate for the difference: running an experiment with a PBS-measurement set to select for each eigenstate of a given observable for equal periods of time t and counting the number of clicks generated for each eigenvalue, one obtains the same statistics (on average) as by running an experiment with an ideal measurement for a single period t and counting how many times the apparatus produces each output. In order to implement a measurement that leaves the photon in the same state as before the interaction, as is required for an observational scheme, one must include not only one set of wave-plates before the PBS, which transforms whatever incoming state $|\psi\rangle$ one wishes to select for to the horizontal state so that the photon can pass the PBS, but also a second set after that PBS, which transforms all horizontally polarized photons that passed the PBS back to the same state $|\psi\rangle$. Note that this is not an *independent* repreparation, making it fundamentally different from an interventionist scheme, and therefore insufficient for full causal tomography.

Data are collected for 36 combinations of measurement settings, ranging over the six Pauli observable eigenstates in measurements on A and B and counting coincidences during $5s$ for each combination of settings. Following the notation introduced in Fig. 2.5, the eigenstate found in the measurement on A is indexed by the choice of Pauli observable $s \in \{1,2,3\}$ and the eigenvalue $c \in \{\pm 1\}$ and denoted $\Pi^{s,c}$, and similarly $u \in \{1,2,3\}$ and $b \in \{\pm 1\}$ specify a projector on B. We denote the count numbers observed for each combination $cbsu$ by $\tilde{P}^{obs}(cbsu)$, noting that they are proportional to joint probabilities of outcomes cb and settings su, due to the way the measurement is realized.

The data can be analysed by a straightforward least-squares fit to a probabilistic mixture of CC and CE terms. In the model, the CC component is parametrized by the positive-semidefinite, trace-one operator ρ_{CB}, which is simply a joint state. The CE component is parametrized by the marginal ρ_C that enters the measurement apparatus together with a PPT conditional $\tau_{B|D}$ that describes how B depends on the earlier variable. The probability of realizing a CC structure, which is what we seek to determine, is encoded in a weight $0 \leq q \leq 1$. The probability of obtaining a click for a combination of settings $cbsu$, according to this model, takes the form

$$P^{mod}(cbsu) = qTr\left[\Pi_B^{u,b} \otimes \Pi_C^{s,c} \rho_{BC}\right] + (1-q)\, Tr\left[\Pi_C^{s,c} \rho_C\right] Tr\left[\Pi_B^{u,b} \otimes \Pi_D^{s,c} \tau_{B|D}\right]. \tag{4.20}$$

In order to predict the observed count numbers, we introduce an additional model parameter N, which corresponds to the number of entangled pairs generated during the period of data collection for each combination of wave-plate settings $cbsu$. This is assumed to be equal for all $cbsu$, based on the facts that we collect counts for each $cbsu$ for the same amount of time and that entangled pairs are generated at an approximately constant rate, when averaged over the relevant time-scale. The count numbers predicted by the model can therefore be written as

$$\tilde{P}^{mod}(cbsu) = NqTr\left[\Pi_B^{u,b} \otimes \Pi_C^{s,c} \rho_{BC}\right] + N(1-q)\, Tr\left[\Pi_C^{s,c} \rho_C\right] Tr\left[\Pi_B^{u,b} \otimes \Pi_D^{s,c} \tau_{B|D}\right] \tag{4.21}$$

and one seeks parameters of the model that minimize the residue

$$\chi^2 \equiv \sum_{cbsu} \frac{\left[\tilde{P}^{mod}(cbsu) - \tilde{P}^{obs}(cbsu)\right]}{\tilde{P}^{mod}(cbsu)}. \tag{4.22}$$

Note that we do not assume in the model that ρ_{CB} is a maximally entangled state, nor that $\tau_{B|D}$ represents a unitary transformation. We only made these assumptions in the theoretical analysis of the scenario in order to show that they ensure a unique solution, but do not expect them to hold exactly in any realistic description of the experiment.

Since normalization of ρ_{CB}, ρ_C and $\tau_{B|D}$ can only be enforced in the fit by adding penalty terms to the target function, it is more efficient to absorb the parameters N and q into ρ_{CB} and ρ_C, replacing them by unnormalized operators $\tilde{\rho}_{CB} \equiv qN\rho_{CB}$ and $\tilde{\rho}_C \equiv (1-q)N\rho_C$:

$$\tilde{P}^{mod}(cbsu) = Tr\left[\Pi_B^{u,b} \otimes \Pi_C^{s,c} \tilde{\rho}_{BC}\right] + Tr\left[\Pi_C^{s,c} \tilde{\rho}_C\right] Tr\left[\Pi_B^{u,b} \otimes \Pi_D^{s,c} \tau_{B|D}\right], \tag{4.23}$$

in terms of which the relevant parameter q takes the form

$$q = \frac{Tr\tilde{\rho}_{CB}}{Tr\tilde{\rho}_{CB} + Tr\tilde{\rho}_C}. \tag{4.24}$$

A convenient parametrization of positive-semidefinite operators is provided in [48]: for a 4×4 operator, define the lower-triangular matrix

$$J = \begin{pmatrix} j_1 & 0 & 0 & 0 \\ j_2 + \mathrm{i}j_3 & j_4 & 0 & 0 \\ j_5 + \mathrm{i}j_6 & j_7 + \mathrm{i}j_8 & j_9 & 0 \\ j_{10} + \mathrm{i}j_{11} & j_{12} + \mathrm{i}j_{13} & j_{14} + \mathrm{i}j_{15} & j_{16} \end{pmatrix}, \tag{4.25}$$

specified by the 16-component real vector $\vec{j}$, and take

$$\tilde{\rho}_{CB} = J^\dagger J. \tag{4.26}$$

This form is positive-semidefinite by construction, requiring no further constraints to enforce the desired property. Furthermore, the trace of the resulting operator is simply

$$Tr\left(J^\dagger J\right) = \sum_i j_i^2 = \left|\vec{j}\right|^2. \tag{4.27}$$

A similar parametrization can be used for $T_D\left(\tau_{B|D}\right)$, to ensure that the conditional $\tau_{B|D}$ itself is PPT, and for the 2×2 matrix $\tilde{\rho}_C$, which requires only four real parameters. In order to assess how well the data fit a model with a fixed probability q^{test}, we add a suitable penalty term to the function χ^2 to be minimized,

$$\lambda_q \left(\frac{\left|\vec{j}_{CB}\right|^2}{\left|\vec{j}_{CB}\right|^2 + \left|\vec{j}_C\right|^2} - q^{test} \right)^2 . \tag{4.28}$$

The remaining constraint, the normalization requirement on quantum conditionals, $Tr_B \tau_{B|D} = \mathbb{I}_D$, is also enforced with a Lagrange multiplier,

$$\lambda_D \sum_{ij} \left| \left(Tr_B \tau_{B|D} - \mathbb{I}_D \right)_{ij} \right|^2 . \tag{4.29}$$

The values of the multipliers are chosen heuristically so as to make the typical value of the penalty term similar in magnitude to the residue χ^2, since smaller values of λ make the penalty terms ineffective at enforcing the constraints; whereas larger λ allow the penalty terms to dominate over the residue χ^2, compromising the sensitivity of the fit.

Fig. 4.6 shows the results from 21 experimental runs implementing different probabilities of common-cause relation, q^{exp}: the data from each run are fitted to the model of a generic probabilistic mixture, ranging over different values of the CC probability q^{test} and tracking the quality of the fit. This reveals a well-defined band of relatively good fits when $q^{test} \approx q^{exp}$, contrasting with much larger residues χ^2 – by several orders of magnitude – when q^{test} is far from q^{exp}. Our analysis therefore allows a conclusive inference from experimental data to the probability of CC implemented in the experiment, proving that it is possible to discern CC from CE relations despite the restriction to observational data. The black points in Fig. 4.6 mark the q^{test} that achieved the best fit for each q^{exp}. Over the 21 runs, the results of this method revealed the implemented q^{exp} with a root-mean-square deviation of only 0.043, as estimated by a Monte Carlo simulation. (To appreciate how small this deviation is, recall that the probabilities q are dimensionless quantities in the interval $[0, 1]$.)

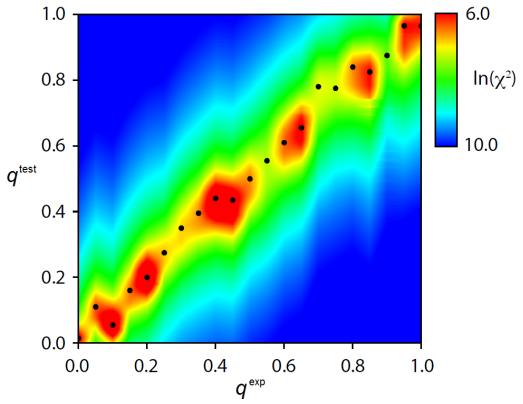

Figure 4.6: Experimental results on distinguishing CC and CE using only observational data. We implement probabilistic mixtures of CC and CE relations, with probability of CC $q^{exp} \in \{0, 0.05, 0.1, ...1.0\}$, and fit the observational statistics to the model (4.23) while enforcing different CC probabilities q^{test}. The residue χ^2 quantifies how much the data conflict with the model. The well-defined band of relatively good fits around $q^{test} \approx q^{exp}$ shows that one can estimate q^{exp} from observational data. Black points mark the best-fitting q^{test} for each value of q^{exp}.

One may note that the residue χ^2 when fitting the experimental data to a probabilistic mixture was generally rather high, considering that the number of free parameters is 5. (Each run yields $6 \times 6 = 36$ count numbers, while the model has 31 independent parameters after enforcing constraints.) The relatively poor fit can be traced to the implementation of the probabilistic swap: Recall that the LCR is switched on and off repeatedly while we collect counts for each combination of wave-plate settings *cbsu*. It is switched at fixed intervals, 25 times during each collection window, with probability q^{exp} of being switched on during any one interval. The overall fraction of each collection window during which the LCR is switched on is therefore not exactly q^{exp}, but rather sampled from a binomial distribution with 25 samples and mean q^{exp}. This implies that each of the 36 count numbers is obtained from a probabilistic mixture with generally slightly different fractions of CC and CE. This explains why the complete dataset for a given q^{exp} produces a surprisingly large χ^2 when fitting to a model with a single probability of CC. Despite this background noise, the high contrast between relatively good fits near $q^{test} \approx q^{exp}$ and worse fits when q^{test} is far from q^{exp} proves that the uniquely quantum causal inference criterion can distinguish CC from CE clearly and decisively.

4.3 Causal analysis of open systems dynamics

The techniques described earlier in this chapter, along with the conceptual framework of quantum causal models, may also prove useful in the study of open quantum systems. Specifically, the causal perspective suggests a definition of non-Markovianity that extends naturally to quantum systems, as well as tools for detecting the phenomenon. Quantum causal models also offer a clear physical interpretation and a suitable mathematical formalism for explaining not completely positive maps, which are often dismissed as unphysical in the conventional framework, even though they have been known to arise both in theory and in experiment [49, 50, 51, 52, 53, 54]. Before we turn to the contributions of quantum causal models, we begin with a brief review of Markovianity and complete positivity in the conventional framework.

The scenario is the following: a system S evolves while interacting with its environment, E. Following the general formalism used in this thesis, the labels S_i and E_i refer to quantum systems at individual points in time, specifically the principal system and the environment with which it interacts, respectively, at various times t_i. To each of these quantum variables we can assign a separate Hilbert space and a static quantum state that encodes our information about the system, just as a classical variable has a set of possible values and one can express one's knowledge about the variable as a probability distribution over this set, with the state or probability distribution for a given variable generally depending on its causal parents, i.e. the variables of the preceding time step. The objects of interest are the discrete time series of states on $\{S_i\}$, for the discussion of Markovianity, and the maps representing individual time-steps, in particular how S_2 depends on S_1, for the discussion of complete positivity.

4.3.1 The Markov condition and complete positivity

We being by considering the definition from classical statistics:

Definition 55. A time sequence of (classical) variables $\{S_i\}$ is termed a *Markov chain* if each element is explicitly dependent only on its direct ancestor: $P(S_{i+1}|S_i, S_{i-1}, ...) = P(S_{i+1}|S_i)$ [55].

A Markov chain is often described as having "no memory" in the sense that its previous history, $\{S_{i-1}, S_{i-2}, ...\}$, provides no additional information for predicting its next step, S_{i+1}, beyond what is encoded in the current S_i. Markov chains are the preferred model for a broad range of classical phenomena, from chemistry and economics to natural language modelling and game theory. There have been a number of attempts to generalize this notion to time sequences of quantum states, which constitute snapshots of the dynamics of a quantum system that interacts with its environment. The proposals include indicators of non-Markovianity based on effects that cannot occur

under CP maps, such as increasing distinguishability between two quantum states [56] or increasing entanglement with an ancilla [57], and formal conditions on the Lindblad master equation (for the case of continuous-time evolution) [58, 59, 60], among others. However, the definitions that have been proposed are generally not equivalent, and no single notion of quantum non-Markovianity is forthcoming [61, 62, 57, 63]

Another common simplifying assumption about the dynamics of open quantum systems, which is related to the Markov property, is that the map S_1 to S_2 is completely positive, obeying definition 14. In physical terms, this ensures that, even when the input into the channel is part of an entangled state with an ancilla, the resulting output is still a positive-semidefinite operator, so that the probabilities it predicts for any measurement outcomes are strictly non-negative. A model that predicts negative probabilities is deemed unphysical, hence all maps representing quantum channels are required to be completely positive. This assumption plays a large role in the experimental characterization of unknown quantum channels (quantum process tomography, abbreviated QPT), which is essential for benchmarking quantum information processing devices. Among other implications, demanding that the reconstructed map be CP makes the fitting of experimental data more robust [64]. Complete positivity also makes many analytical manipulations (for example in quantum information theory) more tractable, since every CP map can be put in the Kraus form [65].

In order to ensure that the map from S_1 to S_2 is completely positive, one normally assumes that the initial state of the system, S_1, is uncorrelated with any environmental degrees of freedom E_1 that may affect its subsequent evolution [66, 49, 50, 51]. There are also other ways to ensure complete positivity of the map: for instance, if the correlation time of the environment is much shorter than the characteristic time-scale of its interaction with the system, so that any correlations between them decay before causing a significant back-action, or if there is some other mechanism preventing the record of the system's initial state that is stored in the environment from affecting the subsequent evolution and thereby influencing the final state of the system, then one also recovers CP dynamics[22].

In many scenarios of practical relevance, one can ensure that at least one of these sufficient conditions holds to good approximation. However, in general, there is no reason to assume that a quantum system is, at any given time, not correlated with its environment, especially if one considers how challenging it is to isolate quantum systems effectively. Similarly, it is extremely unlikely that all information stored in the environment either decays sufficiently quickly or is completely irrelevant to the subsequent evolution of the system due to some other mechanism. Indeed, numerous experiments in QPT have yielded reconstructed maps that are not quite CP, and in several cases the seemingly unphysical results can be traced back to initial correlations [53, 54, 52, 68, 69].

A realistic study of general open quantum systems dynamics will therefore require a more robust framework, including a mathematical formalism and the necessary conceptual underpinnings, for describing dynamics beyond completely positive maps. Before that, however, we seek a succinct characterization of the conditions that lead to complete positivity. This characterization takes a particularly simple form in terms of causal relations and, as we will show, is closely related to a quantum version of the Markov condition.

4.3.2 The causal perspective on Markovianity and complete positivity

We will argue that the language of causal models provides a concise statement of the conditions that ensure Markovianity and complete positivity of the map from S_1 to S_2. To show this, we will begin by recasting the conventional statement of the Markov condition in terms of causal structure:

[22]The so-called Markov approximation, which states that the "memory" about the principal system S that is encoded in its environment decays quickly compared to time-scale of S, is instrumental in deriving the Lindblad form of the master equation governing the continuous-time evolution of an open quantum system. The Lindblad form, in turn, ensures that the map relating the states of the principal system at any two times is completely positive. A pedagogical discussion in given in [67].

Lemma 56. *A causally ordered set of variables $\{S_i\}$ forms a Markov chain if and only if they admit a causal model wherein each S_i is (directly) influenced only by the immediately preceding S_{i-1}.*

Proof. Considering that the variables are causally ordered, and assuming that the indices reflect this order, we know that the joint probability distribution over the set of observed variables $\{S_i\}$ must factorize as

$$P\left(\{S_i\}_i\right) = \Pi_i P\left(S_i|S_{i-1}, S_{i-2}, ...\right), \tag{4.30}$$

with each variable potentially depending on all the variables that precede it. There are two ways in which this scenario can violate the conventional statement of the Markov condition (definition 55), which are illustrated in Fig. 4.7: (a) if the next S_{i+1} is directly influenced not only by the present S_i, but also by some earlier S_{i-n} $(n > 0)$; and (b) if the $\{S_i\}$ do not admit a faithful causal model at all, because two or more S_i are directly influenced by an external E_j that is not part of the observed set $\{S_i\}$. Both instances of non-Markovianity can be traced back to the existence of common causes between different S_i: in one case, the element S_{i-n} of the sequence is a common cause of S_{i+1} and of its own immediate successor, S_{i-n+1}; in the other case, the external (unobserved) E_j is a common cause of two distinct variables, say S_{i-n} and S_i. The converse, that a causal structure that is a purely CE chain implies that the $\{S_i\}$ form a Markov chain, follows directly. □

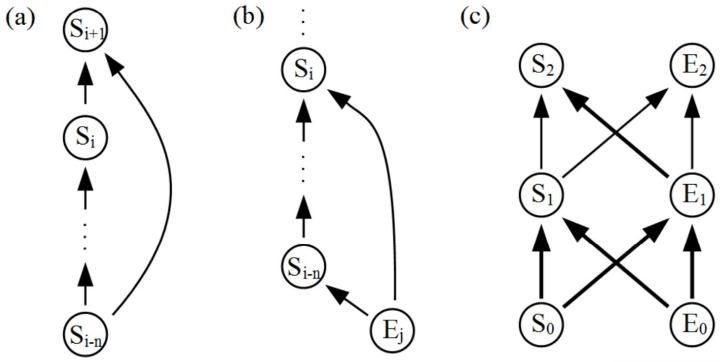

Figure 4.7: Non-Markovianity and not complete positivity in terms of causal structure. Given a time series of variables $\{S_i\}$, there are two features in the causal structure that can lead to a violation of the Markov condition: (a) a variable S_{i-n} acts as a common cause of two later variables, namely its own child S_{i-n+1} and a later S_i, or (b) an unobserved variable E_j, which is not part of the sequence $\{S_i\}$, acts as a common cause of two distinct S_i. (c) The interaction between a system S and its environment E over the course of discrete time steps (denoted by sub-indices) can be represented by the causal structure shown here. If the causal relation between S_1 and S_2 is purely CE, then the map that represents it is CP. However, if there are mechanisms that realize some amount of CC relation, such as correlations in the initial state of S_1 and E_1 (due to the common causes S_0 and E_0) and a subsequent influence of E_1 on S_2 (bold), then the map from S_1 to S_2 may be not completely positive.

Both of the causal structures that violate the Markov condition can be interpreted as evidence of memory effects, since they require some external memory – outside the sequence $\{S_i\}$ itself – to store the information about how the common cause will affect later variables. Once the Markov condition has been cast in terms of causal structure, the generalization from classical stochastic processes to quantum dynamics is straightforward: if the $\{S_i\}$ denote a time series of *quantum* variables, it still holds that the sequence exhibits non-Markovianity if and only if it cannot be modelled as a chain of purely CE relations, but exhibits some evidence of CC relations as well.

The perspective of causal models also offers a much more concise statement of the various effects involved in violating complete positivity. Fig. 4.7c shows the general relation between S_1 and S_2 as part of a larger causal structure capturing the ongoing interaction with the environment. One can see that all the effects that lead to a violation of complete positivity – correlations between S_1 and E_1 due to a common cause in their shared past, the persistence of this information in E_1 for a sufficiently long time and its back-action on the system at a later time, S_2 – together form a single causal path which connects S_1 to S_2 via the common causes S_0E_0 shared by S_1 and E_1 and the subsequent influence of E_1 on S_2. In short, if the causal relation between S_1 and S_2 is purely CE, then the map is CP. It follows that not completely positive maps and the complications in conventional QPT schemes arise only if the causal relation between S_1 and S_2 is not simply CE, but has a non-trivial CC component as well.

4.3.3 Representation in the quantum causal models framework

The above discussion suggests that a general representation of the relation between the states of a quantum system at two times, S_1 and S_2, must capture not only the CE influence, but also the inference along a possible CC path connecting S_1 and S_2. It follows that there may not generally exist an inference map of the form $\mathcal{E}_{S_2|S_1}$, as one might have expected. Instead, a full description of the relation between S_1 and S_2 requires one to distinguish pre- and post-intervention versions of S_1, which we denote C and D respectively, and takes the form of a causal map $\mathcal{E}_{S_2C|D}$.

A formally similar solution to the problem of QPT in the presence of initial correlations was proposed by Modi [39], who describes the circuit relating S_1 to S_2 as a map from instruments on S_1 to states on S_2. In our notation, instruments for probing S_1 are sets of maps $\left\{\mathcal{E}^{xy}_{D|C}\right\}_y$, indexed by outcomes y, and a map from such instruments to outputs S_2 takes precisely the form of a causal map, although the physical interpretation is slightly different.

It is also possible to define a map from simple quantum states on S_1, rather than instruments relating C and D, to the corresponding states on S_2, if one fixes a so-called assignment map: this map extends each ρ_{S_1} to a particular bipartite state $\rho_{S_1E_1}$ (such that $Tr_{E_1}\rho_{S_1E_1} = \rho_{S_1}$), which encodes the initial correlations with the environment and subsequently evolves under the global unitary. This approach to modelling quantum dynamics in the presence of initial correlations has been explored in several works [70, 51, 71, 72], but it was recognized from the outset that the assignment maps that give rise to non-CP maps must violate at least one of a list of reasonable properties, such as linearity [49]. This undesirable effect likely arises because assignments maps are intended to map states on S_1 to joint states with E_1, without taking into account how one came to assign a given state to S_1 in the first place. Since learning about a quantum variable generally disturbs it, in the operational sense that one comes to assign different beliefs to the pre- and post-intervention versions, the information about how the variable was probed should not be ignored. The framework of causal models suggest that we represent the implications of a shared common cause by an inference map rather than an assignment map and provides a physical, operational interpretation of these maps as well as compelling grounds for their defining mathematical properties.

To underscore this point, let us explore how the causal description of open system dynamics circumvents the issue of negative probabilities, which is the reason why not completely positive maps are forbidden in the conventional framework. Suppose that a purely CE influence of S_1 on S_2 was described by a positivity-preserving but not completely positive map. In this scenario, negative probabilities would arise if the input S_1 was prepared in a particular[23] entangled state with an ancilla N. However, if S_1 is at least partially determined by the common cause it shares with E_1, which is precisely what gives rise to the non-CP map, then one cannot prepare S_1 in a state that is also generically entangled with the ancilla N, and in particular any ρ_{NS_1} that would lead to

[23]It follows from the very definition of complete positivity that for every positivity-preserving but non-CP map $\mathcal{E}_{B|A}$, there exists an input state ρ_{AN} that encodes entanglement with an ancilla N such that $\left(\mathcal{E}_{B|A} \otimes \mathcal{I}_N\right)(\rho_{AN})$ is not a positive-semidefinite operator.

unphysical predictions. One can see this by noting that a full description of the relevant systems begins with a tripartite operator $\rho_{NS_1E_1}$, which must be a valid quantum state, and follows a global unitary evolution, ensuring that the result at time t_2 is also a valid density operator. This analysis suggests a different take on the problem of non-CP maps: rather than forbidding all non-CP maps on the grounds that they produce unphysical outputs when applied to certain input states that are entangled with an ancilla, one could simply characterize the class of entangled inputs that are problematic for a given map, bearing in mind that they cannot be realized due to the way in which the non-CP map arose in the first place [50, 73].

Tests of non-Markovianity based on the quantum causal framework. Since we have reduced the necessary and sufficient conditions for complete positivity of the map and the applicability of conventional QPT to simply demanding that the causal relation be purely CE, we can now use previously developed tools to test when this condition is violated. Some of these tools formally reduce to known conditions for non-Markovianity and merely provide a physical interpretation for them. For example, one indicator of not purely CE relations is if there exists a linear map of inferences from S_1 to S_2 of the simple form $\mathcal{E}_{S_2|S_1}$, but the conditional $\tau_{S_2|S_1}$ is NPT. This is equivalent to saying that finding a non-CP map heralds a violation of circumstances assumed in conventional QPT. While this fact had been established mathematically, we hope to have provided a clearer physical account of what circumstances can explain the occurrence of a non-CP map.

However, building on previous chapters of this work, we can also propose new indicators of non-Markovianity. For example, in general there need not exist a linear map $\mathcal{E}_{S_2|S_1}$ at all. Evidence of this can be easily detected, even if one is restricted to an observational scheme, by ranging over an over-complete basis of projectors on S_1 in order to verify whether the mapping to states on S_2 is linear. If one has the power to intervene, probing S_1 with general instruments, one can pursue the more high-level strategy of performing full tomography and analysing the resulting causal map (see chapter 5.3 for some criteria). Unlike the tests mentioned above, which are only sufficient conditions for ruling out Markovianity, full causal tomography yields a complete characterization of the inference map from S_1 to S_2 – including any memory effects due to the environment – and therefore allows one to determine conclusively whether this step in the dynamics admits a Markovian model or not.

These tests allow one to characterize the dynamics of an open quantum system without requiring the usual assumptions embodied in QPT, but they do rely on a different assumption: both observational and interventionist schemes require one to probe S_1 without affecting the relevant environmental degrees of freedom. However, the problem of initial correlations in QPT arises precisely because one generally cannot prepare (a complete set of states on) S_1 without also affecting the environment E_1. In order to make the above results more applicable, one should therefore combine the insight that general relations between two time-ordered variables must be represented by causal maps, which distinguish two copies of the early variable, with a detailed model of the instruments that can be used to probe S_1 in a given implementation, including how they affect the environment. One can then derive what one can learn about the causal map – and consequently deduce about the causal relation – subject to these constraints, similarly to what we have done for the restriction to observational probing in 3.3.

4.4 An operational interpretation of PPT states

In the field of quantum information theory and its applications to information processing, many key questions concern the resource of entanglement: for example, how to verify its presence or absence in experiments, how to construct new families of entangled states and how to classify different types of entanglement, for example. Operators that are both positive-semidefinite and positive under partial transposition – that is, PPT states – play a central role in many of these problems, even though the exact physical implications of the PPT property are not clear. The framework of quantum causal

models suggests an interpretation of these states in terms of the causal structures that can give rise to them; a perspective which is naturally suited to addressing problems in communication and information processing, as well as deeper questions about entanglement and non-locality in quantum theory [74].

The most prominent role of PPT states in entanglement theory is the Peres-Horodecki criterion [75, 76], which establishes that a state of two systems A and B with dimensions $d_A \times d_B = 2 \times 2$ or 2×3 is separable if and only if it is PPT. In higher dimensions, while it is easy to show that every separable state is PPT, the converse does not hold: there are explicit examples of states that are PPT but possess non-zero entanglement; see for example [77, 78]. Separable states are therefore a proper subset of PPT states. PPT can instead be shown to imply zero *distillable* entanglement [79]; the existing entanglement in such states is termed "bound entanglement" by contrast. Entanglement distillation is relevant to practical quantum communication tasks because it allows two parties to use several imperfectly entangled pairs, such as might be distributed over a noisy channel, along with classical communication, in order to distil a smaller number of maximally entangled pairs, which can then be used to perform e.g. quantum teleportation with high fidelity [80, 81]. However, it is not clear whether PPT is a necessary and sufficient condition for ruling out distillable entanglement: it is an open question whether there exist states that are negative under partial transposition (NPT) but possess zero distillable entanglement. For a review of progress on this topic, we refer the reader to section XII.H of [82] and references therein, in particular [83, 84, 85].

One can formulate a necessary and sufficient condition for separability in terms of entanglement witnesses: a state ρ_{AB} is separable if and only if its image under $\mathcal{I}_A \otimes \mathcal{E}_{B'|B}$ is positive-semidefinite for all positivity-preserving but not completely positive maps $\mathcal{E}_{B'|B}$ [76]. Instead of characterizing the set of separable states directly, one can therefore characterize the set of positivity-preserving but not CP maps. The study of these maps, in turn, in particular of relevant properties such as decomposability, again relies heavily on PPT states [86].

Given the interest in operators that are both positive-semidefinite and PPT, it is worthwhile to consider the implications of these properties in the context of causal models:

Theorem 57. *A quantum conditional $\tau_{B|A} \in \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_A)$, which can be interpreted as a Jamiołkowski operator representing a trace-preserving inference map from A to B, is both positive-semidefinite and PPT if and only if it can be explained both by a purely CC and by a purely CE causal relation.*

Proof. This follows directly from propositions 18 and 25. □

Before we give an operational interpretation of separable states, let us first introduce the following:

Definition 58. The causal relation between two quantum variables A and B is said to be *classical cause-effect* if and only if it can be modelled by a directed causal path from one to the other that is blocked by a classical variable X, e.g. $A \to X \to B$. Similarly, the relation is termed *classical common-cause* if and only if it can be modelled as A and B sharing a classical variable X as their (sole) common cause.

Fig. 4.8 illustrates the two scenarios. A classical cause-effect relation can be realized as a measurement on A that yields a (classical) outcome X, followed by a preparation on B that depends on X. Note how the restriction to a classical variable at one point along the causal path precludes any quantum coherence between its endpoints: the channel from A to B that describes such a cause-effect relation is completely decohering. Similarly, if A and B share a classical common cause, then any inference between them must follow the causal path, which is blocked by the classical variable X. Consequently, the map $\mathcal{E}_{B|A} = \mathcal{E}_{B|X} \circ \mathcal{E}_{X|A}$ that represents such inferences is also completely decohering. It follows that the joint states ρ_{BA} that can arise from such a structure cannot exhibit any entanglement between A and B, but only classical correlations. Building on these observations, one can reach the following conclusion:

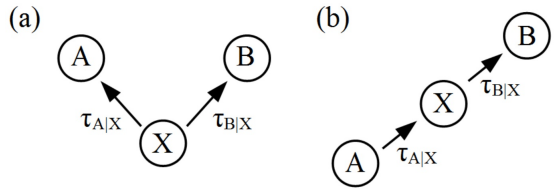

Figure 4.8: (a) Classical cause-effect and (b) classical common-cause relations arise if two quantum variables A and B are connected by a single causal path and this path is blocked by a classical variable X. The canonical example of a classical cause-effect relation is realized by a measure-and-reprepare channels. A classical common-cause structure is the key assumption in deriving Bell-type inequalities.

Theorem 59. *A quantum conditional $\tau_{B|A} \in \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_A)$, which can be interpreted as a Jamiołkowski operator representing a trace-preserving inference map from A to B, is separable if it can be explained either by a classical common-cause relation or by a classical cause-effect relation. Moreover, if $\tau_{B|A}$ is separable, then it can be explained both by a classical common-cause relation and by a classical cause-effect relation.*

Proof. A proof is provided in appendix A.5. □

As we pointed out above, purely mathematical considerations are sufficient to establish that separable states are a strict subset of PPT states. However, considering the causal structures that can give rise to the various classes of conditionals – positive-semidefinite, PPT, separable – offers an operational interpretation of the distinctions between them. The three-way relations between mathematical properties of bipartite operators, their correlations and the causal explanations that can account for them are summarized in Table 4.1.

state	correlations
product	uncorrelated
separable (PPT)	LOCC preparable
non-separable PPT	bound entanglement
NPT*	bound entanglement*
NPT	distillable entanglement

conditional	possible causal explanations
product (Pos and PPT)	disconnected
separable (Pos and PPT)	classical CC or classical CE
non-separable Pos and PPT	quantum CC or quantum CE
Pos and NPT	not purely CE, but quantum CC
Neg and PPT	not purely CC, but quantum CE

Table 4.1: The mathematical properties of bipartite quantum states have operational implications in terms of the correlations that the states can exhibit. For the most part, certain mathematical properties are necessary and sufficient conditions for the corresponding type of correlations, but it remains an open question whether there exist bound entangled NPT states (*). We study instead what the mathematical properties of any bipartite Hermitian operator $\tau_{B|A}$ with $Tr_B \tau_{B|A} = \mathbb{I}_A$ (which makes it Jamiołkowski isomorphic to trace-preserving inference maps) imply about the causal structures that can give rise to it. This approach reveals a strict one-to-one mapping between mathematical properties and operational statements that are directly relevant to tasks in communication and information processing.

5 Classical and non-classical causal structures

The previous chapters illustrate how even a simple modification to classical causal models, of replacing classical random variables with quantum variables and extending conditional probability distributions to suitable operators, produces a number of interesting effects. For example, even if the causal relation between two variables is known to be either purely CE or purely CC, one can meaningfully ask whether the causal relation in each case is effectively classical, in the sense introduced in section 4.4 (definition 58). If the conditionals relating the two variables do exhibit distinctly quantum effects, these can be exploited operationally as detailed in sections 4.1 and 4.2. So far, however, we have maintained the assumption that the overall causal *structure*, i.e. the qualitative specification of which variables influence which others, is no different from what one could have in classical causal models.

Definition 60. The term "classical causal structure" refers to a pattern of causal influences that constitutes a valid causal structure in the framework of classical causal models; that is, such that the causal influences between individual variables are represented by the edges of a single directed graph with no causal cycles.

Giving up this assumption, allowing for more exotic causal structures, constitutes a far more radical generalization of causal models.

The most fundamental motivation for exploring this possibility is the search for a theory of quantum gravity, as discussed in section 1.2. Such a future theory must allow one to recover conventional causal models and quantum mechanics as limiting cases, but as we go beyond these cases and begin to explore novel physical effects, it is only to be expected that causal structure itself becomes ill-defined and subject to typically quantum effects. It has also been predicted that novel causal structures would constitute a resource for computation [87], for instance allowing one to distinguish certain channels perfectly that are otherwise indistinguishable [88]. For the task of determining whether a set of unitary gates commute or anti-commute, undefined causal structures provide an exponential advantage in query complexity, as was verified in a recent experiment [89, 90].

A general mathematical representation of how two quantum variables can be related if they are not constrained to a conventional causal structure is derived by Oreshkov et al. in [38]. The so-called process matrix introduced in [38] is formally similar to the operator $\tau_{\mathcal{C}|\mathcal{D}}$ from proposition 28 and a more general case of the causal map (definition 30). Oreshkov et al. consider two parties, Alice and Bob, each of whom receives a quantum system in their laboratory and is allowed to interact with it according to the rules of standard quantum mechanics before sending the system out again. Crucially, they do not assume that the relation between the two laboratories is captured by a causal structure with the properties usually demanded in classical causal models, such as a well-defined global causal ordering of all variables. In the model of Oreshkov et al., there is no global background time that would distinguish whether Alice interacts with the system before or after Bob does. This opens the door to causal cycles and grandfather paradoxes:

Example 61. Suppose that Alice receives a classical bit, observes its value and sends it out again, unchanged. The bit reaches Bob, who also observes its value but then flips the bit before it leaves his lab. If the model allows causal cycles, the bit can then once again appear at the incoming port of Alice's lab, leading to a logical contradiction: the value that Alice observes the bit to have should be both the opposite of what Bob observed, since Bob flipped the bit, and the same, since Alice did not flip the bit before sending it to Bob.

Such paradoxes can arise regardless of whether the carriers of information are classical or quantum systems. In order to explore more exotic causal structures without risking contradictory predictions, Oreshkov et al. therefore impose the following requirement on the process matrix: for any experiments that Alice and Bob may choose to perform in their individual labs, the statistics predicted

by the process matrix must define a valid probability distribution over measurement outcomes, consisting of non-negative real numbers whose sum over all possible outcomes is one. In the case of a causal cycle, for example, there is always some choice of the actions in the local labs that gives rise to a grandfather paradox, as described in the example above. In this case one cannot even assign probabilities to combinations of outcomes observed by Alice and Bob in a self-consistent manner. The constraint proposed by Oreshkov et al. is therefore sufficient to exclude those causal structures that inevitably lead to logical contradictions. However, there remain a number of causal structures that satisfy the consistency constraint and yet could not be accommodated in a (single) classical causal model, for instance, convex combinations of one term in which A causally precedes B and another with the opposite ordering.

The constraint on the process matrix implicitly defines a constraint on the causal structures that can be permitted without risking logical paradoxes. However, it would be preferable to characterize the space of allowed scenarios directly in terms of causal structures rather than process matrices, since the causal structure provides a much clearer representation of how information flows between events and may even be the underlying physical element that gives rise to an emergent space-time. It is clear that the consistency condition is satisfied by any causal structure that is allowed in classical causal models. Moreover, the simplest classical causal structures, consisting of just a single causal influence connecting two variables, constitute a natural building block of more complex patterns of causal influences. We are therefore concerned with ways in which one can *combine* individual causal influences, or, more generally, various causal structures that are (separately) allowed in classical causal models.

One example of considerable interest involves a combination of opposite causal orderings: one component has A influencing B, while the other has B influencing A. This scenario provides the advantages mentioned in the introduction, such as gate discrimination or distinguishing commuting and anti-commuting unitaries [88, 89, 90]. However, considering that most current theories take it as an implicit and fundamental assumption that events (such as interacting with a system to prepare or measure it) take place in a well-defined causal order, realizing a combination of two conflicting causal orderings would require either post-selection or exotic new physics. It has been suggested that the desired effect could be achieved by placing a large mass in a superposition of position eigenstates, which would produce a coherent combination of speeding up and slowing down the proper time at each location, and consequently a coherent combination of the causal ordering of two events at those locations [91, 92]. However, maintaining coherence for sufficiently large masses and distances to produce a noticeable effect presents its own series of challenges.

In this work, we choose instead a different example to explore the hierarchy of possible combinations of causal structures: building on the results described in the previous chapters, we consider two time-ordered variables, A and B, whose causal relation is generally a combination of CC and CE. Since both causal structures are compatible with the same global causal order, this example allows for a straightforward experimental realization, which is discussed in more detail in section 5.4. Despite this simplifying property, the spectrum of ways of combining CC and CE relations is already much richer if A and B are quantum variables than if they were classical variables. We derive a rigorous sufficient condition for witnessing non-classicality in the way the two causal relations are combined. This kind of insight into what it means - both conceptually and mathematically - to combine two causal structures in a non-classical manner is expected to carry over to the study of more exotic combinations.

5.1 Classical ways of combining common-cause and cause-effect relations

The following discussion begins by focusing on the case of classical variables, since they are more accessible to our intuition. The hierarchy of combinations of CC and CE that can be established in the classical limit is then applied to the case where A and B are quantum variables, where it admits a natural extension.

The simplest combination of CC and CE relations, both mathematically and conceptually, is arguably a convex or probabilistic mixture. Such a structure arises if the late variable, B, is influenced either directly by A or by the common cause it shares with A: in terms of the conditionals introduced in 2.3c,

$$P^{prob}(B|DE) = qP(B|E) + (1-q)P(B|D), \tag{5.1}$$

with some weight $0 \leq q \leq 1$. This can be realized by a causal mechanism with the property that, in any given run of the experiment, B is influenced either only by D or only by E. A more general combination of CC and CE arises if $P(B|DE)$ does not take this form, for example

$$P(B|DE) = \delta(B, D \oplus E). \tag{5.2}$$

This cannot be expressed as a probabilistic mixture of B depending on either D or E because B is determined by a combination of D and E; that is, in any given run of the experiment, both D and E play a role in determining B.

Definition 62. Any combination of CC and CE that cannot be written as a probabilistic mixture is termed a *physical* mixture for contrast.

It is interesting to note that, even in the completely classical limit, the possibilities for combining two types of causal relation between A and B go beyond mere probabilistic mixtures. In fact, probabilistic mixtures of CC and CE are often rather contrived. A good example is again the relation between taking a medication (A) and the patient's recovery (B), which could be either directly cause and effect or related via a common cause, such as gender. In this scenario, one normally expects that the medication and, say, gender-specific hormones *interact* to determine the odds of recovery – that is, B depends on a non-trivial combination of its two causes, realizing a physical mixture of CC and CE relations. One could, however, imagine that there are in fact two different forms of the disease, which are hard to distinguish at the time of initial diagnosis. One of them responds only to medication, but not gender-specific hormones, whereas the other form of the disease runs its course unaffected by the medication, depending only on the patient's gender. In this case, the odds of recovery of a newly diagnosed patient, who could suffer from either variety of the disease, will depend either on whether they receive medication or on their gender. This constitutes a probabilistic mixture of CC and CE relations between medication and recovery. The existence of non-probabilistic mixtures of causal relations stands in contrast with the rules governing individual classical variables: if a classical variable X can take the values x_1 or x_2, then the only way to combine these two possibilities is to have a probabilistic mixture, with probabilities $\{P(X = x_1), P(X = x_2)\}$. Causal relations, on the other hand, can be combined in a greater variety of ways, even if the variables are only classical. By extension, the space of possible causal relations between quantum variables promises an even richer structure.

Berkson's paradox. One can devise a number of statistical tests to determine whether two variables A and B are related by a probabilistic mixture of CC and CE or by a physical mixture; some examples are described in section 5.3 below. The present section focuses on one test that offers both interesting conceptual underpinnings and a natural generalization that captures quantum coherence. The test is based on a phenomenon of classical statistics known as Berkson's paradox, which was first observed in the analysis of clinical studies [93], but the effect is equally well illustrated by an example from the academic world, depicted in Fig. 5.1a. Suppose, for simplicity, that the abilities at teaching and at research within the pool of applicants for a faculty position are statistically independent. At the same time, if one considers current faculty members, one may find a negative correlation between teaching and research abilities: good researchers tend to be bad teachers and vice-versa. This was deemed so counter-intuitive as to be termed a paradox. The explanation lies in the post-selection: in order to become a faculty member, a candidate must display competitive skills in at least one area, teaching and/or research. Conversely, knowing that someone was hired as faculty allows one

to retrodict that they were not incompetent at both teaching and research, thereby inducing negative correlations between otherwise independent variables. In short, post-selection on a common effect can induce correlations between its causal parents. In the causal structure considered here, post-selection on B can induce correlations between its parents D and E.

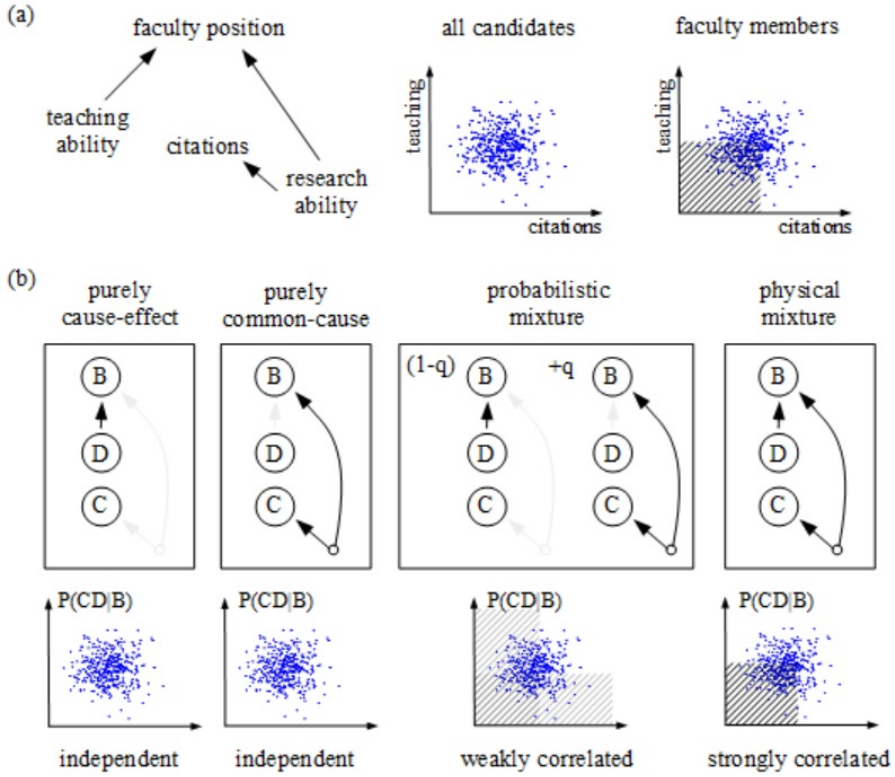

Figure 5.1: Berkson's paradox in academia and in an abstract causal structure. (a) We assume that research ability – and consequently the number of citations – is uncorrelated with teaching ability in principle. Post-selecting on a common effect, such as getting a faculty position, excludes some candidates (hatched) and thereby induces correlations. (b) The strength of the induced correlations between C and D reflects different combinations of causal relations between A and B. In the case of a probabilistic mixture, post-selection on B has either implications for D (with probability $1-q$) or for C (with probability q), for example eliminating low values of either D or C with the respective probabilities (lightly hatched, overlapping half-planes in the scatter plot). In the case of a physical mixture, on the other hand, post-selection on B may have simultaneous implications for D and C, for example eliminating the quandrant in which both D and C are small.

In order to apply this effect to the classification of causal relations, we recast the phenomenon in terms of B, D and C, which are the variables accessible in a tomographic characterization. Instead of two causal parents of B, we now have one parent, D, and one causal sibling, C, which descends from a second parent of B. By the same mechanism as before, post-selection on B can induce correlations between C (via E) and D. To continue with the example of an academic hiring process: the committee might decide that research ability is difficult to gauge directly and instead use the number of citations, which is a causal child of research ability, but not teaching ability.

Crucially, post-selection on B can only induce correlations between C and D if B is causally influenced by both D and E, in other words, if the relation between A and B has both CC and CE elements. For example, if a college bases its hiring decisions solely on teaching ability, then

the teaching ability of successful candidates will have no correlation with their publication record. Moreover, the *strength* of the correlations between C and D induced by post-selection reflects how the two influences are combined, thereby allowing us to classify combinations of CC and CE relations.

Example 63. To see this heuristically, consider the example of a physical mixture introduced in eq. (5.2), where B is given deterministically by the binary sum $P(B|DE) = \delta(B, D \oplus E)$, and assume furthermore that C is also binary and perfectly correlated with E, $P(CE) = \delta(C, E)\,\mu(C)$ where $\mu(C)$ denotes the uniform distribution. In this case, post-selecting on $B = 0$ implies that $D = E$, and consequently $D = C$, therefore inducing perfect correlations. In the case of a probabilistic mixture, by contrast, knowing B only allows one to make either a non-trivial inference about C (by means of E) or a non-trivial inference about D, but never about both variables at the same time. This is sufficient for inducing non-trivial correlations between C and D, but not perfect correlations.

Sections 5.3.3 and 5.3.4 below derive quantitative bounds on the strength of induced correlations in the general case of probabilistic mixtures.

Calculating induced correlations. For a formal derivation of the induced correlations between C and D, consider first the simpler problem, of quantifying the induced correlations between two causal parents, D and E. The relationship between D, E and their common effect B is in principle specified by the conditional $P(B|DE)$. In order to assess the correlations between D and E conditional on B, encoded in the Bayesian inverse

$$P(DE|B) = \frac{P(B|DE)\,P(DE)}{P(B)}, \tag{5.3}$$

one must also specify a prior distribution $P(DE)$, which may affect the correlations encoded in the resulting $P(DE|B)$. As an extreme example of the impact of the prior on $P(DE|B)$, suppose that the prior over D and E is a product of delta distributions, $P(DE) = \delta(d, 0)\,\delta(e, 0)$. Generally, $P(DE|B)$ will be zero for all (d, e) except the support of $P(DE)$ (i.e. those values for which $P(DE) \neq 0$), which in this case implies that $P(DE|B) = \delta(d, 0)\,\delta(e, 0)$: D and E are statistically independent when we condition on B. Any correlations that could have been induced by the post-selection on B are hidden due to the poor choice of priors. On the other hand, if the prior $P(DE)$ already contains correlations between D and E, then the post-selected distribution $P(DE|B)$ can contain the same correlations even if B is completely independent of D and E – the correlations in this case are not induced by the act of post-selection. In order to avoid these issues and isolate precisely the correlations that are due to post-selection, we will assume that we have no prior information about D and E, assigning the uniform prior, $P(DE) = \mu(D)\,\mu(E)$. That is, for the purpose of testing whether post-selection on B induces correlations, we consider specifically the distribution

$$P(DE|B) = \frac{P(B|DE)\,\mu(D)\,\mu(E)}{P(B)}, \tag{5.4}$$

where the marginal on B is obtained under the assumption of uniform priors,

$$P(B) = \sum_{DE} P(B|DE)\mu(D)\,\mu(E)\,. \tag{5.5}$$

The induced correlations between D and C, a causal parent and a causal sibling, follow analogously: in principle, the causal mechanisms relating C, D and B are represented by the joint distribution $P(CE)$, which encodes how C and E are correlated due to their common cause, and the conditional $P(B|DE)$, which encodes how D and E influence B. Since E is not observed, we combine them to obtain

$$P(CB|D) = \sum_{E} P(CE)\,P(B|DE)\,. \tag{5.6}$$

For the same reasons as before, we assign[24] a uniform prior to D, obtaining $P(CBD) = P(CB|D)\,\mu(D)$ and its marginal $P(B) = \sum_{CD} P(CBD)$, and use the conditional

$$P^b(CD) \equiv P(CD|B=b) = \frac{P(C, B=b|D)\,\mu(D)}{P(B=b)} \tag{5.7}$$

to evaluate induced correlations.

For future reference, let us also establish the following property: the marginal on D of the three-variable joint distribution $P(CBD) = P(CB|D)\,\mu(D)$ is uniform by design,

$$\sum_{CB} P(CB|D)\,\mu(D) = \mu(D), \tag{5.8}$$

but the same is generally not true once we condition on $B = b$:

$$\sum_{C} P^b(CD) = \sum_{C} \frac{P(CB=b|D)\,\mu(D)}{P(B=b)} \neq \mu(D) \text{ (generally)}. \tag{5.9}$$

As an intuitive example, suppose that the conditional $P(CB|D)$ simply sets B equal to D while C is sampled according to a uniform distribution, $P(CB|D) = \mu(C)\,\delta(B, D)$. The joint distribution $P(CBD)$ encodes perfect correlations between D and B, and post-selecting on $B = b$ leaves $\sum_C P^b(CD) = \delta(D, b)$. That is, knowing that $B = b$, one can retrodict that $D = b$ with certainty.

We illustrate the calculations for assessing Berkson-type induced correlations with the following example:

Example 64. Let us analyse the physical mixture scenario introduced in example 63 in terms of induced correlations between C and D: we find

$$P(CB|D) = \delta(B, D \oplus C)\,\mu(C), \tag{5.10}$$

so that the post-selected distribution becomes

$$P^b(CD) = \delta(C, D \oplus b)\,\mu(D), \tag{5.11}$$

which implies perfect positive correlations if $B = 0$, and perfect negative correlations if $B = 1$. Now consider a probabilistic mixture of B being either set equal to E or to D,

$$P(B|DE) = q\delta(B, E) + (1-q)\,\delta(B, D), \tag{5.12}$$

while C is again perfectly correlated with E, $P(CE) = \delta(C, E)\,\mu(C)$. (Assume for simplicity that all variables range over the same set of values, with cardinality d.) This implies

$$P(CB|D) = q\delta(B, C)\,\mu(C) + (1-q)\,\delta(B, D)\,\mu(C), \tag{5.13}$$

and, assuming a uniform prior on D,

$$P(CBD) = q\delta(B, C)\,\mu(C)\,\mu(D) + (1-q)\,\delta(B, D)\,\mu(C)\,\mu(D). \tag{5.14}$$

In this case, the marginal over B is also uniform, $\sum_{CD} P(CBD) = \mu(B)$, and the conditional distribution becomes

$$P^b(CD) = q\delta(b, C)\,\mu(D) + (1-q)\,\mu(C)\,\delta(b, D). \tag{5.15}$$

Any instance of $D \neq b$ implies that $C = b$, and similarly any instance of $C \neq b$ implies that $D = b$. This provides some measure of negative correlation in the induced state. However, it is also possible to find $C = b = D$ (which occurs with probability $1/h$), hence the induced state does not encode perfect (negative) correlations.

[24] If the probing scheme used on A realizes perfectly randomized preparations of D, then it yields directly the joint distribution $P(CBD) = P(CB|D)\,\mu(D)$ generated by a uniform prior on D. But even if the randomization is not perfect, one can determine the conditional $P(CB|D)$ from experimental data and from there *compute* the desired joint state.

Once the induced probability distribution has been obtained, one can quantify the correlations between C and D that it encodes and draw conclusions about the causal structure. This step is further developed in section 5.3. Before that, however, we will introduce a quantum version of Berkson's paradox, which, as we will see, admits a closely related analysis, but with broader implications.

5.2 The quantum Berkson effect

If A and B are quantum variables, they also exhibit a Berkson effect: if one distinguishes two copies of A, denoted C and D, with C having a purely CC relation to B while D is purely CE-related, then post-selection on B can induce correlations between C and D, and the strength of the induced correlations heralds how the structure combines CC and CE relations. In the case of quantum variables, it is not clear what it means to post-select on B directly, but for the purpose of observing induced correlations and using them to classify causal structures, it is sufficient to post-select on the outcome of a measurement performed on B instead. Similarly, correlations between two quantum variables C and D manifest themselves, at an operational level, as correlations between the outcomes of measurements on C (i.e. which element of a given POVM is obtained) and the settings of preparation procedures on D (i.e. which one of an ensemble of states is prepared). The correlations observed when measuring different POVMs and preparing elements of different ensembles can be summarized in a single operator, and by repeating the experiment many times (always post-selecting on a particular measurement outcome at B, in our case) and ranging over informationally complete sets of preparations on D and measurements on C, one can reconstruct this operator.

Formally, the circuit relating C, D and B is represented by the Choi state χ_{CBD} that is isomorphic to the causal map $\mathcal{E}_{CB|D}$. It is analogous to the classical joint distribution $P(CBD) = P(CB|D)\,\mu(D)$ obtained under the assumption of a uniform prior on D: indeed, the marginal of the Choi state is

$$Tr_{CB}\chi_{CBD} = \frac{1}{h_D}\mathbb{I}_D, \tag{5.16}$$

since $\mathcal{E}_{CB|D}$ is a trace-preserving map. Now consider a POVM measurement $\left\{M^b\right\}_b$ on B, whose outcome probabilities can be obtained as $P(b) = Tr\left[M_B^b\chi_{CBD}\right]$. This is the probability of finding the outcome b in the measurement on B assuming that the circuit's input at D was the maximally mixed state. Post-selecting on this outcome singles out a component of the Choi state,

$$\chi_{CD}^b = \frac{1}{P(b)}Tr_B\left[M_B^b\chi_{CBD}\right]. \tag{5.17}$$

This is by design a positive-semidefinite operator with unit trace – a valid quantum state, analogous to the probability distribution $P^b(CD)$. By the Choi isomorphism, this corresponds to a completely positive map $\mathcal{E}_{C|D}^b : \mathcal{L}(\mathcal{H}_D) \to \mathcal{L}(\mathcal{H}_C)$. We stress that the map from D to C is unlike most maps encountered in standard quantum mechanics in that its input, D, lies in the causal future of its output, C, hence D cannot causally influence C. Instead, $\mathcal{E}_{C|D}^b$ represents only an inference: given that one prepared a certain state on D, and knowing that the common effect B was found to be in a certain state, one can retrodict what state one would have found on C.

For completeness, we note that, as in the classical case, the marginal of χ_{CD}^b on D is generally not the maximally mixed state:

$$Tr_C\chi_{CD}^b = \frac{1}{P(b)}Tr_{CB}\left[M_B^b\chi_{CBD}\right] \neq \frac{1}{h_D}\mathbb{I}_D \text{ (generally)}. \tag{5.18}$$

The map $\mathcal{E}_{C|D}^b$ that is isomorphic to χ_{CD}^b is therefore not necessarily trace-preserving. This is reasonable, since $\mathcal{E}_{C|D}^b$ is derived from the causal map $\mathcal{E}_{CB|D}$ by post-selection. To see how $\mathcal{E}_{C|D}^b$ can fail to be trace-preserving, consider a purely CE relation between two qubits, encoded in the Choi

state $\chi_{CBD} = \rho_C \otimes |\Phi^+\rangle\langle\Phi^+|_{BD}$, and the effects of post-selecting on finding the state $|0\rangle$ on B. The outcome b occurs with probability $P(b) = \frac{1}{2}$, and the induced Choi state is $\chi^b_{CD} = \rho_C \otimes |0\rangle\langle 0|_D$, whose marginal on D is manifestly not uniform. Intuitively, if we input the state $|1\rangle$ on D in this scenario, the state on B on which we are conditioning, $|0\rangle$, will not occur.

Like any quantum state, χ^b_{CD} can encode varying degrees of correlation between C and D, and correlations above a certain threshold rule out a mere probabilistic combination of CC and CE relations between A and B. Section 5.3.4 below establishes an upper bound on the mutual information between C and D that can be achieved by a probabilistic mixture. Crucially, however, a quantum state χ^b_{CD} can also encode correlations that are strictly stronger than in any classical probability distribution: if C and D become entangled under post-selection on a measurement outcome on B, this bears witness to an unusual way of combining CC and CE relations:

Definition 65. If two variables are related by a combination of a CC mechanism and a CE mechanism, and the two mechanisms are combined in a way that cannot be accounted for classically, then the causal relation between the variables is termed an *intrinsically quantum physical* mixture of CC and CE.

Physically, entanglement in the Choi state χ^b_{CD} means that the associated inference map $\mathcal{E}^b_{C|D}$ is not entanglement-breaking, that is, it preserves some coherence. In other words, for a given input state prepared on D, the corresponding output state one should expect to find on C (under post-selection on a certain measurement outcome on B) cannot be obtained by applying a measurement on D and taking the appropriately weighted convex combination of a set of states on C.

Example 66. An intrinsically quantum physical mixture of CC and CE relations between two qubits. We do not give the causal map explicitly, since it is rather cumbersome, but instead specify the circuit elements that can be used to realize it. The example is a generalization of the probabilistic mixture of a maximally entangled state (CC) and a unitary channel (CE) between A and B given in section 4.2, which used the probabilistic swap gate,

$$\mathcal{E}_{BF|DE} = (1-q)\,\mathcal{I}_{B|D} \otimes \mathcal{I}_{F|E} + q\mathcal{I}_{B|E} \otimes \mathcal{I}_{F|D}, \tag{5.19}$$

followed by a partial trace over qubit F. As a natural generalization, consider a gate that combines two-qubit identity and swap operations coherently:

$$\mathcal{E}_{BF|DE}(\rho_{DE}) = U_{BF|DE}\rho_{DE}U^\dagger_{BF|DE} \tag{5.20}$$

with the two-qubit unitary

$$U_{BF|DE} = \cos\frac{\theta}{2}\,\mathbb{I}_{B|D} \otimes \mathbb{I}_{F|E} + \mathrm{i}\sin\frac{\theta}{2}\,\mathbb{I}_{\mathrm{B|E}} \otimes \mathbb{I}_{\mathrm{F|D}}, \tag{5.21}$$

known as the partial swap. As in section 4.2, the initial state prepared on C and E is the maximally entangled state defined in eq. (2.14),

$$\rho_{CE} = \left|\Phi^+\right\rangle\left\langle\Phi^+\right|, \tag{5.22}$$

and the partial swap gate is followed by a partial trace over F.

The partial swap in this circuit allows Berkson-type induced correlations to arise: by post-selecting on finding the state $|0\rangle$ on B, one induces the state

$$\chi^b_{CD} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}\left[\cos\frac{\theta}{2}\,|01\rangle - \mathrm{i}\sin\frac{\theta}{2}\,|10\rangle\right]\left[\cos\frac{\theta}{2}\,\langle 01| + \mathrm{i}\sin\frac{\theta}{2}\,\langle 10|\right]. \tag{5.23}$$

Note that one does not obtain a pure state; when tracing over the second output qubit, F, purity is lost. However, the resulting mixed state is still entangled for any $\theta \neq 0, \pi$, thereby witnessing

the non-classical combination of CC and CE. One can also see that B is influenced by both inputs, D and E, simultaneously and in fact coherently with one another, in the sense that the output B depends on coherence terms between D and E in the joint input state: for example, if one inputs either $|\Psi^+\rangle_{DE}$ or $|\Psi^-\rangle_{DE}$, the resulting outputs on B are different.

To avoid confusion, let us recall a distinction introduced in Fig. 3.1: we stress that post-selection on B induces a map with input D and output C, which represents inferences. It is derived from the circuit inside the dashed box in Fig. 3.1b, which connects the quantum variables A and B and is therefore the object of interest for the purpose of our discussion. By contrast, a map with input C and output D is normally part of an instrument that is plugged into the unknown circuit in order to probe its functionality, depicted in Fig. 3.1a. The latter map may also exhibit correlations and entanglement in the Choi state, for example if the experimenter chooses to use an instrument such that the map from C to D is a unitary transformation. However, for the purpose of classifying the causal relations realized by the circuit in the dashed box, we focus solely on the circuit itself, viz. the causal map $\mathcal{E}_{CB|D}$ and the various $\mathcal{E}^b_{C|D}$ induced by post-selection, independently of the instruments used to probe it. The distinction is illustrated by a simple example: if the variable D is trivial (Hilbert space dimension 1), then the object of interest to the present discussion is the state ρ_C which is prepared by the circuit, not the POVM on C with which some agent probes this state.

5.3 Indicators of different combinations of CC and CE

Causal tomography provides a complete specification of the relation between two causally-ordered quantum variables, in the form of the causal map. With the causal map in hand, one can in principle compute how far it is, by some suitable metric, from the closest map of a particular form, such as a classical physical or probabilistic mixture. However, between the full tomography and the optimization (bearing in mind that the set of all causal maps that can be expressed as classical physical mixtures, for example, may not admit a straightforward parametrization), this procedure is highly inefficient if one only wants to determine, with some reasonable confidence, whether a given circuit has a particular causal structure. This section introduces several tests and quantitative indicators of causal structure that are more efficient than a full tomographic reconstruction.

For comparison, consider first a simple brute-force approach: a least-squares fit of experimental statistics (from informationally complete sets of preparations and measurements) to a model with the desired causal relation. This method is effective if one wants to compare the likelihood of several models of the same form, such as probabilistic mixtures of CC and CE elements with different mixing parameters, as in section 4.2. However, if the goal is to compare models of slightly different forms, for instance a probabilistic mixture and an intrinsically quantum physical mixture, then it becomes more challenging to compare the quality of the fit. In order to obtain reliable quantitative results, one must tally the number of free parameters in the different types of models and estimate the effects of the most likely sources of noise. Even then, collecting informationally complete statistics and fitting them to various alternative models is relatively expensive in terms of experimental and computational resources.

A slightly more efficient solution is to collect complete tomographic data, but reconstruct only one best-fitting causal map $\mathcal{E}_{CB|D}$, without imposing any constraints on the underlying causal structure. The causal structure can then be assessed by computing suitable indicators from $\mathcal{E}_{CB|D}$. In order to reduce the quantity of data to be collected in the first place, one can instead resort to an observational scheme. As discussed in section 4.2, this can be sufficient for ruling out purely CC and purely CE relations or for determining the probability of CC under the assumption that one is presented with a probabilistic mixture. Observational data can also provide evidence that the relation between A and B is not a mere probabilistic mixture of CC and CE: since both CC and CE imply that there exists a linear map from the states on A, found by an observational scheme, to states on B, any non-linearity in such a map implies a combination of CC and CE that is not a probabilistic mixture. Note that, in order to detect this effect, one must range over an over-complete basis of $\mathcal{L}(\mathcal{H}_A)$. Once

the non-linearity has been characterized, it may be possible to derive quantitative indicators of the causal relation in terms of it. This avenue is not explored here.

In the following sections, we introduce several indicators of causal relations, some of which have been applied successfully in the experiment detailed in section 5.4: indicators that can be computed given a complete reconstruction of the causal map, witnesses that require only a small set of count numbers to be evaluated and can rule out probabilistic mixtures, and two indicators of classical and intrinsically quantum physical mixtures which are based on the Berkson effect and only require data about one or two post-selected states. For simplicity, part of the discussion is restricted to qubits.

5.3.1 Indicators based on the full causal map

Once we have collected complete tomographic data, questions about the causal structure can be answered by finding the best-fitting causal map without imposing any constraints on its form and then computing indicators from it. For example, given the Choi state χ_{CBD}, one may compute the (quantum) mutual information between B and C and between B and D in the respective marginals. Non-zero mutual information implies that there is a non-trivial causal connection: a CC relation in the case of B and C, and a CE influence in the case of B and D. Moreover, if B and C both have Hilbert space dimensions h but the mutual information between C and B is greater than $\log h$ bits, this shows that the CC relation between C and B is not effectively classical (per definition 58), and similarly for the CE path from D to B. However, many CC or CE relations that are not effectively classical in the sense of definition 58 do not reveal this fact with large mutual information.

Instead, we propose the following construction:

Definition 67. Given a causal map $\mathcal{E}_{CB|D}$ that describes the general causal relation between two causally-ordered quantum variables, let $\tau_{CB|D}$ denote its Jamiołkowski representation and let $\rho_C \equiv Tr_{BD}\tau_{CB|D}$. Based on the operator

$$\tau_{B|CD} \equiv \rho_C^{-\frac{1}{2}} \tau_{CB|D} \rho_C^{-\frac{1}{2}} \tag{5.24}$$

we define the following indicators of *non-classical* CC or CE relations, respectively:

$$\begin{cases} u_{cc} \equiv -\lambda_{min}\left(T_{CD}\tau_{B|CD}\right) \\ u_{ce} \equiv -\lambda_{min}\left(\tau_{B|CD}\right), \end{cases} \tag{5.25}$$

where λ_{min} denotes the smallest eigenvalue of a Hermitian operator.

The operator $\tau_{B|CD}$ is Jamiołkowski isomorphic to a map of inferences from CD to B obtained by Bayesian inversion of the original causal map, and we use this operator instead of the more straightforward $\tau_{CB|D}$ in order to put C and D on an equal footing, as inputs into the inference map. For the purpose of the following discussion, we note the following: since the causal map $\mathcal{E}_{CB|D}$ is completely positive, it holds that $T_D\tau_{CB|D} \geq 0$, and consequently $T_D\tau_{B|CD} \geq 0$; that is, $\tau_{B|CD}$ is positive under partial transposition on the CE input, D.

By design, the indicators defined in eq. (5.25) reflect the underlying causal relations: If the causal structure is purely CC, then the Jamiołkowski operator factorizes as $\tau_{B|CD} = \tau_{B|C} \otimes \mathbb{I}_D$, and consequently partial transposition on D leaves the spectrum of eigenvalues unchanged, hence $\tau_{B|CD} \geq 0$ and therefore $u_{ce} \leq 0$. It follows that $u_{ce} > 0$ bears witness to a non-trivial CE component. However, even if the causal structure is purely CE, so that $\tau_{B|CD} = \mathbb{I}_C \otimes \tau_{B|D}$, it is not necessarily the case that $u_{ce} > 0$. Indeed, if D and B have a *classical* cause-effect relation, this implies (by theorem 59) that $\tau_{B|D}$ is separable, hence both PPT and positive-semidefinite, so $\tau_{B|D} \geq 0$, from which it follows that $u_{ce} \leq 0$. The indicator u_{ce} defined above therefore witnesses a CE relation that is non-trivial and non-classical. Note that $u_{ce} > 0$ does not guarantee a *purely* CE relation; it only rules out the possibility that the relation is purely CC or classical CE. Similarly, $u_{cc} > 0$ heralds a non-trivial, non-classical CC relation.

Bounds on (u_{ce}, u_{cc})**.** Any causal map $\mathcal{E}_{CB|D}$ can be reduced in this way to a point in a two-dimensional diagram of (u_{ce}, u_{cc}), and finding the point in the region $u_{ce} > 0$ heralds a non-trivial, non-classical CE relation, while the half-plane $u_{cc} > 0$ implies a non-trivial, non-classical CC relation. Beyond these two simple cases, there are other points and regions of the (u_{ce}, u_{cc})-plane that imply simple conclusions about the causal relation, as shown in Fig. 5.2. The interpretation of the values of (u_{ce}, u_{cc}) is based on several bounds which we establish in the following:

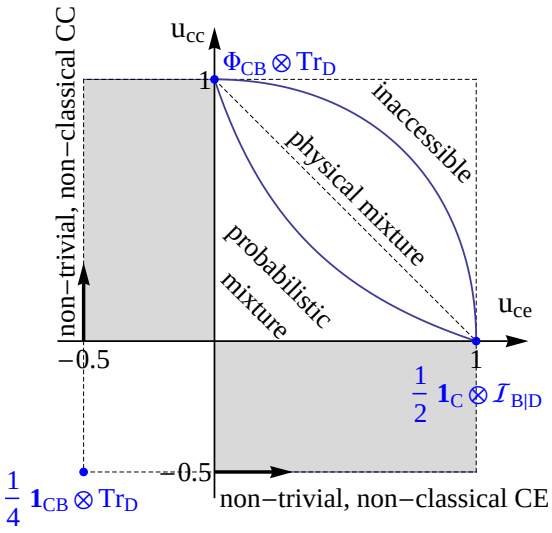

Figure 5.2: Representation of causal relations between two qubits by the witnesses u_{cc} and u_{ce} defined in (5.25). Black captions identify regions associated with certain causal structures: the first quadrant is formed by combinations of non-classical CC and CE relations, whereas the shaded area on the lower right ($u_{ce} > 0$, $u_{cc} \leq 0$) heralds a non-classical CE influence, and vice versa for the upper left. For points in the lower left quadrant, one can rule out neither CC nor CE explanations. Blue points identify special cases of interest: $\mathcal{E}_{CB|D} = \frac{1}{2}\mathbb{I}_C \otimes \mathcal{I}_{B|D}$ realizes the identity channel from D to B and $\mathcal{E}_{CB|D} = |\Phi^+\rangle\langle\Phi^+|_{CB} \otimes Tr_D$ prepares a maximally entangled state between C and B. Both are noiseless, maximally non-classical and purely CE and purely CC, respectively. Finally, $\mathcal{E}_{CB|D} = \frac{1}{4}\mathbb{I}_C \otimes \mathbb{I}_B \otimes Tr_D$ implies neither a CC nor a CE connection, but rather total causal disconnection.
Probabilistic mixtures of CC and CE maps must remain below the dashed straight line by theorem 69, but numerical simulations show that in fact they only reach the concave curve eq. (5.29). All points beyond this curve must be due to physical mixtures of non-classical CC and CE. The upper limit on (u_{ce}, u_{cc}) for any causal map is the convex curve eq. (5.30), which is reached by the intrinsically quantum physical mixture from example 66.

Theorem 68. *The witnesses defined in eq.(5.25) are lower-bounded by*

$$u_{ce}, u_{cc} \geq -\frac{1}{h_B}, \tag{5.26}$$

where h_B is the dimension of the Hilbert space of variable B. These lower bounds are unique in the following sense: if either of them is saturated, then so is the other, and the inference map is uniquely determined to be

$$\tau_{B|CD} = \frac{1}{h_B}\mathbb{I}_B \otimes \mathbb{I}_{CD}, \tag{5.27}$$

whereby C and D provide no information about B at all.

Proof is provided in appendix A.6.

Theorem 69. *If $\tau_{B|CD}$ represents a non-trivial probabilistic mixture of CC and CE relations (excluding purely CC or CE relations), then the pair (u_{ce}, u_{cc}) lies in the half-plane delimited by*

$$u_{ce} + u_{cc} < \frac{h_A}{2}. \tag{5.28}$$

By the contrapositive, larger values of the sum herald a physical mixture.

Proof is again provided in appendix A.6. Using numerical simulations to test which values of (u_{ce}, u_{cc}) can be reached by randomly chosen probabilistic mixtures, one finds that, at least in the case of qubits, this bound is not saturated: all points (u_{ce}, u_{cc}) generated by probabilistic mixtures of random unitaries and maximally entangled states lie within a strictly smaller region, and noise in either component is only expected to reduce u_{ce} and u_{cc}. Rather than derive an expression for the boundary of this region analytically, we note that the appropriate curve in (u_{ce}, u_{cc}) is likely to be generated by probabilistic mixtures of extremal cause-effect and common-cause relations, such as the identity channel and the $|\Phi^+\rangle$ state. Probabilistic mixtures of this particular pair trace out the curve

$$u_{ce} + u_{cc} + 2u_{ce}u_{cc} \leq 1, \tag{5.29}$$

and numerical results confirm that this in indeed a tight bound on the region of (u_{ce}, u_{cc}) that can be reached by probabilistic mixtures.

By contrast, an intrinsically quantum physical mixture of the identity channel and $|\Phi^+\rangle$, as introduced in example 66, traces out the curve

$$3u_{ce}^2 + 3u_{cc}^2 - 2u_{ce}u_{cc} + 2u_{ce} + 2u_{cc} \leq 5, \tag{5.30}$$

which puts it strictly above the bound for probabilistic mixtures. For the same reason as before, one would expect this curve to form the boundary of the region of (u_{ce}, u_{cc}) that can be reached by quantum physical mixtures, and consequently by any causal map. Numerical tests suggests that, at least for qubits, this is indeed the case, though an analytical proof is still outstanding. A physical mixture of classical CC and CE conditionals, meanwhile, will not violate even the bound of theorem 69, since effectively classical relations between C and B and between D and B imply that $u_{ce}, u_{cc} \leq 0$.

5.3.2 Witness of physical mixture based on observational data

This section introduces an operator V_{CBD} whose expectation value[25] for the Jamiołkowski operator $\tau_{CB|D}$,

$$\langle V \rangle_\tau \equiv Tr\left[V_{CBD}\tau_{CB|D}\right], \tag{5.31}$$

can be evaluated using only observational data, and with the property that a non-zero expectation value rules out a probabilistic mixture of CC and CE. The derivation of a family of witnesses with this property relies on simple algebraic arguments, but it holds only under the assumption that the marginal on C is the maximally mixed state, $\rho_C = \frac{1}{2}\mathbb{I}_C$. That is, a non-zero result rules out Jamiołkowski operators of the form

$$\tilde{\tau}_{CB|D} = q\rho_{CB} \otimes \mathbb{I}_D + (1-q)\,\frac{1}{2}\mathbb{I}_C \otimes \tau_{B|D}, \tag{5.32}$$

[25]The expression for $\langle V \rangle_\tau$ given in the text is not strictly an expectation value, since the Jamiołkowski operator $\tau_{CB|D}$ is not generally a valid quantum state. However, the same quantity can also be expressed as $\langle V \rangle_\tau = Tr\left[T_D\left(2V_{CBD}\right)\chi_{CBD}\right]$, that is, the expectation value of the observable $T_D\left(2V_{CBD}\right)$ in the Choi state χ_{CBD}, which is a valid quantum state that encodes the same information as the causal map $\mathcal{E}_{CB|D}$ or the operator $\tau_{CB|D}$.

where ρ_{CB} is a quantum state with $Tr_B \rho_{CB} = \frac{1}{2}\mathbb{I}_C$ and $\tau_{B|D}$ is a CE conditional, which implies $Tr_B \tau_{B|D} = \mathbb{I}_D$. If one wants to rule out generic probabilistic mixtures, which do not exhibit this symmetry between C and D, one must resort to the more general witness introduced in the next section.

Based on section 3.3, we know that the restriction to an observational scheme on A implies that one can only determine inner products of the form

$$Tr\left[(M_B \otimes V_{CD}) \cdot \tau_{CB|D}\right], \tag{5.33}$$

where M can be any observable on B, but V_{CD} remains constrained to the accessible subspace $\mathcal{A}$, which is generated by products of identical operators, $\Pi_C^{s,c} \otimes \Pi_D^{s,c}$. Equivalently, any V_{CD} that can be evaluated by an observational scheme must be a linear combination of the G^k with $k = 1, ...9$ introduced in definition 38, which we reproduce here for reference:

$$\begin{cases} G^{+j} \equiv \sigma_0 \otimes \sigma_j + \sigma_j \otimes \sigma_0 & [j = 1, 2, 3] \\ G^{+(i+j+1)} \equiv \sigma_i \otimes \sigma_j + \sigma_j \otimes \sigma_i & [(i, j) = (1, 2), (1, 3), (2, 3)]. \\ G^7 = \sigma_0 \otimes \sigma_0 + \frac{1}{3}\sum_{j\neq 0} \sigma_j \otimes \sigma_j \\ G^8 = \sigma_1 \otimes \sigma_1 + \sigma_2 \otimes \sigma_2 - 2\sigma_3 \otimes \sigma_3 \\ G^9 = \sigma_1 \otimes \sigma_1 - \sigma_2 \otimes \sigma_2. \end{cases} \tag{5.34}$$

We will show the following:

Claim 70. One can design a witness V_{CBD} with the desired properties – accessible by an observational scheme and such that $Tr\left[V_{CBD}\tilde{\tau}_{CB|D}\right]$ is zero for probabilistic mixtures of the form (5.32) – by taking $V_{CBD} = M_B \otimes V_{CD}$, where M can be any observable on B and V_{CD} is a linear combination of the basis elements G^k with $k \in \{4, 5, 6, 8, 9\}$ above.

Proof. The requirement that V_{CD} be accessible using observational data can be met by design if we explicitly define V_{CBD} as a linear combination of the accessible G^k. At the same time, we want to ensure that $Tr\left[(M_B \otimes V_{CD})\,\tilde{\tau}_{CB|D}\right] = 0$ for any $\tilde{\tau}_{CB|D}$ of the form (5.32). To this end, note that, if the causal relation is a probabilistic mixture of the form (5.32), then the inner product for $k \in \{4, 5, 6\}$ is

$$Tr\left[M_B \otimes G^{4,5,6}\tilde{\tau}_{CB|D}\right] = 0, \tag{5.35}$$

as one can see be substituting

$$Tr\left[M_B \otimes G^{4,5,6}\tilde{\tau}_{CB|D}\right] = Tr\left[M_B \otimes \left(\sigma_i^C \otimes \sigma_j^D + \sigma_j^C \otimes \sigma_i^D\right) \cdot \left(q\rho_{CB} \otimes \mathbb{I}_D + (1-q)\,\frac{1}{2}\mathbb{I}_C \otimes \tau_{B|D}\right)\right] = 0 \tag{5.36}$$

and noting that $\tilde{\tau}_{CB|D}$ contains one identity operator in each term. Similarly, for $k \in \{8, 9\}$, we have

$$Tr\left[M_B \otimes G^{8,9}\tilde{\tau}_{CB|D}\right] = 0, \tag{5.37}$$

since, for $j \in \{1, 2, 3\}$

$$Tr\left[M_B \otimes \left(\sigma_j^C \otimes \sigma_j^D\right) \cdot T_D\left(q\rho_{CB} \otimes \mathbb{I}_D + (1-q)\,\frac{1}{2}\mathbb{I}_C \otimes \tau_{B|D}\right)\right] = 0. \tag{5.38}$$

Therefore, if any of these five inner products are non-zero, then the causal relation is not a mixture of the form (5.32). Meanwhile, the other four inner products, $Tr\left[M_B \otimes G^k\tilde{\tau}_{CB|D}\right]$ with $k \in \{1, 2, 3, 7\}$, are not necessarily zero even for probabilistic mixtures, and therefore cannot serve as witnesses. It follows that any linear combination of G^k with $k \in \{4, 5, 6, 8, 9\}$, tensored with any observable on B, serves as a witness with the desired properties. □

Example 71. As an explicit example of an operator V_{CD} that requires only four different projectors on C, consider the following:

$$\begin{aligned}\tilde{V}_{CD} = \quad & +|\psi\psi\rangle\langle\psi\psi| - |\psi_\perp\psi_\perp\rangle\langle\psi_\perp\psi_\perp| \\ & +|\phi\phi\rangle\langle\phi\phi| - |\phi_\perp\phi_\perp\rangle\langle\phi_\perp\phi_\perp|,\end{aligned} \tag{5.39}$$

for any two orthogonal bases $\{|\psi\rangle, |\psi_\perp\rangle\}$ and $\{|\phi\rangle, |\phi_\perp\rangle\}$. By design, $\tilde{V}_{CD}$ lies in the subspace spanned by products of the form $\Pi_C^{s,c} \otimes \Pi_D^{s,c}$, hence it can be evaluated using an observational scheme. Furthermore, one can show that its inner product with G^k, $k = 1, 2, 3, 7$ is zero, i.e. it is generated only by G^k with $k \in \{4, 5, 6, 8, 9\}$, and consequently a non-zero expectation value does rule out a probabilistic mixture of the form (5.32). To see this, note that

$$Tr\left[\left(|\psi\psi\rangle\langle\psi\psi| - |\psi_\perp\psi_\perp\rangle\langle\psi_\perp\psi_\perp|\right)(\sigma_0 \otimes \sigma_j)\right] = \langle\psi|\sigma_j|\psi\rangle - \langle\psi_\perp|\sigma_j|\psi_\perp\rangle = 0 \ (j \neq 0), \tag{5.40}$$

and similarly for $\sigma_j \otimes \sigma_0$, which implies that

$$Tr\left[\tilde{V}_{CD}G^k\right] = 0 \ \ (k = 1, 2, 3). \tag{5.41}$$

Similarly, letting $\vec{v}$ denote the Bloch vector of the (pure) state $|\psi\rangle\langle\psi| = \frac{1}{2}(\sigma_0 + \vec{v}\cdot\vec{\sigma})$, one finds

$$Tr\left[|\psi\psi\rangle\langle\psi\psi|\left(\sigma_0 \otimes \sigma_0 + \frac{1}{3}\sum_{j\neq 0}\sigma_j \otimes \sigma_j\right)\right] = 1 + \frac{1}{3}\vec{v}\cdot\vec{v} = \frac{4}{3} \ \forall\,|\psi\rangle, \tag{5.42}$$

so that the contributions from the four terms in $\tilde{V}_{CD}$ cancel and one recovers

$$Tr\left[\tilde{V}_{CD}G^7\right] = 0. \tag{5.43}$$

Example 72. As a second example, which is a special case of the more general witness introduced in the next section, one can take $V_{CD} = \sigma_i \otimes \sigma_j$ with $i \neq j$ both non-zero. By furthermore choosing the observable on B to be $\sigma_{k\neq 0}$, the third non-trivial Pauli ($i \neq k \neq j$), the relevant expectation value takes the simple form

$$\langle V\rangle_\tau = Tr\left[\left(\sigma_i^C \otimes \sigma_j^D \otimes \sigma_k^B\right)\tau_{CB|D}\right] \equiv W_0. \tag{5.44}$$

If this is non-zero, one can exclude a probabilistic mixture with maximally mixed marginals on C.

5.3.3 Witness of physical mixture based on the Berkson effect

This section introduces a witness W, in the broader sense of a function computed from an informationally incomplete subset of experimental statistics, such that $W \neq 0$ rules out a probabilistic mixture of CC and CE relations between two qubits,

$$\tau_{CB|D}^{prob} = q\rho_{CB} \otimes \mathbb{I}_D + (1-q)\rho_C \otimes \tau_{B|D}, \tag{5.45}$$

with a generic marginal $\rho_C = Tr_B\rho_{CB}$ on C. (Note that the two terms must have identical marginals on C due to the structure of the circuit in Fig. 2.3d: if the causal relation is CE, then the gate must take the form $\mathcal{E}_{B|DE} = \mathcal{E}_{B|D} \otimes Tr_E$, so that $\rho_C = Tr_E\rho_{CE}$. On the other hand, if the causal relation is CC, the gate must be $\mathcal{E}_{B|DE} = \mathcal{E}_{B|E} \otimes Tr_D$, hence the circuit produces $\rho_{CB} = \left(\mathbb{I}_C \otimes \mathcal{E}_{B|E}\right)(\rho_{CE})$, which also leads to $Tr_B\rho_{CE} = Tr_E\rho_{CE} = \rho_C$.)

The experimental data from which the witness is computed is obtained by measuring fixed Pauli observables σ_s and σ_u on C and B, respectively, with outcomes $c, b \in \{\pm 1\}$, while preparing D in

eigenstates of a fixed σ_t that are indexed by the corresponding eigenvalues, $d \in \{\pm 1\}$. (Each choice of s, t, u gives rise to a different witness of this family.) Conditional on the eigenvalue d that was chosen for the preparation on D, one finds

$$P(cb|d) = \mathrm{Tr}\left[\tau_{CB|D}\Pi_C^{s,c} \otimes \Pi_B^{u,b} \otimes \Pi_D^{t,d}\right]. \tag{5.46}$$

The witness is intended to capture Berkson-type induced correlations. To this end, following (5.7), we assume a uniform prior, i.e. that the eigenvalues $d = \pm 1$ are chosen with uniform probability, $P(d) = \frac{1}{2}$. This gives rise to a joint distribution

$$P(cdb) = \frac{1}{2}\mathrm{Tr}\left[\tau_{CB|D}\Pi_C^{s,c} \otimes \Pi_B^{u,b} \otimes \Pi_D^{t,d}\right], \tag{5.47}$$

from which one can derive $P(cd|b)$, the post-selected distribution. As a measure of the induced correlations, consider the *covariance*,

$$\begin{aligned}\mathrm{cov}(c,d) &\equiv \langle cd\rangle - \langle c\rangle\langle d\rangle \\ &= \sum_{cd} cdP(cd) - \left[\sum_{cd} cP(cd)\right]\left[\sum_{cd} dP(cd)\right],\end{aligned} \tag{5.48}$$

associated with a generic probability distribution over c, d. In our particular case, since the variables are binary and take the values ± 1, one can show[26] that the expression simplifies to

$$\mathrm{cov}(c,d) = 4\left[P(++)P(--) - P(+-)P(-+)\right]. \tag{5.49}$$

If c and d follow a conditional distribution induced by post-selection on b, the covariance becomes

$$\mathrm{cov}(c,d|b) = 4\left[P(++|b)P(--|b) - P(+-|b)P(-+|b)\right]. \tag{5.50}$$

Note that $\mathrm{cov}(c,d|b)$ need not be zero for probabilistic mixtures, since any combination of CC and CE can induce some measure of correlation under post-selection (as illustrated by the examples in section 5.1). Instead, we define the witness to be a suitably weighted difference of the covariances induced by $b = \pm 1$:

$$W \equiv 2\sum_b bP(b)^2\mathrm{cov}(cd|b). \tag{5.51}$$

$$= 8\sum_{b=\pm 1} b\left[P(++b)P(--b) - P(+-b)P(-+b)\right]. \tag{5.52}$$

One can show that this expression is zero for any probabilistic mixture, i.e. any conditional of the form (5.45). The proof is rather technical and is relegated to appendix A.7. The normalization factor 8 is introduced so that the maximal value of W is $+1$. This maximum is achieved when $b = +1$ implies perfect (positive) correlation and $b = -1$ implies perfect anti-correlation. Unlike conventional entanglement witnesses, which are positive for separable states (the "boring" case) and herald entanglement by turning negative, the witness introduced here is exactly zero for probabilistic mixtures, whereas both strictly positive and strictly negative values of W herald a physical mixture.

Simpler form in a limiting case. In the limit where the marginals over CB and DB are both uniform, we will show that the witness W reduces to a special case of the algebraically motivated witnesses that was introduced in eq. (5.44) in the previous section:

[26] In order to see this, it is helpful to multiply the first term in the definition of $\mathrm{Cov}(c,d)$ by the resolution of identity $1 = \sum_{cd} P(cd)$.

Claim 73. If $Tr_D \tau_{CB|D} = \mathbb{I}_{CB}$ and $Tr_C \tau_{CB|D} = \mathbb{I}_{BD}$, then the witness W reduces to

$$W = \sum_{cdb} cdbP(cdb) = Tr\left[\left(\sigma_C^s \otimes \sigma_B^u \otimes \sigma_D^t\right) \cdot \tau_{CB|D}\right] = W_0. \tag{5.53}$$

Proof. In terms of the observed statistics $P(cdb)$, the hypothesis of uniform marginals takes the form

$$P(cb) \equiv \sum_d P(cdb) = \frac{1}{4} \ \forall c, b, \tag{5.54}$$

$$P(db) \equiv \sum_c P(cdb) = \frac{1}{4} \ \forall d, b, \tag{5.55}$$

This ensure that each outcome b is equally probable, $P(b) = \frac{1}{2}$, and therefore the conditional distributions also satisfy

$$P(c|b) \equiv \sum_d P(cd|b) = \frac{1}{2} \ \forall c, b \tag{5.56}$$

$$P(d|b) \equiv \sum_c P(cd|b) = \frac{1}{2} \ \forall d, b \tag{5.57}$$

Therefore the conditional distribution $P(cd|b)$ has uniform marginals on c and d. The expectation values $\langle c \rangle$ and $\langle d \rangle$ are zero, so that the covariance (5.48) simplifies to

$$\mathrm{cov}(c, d) = \langle cd \rangle = \sum_{cd} cdP(cd), \tag{5.58}$$

and therefore

$$W = 2 \sum_b bP(b)^2 \sum_{cd} cdP(cd|b) \tag{5.59}$$

$$= \sum_{cdb} cdbP(cdb) = W_0 \tag{5.60}$$

□

In this sense, the witness W is a generalization of the simple product of Paulis that defines W_0, and which is easily seen to be zero for probabilistic mixtures if $\rho_C = \frac{1}{2}\mathbb{I}_C$.

Example 74. Application of the witness W to the intrinsically quantum physical combination of CC and CE using the partial swap unitary. Consider the causal relation introduced in example 66, setting $\theta = \frac{\pi}{4}$ to generate an equal mixture of CC and CE:

$$\rho_{CE} = \left|\Phi^+\right\rangle \left\langle\Phi^+\right| \tag{5.61}$$

$$U_{BF|DE} = \frac{1}{\sqrt{2}} \mathcal{I}_{B|D} \otimes \mathcal{I}_{F|E} + \frac{\mathrm{i}}{\sqrt{2}} \mathcal{I}_{B|E} \otimes \mathcal{I}_{F|D}, \tag{5.62}$$

and measurements of σ_1 on C, σ_3 on B and preparations of σ_2-eigenstates on D. The experimental statistics take the form

$$\begin{aligned} P(cdb) &= \frac{1}{2} \mathrm{Tr}\left[Tr_E\left(\rho_{CE} \tau_{B|DE}\right) \Pi_C^{1,c} \otimes \Pi_B^{3,b} \otimes \Pi_D^{2,d}\right] \\ &= \frac{1}{2} \cdot \frac{1}{8} + \frac{1}{2} \cdot \frac{1}{4} \left[\delta(b, -1) \delta(c, d) + \delta(b, +1) (1 - \delta(c, d))\right], \end{aligned} \tag{5.63}$$

that is, with probability $\frac{1}{2}$, the eigenvalues *cbd* are entirely uncorrelated with each other, but with the remaining probability $\frac{1}{2}$, it holds that $b = -1$ if and only if $c = d$, whereas $b = +1$ if and only if $c \neq d$. The second term reproduces exactly the example of a classical physical mixture discussed in section 5.1. Even with the admixture of an entirely uncorrelated distribution, the covariance $\mathrm{cov}\,(c, d|b)$ under post-selection is still

$$\mathrm{cov}\,(c, d|b = \pm 1) = \mp\frac{1}{2}, \tag{5.64}$$

hence the witness takes the value

$$W = -\frac{1}{2}. \tag{5.65}$$

The fact that this is non-zero rules out any probabilistic mixture of the form (5.45).

5.3.4 Bounds on induced mutual information for probabilistic mixtures

This section derives a quantitative criterion for ruling out probabilistic mixtures based on Berkson-type induced correlations, namely a tight upper bound on the quantum mutual information $I\,(C : D)$ in the post-selected state χ^{b}_{CD} introduced in eq. (5.17). The main part of the derivation is given in terms of classical variables and proceeds by first deriving a bound on the mutual information between the two causal parents, D and E, conditional on B. A bound on the mutual information between C and D then follows directly by noting that any correlations between C and D must be mediated by E, hence $I\,(C : D) \leq I\,(E : D)$ (with equality when C and E are perfectly correlated). The extension to truly quantum variables, i.e. beyond the classical limit, is discussed at the end of this section.

For reference, we define the mutual information of two quantum variables:

Definition 75. Given a quantum state on two variables, ρ_{DE} , let S denote the von Neumann entropy and let $S\,(DE) \equiv S\,(\rho_{DE})$, $S\,(D) \equiv S\,(Tr_E \rho_{DE})$ and $S\,(E) \equiv S\,(Tr_D \rho_{DE})$. Then the quantum mutual information between D and E is

$$I\,(D : E) \equiv S\,(D) + S\,(E) - S\,(DE)\,. \tag{5.66}$$

Note that, in the limit where ρ_{DE} is effectively classical, as discussed in section 2.3.4, and reduces to a joint probability distribution over two classical variables D and E, the quantum mutual information reduces to the classical (Shannon) mutual information. We will begin by considering this limit.

The task, now, is to upper-bound the mutual information $I\,(D : E)$ in the state induced by post-selecting on $B = b$, when the relation between D, E and B is a probabilistic mixture,

$$P\,(B|DE) = qP_E\,(B|E) + (1 - q)\,P_D\,(B|D)\,. \tag{5.67}$$

We introduce subscripts to distinguish the conditionals that form the mixture, $P_E\,(B|E)$ and $P_D\,(B|D)$, from conditionals of the same form that one can construct by combining the overall $P\,(B|DE)$ with suitable marginals, such as

$$P\,(B|E) \equiv \sum_D P\,(B|DE)\,P\,(D) = qP_E\,(B|E) + (1 - q)\sum_D P_D\,(B|D)\,P\,(D) \neq P_E\,(B|E)\,. \tag{5.68}$$

In keeping with the previous discussion of the Berkson effect, in particular eq. (5.4), we will take priors on D and E that are uniform distributions over all possible values of D and E, respectively. The post-selected distribution then inherits the probabilistic form,

$$P\,(DE|B) = p_b \mu\,(D)\,P_E\,(E|B) + (1 - p_b)\,P_D\,(D|B)\,\mu\,(E) \tag{5.69}$$

where we introduce the shorthand

$$P_D\left(D|B\right) = P_D\left(B|D\right)\mu\left(D\right) / \left[\sum_D P_D\left(B|D\right)\mu\left(D\right)\right], \tag{5.70}$$

$$P_E\left(E|B\right) = P_E\left(B|E\right)\mu\left(E\right) / \left[\sum_E P_E\left(B|E\right)\mu\left(E\right)\right], \tag{5.71}$$

$$p_b = \left[q\sum_E P_E\left(B|E\right)\mu\left(E\right)\right] / \left[q\sum_E P_E\left(B|E\right)\mu\left(E\right) + (1-q)\sum_D P_D\left(B|D\right)\mu\left(D\right)\right]. \tag{5.72}$$

(The Bayesian inverse $P_D\left(D|B\right)$ encodes what one can retrodict about D given B in the scenario where B only depends on D by the conditional $P_D\left(B|D\right)$, rather than depending on both D and E by the probabilistically mixed $P\left(B|DE\right)$, and similarly for $P_E\left(E|B\right)$.)

One can see that $I\left(D:E\right)$ is at most one bit, regardless of the cardinality of the variables themselves; that is, when B is already known, learning D provides at most one bit of new information about E. To show this, we introduce a classical, binary switch variable S that determines, in each run, whether B is influenced by D or by E. The conditional distribution then takes the form

$$P\left(DE|BS\right) = \delta\left(S,0\right)\mu\left(D\right)P_E\left(E|B\right) + \delta\left(S,1\right)P_D\left(D|B\right)\mu\left(E\right). \tag{5.73}$$

Note that, once S is known, D and E become statistically independent despite post-selection on B: either D follows the uniform distribution and E can be inferred from B by $P_E\left(E|B\right)$, or E is uniformly distributed and D follows $P_D\left(D|B\right)$. Using D to infer E therefore involves two distinct steps: first one uses D to infer S (exploiting the difference between the distributions $\mu\left(D\right)$ and $P_D\left(D|B\right)$ that arise for different values of S), and then one can determine whether E follows the distribution $\mu\left(E\right)$ or $P_E\left(E|B\right)$ from the value of S. Crucially, after the first step, the entire new information that one has acquired is compressed in a single bit, S. Due to this bottleneck in the process of inference, learning D cannot provide more than one bit of new information about E, under the assumption that B is known. Note that this argument holds in the same form if D and E are quantum variables: even though they could in general exhibit non-classical correlations, negative mutual information etc., the constraints in this particular situation imply that any inference from D to E is entirely mediated by the classical switch variable S, and therefore limited to one classical bit of information.

A more stringent upper bound on the induced mutual information, especially for variables with low cardinality, can be derived analytically. The argument is briefly outlined here, with a detailed derivation given in appendix A.8. Firstly, the mutual information $I\left(D:E\right)$ in eq. (5.69) is maximized when $P_D\left(D|B\right)$ and $P_E\left(E|B\right)$ are sharply peaked, producing one particular value with certainty. Intuitively, this makes it easiest to distinguish $P_D\left(D|B\right)$ from $\mu\left(D\right)$, thereby learning the value of the switch variable, and maximizes what one can subsequently deduce about E. Formally, it follows from the convexity of the mutual information $I\left(D:E\right)$ as a function of the conditional $P\left(D|E\right)$, for a fixed marginal $P\left(E\right)$, and vice-versa. In order to determine the overall maximum of $I\left(D:E\right)$, we must also find the optimal mixing weight p_b. In general, this depends on the cardinality of D and E, but if both variables take an equal number of values h, then the highest mutual information is achieved when $p_b = \frac{1}{2}$. This is a matter of symmetry between D and E, but also follows from the observation that one can learn the most about the switch variable, i.e. whether B is influenced by D or by E, if it is previously completely unknown. This completely determines the distribution $P\left(DE|B\right)$ that maximizes the mutual information, and allows us to explicitly compute the maximal induced mutual information for a probabilistic mixture, which leads to the following result:

Theorem 76. *Given two classical variables D and E of equal cardinality, h, which produce a common effect B by a probabilistically mixed conditional,*

$$P\left(B|DE\right) = qP_E\left(B|E\right) + \left(1-q\right)P_D\left(B|D\right), \tag{5.74}$$

the mutual information in the distribution $P(DE|B)$ induced by post-selection on any given value of B (assuming uniform priors on D and E) is tightly upper-bounded by

$$I^{prob}(D:E) \leq \left(1+\frac{1}{h}\right) + \log_2 h - \left(1+\frac{1}{h}\right)\log_2(h+1). \tag{5.75}$$

Therefore induced mutual information in excess of this bound heralds a physical mixture, i.e. a conditional $P(B|DE)$ that cannot be decomposed as a probabilistic mixture of the form above.

It is interesting to note that the heuristic bound that we derived above, which allows at most one bit of mutual information, can only be saturated in the limit of infinitely large h. If D and E have small cardinality, then the quantitative bound is much more stringent. For binary variables, for example, it is $\frac{5}{2} - \frac{3}{2}\log_2 3 \approx .12$ bits. By contrast, the maximal mutual information induced by post-selection if there are no constraints on the causal structure is simply $\log h$.

If D and E are quantum instead of classical variables, their state under post-selection on some measurement outcome on B can be written

$$\rho^b_{DE} = p_b\, \mathbb{I}_D \otimes \rho^b_E + (1-p_b)\, \rho^b_D \otimes \mathbb{I}_E, \tag{5.76}$$

where ρ^b_D and ρ^b_E generally depend on the choice of measurement and outcome. As before, the prior over D and E used to derive the induced state is uniform, i.e. the maximally mixed states. This implies that there is in fact only one non-trivial density operator on each $\mathcal{H}_D$ and $\mathcal{H}_E$ in the problem, and consequently there exists a preferred basis of Hilbert space in which all density operators of interest are diagonal. By using this basis to evaluate von Neumann entropies, the problem is reduced to the classical case and all previous results carry over, including the tight upper bound on $I(D:E)$ as a function of the dimension of the Hilbert spaces, $h = \dim \mathcal{H}_D = \dim \mathcal{H}_E$.

5.3.5 Indicator of intrinsically quantum physical mixture

In order to detect an intrinsically quantum combination of CC and CE relations between two variables, it is sufficient to characterize the Choi state χ^b_{CD} induced by post-selection on a single POVM element M^b on B, defined in eq. (5.17), and test whether it contains any form of non-classical correlations. (We stress that the converse is not true: post-selecting on any one POVM element on B and finding a state that is only classically correlated does not imply that the CC and CE mechanisms were combined in an incoherent manner.) That is, one collects statistics from measurements on C and preparations on D that range over tomographically complete sets, but only for a single POVM element on B, uses these statistics to reconstruct the map $\mathcal{E}^b_{C|D}$ and applies standard tests of correlation to the associated Choi state. This is much more efficient in terms of both experimental and computational resources than a full tomographic reconstruction of the causal map $\mathcal{E}_{CB|D}$, and focuses directly on the classification of the causal relation.

One possible indicator of non-classical correlations in the induced state, extending the ideas from the previous section, is if the mutual information exceeds the classical bound: if two variables with Hilbert space dimension h are only classically correlated, their mutual information is at most $\log h$. However, this is a relatively strong criterion for verifying a non-classical combination of CC and CE. Two qubits connected by a partial swap (example 66), for instance, only achieves a mutual information of at most $3 - \frac{3}{2}\log_2 3 \approx 0.62$ bits (when $\theta = \frac{\pi}{2}$), which is well within the classical bounds. A more sensitive test is based on the negativity, a standard measure of entanglement, evaluated on the induced Choi state:

$$\mathcal{N}\left(\chi^b_{CD}\right) \equiv \frac{1}{2}\left[Tr\left|T_D \chi^b_{CD}\right| - 1\right]. \tag{5.77}$$

As pointed out above, a non-classical combination of CC and CE relations does not necessarily manifest itself in the Choi state induced by post-selecting on a given POVM element. However,

one can reduce the number of measurements on B for which one must test χ^b_{CD} by noting the following: any useful measure of non-classical correlations, such as entanglement, must not increase when taking convex combinations of states, since these can be achieved by local operations and classical communication (LOCC). At the same time, generic POVM elements can be obtained by taking convex combinations of rank-one projective measurements, also known as sharp measurements. Consequently, in order to rule out non-classical correlations in the induced states χ^b_{CD} for all POVM measurements on B, it is sufficient to exclude them for all sharp measurements. However, testing only with projectors onto, say, the six Pauli eigenstates, which may be more easily accessible than more general measurements in a given experimental architecture, is generally not sufficient. Indeed, one can construct pathological examples of causal maps that only produce non-zero negativity in χ^b_{CD} under post-selection on a single pure state on B. This shows that it is in general necessary to test all sharp measurements in order to determine whether any of them induce non-classical correlations.

A strictly stronger signature of non-classical correlations in χ^b_{CD} is the violation of a Bell-type inequality. Recall that entanglement in a bipartite state means that the measurement statistics it generates cannot be explained by a state of the form $\sum_j P(J=j)\,\rho^j_C \otimes \rho^j_D$, with quantum systems C and D *of a given dimension*. Such a state could be prepared by selecting a classical setting J according to some probability distribution, broadcasting it to the two parties and having them prepare their quantum systems according to the classical setting. A Bell-inequality violation, on the other hand, rules out any explanation in terms of a classical hidden variable being broadcast to the two parties and influencing their measurement statistics, without making any assumptions about the dimension of the systems or whether they are even described by quantum mechanics. In this sense, Bell-inequality violations allow one to draw explicit conclusions about the kinds of causal model that can account for the observed correlations. However, such violations are strictly harder to generate than entanglement: while it is easy to see that every state that violates a Bell inequality must be entangled and the converse holds for pure states, there exist explicit examples of mixed entangled states that do not violate any Bell-type inequality [94]. In particular, the induced states χ^b_{CD} that arise when A and B are both qubits are apparently too mixed to violate any Bell inequalities (see example below), though it may be possible to construct such scenarios in higher dimensions.

If one wants to further minimize the number of preparations and measurements required to detect entanglement in χ^b_{CD} – especially if one knows the expected form of χ^b_{CD} – one can also adapt existing results on entanglement witnesses for bipartite states. This is essentially a problem in Bayesian inference between quantum variables: in the conventional setting for entanglement witnesses, one is given a state ρ_{CD} prepared by an unknown circuit and seeks a POVM, that is, a conditional of the form $W_{Y|XCD}$, which produces a classical outcome Y given a setting X, such that the resulting statistics $Tr\left[W_{Y=y|X=x,CD}\rho_{CD}\right] = P(Y=y|X=x)$ bear witness to entanglement between C and D. In our case, C is prepared by the circuit fragment inside the dashed box in Fig. 2.3d, but D is an input that is sent back into that circuit after being probed, and their relation is encoded in the map $\mathcal{E}^b_{C|D}$ that is Choi isomorphic to the state χ^b_{CD}. Accordingly, the operator W for this scenario must represent a set of quantum instruments from C to D, with classical setting X and outcome Y: it takes the form $W_{YD|XC}$. Given an entanglement witness $W_{Y|XCD}$, one can derive the corresponding set of instruments, which form the conditional $W_{YD|XC}$. The details of the translation are a lengthy exercise in Bayesian inversion, but since they do not add any new insights to the present discussion, we do not reproduce them here.

Example 77. Induced quantum mutual information and negativity for the intrinsically quantum physical combination of CC and CE introduced in example 66. Under post-selection on finding the state $|0\rangle$ on B, that scenario gives rise to the induced Choi operator

$$\chi^b_{CD} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}\left[\cos\frac{\theta}{2}\,|01\rangle - \mathrm{i}\sin\frac{\theta}{2}\,|10\rangle\right]\left[\cos\frac{\theta}{2}\,\langle 01| + \mathrm{i}\sin\frac{\theta}{2}\,\langle 10|\right]. \tag{5.78}$$

The quantum mutual information in this state is less than 1 bit for all θ, which is not sufficient to

witness non-classicality, nor does the state violate any Bell inequality. The reason in both cases is the admixture of $|00\rangle$: if not for this term, χ^b_{CD} would be pure and maximally entangled, with a full 2 bits of mutual information and a maximal Bell-inequality violation. (The mixedness arises because we choose A and B to have the same dimensionality: this implies that $h_D h_E = h_B^2$, and consequently any gate from DE to B must also produce an ancillary output, F, with $h_F = h_B$, which is traced out, incurring a loss of information and therefore purity.) For this reason, both mutual information and Bell inequality violations are generally not promising as sufficiently sensitive tests of intrinsically quantum mixtures of causal relations between two variables of equal dimension. Negativity, on the other hand, can detect the non-classicality in this case despite the mixedness of χ^b_{CD}:

$$\mathcal{N}\left(\chi^b_{CD}\right) = \frac{1}{4}\left[\sqrt{1+\sin^2\theta} - 1\right], \tag{5.79}$$

which is non-zero for all but the extremes $\theta = 0, \pi$.

5.4 Experiment

The following section describes a quantum optics experiment[27] that realizes an intrinsically quantum combination of CC and CE relations between two qubits, as well as classical physical and probabilistic mixtures of CC and CE for contrast, and demonstrates how the indicators introduced in the previous section herald different types of causal relations. The experiment proceeds much as described in section 4.2: the abstract structure of the circuit that relates A and B is as shown in Fig. 2.3d, preparing an initial bipartite state

$$\rho_{CE} = \left|\Phi^+\right\rangle\left\langle\Phi^+\right| \tag{5.80}$$

followed by different choices of the two-qubit gate $\mathcal{E}_{BF|DE}$ (after which F is traced out), which imply different causal relations between A and B. The qubits are encoded in the polarization of photons, which is measured using polarizing beam-splitters combined with wave-plates so as to only transmit one eigenstate and dump the photon if it is in the orthogonal state. Here we do not restrict ourselves to observational probing of A, but allow independent measurements on C and preparations on D. We observe the correlations between these and the outcomes of measurements on B. All three range over the six eigenstates of Pauli observables. The choices of Pauli observable are denoted by $s, t, u \in \{1, 2, 3\}$ for C, D and B, respectively, and the resulting eigenvalues by $c, d, b \in \{\pm 1\}$, as shown in Fig. 2.5. The experimental setup is shown in Fig. 5.3.

The central element is the partial swap gate, introduced in eq. (5.21):

$$U_{BF|DE} = \cos\frac{\theta}{2}\,\mathbb{I}_{B|D}\otimes\mathbb{I}_{F|E} + \mathrm{i}\sin\frac{\theta}{2}\,\mathbb{I}_{\mathrm{B|E}}\otimes\mathbb{I}_{\mathrm{F|D}}. \tag{5.81}$$

The term incorporating $\mathbb{I}_{B|D}$ realizes a CE path from D to B, whereas the term incorporating $\mathbb{I}_{B|E}$ realizes a CC connection between C and B, since E is prepared in a maximally entangled state with C. In the partial swap unitary, the two terms are combined in a manner which is, as various indicators will show, distinctly non-classical. The experimental implementation follows Černoch et al. [95]: an interferometer is modified such that two indistinguishable photons arriving simultaneously along different paths will only exit along different paths, and therefore be able to produce a coincidence count in the subsequent detection, if (a) their polarization state lies in the triplet subspace and both follow one particular path through the interferometer, or (b) their polarization state is the singlet and they follow different paths through the interferometer. By adding a phase shift of $\frac{\theta}{2}$ in one path, one can introduce a phase difference $-\theta$ between the triplet and singlet subspaces of polarization states. Recalling that the projectors onto these subspaces can be expressed as $\frac{1}{2}(\mathbb{I} \pm S)$, where $\mathbb{I} = \mathbb{I}_{B|D}\otimes\mathbb{I}_{F|E}$ and $S = \mathbb{I}_{B|E}\otimes\mathbb{I}_{F|D}$ denote the two-qubit identity and swap gates, respectively, one

[27] The experiment was designed, realized and analysed in collaboration with J.-P. MacLean, R. W. Spekkens and K. J. Resch (see statement of contributions in the front matter).

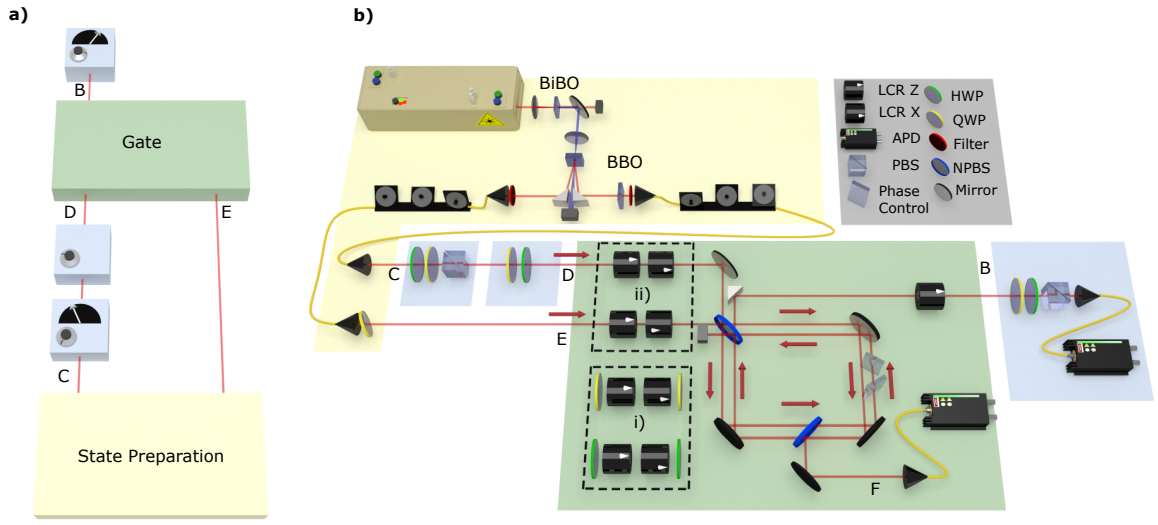

Figure 5.3: Experimental setup implementing an intrinsically quantum combination of CC and CE relations between the polarization degrees of freedom of two photons. (a) Coloured boxes distinguish functional components of the setup that correspond to the different elements of the abstract circuit (Fig. 2.3d): preparation of an entangled state (yellow), measurement of photon A (grey), partial swap gate combined with dephasing (green) and measurement of photon B (grey). (b) Experimental setup. Notation for optical components: Bismuth-Borate (BiBO), β-Barium-Borate (BBO), half-wave plate (HWP), quarter-wave plate (QWP), liquid-crystal retarder (LCR), polarizing beam splitter (PBS), non-polarizing beam splitter (NPBS), avalanche photo diode (APD).

finds that the unitary implemented by the interferometer is indeed the partial swap (up to a global phase),

$$\frac{1}{2}\left(\mathbb{I}+S\right)+e^{-\mathrm{i}\theta}\frac{1}{2}\left(\mathbb{I}-S\right)=e^{-\mathrm{i}\theta/2}\left[\frac{e^{+\mathrm{i}\theta/2}+e^{-\mathrm{i}\theta/2}}{2}\mathbb{I}_{B|D}\otimes\mathbb{I}_{F|E}+\frac{e^{+\mathrm{i}\theta/2}-e^{-\mathrm{i}\theta/2}}{2}\mathbb{I}_{B|E}\otimes\mathbb{I}_{F|D}\right]. \tag{5.82}$$

In order to explore the transition from intrinsically quantum to effectively classical causal relations, we combine the partial swap with dephasing noise of variable intensity on D, E, B and F. (Note that dephasing on F becomes irrelevant when we subsequently trace out that variable, and the experimental realization does not apply dephasing on F in order to simplify the setup. However, in order to make the present derivation more even-handed, we consider dephasing on all four variables.) Let

$$\Delta_{p,\hat{n}}\left(\rho\right)\equiv\left(1-\frac{p}{2}\right)\rho+\frac{p}{2}\left(\hat{n}\cdot\vec{\sigma}\right)\rho\left(\hat{n}\cdot\vec{\sigma}\right) \tag{5.83}$$

denote the channel that, with probability p, applies dephasing on the eigenbasis of $\hat{n}\cdot\vec{\sigma}$, defined by the unit vector $\hat{n}$. If $p=0$, it reduces to the identity channel; if $p=1$, the effect is that of a (non-destructive) projective measurement of $\hat{n}\cdot\vec{\sigma}$, which destroys coherences, leaving only probabilistic mixtures of the two eigenstates of $\hat{n}\cdot\vec{\sigma}$. Total dephasing ($p=1$) effectively reduces the qubit to a classical bit specifying which of the two eigenstates is realized. When a gate $\mathcal{E}_{BF|DE}$ is combined with total dephasing on both inputs and outputs, it can be effectively described as a classical conditional relating the eigenvalues d, e, b and f, which, in the case of qubits, we take to be $d,e,b,f\in\{\pm1\}$. That is, the map defined by

$$\tilde{\mathcal{E}}_{BF|DE}\left(\rho_{DE}\right)\equiv\left(\Delta^{B}_{1,\hat{n}_B}\otimes\Delta^{F}_{1,\hat{n}_F}\right)\circ\mathcal{E}_{BF|DE}\circ\left(\Delta^{D}_{1,\hat{n}_D}\otimes\Delta^{E}_{1,\hat{n}_E}\right)\left(\rho_{DE}\right) \tag{5.84}$$

takes the form

$$\tilde{\mathcal{E}}_{BF|DE}(\rho_{DE}) = \sum_{bfde\in\{\pm1\}} |b_{\hat{n}_B}\rangle\langle b_{\hat{n}_B}| \otimes |f_{\hat{n}_F}\rangle\langle f_{\hat{n}_F}| P(bf|de) \langle d_{\hat{n}_D}|\langle e_{\hat{n}_E}|\rho_{DE}|d_{\hat{n}_D}\rangle|e_{\hat{n}_E}\rangle, \quad (5.85)$$

where $|b_{\hat{n}_B}\rangle$ denotes an eigenstate of $\hat{n}_B\cdot\vec{\sigma}$ with eigenvalue b, and similarly for F, D and E, and $P(bf|de)$ is a conditional probability distribution. For example, the effect of the partial swap unitary in the $\hat{z}\cdot\vec{\sigma}=\sigma_3$ eigenbasis is

$$\begin{cases} U|00\rangle = |00\rangle \\ U|01\rangle = \cos\frac{\theta}{2}\,|01\rangle + \mathrm{i}\sin\frac{\theta}{2}\,|10\rangle \\ U|10\rangle = \mathrm{i}\sin\frac{\theta}{2}\,|01\rangle + \cos\frac{\theta}{2}\,|10\rangle \\ U|11\rangle = |11\rangle. \end{cases} \quad (5.86)$$

By combining the associated gate with total dephasing in the eigenbasis of σ_3 on D, E, B and F, one obtains a map $\tilde{\mathcal{E}}_{BF|DE}$ that eliminates all coherences between eigenstates of σ_3; making, for example, $\tilde{\mathcal{E}}_{BF|DE}(|01\rangle\langle 01|) = \cos^2\frac{\theta}{2}|01\rangle\langle 01| + \sin^2\frac{\theta}{2}|10\rangle\langle 10|$. It can be effectively described by specifying only the conditional probability distribution over eigenvalues of σ_3 found on the four qubits,

$$\begin{cases} P(bf|de=++) = \delta(bf,++) \\ P(bf|de=+-) = \cos^2\frac{\theta}{2}\,\delta(bf,+-) + \sin^2\frac{\theta}{2}\,\delta(bf,-+) \\ P(bf|de=-+) = \sin^2\frac{\theta}{2}\,\delta(bf,+-) + \cos^2\frac{\theta}{2}\,\delta(bf,-+) \\ P(bf|de=--) = \delta(bf,--). \end{cases} \quad (5.87)$$

If we subsequently trace out qubit F, the conditional probability distribution takes the simple form

$$P(b|de) = \cos^2\frac{\theta}{2}\,\delta(b,d) + \sin^2\frac{\theta}{2}\,\delta(b,e), \quad (5.88)$$

which is manifestly a probabilistic mixture of B depending either only on D or only on E. Since the outcomes c and e of σ_3-measurements on C and E are perfectly correlated, we furthermore expect

$$P(cd|b) = \cos^2\frac{\theta}{2}\,\delta(b,d) + \sin^2\frac{\theta}{2}\,\delta(b,c). \quad (5.89)$$

Consequently, the causal relation between qubits A and B in this scenario can be described as a probabilistic mixture of CE and CC, and because of the total dephasing, it is also effectively classical.

In order to implement a causal relation that is effectively classical but a physical mixture, we combine the partial swap with total dephasing on different bases. The construction of this example builds on the definition of the witness of physical mixture given in eq. (5.52). Recall that, for the partial swap alone, without any dephasing, we have $W\neq 0$, which rules out a probabilistic mixture (as shown in the example at the end of section 5.3.3). Notably, the witness is defined in terms of the outcomes of measurements of fixed (Pauli) observables, namely, σ_1 on C, σ_2 on D and σ_3 on B, and therefore its value will remain unchanged if one adds dephasing on the respective eigenbases. Furthermore, the particular state prepared on CE in this experiment takes a particular form in the eigenbasis of σ_1,

$$\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|{+}{+}\rangle + |{-}{-}\rangle), \quad (5.90)$$

and in fact on the eigenbasis of any $\hat{n}\cdot\vec{\sigma}$ with $\hat{n} = \cos\eta\hat{x} + \sin\eta\hat{z}$. This ensures that dephasing on any such eigenbasis at C generates the same state on CE as if the same dephasing had been applied on E: for the case of $\hat{n}=\hat{x}$,

$$\left(\Delta_C^{1,\hat{x}}\otimes\mathcal{I}_E\right)\frac{1}{2}(|{+}{+}\rangle + |{-}{-}\rangle)(\langle{+}{+}| + \langle{+}{+}|) = \left(\mathcal{I}_C\otimes\Delta_E^{1,\hat{x}}\right)\frac{1}{2}(|{+}{+}\rangle + |{-}{-}\rangle)(\langle{+}{+}| + \langle{+}{+}|), \quad (5.91)$$

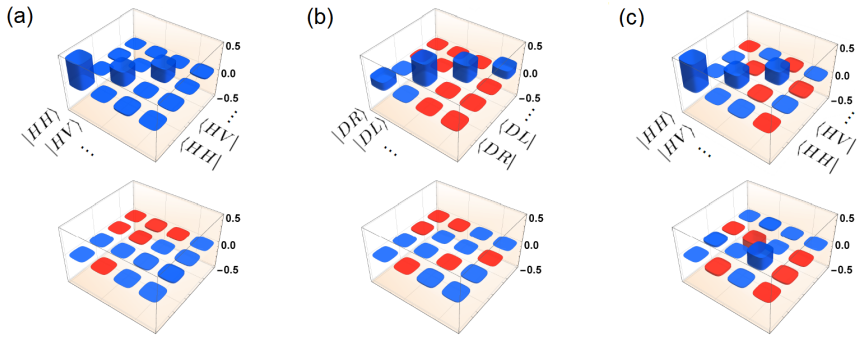

Figure 5.4: Choi states χ^b_{CD} induced by post-selection on the $|0\rangle$-state on B from experiments aiming to implement (a) a classical probabilistic mixture, (b) a classical physical mixture, and (c) an intrinsically quantum physical mixture; specifically the standard examples of each category introduced in the text, with $\theta = \frac{\pi}{2}$. The top (bottom) panels show real (imaginary) components; blue (red) represents positive (negative) values. Panels (a) and (b) are shown in the eigenbases of $\sigma_3^C \otimes \sigma_3^D$ and $\sigma_1^C \otimes \sigma_2^D$, respectively, which diagonalize the operators we aim to implement, in order to highlight their effectively classical nature. The diagonal matrix elements reduce to the elements of the effective classical probability distributions derived in eqs. (5.89) and (5.94), respectively. Panel (c) is close to the expected state induced by an intrinsically quantum combination, eq. (5.23) with $\theta = \frac{\pi}{2}$, and is manifestly non-classical (indeed, its negativity is $\mathcal{N} = 0.083 \pm 0.003$).

and similarly for generic $\hat{n} = \cos\eta\hat{x} + \sin\eta\hat{z}$ instead of $\hat{x}$. (From the equality for total dephasing one can easily derive a similar equality for any probability of dephasing p.) In summary, by combining the partial swap unitary with total dephasing on the eigenbasis of σ_2 on D, σ_3 on B and σ_1 on E, while keeping $\rho_{CE} = |\Phi^+\rangle\langle\Phi^+|$, one implements a causal map that produces $W = -\frac{1}{2}$, and therefore cannot be explained as a mere probabilistic mixture of CC and CE. At the same time, all variables are effectively classical: by tracing over F in eq. (5.85) (and substituting $\hat{n}_B = \hat{z}$, $\hat{n}_D = \hat{y}$ and $\hat{n}_E = \hat{x}$), we find that the map from DE to B takes the form

$$\mathcal{E}_{B|DE}(\rho_{DE}) = \sum_{bde\in\{\pm1\}} |b_{\hat{z}}\rangle\langle b_{\hat{z}}| P(b|de) \langle d_{\hat{y}}| \langle e_{\hat{x}}| \rho_{DE} |d_{\hat{y}}\rangle |e_{\hat{x}}\rangle , \tag{5.92}$$

and the classical eigenvalues that specify the states of E, D and B are related by

$$P(b|de) = \frac{1}{2}\mu(b) + \frac{1}{2}\delta(b, -de) . \tag{5.93}$$

One can see that, despite the admixture of the uniform distribution over b (due to the information loss incurred when tracing out F), the conditional $P(b|de)$ makes b depend on a combination of d and e, and therefore represents a non-trivial physical mixture of CC and CE. By extension, the overall causal map $\mathcal{E}_{CB|D}$ is also effectively classical, and its effects are completely specified by a classical conditional probability distribution relating the eigenvalues c, d and b, which takes the form

$$P(cd|b) = \frac{1}{2}\mu(c)\mu(d) + \frac{1}{2}\delta(c, -d) . \tag{5.94}$$

The experiment focuses on these three standard examples of combinations of CC and CE: pure partial swap (intrinsically quantum physical mixture), partial swap with total dephasing in the eigenbases of σ_1 on C, σ_2 on D and σ_3 on B (classical physical) and partial swap with total dephasing in the σ_3 eigenbasis on C, D and B (classical probabilistic mixture). The experimental data collected in each scenario, denoted $\tilde{P}^{obs}(cdbstu)$, is the set of count numbers obtained for combinations of

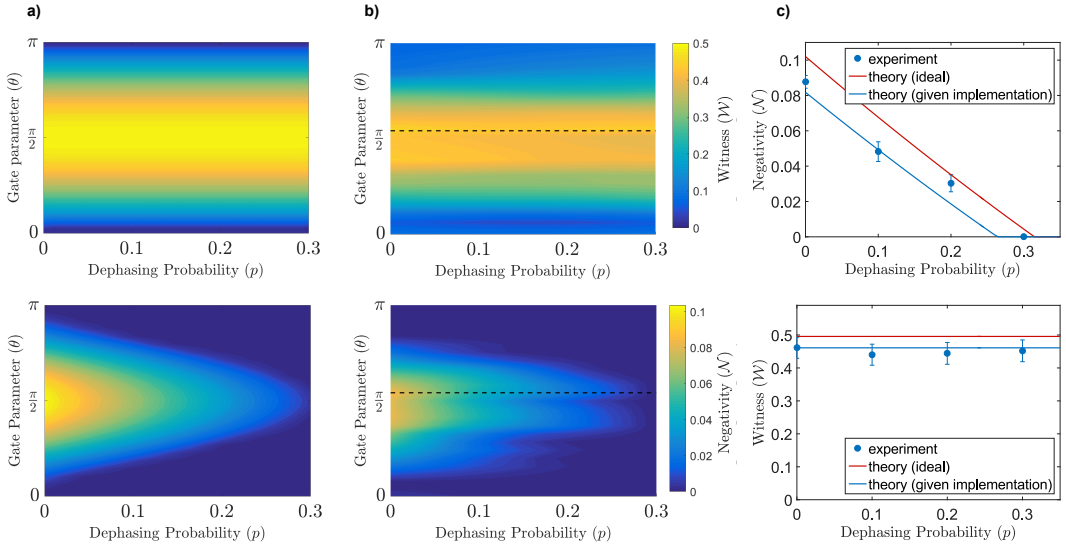

Figure 5.5: Measures of induced correlation for a family of causal relations: (a) theory, (b) experiment, (c) cross-section of contour plot at dashed line, $\theta = 97.4°$. The witness $W \neq 0$ heralds a physical mixture without distinguishing classical from quantum, while the negativity $\mathcal{N} > 0$ attests to an intrinsically quantum physical mixture. The partial swap parameter θ introduced in eq. (5.21) interpolates between purely CE ($\theta = 0$) and purely CC ($\theta = \pi$) relations, ranging over intrinsically quantum combinations. The dephasing probability p controls the transition to the classical limit; as it reaches $p = 0.3$, all evidence of non-classical behaviour is lost.

settings s, t, u and eigenvalues c, b, d. As one ranges over six Pauli eigenstates on each qubit, this produces $6^3 = 216$ count numbers. They are more than sufficient for reconstructing the full causal map by a least-squares fit, similar to the fitting described in section 4.2: the model takes the form

$$\tilde{P}^{mod}\left(cdbstu\right) = NTr\left[\Pi_C^{s,c} \otimes \Pi_B^{u,b} \otimes \Pi_D^{t,d} \tau_{CB|D}\right], \tag{5.95}$$

where we assume that each combination $cdbstu$ is tested with an equal number of entangled pairs N. The least-squares fit minimizes the residue

$$\chi^2 \equiv \sum_{cdbstu} \frac{\left[\tilde{P}^{mod}\left(cdbstu\right) - \tilde{P}^{obs}\left(cdbstu\right)\right]}{\tilde{P}^{mod}\left(cdbstu\right)}, \tag{5.96}$$

subject to the constraints that $T_D \tau_{CB|D}$ must be positive-semidefinite, which can be ensured by parametrizing it following [48], and that marginalizing over B gives rise to a map that traces out D and prepares a fixed state on C, which can be enforced with a penalty term of the form

$$\lambda \sum_{ij} \left| \left(Tr_B \tau_{CB|D} - Tr_{BD} \tau_{CB|D} \otimes \mathbb{I}_D\right)_{ij} \right|^2. \tag{5.97}$$

Once the experimental causal map has been reconstructed, one can quantify how close it is to the intended target by computing the fidelity between the reconstructed and intended Choi states, χ_{CBD}. By this measure, the experiment realized the classical probabilistic mixture with $(98.1 \pm 0.2)\,\%$ fidelity, the classical physical mixture with $(98.06 \pm 0.08)\,\%$ fidelity and the intrinsically quantum physical mixture with $(93.7 \pm 0.3)\,\%$ fidelity, with uncertainties estimated by a Monte Carlo simulation that added Poisson noise to the observed count numbers.

The indicators W and $\mathcal{N}$ introduced above are evaluated from different subsets of this data. Fixing $s = 1$, $t = 2$, $u = 3$ and using the shorthand $\tilde{P}^{obs}(cdb) \equiv \tilde{P}^{obs}(cdb, s = 1, t = 2, u = 3)$, we evaluate

$$W = 8 \sum_{b=\pm 1} b \left[\tilde{P}^{obs}(++b)\tilde{P}^{obs}(--b) - \tilde{P}^{obs}(+-b)\tilde{P}^{obs}(-+b) \right] / \left(\sum_{cdb} \tilde{P}^{obs}(cdb) \right)^2 . \quad (5.98)$$

In order to assess the negativity $\mathcal{N}$ induced by post-selecting on the state $|0\rangle$ on B, we fix instead $u = 3$, $b = +1$ and fit the resulting 36 count numbers $\tilde{P}^{obs}(cdst) \equiv \tilde{P}^{obs}(cd, b = +1, st, u = 3)$ to a model of the form

$$\tilde{P}^{mod}(cdst) = NTr \left[\Pi_C^{s,c} \otimes \Pi_B^{3,+} \otimes \Pi_D^{t,d} \tau_{CB|D} \right] \quad (5.99)$$

where $T_D \tau_{C|D}^b$ is once again a positive-semidefinite operator. Note that we do not enforce any other constraints; in particular, one should not assume that the marginal on the input variable $Tr_C \tau_{C|D}^b$ is the identity operator because the map $\mathcal{E}_{C|D}^b$ induced by post-selection need not be trace-preserving. The best-fitting induced Choi states $\chi_{CD}^b \equiv \frac{1}{2} T_D \tau_{C|D}^b$ for the three scenarios are shown in Fig. 5.4. The induced states from the experiments targeting classical probabilistic and physical mixtures are, to good approximation, diagonal in the preferred bases, and their diagonal matrix elements reproduce the expected joint probability distributions eqs. (5.89) and (5.94), respectively. In the experiment that targeted an intrinsically quantum physical mixture, we find that χ_{CD}^b shows a clear signature of entanglement, with negativity $\mathcal{N} = 0.083 \pm 0.003$. This shows that we successfully realized an intrinsically quantum physical mixture of common-cause and cause-effect relations between two variables.

Finally, we track the witness of physical mixture W and the signature of non-classical combinations $\mathcal{N}$ for a two-parameter family of causal relations, with the partial swap parameter ranging from $\theta = 0$ (identity channel, purely CE) to π (maximally entangled state $|\Phi^+\rangle$, purely CC) and adding dephasing with probability p ranging from 0 to 1, on the bases defined by $\hat{n}_C = \hat{x}$, $\hat{n}_D = \hat{y}$, $\hat{n}_B = \hat{z}$. The results are shown in Fig. 5.5 for the interval $p \leq 0.3$, since above this value all evidence of non-classical behaviour is lost and the data do not exhibit any noteworthy features. Due to the choice of dephasing bases, the witness W is unaffected by the dephasing; it attests to a physical mixture of CC and CE for all values of the parameters except for the extremes $\theta = 0, \pi$, as expected. The indicator of intrinsically quantum physical mixture $\mathcal{N}$, on the other hand, is reduced to zero as the dephasing probability reaches $p = 0.3$. (This does not imply that all data beyond this point can be explained by a classical physical mixture, since $\mathcal{N} > 0$ is only a sufficient condition for detecting an intrinsically quantum physical mixture, but it is consistent with this hypothesis. We have only given an explicit effectively classical model in the limit $p = 1$.) For all values of p below 0.3, however, the non-zero values of $\mathcal{N}$ verify that we have indeed realized an intrinsically quantum combination of CC and CE relations between the two qubits.

6 Conclusions and outlook

6.1 Quantum variables on simple causal structures

The first goal of this thesis was to investigate the principles that govern information and knowledge in the quantum sphere – specifically from the point of view of causal models – and to pin down how they differ from their classical counterparts, in hopes of gaining a deeper understanding of quantum theory. For this purpose, we replaced classical random variables by a quantum generalization, but considered only simple causal structures of the same types allowed in classical causal models: two variables related either as cause and effect or by a common cause, or at most a combination of one mechanism of each type. In this setting, one can define generalizations of joint probability distributions, belief propagation and Bayesian inversion, to name a few, that reproduce the desirable properties of their classical counterparts. Many elements of the conventional formalism of quantum mechanics, such as preparations and different types of measurement, can be represented as part of this unified framework.

This approach provides a new perspective on the fundamental question of what one can know and learn about a quantum system, considering that any attempt to gain information about the system tends to alter its subsequent behaviour. The framework of causal models accommodates this fact by splitting each quantum variable A into two versions, C and D – one before, the other after the intervention –, and providing distinct analyses of what one learns about each of them. The necessity of distinguishing two versions of A becomes apparent if one considers the map of inferences from A to a second variable, B, related by a combination of common-cause and cause-effect: this map can exhibit explicit interference[28] between the inferences based on C and D.

When splitting a quantum variable A, one must modify the causal structure so as to ensure that C has no causal children and D has no causal parents among the remaining variables. This leaves both versions accessible for an agent to acquire information about them (using measurements on C and preparations on D) and therefore provides an operationally meaningful way to consider "what an agent knows about the variable". Based upon this, one can generalize joint and conditional probability distributions to operators over any number of quantum variables.

In most applications of quantum measurements, the disturbance that inevitably accompanies them is treated as a kind of noise, which obscures the information that one seeks to extract from the system. In the context of causal models, however, the fact that measuring a quantum variable does not simply reveal a single, fixed piece of information about it turns out to be a useful resource: it allows for a difference between what one learns about C and D, which is essential in obtaining complete information about the variable A and its relation to other variables. This situation is somewhat reminiscent of the cases of the no-cloning theorem and the fundamentally probabilistic nature of quantum mechanics: once viewed as constraints that hamper one's abilities, they were later recognized as powerful resources that provide secure communication and strong randomness[29], respectively. (As an aside, we note that this strong randomness provided by quantum mechanics may in turn simplify our understanding of causality, since it resolves the philosophical problem of free will that limits the interventionist definition of causation.) The insight that the violation of informational symmetry can be a useful resource provides both motivation and a new context for further study of quantum measurements, focusing on the origin and implications of informational asymmetry.

Another long-standing puzzle is that the quantum state, despite being the optimal operational characterization of a quantum variable, does not provide a complete description of that variable – unlike the value of a classical variable, which specifies it completely. In classical causal models, knowing the value of one variable can make two others statistically independent, depending on the

[28] That is, the probability amplitudes predicted by each piece of information alone are not combined by simple addition or convex combination.

[29] There is ongoing debate in computer science and philosophy about whether it is possible to even verify true randomness. However, by any testable standard, quantum mechanics can provide very strong randomness.

causal structure relating them. This screening of inferences plays an important role throughout the framework. In quantum causal models, the fact that one cannot give a complete description of a quantum variable manifests itself as a breakdown of this screening. In some cases, the issue can be resolved by splitting the variable in question, such as A in the chain $F \to A \to B$, where the pre-intervention version of A is only related to F and the post-intervention version only to B. However, the same approach does not work for common-cause relations, such as $F \leftarrow A \to B$, because the degrees of freedom related to F and B are not separated by the distinction between pre- and post-intervention versions. More research is needed to determine whether one can recover the notion of conditional independence in the case of a quantum common cause – either by modifying the splitting prescription or by some other approach – or whether the failure of screening in this scenario reflects a more fundamental effect than in the cause-effect chain.

Finally, let us address the question of whether the indeterminacy in a quantum state is merely a matter of ignorance of the agent studying the system (epistemic) or an intrinsic feature of some underlying reality (ontic), or some non-trivial combination of the two. While it is not clear how this difference would manifest itself in a quantum causal model, the causal framework does make a very clear distinction between a closely related pair of concepts, namely inference and influence. One can identify epistemic states are those that can be updated by inference, whereas ontic states only change in response to a (causal) influence. With further development, the framework of quantum causal models is therefore a promising tool for gaining some insight into the nature of the quantum state.

6.2 Quantum conditionals

Going beyond the effects that essentially concern just a single variable, we also considered the properties of conditionals that relate quantum variables. One of the most striking observations in this context is that quantum conditionals have different mathematical forms depending on the underlying the causal structure. This contrasts with classical conditionals, which are independent of the causal structure. As a pedagogical example of these different forms, consider an inference map between two variables related as cause and effect: this map must be completely positive, since one is free to input any state, including one that is entangled with an ancilla. If the two variables have a common cause, on the other hand, then the map must be completely co-positive. The partial transposition that differentiates common-cause from cause-effect conditionals can be traced back to the intrinsic asymmetry of the causal arrow, which distinguishes parents from children and the pre-intervention version from the post-intervention version of a variable. Indeed, in the case of a general causal model, whose variables can be split into a set $\mathcal{C}$ of pre-intervention versions and a set $\mathcal{D}$ of post-intervention versions, the conditional $\tau_{\mathcal{C}|\mathcal{D}}$ must be positive under partial transposition on the set of post-intervention variables, $\mathcal{D}$.

Positivity under partial transposition is the simplest consistency condition imposed on a conditional relating several quantum variables, but it is not the only one. This is illustrated by the case of two time-ordered variables, A and B: in this case, we also demand that the marginal of the conditional factorize, $\tau_{BC|D} = \rho_C \otimes \mathbb{I}_D$, because the input, D, is causally posterior to the output, C. Similar constraints will apply to more general conditionals in order to prevent causal cycles. A second type of constraint applies if two variables, B_1 and B_2, share a common cause A: in this case, the conditionals by which each B_i depends on A are subject to a monogamy-type constraint, which prevents them from both obtaining perfect copies of A. More exotic types of constraints may exist, and the task of completely characterizing the set of quantum conditionals that are compatible with a given causal structure remains for future work. It ties into the study of how certain signatures in the joint statistics can reflect properties of the underlying causal structure, which is an active topic in classical causal modelling.

The role of partial transposition in causal models provides an operational interpretation of the class of PPT states. We have shown that the classes of separable states, non-separable PPT states,

NPT states and PPT negative operators can be characterized exactly in terms of the classes of causal structures from which they can arise. At the same time, the mathematical properties of states have implications for the correlations that these states can exhibit, such as bound or distillable entanglement, which play a central role in quantum communication. Our result therefore allows us identify classes of correlations with the classes of causal structures that can give rise to them, providing a natural representation for communication problems that involve several parties signalling each other through pre-defined channels. In particular, by analysing whether a quantum cause-effect structure can give rise to bound entanglement, one could settle the open question of whether there exist bound entangled NPT states.

Aside from these more theoretical implications, the fact that quantum conditionals witness the underlying causal structure can be exploited for causal discovery. To make this claim rigorous, we began by exploring what one can learn about the versions C and D by different ways of probing the quantum variable, and in particular how this compares to classical passive observation. Although passive observation itself does not admit an exact quantum analogue, we defined a class of probing schemes that are like passive observation in the way that is relevant for causal inference; namely, in that they reveal the same information about C and D. This class of so-called observational schemes does extend to quantum variables, and we show that, for a particular probing scheme in this class, one can characterize not the entire operator space $\mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_D)$, but only a subspace. Notably, this information is sufficient to distinguish whether the variable is related to a second, later one as cause and effect or via an unobserved common cause. Classically, observational probing schemes cannot distinguish the causal structure in this case. The effect we have found is therefore uniquely quantum, and it adds a new paradigm to the toolbox of causal discovery.

The observation that not all quantum conditionals must correspond to CP maps promises a better understanding of how such maps can arise in the dynamics of open quantum systems, as well as the closely related phenomenon of quantum non-Markovianity. An analysis in terms of causal structure reveals that violations of complete positivity and Markovianity can both be traced to common-cause relations. Indeed, we know that common-cause relations are represented by maps that are completely co-positive, but not necessarily completely positive. This explains why CP maps alone are inappropriate for such scenarios, requiring cumbersome extensions or breaking down in contradiction altogether. The formalism of quantum conditionals (and causal maps in particular) offers a more appropriate representation of the relation between the states of a system at different times under non-Markovian dynamics, along with the corresponding physical interpretation.

By extension, one can use several tests derived within our formalism to detect and quantify violations of Markovianity. However, in order to design tests that can be applied in a realistic setting, one should not assume that one can apply even an observational probing scheme on the principal system alone, since the fact that it is hard to measure or prepare the principal system without also affecting the environment lies at the root of the entire problem. Instead, one should combine the insights and mathematical formalism of quantum causal models with a detailed description of the instruments that *can* be used to probe S_1 in a given implementation, including how they affect the environment. One can then derive what one can learn about the causal map – and consequently deduce about the causal relation – subject to those constraints, similarly to what we have done for the restriction to observational probing in section 3.3.

6.3 Causal structure in a quantum world

The more ambitious goal in studying quantum causal models is to understand the possible modifications to the causal structure that arise when one takes into account quantum theory. By contrast, note that the previous section considered only the implications of non-classical features of the *conditionals*, which encode how each variable depends on its parents within a given causal structure. For example, in conventional quantum mechanics, quantum channels that preserve entanglement allow the input (cause) to affect the output (effect) in a way that cannot be simulated by a classical

channel. Non-classical features of the causal *structure*, however, are a different class of effects.

As mentioned earlier, the general causal relations among a set of quantum variables can be summarized in the form of a single overall conditional $\tau_{\mathcal{C}|\mathcal{D}}$. But, although this is a complete characterization of the causal relations, it does not encode the causal structure in a particularly accessible way. The appeal of causal models – both their philosophical elegance and their usefulness for practical applications – stems from their modularity; the fact that they decompose a network of causal relations into individual, autonomous mechanisms. In the context of classical causal models, this motivation gave rise to the entire field of causal discovery, tasked with analysing observed joint probability distributions in order to determine which combinations of causal mechanisms can explain them. The problem of causal discovery on a set of quantum variables can certainly benefit from the classical techniques, but it will most likely also face uniquely quantum features of the causal structure which cannot be analysed using existing techniques.

In order to study such features, we pursue a bottom-up approach and consider how certain elementary components that form the causal structure can be combined in non-classical ways. It is not yet clear what one should consider as elementary components for this purpose: whether it should be causal mechanisms (giving one variable in terms of its parents), different paths that connect two variables within a single larger classical causal structure, or even distinct causal structures that are considered incompatible in the classical framework, such as $A \to B$ and $B \to A$. In theory, one should use the extremal elements of the space of causal relations between sets of quantum variables (in the sense that suitable combinations of them can generate all other elements), but since we are only beginning to explore the mathematical structure of this space, it is not clear what its extremal elements are. As a first example, we therefore consider a scenario wherein two possible definitions coincide: in the case of two causally ordered variables, there are two distinct causal relations – purely common-cause and purely cause-effect –, which are also realized by different causal mechanisms. We consider combinations of these two causal relations.

Before turning to the classification of quantum combinations of causal relations, we first consider the same scenario with classical variables. Specifically, we contrast the space of causal relations between classical variables with the space of states of a single classical variable: the latter is the space of probability distributions, whose extremal elements are delta distributions, while all other elements are probabilistic mixtures thereof. The space of causal relations between two classical variables, by contrast, was found to include other possibilities beyond probabilistic mixtures of purely common-cause and purely cause-effect. That is, scenarios with different causal relations can be combined in more general ways than states of knowledge about a single variable. In the case of quantum variables, the space of states of a single variable includes both probabilistic mixtures and coherent superpositions of a given set of basis states. By extension, we expect that scenarios with different causal relations between quantum variables can be combined in an even richer set of ways.

A central result of chapter 5 is that one can realize a non-classical combination of common-cause and cause-effect without requiring exotic physics, post-selection or other assumptions. (We do use post-selection to assess the Berkson-type induced correlations and thereby witness the intrinsically quantum causal relation, but not to realize it in the first place.) We proved this by providing an explicit example and realizing it in a tabletop quantum optics experiment.

However, while we have proposed a sufficient condition for detecting one type of non-classical combination of causal relations, a comprehensive definition of this phenomenon is still being developed. As one considers more variables, which support a larger and more diverse set of causal relations, the ways in which these can be combined will quickly become more complex, as it is the case, for example, in multipartite entanglement. Indeed, the theory of non-classical causal structure is expected to be at least somewhat related to the theory of entanglement, since both ultimately manifest themselves in (non-classical) correlations. As with entanglement, exploring the ways in which these more general causal structures can be non-classical is likely to be a gradual process, but it promises to reveal both a rich mathematical structure and physical implications that we cannot yet foresee.

References

[1] Leifer, M. & Spekkens, R. W. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Phys. Rev. A* **88**, 052130 (2013).

[2] Ried, K. *et al.* A quantum advantage for inferring causal structure. *Nat. Phys.* **11**, 414–420 (2015).

[3] Hume, D. *An enquiry concerning human understanding* (1748).

[4] Pearson, K. *The grammar of science* (Adam and Charles Black, London, 1911), 3rd edn.

[5] Pearl, J. *Causality: models, reasoning and inference* (Cambridge Univ. Press, New York, 2000).

[6] Spirtes, P., Glymour, C. & Scheines, R. *Causation, prediction, and search* (MIT Press, Cambridge, 2000).

[7] Woodward, J. F. *The Oxford handbook of causation*, chap. Agency and interventionist theories, 234–264 (Oxford University Press, 2009).

[8] Healey, R. *The Oxford handbook of causation*, chap. Causation in quantum mechanics (Oxford University Press, 2009).

[9] Bell, J. S. *Speakable and unspeakable in quantum mechanics*, chap. La nouvelle cuisine, 232–248 (Cambridge University Press, 2004), 2nd edn.

[10] Feynman, R. P. *The character of physical law* (MIT Press, 1967).

[11] Hardy, L. Quantum theory from five reasonable axioms (2001). URL `quant-ph/0101012`.

[12] Fuchs, C. A. Quantum mechanics as information, mostly. *J. Mod. Opt.* **50**, 987–1023 (2003).

[13] Dakic, B. & Brukner, Č. *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, 365–392 (Cambridge Univ. Press, 2009).

[14] Masanes, L. & Müller, M. P. A derivation of quantum theory from physical requirements. *New J. Phys.* **13**, 063001 (2011).

[15] Chiribella, G., D'Ariano, G. M. & Perinotti, P. Informational derivation of quantum theory. *Phys. Rev. A* **84**, 012311 (2011).

[16] Fivel, D. I. Derivation of the rules of quantum mechanics from information-theoretic axioms. *Foundations of Physics* **42**, 291–318 (2011).

[17] Zaopo, M. Information theoretic axioms for quantum theory (2012). URL `arXiv:1205.2306`.

[18] Robb, A. A. *A theory of time and space* (Cambridge University Press, Cambridge, 1914).

[19] Robb, A. A. *The geometry of time and space* (Cambridge University Press, Cambridge, 1936).

[20] Hawking, S., King, A. & McCarthy, P. A new topology for curved space-time which incorporates the causal, differential, and conformal structures. *J. Math. Phys.* **17**, 174 (1976).

[21] Malament, D. The class of continuous time-like curves determines the topology of spacetime. *J. Math. Phys.* **18**, 1399 (1977).

[22] Bombelli, L., Lee, J., Meyer, D. & Sorkin, R. Space-time as a causal set. *Phys. Rev. Lett.* **59**, 521 (1987).

[23] Hardy, L. Probability theories with dynamic causal structure: A new framework for quantum gravity (2005). URL `arXiv:gr-qc/0509120`.

[24] Hardy, L. Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure. *J. Phys. A* **40**, 3081 (2007).

[25] Surya, S. Causal set topology. *Theoretical Computer Science* **1–2**, 1880–197 (2008).

[26] Wallden, P. Causal sets dynamics: Review & outlook. *Journal of Physics: Conference Series* **453**, 012023 (2013).

[27] Reichenbach, H. *The direction of time* (University of California Press, 1956).

[28] Choi, M. D. Completely positive linear maps on complex matrices. *Linear Algebra and Applications* **10**, 285–290 (1975).

[29] Jamiołkowski, A. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.* **3**, 275–278 (1972).

[30] D'Ariano, G. M. & Lo Presti, P. Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. *Phys. Rev. Lett.* **86**, 4195–4198 (2001).

[31] Leung, D. W. Choi's proof as a recipe for quantum process tomography. *J. Math. Phys.* **44**, 528 (2003).

[32] Leifer, M. S. Quantum dynamics as an analog of conditional probability. *Phys. Rev. A* **74**, 042310 (2006).

[33] Cerf, N. J. & Adami, C. Negative entropy and information in quantum mechanics. *Phys. Rev. Lett.* **79**, 5194–5197 (1997).

[34] Cerf, N. J. & Adami, C. Information theory of quantum entanglement and measurement. *Physica D: Nonlinear Phenomena* **120**, 62 – 81 (1998). Proceedings of the Fourth Workshop on Physics and Consumption.

[35] Pegg, D. T., Barnett, S. M. & Jeffers, J. Quantum retrodiction in open systems. *Phys. Rev. A* **66**, 022106 (2002).

[36] Schrödinger, E. Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society* **31**, 555–563 (1935).

[37] Wiseman, H. M., Jones, S. J. & Doherty, A. C. Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox. *Phys. Rev. Lett.* **98**, 140402 (2007).

[38] Oreshkov, O., Costa, F. & Brukner, C. Quantum correlations with no causal order. *Nat. Commun.* **3**, 1092 (2012).

[39] Modi, K. Operational approach to open dynamics and quantifying initial correlations. *Nat. Sci. Rep.* **2**, 581 (2012).

[40] Chiribella, G., D'Ariano, G. M. & Perinotti, P. Theoretical framework for quantum networks. *Phys. Rev. A* **80**, 022339 (2009).

[41] Hardy, L. The operator tensor formulation of quantum theory. *Philos. T. Roy. Soc. A* **370**, 3385–3417 (2012).

[42] Glymour, P. S. & Scheines, R. *Evolving knowledge in the natural and behavioral sciences*, chap. Causality from probability, 181 – 199 (Pitman, 1990).

[43] Janzing, D. *et al.* Information-geometric approach to inferring causal directions. *Artificial Intelligence* **182–183**, 1–31 (2012).

[44] Verstraete, F. *A study of entanglement in quantum information theory.* Ph.D. thesis, Katholieke Universiteit Leuven (2002).

[45] Jevtic, S., Pusey, M., Jennings, D. & Rudolph, T. Quantum steering ellipsoids. *Phys. Rev. Lett.* **113**, 020402 (2014).

[46] Bengtsson, I. & Życzkowski, K. *Geometry of quantum states* (Cambridge University Press, 2008).

[47] Ruskai, M. B., Szarek, S. & Werner, E. An analysis of completely-positive trace-preserving maps on 2x2 matrices. *Linear Algebra and Applications* **347**, 159–187 (2002).

[48] James, D. F. V., Kwiat, P. G., Munro, W. J. & White, A. G. Measurement of qubits. *Phys. Rev. A* **64**, 052312 (2001).

[49] Pechukas, P. Reduced dynamics need not be completely positive. *Phys. Rev. Lett.* **73**, 1060 (1994).

[50] Stelmachovic, P. & Buzek, V. Dynamics of open quantum systems initially entangled with environment: Beyond the Kraus representation. *Phys. Rev. A* **64**, 062106 (2001).

[51] Carteret, H., Terno, D. R. & Życzkowski, K. Physical accessibility of non-completely positive maps. *Phys. Rev. A* **77**, 042113 (2008).

[52] Altepeter, J. B. *et al.* Ancilla-assisted quantum process tomography. *Phys. Rev. Lett.* **90**, 193601 (2003).

[53] Weinstein, Y. S. *et al.* Quantum process tomography of the quantum Fourier transform. *J. Chem. Phys.* **121**, 6117 (2004).

[54] Howard, M. *et al.* Quantum process tomography and Lindblad estimation of a solid-state qubit. *New J. Phys.* **8**, 33 (2006).

[55] Markov, A. A. Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.* **15**, 135–156 (1906).

[56] Laine, E.-M., Piilo, J. & Breuer, H.-P. Measure for the non-Markovianity of quantum processes. *Phys. Rev. A* **81**, 062115 (2010).

[57] Ángel Rivas, Huelga, S. F. & Plenio, M. B. Quantum non-Markovianity: characterization, quantification and detection. *Reports on Progress in Physics* **77**, 094001 (2014).

[58] Mohseni, M. & Rezakhani, A. T. Equation of motion for the process matrix: Hamiltonian identification and dynamical control of open quantum systems. *Phys. Rev. A* **80**, 010101 (2009).

[59] Wolf, M. M., Eisert, J., Cubitt, T. S. & Cirac, J. I. Assessing non-Markovian quantum dynamics. *Phys. Rev. Lett.* **101**, 150402 (2008).

[60] Chruściński, D. & Kossakowski, A. Non-markovian quantum dynamics: Local versus nonlocal. *Phys. Rev. Lett.* **104**, 070406 (2010).

[61] Vacchini, B., Smirne, A., Laine, E.-M., Piilo, J. & Breuer, H.-P. Markovianity and non-Markovianity in quantum and classical systems. *New J. Phys.* **13**, 093004 (2011).

[62] Breuer, H.-P. Foundations and measures of quantum non-Markovianity. *J. Phys. B* **45**, 154001 (2012).

[63] Chruściński, D. & Maniscalco, S. Degree of non-Markovianity of quantum evolution. *Phys. Rev. Lett.* **112**, 120404 (2014).

[64] Boulant, N., Havel, T. F., Pravia, M. A. & Cory, D. G. Robust method for estimating the Lindblad operators of a dissipative quantum process from measurements of the density operator at multiple time points. *Phys. Rev. A* **67**, 042322 (2003).

[65] Kraus, K. *States, Effects, and Operations: Fundamental Notions of Quantum Theory* (Springer, Berlin, 1983).

[66] Nielsen, M. A. & Chuang, I. L. *Quantum computation and quantum information* (Cambridge Univ. Press, New York, 2010).

[67] Benenti, G., Casati, G. & Strini, G. *Principles of quantum computation and information*, vol. 2 (World Scientific, 2007).

[68] Tang, J.-S. *et al.* Measuring non-Markovianity of processes with controllable system-environment interaction. *EPL (Europhysics Letters)* **97**, 10002 (2012).

[69] Liu, B.-H. *et al.* Experimental control of the transition from Markovian to non-Markovian dynamics of open quantum systems. *Nat. Phys.* **7**, 931–934 (2011).

[70] Kuah, A.-M., Modi, K., Rodríguez-Rosario, C. A. & Sudarshan, E. C. G. How state preparation can affect a quantum experiment: Quantum process tomography for open systems. *Phys. Rev. A* **76**, 042113 (2007).

[71] Modi, K., Rodríguez-Rosario, C. A. & Aspuru-Guzik, A. Positivity in the presence of initial system-environment correlation. *Phys. Rev. A* **86**, 064102 (2012).

[72] Brodutch, A., Datta, A., Modi, K., Rivas, A. & Rodríguez-Rosario, C. A. Vanishing quantum discord is not necessary for completely positive maps. *Phys. Rev. A* **87**, 042301 (2013).

[73] Jordan, T. F., Shaji, A. & Sudarshan, E. C. G. Dynamics of initially entangled open quantum systems. *Phys. Rev. A* **70**, 052110 (2004).

[74] Wood, C. J. & Spekkens, R. W. The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning. *New J. Phys.* **17**, 033002 (2015).

[75] Peres, A. Separability criterion for density matrices. *Phys. Rev. Lett.* **77**, 1413–1415 (1996).

[76] Horodecki, M., Horodecki, P. & Horodecki, R. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A* **223**, 1 – 8 (1996).

[77] Horodecki, P. Separability criterion and inseparable mixed states with positive partial transposition. *Phys. Lett. A* **232**, 333 (1997).

[78] Piani, M. & Mora, C. E. Class of positive-partial-transpose bound entangled states associated with almost any set of pure entangled states. *Phys. Rev. A* **75**, 012305 (2007).

[79] Horodecki, M., Horodecki, P. & Horodecki, R. Mixed-state entanglement and distillation: Is there a "bound" entanglement in nature? *Phys. Rev. Lett.* **80**, 5239–5242 (1998).

[80] Bennett, C. H., Bernstein, H. J., Popescu, S. & Schumacher, B. Concentrating partial entanglement by local operations. *Phys. Rev. A* **53**, 2046 (1996).

[81] Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed-state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3824 (1996).

[82] Horodecki, R., Horodecki, P., Horodecki, M. & Horodecki, K. Quantum entanglement. *Rev. Mod. Phys.* **81**, 865–942 (2009).

[83] DiVincenzo, D. P., Shor, P. W., Smolin, J. A., Terhal, B. M. & Thapliyal, A. V. Evidence for bound entangled states with negative partial transpose. *Phys. Rev. A* **61**, 062312 (2000).

[84] Dür, W., Cirac, J. I., Lewenstein, M. & Bruß, D. Distillability and partial transposition in bipartite systems. *Phys. Rev. A* **61**, 062313 (2000).

[85] Clarisse, L. *Entanglement Distillation; A Discourse on Bound Entanglement in Quantum Information Theory.* Ph.D. thesis, University of York (2006). URL `arXiv:quant-ph/0612072`.

[86] Chruściński, D. & Kossakowski, A. Class of positive partial transposition states. *Phys. Rev. A* **74**, 022308 (2006).

[87] Hardy, L. Quantum gravity computers: On the theory of computation with indefinite causal structure (2007). URL `arXiv:quant-ph/0701019`.

[88] Chiribella, G. Perfect discrimination of no-signalling channels via quantum superposition of causal structures. *Phys. Rev. A* **86**, 040301 (2012).

[89] Araújo, M., Costa, F. & Brukner, u. Computational advantage from quantum-controlled ordering of gates. *Phys. Rev. Lett.* **113**, 250402 (2014).

[90] Procopio, L. M. *et al.* Experimental superposition of orders of quantum gates. *Nat. Commun.* **6**, 7913 (2015).

[91] Zych, M., Costa, F., Pikovski, I. & Časlav Brukner. Quantum interferometric visibility as a witness of general relativistic proper time. *Nat. Commun.* **2**, 505 (2011).

[92] Pikovski, I., Zych, M., Costa, F. & Časlav Brukner. Universal decoherence due to gravitational time dilation. *Nat. Phys.* **11**, 668–672 (2015).

[93] Berkson, J. Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin* **2**, 47–53 (1946).

[94] Werner, R. F. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Phys. Rev. A* **40**, 4277–4281 (1989).

[95] Černoch, A., Soubusta, J., Bartůšková, L., Dušek, M. & Fiurášek, J. Experimental realization of linear-optical partial swap gates. *Phys. Rev. Lett.* **100**, 180501 (2008).

[96] Knutson, A. & Tao, T. Honeycombs and sums of Hermitian matrices. *Notices of the AMS* **48**, 175–186 (2001). URL `arXiv:math/0009048`.

[97] Rana, S. Negative eigenvalues of partial transposition of arbitrary bipartite states. *Phys. Rev. A* **87**, 054301 (2013).

[98] Cover, T. M. & Thomas, J. A. *Elements of information theory* (Wiley, 2006).

A Appendix

A.1 Limiting cases of causal tomography

This appendix shows how causal tomography reduces to conventional tomography of quantum processes and bipartite states in the limiting cases where the causal relation between A and B is either purely cause-effect or purely common-cause. For concreteness, we continue with the example of qubits discussed in the main text, but the generalization to higher-dimensional variables is straightforward.

If the causal relation is purely cause-effect, the Jamiołkowski operator representing the causal map takes the form

$$\tau^{ce}_{CB|D} = \rho_C \otimes \tau_{B|D}, \tag{A.1}$$

hence the probability distribution relating settings and outcomes in the experiment factorizes as

$$\mathrm{Tr}_{CBD}\left[\tau^{ce}_{CB|D}(\Pi_C^{s'c} \otimes \Pi_B^{u'b} \otimes \Pi_D^{t'd})\right] = \mathrm{Tr}\left[\rho_C \Pi_C^{s'c}\right] \mathrm{Tr}\left[\tau_{B|D}\left(\Pi_B^{u'b} \otimes \Pi_D^{t'd}\right)\right] = P\left(c|s'\right) P\left(b|dt'u'\right). \tag{A.2}$$

One can isolate the correlations that characterize the channel from D to B by marginalizing over c, that is, by considering only the coefficients $\mathcal{C}_{0tu}$ from the construction above: substituting $\tau^{ce}_{CB|D}$, we find for $t', u' \in \{1, 2, 3\}$

$$\mathcal{C}_{0t'u'} = \sum_{c=\pm 1} P\left(c|s'\right) \sum_{b,d=\pm 1} bd\, P\left(b|dt'u'\right) = \mathrm{Tr}\left[\tau_{B|D}\left(\sigma_B^{u'} \otimes \sigma_D^{t'}\right)\right], \tag{A.3}$$

whereas the coefficients $\mathcal{C}_{0tu}$ with $t = 0$ and/or $u = 0$ take the form

$$\mathcal{C}_{00u'} = \sum_{c=\pm 1} P\left(c|s'\right) \sum_{b,d=\pm 1} b\, P\left(b|dt'u'\right) = \mathrm{Tr}\left[\tau_{B|D}\left(\sigma_B^{u'} \otimes \sigma_D^{0}\right)\right] \tag{A.4}$$

$$\mathcal{C}_{0t'0} = \sum_{c=\pm 1} P\left(c|s'\right) \sum_{b,d=\pm 1} d\, P\left(b|dt'u'\right) = \mathrm{Tr}\left[\tau_{B|D}\left(\sigma_B^{0} \otimes \sigma_D^{t'}\right)\right] = \sum_{d} d = 0 \tag{A.5}$$

$$\mathcal{C}_{000} = \sum_{c=\pm 1} P\left(c|s'\right) \sum_{b,d=\pm 1} P\left(b|dt'u'\right) = \mathrm{Tr}\left[\tau_{B|D}\left(\sigma_B^{0} \otimes \sigma_D^{0}\right)\right] = \sum_{d} 1 = 2. \tag{A.6}$$

Using these coefficients, one can explicitly reconstruct

$$\tau_{B|D} = \frac{1}{4} \sum_{tu=0}^{3} \mathcal{C}_{0tu} \sigma_B^u \otimes \sigma_D^t. \tag{A.7}$$

Similarly, marginalizing over b and d yields the data to reconstruct the marginal state ρ_C: we have

$$\mathcal{C}_{s'00} = \sum_{c=\pm 1} cP\left(c|s'\right) \sum_{b,d=\pm 1} P\left(b|dt'u'\right) = 2\mathrm{Tr}\left[\rho_C \sigma_C^{s'}\right] \tag{A.8}$$

and $\mathcal{C}_{000} = 2$, as established above, so that

$$\rho_C = \frac{1}{4} \sum_{s=0}^{3} \mathcal{C}_{s00} \sigma_C^s. \tag{A.9}$$

If, on the other hand, the causal relation is purely common-cause, then the Jamiołkowski operator representing the causal map takes the form

$$\tau^{cc}_{CB|D} = \rho_{CB} \otimes \mathbb{I}_D, \tag{A.10}$$

hence the probability distribution relating settings and outcomes in the experiment factorizes as

$$\mathrm{Tr}_{CBD}\left[\tau^{ce}_{CB|D}(\Pi_C^{s'c}\otimes\Pi_B^{u'b}\otimes\Pi_D^{t'd})\right]=\mathrm{Tr}\left[\rho_{CB}\left(\Pi_C^{s'c}\otimes\Pi_B^{u'b}\right)\right]\mathrm{Tr}\left[\Pi_D^{t'd}\right]=P\left(cb|s'u'\right). \quad \text{(A.11)}$$

One can see that the coefficients with $t'\in\{1,2,3\}$ must be zero, $\mathcal{C}_{st'u}=0$, because the marginal of $\tau^{cc}_{CB|D}$ on D is the identity operator. The coefficients with $t=0$ encode the state ρ_{CB}: for $s',u'\in\{1,2,3\}$

$$\mathcal{C}_{s'0u'}=\sum_{c,b,d=\pm1}P\left(cb|s'u'\right)=2\mathrm{Tr}\left[\rho_{CB}\left(\sigma_C^{s'}\otimes\sigma_B^{u'}\right)\right], \quad \text{(A.12)}$$

whereas the coefficients $\mathcal{C}_{s0u}$ with $s=0$ and/or $u=0$ take the form

$$\mathcal{C}_{00u'}=\sum_{c,b,d=\pm1}b\;P\left(cb|s'u'\right)=2\mathrm{Tr}\left[\rho_{CB}\left(\sigma_C^{0}\otimes\sigma_B^{u'}\right)\right] \quad \text{(A.13)}$$

$$\mathcal{C}_{s'00}=\sum_{c,b,d=\pm1}c\;P\left(cb|s'u'\right)=2\mathrm{Tr}\left[\rho_{CB}\left(\sigma_C^{s'}\otimes\sigma_B^{0}\right)\right] \quad \text{(A.14)}$$

$$\mathcal{C}_{000}=\sum_{c,b,d=\pm1}P\left(cb|s'u'\right)=2\mathrm{Tr}\left[\rho_{CB}\left(\sigma_C^{0}\otimes\sigma_B^{0}\right)\right]=2. \quad \text{(A.15)}$$

We can therefore reconstruct

$$\rho_{CB}=\frac{1}{8}\sum_{su=0}^{3}\mathcal{C}_{s0u}\sigma_C^s\otimes\sigma_B^u. \quad \text{(A.16)}$$

A.2 Bloch sphere representation of prediction and retrodiction

Consider a single qubit that is subjected to a projective measurement. What can one retrodict about the pre-intervention variable, C, if the prior ρ_C is not maximally mixed? For brevity, we fix the setting s and let Π^m denote the projector associated with the measurement outcome $m\in\{\pm1\}$. In this case, the retrodictive state takes the form

$$\rho_C^m=\frac{1}{P\left(m\right)}\left[\rho_C^{\frac{1}{2}}\Pi^m\rho_C^{\frac{1}{2}}\right]. \quad \text{(A.17)}$$

We seek to express this in terms of Bloch vectors, which are related to operators by the rule

$$\rho=\frac{1}{2}\left[\sigma_0+\sum_j v_j\sigma_j\right]\quad\Leftrightarrow\quad v_j=Tr\left[\rho\sigma_j\right]. \quad \text{(A.18)}$$

For simplicity, we will choose the $\hat{z}$-axis of the Bloch sphere such that the Bloch vector representing the prior ρ_C takes the form $\vec{q}=(0,0,\epsilon)$, where ϵ measures the strength of the bias. This implies that ρ_C is diagonal in the eigenbasis of σ_3, and we can write explicitly

$$\rho_C^{\frac{1}{2}}=\begin{pmatrix}\sqrt{\frac{1+\epsilon}{2}} & 0\\ 0 & \sqrt{\frac{1-\epsilon}{2}}\end{pmatrix}. \quad \text{(A.19)}$$

Let $\vec{p}=(p_1,p_2,p_3)$ denote the Bloch vector representing Π^{+1}, noting that the complementary outcome, $m=-1$, is therefore represented by the Bloch vector $-\vec{p}$. Using expression (A.18), we have

$$\Pi^m=\frac{1}{2}\begin{pmatrix}1+mp_3 & mp_1-\mathrm{imp}_2\\ mp_1+\mathrm{imp}_2 & 1-mp_3\end{pmatrix}. \quad \text{(A.20)}$$

Multiplying the matrices, one can show that

$$\begin{cases} Tr\left[\sigma_0 \rho_C^{\frac{1}{2}} \Pi^m \rho_C^{\frac{1}{2}}\right] = \frac{1}{2}\left(1 + m p_3 \epsilon\right) \\ Tr\left[\sigma_{1,2} \rho_C^{\frac{1}{2}} \Pi^m \rho_C^{\frac{1}{2}}\right] = \frac{\sqrt{1-\epsilon^2}}{2} m p_{1,2} \\ Tr\left[\sigma_3 \rho_C^{\frac{1}{2}} \Pi^m \rho_C^{\frac{1}{2}}\right] = \frac{1}{2}\left(m p_3 + \epsilon\right), \end{cases} \tag{A.21}$$

as well as

$$P\left(m\right) = Tr\left[\sigma_0 \rho_C^{\frac{1}{2}} \Pi^m \rho_C^{\frac{1}{2}}\right] = Tr\left[\Pi^m \rho_C\right] = \frac{1}{2}\left(1 + m p_3 \epsilon\right). \tag{A.22}$$

Renormalizing the inner products with $\sigma_{1,2,3}$ by $P\left(m\right)$, we find the elements of the Bloch vector $\vec{c}^m$ that represents the retrodictive state ρ_C^m:

$$\begin{cases} c_{1,2} = Tr\left[\sigma_{1,2} \frac{1}{P(m)} \rho_C^{\frac{1}{2}} \Pi^m \rho_C^{\frac{1}{2}}\right] = \frac{\sqrt{1-\epsilon^2}}{(1+m p_3 \epsilon)} m p_{1,2} \\ c_3 = Tr\left[\sigma_3 \frac{1}{P(m)} \rho_C^{\frac{1}{2}} \Pi^m \rho_C^{\frac{1}{2}}\right] = \frac{(m p_3 + \epsilon)}{(1+m p_3 \epsilon)}. \end{cases} \tag{A.23}$$

A.3 Geometric interpretation of single-qubit inference maps

Theorem. *(45 in the main text) Given a single-qubit conditional $\tau_{B|A}$, the inference ellipsoid representing $\tau_{B|A}$ is the unit sphere if and only if the only causal explanation of $\tau_{B|A}$ is either an extremal CE relation or an extremal CC relation.*

Proof. ($\Leftarrow$) If a quantum channel applies a unitary transformation, then every pure input state results in a pure output. In terms of the Bloch sphere, this implies that the image of the set of all pure states is the unit sphere. Similarly, if ρ_{BA} is a pure, entangled two-qubit state, then it can be written in the Schmidt form as $\sum_{i=0,1} \alpha_i \left|a_i\right\rangle \left|b_i\right\rangle$, with two non-zero coefficients α_i, for some pair of orthonormal bases $\left\{\left|a_i\right\rangle\right\}_i$ of $\mathcal{H}_A$ and $\left\{\left|b_i\right\rangle\right\}_i$ of $\mathcal{H}_B$. The conditional $\tau_{B|A} = \rho_A^{-\frac{1}{2}} \rho_{BA} \rho_A^{-\frac{1}{2}}$ has the Schmidt form $\sum_{i=0,1} \frac{1}{\sqrt{2}} \left|a_i\right\rangle \left|b_i\right\rangle$, and consequently finding any pure state $\cos\theta \left|a_0\right\rangle + \mathrm{e}^{\mathrm{i}\phi} \sin\theta \left|a_1\right\rangle$ on A implies an equally pure state $\cos\theta \left|b_0\right\rangle + \mathrm{e}^{-\mathrm{i}\phi} \sin\theta \left|b_1\right\rangle$ on B. It follows that the inference ellipsoid is again a unit sphere.

($\Rightarrow$) By applying suitable unitaries on A and B before and after a given conditional $\tau_{B|A}$, one can diagonalize the matrix T in the effective conditional,

$$\frac{1}{4}\mathrm{Tr}\left[\left(V_B \otimes U_A\right) \tau_{B|A} \left(V_B \otimes U_A\right)^\dagger \sigma_B^i \otimes \sigma_A^j\right] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ c_1 & t_1 & 0 & 0 \\ c_2 & 0 & t_2 & 0 \\ c_3 & 0 & 0 & t_3 \end{pmatrix}_{ij}, \tag{A.24}$$

and furthermore enforce that both $t_1 \geq 0$ (by applying π rotations about the z axis of the Bloch sphere, inverting the signs of both c_1, t_1 and c_2, t_2, if necessary) and $t_3 \geq 0$ (with a π rotation about the x axis, which inverts both c_2, t_2 and c_3, t_3). Since both the set of unitary channels and the set of pure maximally entangled states are closed under local unitaries, the analysis can be cast in terms of the diagonalized and suitably rotated $\left(V_B \otimes U_A\right) \tau_{B|A} \left(V_B \otimes U_A\right)^\dagger$. If the inference ellipsoid is a unit sphere, its centre lies at the origin, $\vec{c} = \vec{0}$, and the scale factors of the axes are $t_1 = t_3 = +1$ and $t_2 = \pm 1$. This leaves two alternatives: $\vec{t} = (1, 1, 1)$, which corresponds to the identity channel, and $\vec{t} = (1, -1, 1)$, which gives (by explicit reconstruction) $\tau_{B|A} = \left|\Phi^+\right\rangle \left\langle\Phi^+\right|$. The causal explanation of the original $\tau_{B|A}$ can be recovered by applying the inverse unitaries on A and B, which yields generic unitary transformations (if $\tau_{B|A}$ was the identity channel) and generic pure, maximally entangled states (if $\tau_{B|A}$ was $\left|\Phi^+\right\rangle$). If a common-cause conditional $\tau_{B|A}$ is pure and maximally entangled, then the joint state $\rho_{BA} = \rho_A^{\frac{1}{2}} \tau_{B|A} \rho_A^{\frac{1}{2}}$, where the marginal ρ_A can be any positive-semidefinite, trace-one operator, is pure and (not necessarily maximally) entangled.

We prove lemma 51 from the main text before lemma 48, since the proof of the latter contains steps that closely resemble the former. □

Lemma. *(51 in the main text) Given a generic single-qubit conditional $\tau_{B|A}$, whose Pauli basis representation is*

$$\frac{1}{4}Tr\left[\tau_{B|A}\sigma_B^i\otimes\sigma_A^j\right]=\begin{pmatrix}1 & \vec{0}^T\\ \vec{c} & T\end{pmatrix}_{ij}, \tag{A.25}$$

there exist rotations of the Bloch spheres of A and B, viz. a product of local unitary operations on $\mathcal{H}_A$ and $\mathcal{H}_B$, that put the Pauli basis decomposition in the form

$$\frac{1}{4}Tr\left[(V_B\otimes U_A)\,\tau_{B|A}\,(V_B\otimes U_A)^\dagger\,\sigma_B^i\otimes\sigma_A^j\right]=\begin{pmatrix}1 & 0 & 0 & 0\\ c_1 & t_1 & 0 & 0\\ c_2 & 0 & t_2 & 0\\ c_3 & 0 & 0 & t_3\end{pmatrix}_{ij}. \tag{A.26}$$

The diagonal elements $\vec{t}=(t_1,t_2,t_3)$ are the eigenvalues of the correlation matrix T in the Pauli representation of $\tau_{B|A}$.

Proof. Let us first establish that every unitary V on a single-qubit Hilbert space $\mathcal{H}_B$ realizes a rotation in the Pauli-basis representation of $\mathcal{L}(\mathcal{H}_B)$ that involves only the non-trivial Pauli operators. To see this, consider the effect of a unitary V on a generic Hilbert space vector: denoting $v_i\equiv Tr(\sigma_i\rho)$ for a generic initial state ρ and $v'_k=Tr\left[\sigma_k\left(V\rho V^\dagger\right)\right]$, the two are related by

$$v'_k=Tr\left[\sigma_k V\left(\frac{1}{2}\sum_{i=0}^{3}v_i\sigma_i\right)V^\dagger\right]=\sum_{i=0}^{3}\frac{1}{2}Tr\left[\sigma_k V\sigma_i V^\dagger\right]v_i\equiv\sum_{i=0}^{3}V^P_{ki}v_i. \tag{A.27}$$

Note that the elements of V^P are real and that they satisfy

$$\sum_{i=1}^{3}V^P_{ki}\left(V^P\right)^T_{il}=\frac{1}{2}\sum_{i=0}^{3}Tr\left[\sigma_k V\sigma_i V^\dagger\right]\frac{1}{2}Tr\left[V^\dagger\sigma_l V\sigma_i\right]=\frac{1}{2}Tr\left[\sigma_k V\left(V^\dagger\sigma_l V\right)V^\dagger\right]=\frac{1}{2}Tr\left[\sigma_k\mathbb{I}\sigma_l\mathbb{I}\right]=\delta_{kl}, \tag{A.28}$$

characterizing a rotation in Pauli space, and furthermore

$$V^P_{k0}=\frac{1}{2}Tr\left[\sigma_k VV^\dagger\right]=\delta_{k0}=V^P_{0k}, \tag{A.29}$$

that is, the rotation only involves the non-trivial Pauli operators. Conversely, one can see that every rotation in the subspace of $\mathcal{L}(\mathcal{H}_B)$ spanned by $\{\sigma_k\}$ with $k=1,2,3$ is a linear map that takes pure states (positive-semidefinite, rank-one, trace-one operators) to other pure states, and can be realized by a unitary.

Now, in the Pauli basis representation of $\tau_{B|A}$,

$$\frac{1}{4}\mathrm{Tr}\left[\tau_{B|A}\sigma_B^i\otimes\sigma_A^j\right]=\begin{pmatrix}1 & \vec{0}^T\\ \vec{c} & T_{3\times3}\end{pmatrix}_{ij}, \tag{A.30}$$

the real matrix T can be made diagonal by multiplying on the left and right with suitable real rotations V^P and U^P, which only have a non-trivial effect on the basis elements indexed by $i,j\neq0$,

$$\begin{pmatrix}1 & \vec{0}^T\\ \vec{0} & \bar{V}^P_{3\times3}\end{pmatrix}\begin{pmatrix}1 & \vec{0}^T\\ \vec{c} & T_{3\times3}\end{pmatrix}\begin{pmatrix}1 & \vec{0}^T\\ \vec{0} & \bar{U}^P_{3\times3}\end{pmatrix}=\begin{pmatrix}1 & \vec{0}^T\\ \vec{c} & \mathrm{diag}\vec{t}\end{pmatrix}. \tag{A.31}$$

That is, there exist matrices with the properties $V^P_{k0} = V^P_{0k} = \delta_{k0} = U^P_{k0} = U^P_{0k}$ and

$$\sum_{i=1}^{3} V^P_{ki} \left(V^P\right)^T_{il} = \delta_{kl} = \sum_{j=1}^{3} \left(U^P\right)^T_{kj} U^P_{jl}, \tag{A.32}$$

such that

$$\sum_{i,j=1}^{3} V^P_{ki} \left(\mathrm{Tr}\tau_{B|A}\sigma^i_B \otimes \sigma^j_A\right) U^P_{jl} = t_k \delta_{kl} \quad (k,l = 1,2,3)\,. \tag{A.33}$$

Note that V^P acts only on the Pauli representation of the factor space $\mathcal{L}\left(\mathcal{H}_B\right)$, whereas U^P acts only on the Pauli representation of the factor space $\mathcal{L}\left(\mathcal{H}_A\right)$. They are therefore equivalent to particular unitary rotations on the underlying Hilbert spaces, according to the identification

$$\rho'_{BA} \equiv \sum_{kl} \sigma^k_B \left[\sum_{i,j=0}^{3} V^P_{ki} \left(\mathrm{Tr}\rho_{BA}\sigma^i_B \otimes \sigma^j_A\right) U^P_{jl}\right] \sigma^l_A = \left(V_B \otimes U_A\right) \rho_{BA} \left(V_B \otimes U_A\right)^\dagger \quad \forall \rho_{BA} \in \mathcal{L}\left(\mathcal{H}_B \otimes \mathcal{H}_A\right). \tag{A.34}$$

□

Lemma. *(48 in the main text) A unital single-qubit conditional $\tau_{B|A}$ can be represented on the Pauli basis as*

$$\frac{1}{4} Tr\left(\tau_{B|A}\sigma^i_B \otimes \sigma^j_A\right) = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & T_{3\times 3} \end{pmatrix}_{ij}, \tag{A.35}$$

with the eigenvalues of the matrix of correlations T forming a three-vector $\vec{t}$. Then $\tau_{B|A}$ admits a CC explanation if and only if $\vec{t}$ lies within the tetrahedron with vertices

$$\left\{(1,1,-1), (1,-1,1), (-1,1,1), (-1,-1,-1)\right\}, \tag{A.36}$$

and a CE explanation if and only if it lies within the tetrahedron

$$\left\{(1,1,1), (1,-1,-1), (-1,1,-1), (-1,-1,1)\right\}. \tag{A.37}$$

Proof. The fact that $\tau_{B|A}$ is unital implies that, in terms of the Pauli form,

$$\begin{pmatrix} 1 & \vec{0}^T \\ \vec{c} & T_{3\times 3} \end{pmatrix} \begin{pmatrix} 1 \\ \vec{0} \end{pmatrix} = \begin{pmatrix} 1 \\ \vec{0} \end{pmatrix}, \tag{A.38}$$

i.e. the marginal on B is also maximally mixed and the ellipsoid is centred at the origin, $\vec{c} = \vec{0}$. Using the same approach as in lemma 51 proved above, the Pauli form in this case can be diagonalized completely,

$$\begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & V^P_{3\times 3} \end{pmatrix} \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & T_{3\times 3} \end{pmatrix} \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & U^P_{3\times 3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & t_1 & 0 & 0 \\ 0 & 0 & t_2 & 0 \\ 0 & 0 & 0 & t_3 \end{pmatrix}. \tag{A.39}$$

Furthermore, following lemma 52, the product of local local unitaries $V_B \otimes U_A$ does not change whether a given bipartite state is positive-semidefinite or PPT. That is, $\tau_{B|A}$ is positive-semidefinite (PPT) if and only if the Pauli-diagonal form

$$\tilde{\tau}_{B|A} = \left(U_A \otimes V_B\right) \tau_{B|A} \left(U_A \otimes V_B\right)^\dagger = \sigma^0_B \otimes \sigma^0_A + \sum_{k=1}^{3} t_k \sigma^k_B \otimes \sigma^k_A \tag{A.40}$$

is positive-semidefinite (PPT).

The eigenvalues of $\tilde{\tau}_{B|A}$ can be found explicitly, and the condition of positivity becomes

$$\begin{cases} +t_1 + t_2 - t_3 \geq -1 \\ +t_1 - t_2 + t_3 \geq -1 \\ -t_1 + t_2 + t_3 \geq -1 \\ -t_1 - t_2 - t_3 \geq -1. \end{cases} \tag{A.41}$$

The four inequalities are simultaneously satisfied if and only if $\vec{t}$ lies within the tetrahedron with vertices$\{(1,1,-1),(1,-1,1),(-1,1,1),(-1,-1,-1)\}$, which is therefore a necessary and sufficient condition for $\tilde{\tau}_{B|A} \geq 0$, and consequently $\tau_{B|A} \geq 0$, which admits a common-cause explanation. Analogously, a cause-effect explanation is possible if and only if $T_A\tilde{\tau}_{B|A} \geq 0$, which holds if and only if

$$\begin{cases} +t_1 + t_2 - t_3 \geq -1 \\ +t_1 - t_2 + t_3 \geq -1 \\ -t_1 + t_2 + t_3 \geq -1 \\ -t_1 - t_2 - t_3 \geq -1, \end{cases} \tag{A.42}$$

i.e., if and only if $\vec{t}$ lies within the tetrahedron with vertices $\{(1,1,1),(1,-1,-1),(-1,1,-1),(-1,-1,1)\}$. □

Theorem. *(54 in the main text) Let $\tau_{B|A}$ be any single-qubit conditional and assume that $\tau_{B|A}$ is decidable, i.e. the inference ellipsoid does not fit inside a nested tetrahedron. Then $\tau_{B|A}$ admits a CE explanation if and only if the colouring of the inference ellipsoid is right-handed, and a CC explanation if and only if the ellipsoid is left-handed.*

Proof. Using lemma 51, the generic single-qubit quantum conditional $\tau_{B|A}$ can be represented (up to local unitaries) by the parameters $\vec{c}, \vec{t}$; and by lemma 52, local unitary transformation leave the properties of being positive-semidefinite or positive under partial transposition unchanged. We will show that, if $\tau_{B|A}$ is positive-semidefinite and lies in an octant with $t_1t_2t_3 > 0$, then it is also PPT, and consequently undecidable. Similarly, any PPT conditional with $t_1t_2t_3 < 0$ is also positive-semidefinite, and therefore undecidable. By the contrapositive, if a single-qubit conditional is known to be decidable, then it is PPT if and only if $t_1t_2t_3 > 0$ and positive-semidefinite if and only if $t_1t_2t_3 < 0$. This, by definition 43, corresponds to left- and right-handedness, respectively.

Let us now show that any conditional $\tau_{B|A}$ that is positive-semidefinite and has $t_1t_2t_3 > 0$ is also PPT. (The case of PPT operators with $t_1t_2t_3 < 0$ follows by symmetry.) The conditional $\tau_{B|A}$ is represented (up to local unitaries) by the parameters $\vec{c}$ and $\vec{t}$, and we can use suitable unitaries to enforce that $t_1, t_2, t_3 > 0$. (For example, a rotation about the x-axis of the Bloch sphere of A by an angle π inverts the signs of t_2 and t_3.) Since $\tau_{B|A}$ is positive-semidefinite, it follows that its partial transpose $T_A\tau_{B|A}$ is PPT. If the partial transposition is realized on the eigenbasis of σ_A^3, inverting the sign of σ_A^2, then $T_A\tau_{B|A}$ is represented by the parameters $\vec{c}' = \vec{c}$ and $\vec{t}' = (t_1, -t_2, t_3)$. These parameters $\vec{c}'$ and $\vec{t}'$ satisfy the conditions of theorem 53, which we rewrite in terms of the c_i and t_i that represent the original $\tau_{B|A}$ (with slight simplifications due to the fact that $t_3 > 0$) as

H1. $|c_3| + |t_3| = 1$, $c_1 = c_2 = 0$ and $t_1 = \pm t_2$; or

H2. $|c_3| + |t_3| < 1$, and all of the following hold:

1.
$$(t_1 - t_2)^2 \leq (1 + t_3)^2 - c_3^2 - \left(c_1^2 + c_2^2\right)\left(\frac{1 + t_3 - |c_3|}{1 - t_3 - |c_3|}\right) \equiv f_+(t_3, \vec{c}), \tag{A.43}$$

2.
$$(t_1 + t_2)^2 \leq (1 - t_3)^2 - c_3^2 - \left(c_1^2 + c_2^2\right)\left(\frac{1 - t_3 + |c_3|}{1 + t_3 + |c_3|}\right) \equiv f_-(t_3, \vec{c}), \tag{A.44}$$

3.

$$\left[1-\left(t_1^2+t_2^2+t_3^2\right)-\left(c_1^2+c_2^2+c_3^2\right)\right]^2 \geq 4\left[t_1^2\left(c_1^2+t_2^2\right)+t_2^2\left(c_2^2+t_3^2\right)+t_3^2\left(c_3^2+t_1^2\right)+2t_1t_2t_3\right]. \tag{A.45}$$

The thesis we aim to prove is that $\tau_{B|A}$ is also PPT, that is, the parameters $\vec{c}$ and $\vec{t}$ themselves satisfy the conditions of theorem 53 (again simplified using the fact that $t_3 > 0$),

T1. $|c_3| + |t_3| = 1$, $c_1 = c_2 = 0$ and $t_1 = \pm t_2$; or

T2. $|c_3| + |t_3| < 1$, and all of the following hold:

1.
$$(t_1+t_2)^2 \leq (1+t_3)^2 - c_3^2 - \left(c_1^2+c_2^2\right)\left(\frac{1+t_3-|c_3|}{1-t_3-|c_3|}\right) = f_+\left(t_3,\vec{c}\right), \tag{A.46}$$

2.
$$(t_1-t_2)^2 \leq (1-t_3)^2 - c_3^2 - \left(c_1^2+c_2^2\right)\left(\frac{1-t_3+|c_3|}{1+t_3+|c_3|}\right) = f_-\left(t_3,\vec{c}\right), \tag{A.47}$$

3.

$$\left[1-\left(t_1^2+t_2^2+t_3^2\right)-\left(c_1^2+c_2^2+c_3^2\right)\right]^2 \geq 4\left[t_1^2\left(c_1^2+t_2^2\right)+t_2^2\left(c_2^2+t_3^2\right)+t_3^2\left(c_3^2+t_1^2\right)-2t_1t_2t_3\right]. \tag{A.48}$$

Note that the only difference between the hypotheses that follow from $\tau_{B|A}$ being positive-semidefinite and the theses that ensure that $\tau_{B|A}$ is PPT is a sign change in t_2, which reflects the partial transposition. We will show how the hypothesis H1-H3 imply the theses T1-T3.

If $\vec{c}$ and $\vec{t}$ satisfy the hypothesis H1, then the thesis T1 also holds, and $\tau_{B|A}$ is PPT. Now suppose that $\vec{c}$ and $\vec{t}$ satisfy H2 instead. One can see that hypothesis H2.3 implies thesis T2.3, by noting that the left-hand sides are equal and using the fact that $t_1t_2t_3 > 0$ to show that the right-hand sides satisfy

$$t_1^2\left(c_1^2+t_2^2\right)+t_2^2\left(c_2^2+t_3^2\right)+t_3^2\left(c_3^2+t_1^2\right)+2t_1t_2t_3 \geq t_1^2\left(c_1^2+t_2^2\right)+t_2^2\left(c_2^2+t_3^2\right)+t_3^2\left(c_3^2+t_1^2\right)-2t_1t_2t_3. \tag{A.49}$$

Along similar lines, the fact that $t_1, t_2 > 0$ implies that

$$(t_1-t_2)^2 < (t_1+t_2)^2, \tag{A.50}$$

and therefore hypothesis H2.2 ensures that thesis T2.2 holds.

Finally, one can show hypothesis H2.2 also implies thesis T2.1, using the fact that $f_-\left(t_3,\vec{c}\right) < f_+\left(t_3,\vec{c}\right)$. Indeed, consider

$$\begin{aligned} f_+\left(t_3,\vec{c}\right) - f_-\left(t_3,\vec{c}\right) &= 4t_3 - \left(c_1^2+c_2^2\right)\left(\frac{1+t_3-|c_3|}{1-t_3-|c_3|} - \frac{1-t_3+|c_3|}{1+t_3+|c_3|}\right) \\ &= 4t_3 - \left(c_1^2+c_2^2\right)\left[(1+t_3)^2-(1-t_3)^2\right] / \left(1+t_3+|c_3|\right)\left(1-t_3-|c_3|\right) \\ &= \frac{4t_3}{1-\left(t_3+|c_3|\right)^2}\left[1-\left(t_3+|c_3|\right)^2-\left(c_1^2+c_2^2\right)\right]. \end{aligned} \tag{A.51}$$

Recall that we chose unitaries that ensure $t_3 > 0$ and note that $|c_3|+|t_3| < 1$ follows from hypothesis H2, so the first factor is positive. In order to evaluate the sign of the second factor, we use the fact that the map $\mathcal{E}_{B|A}$ that is isomorphic to $\tau_{B|A}$ is positivity-preserving, hence the inference ellipsoid that represents it does not exceed the Bloch sphere. In particular its north and south poles (the images of the eigenstates of σ_z, viz. the Bloch vectors $(0,0,\pm 1)$) have Bloch vectors whose norm does not exceed one,

$$(c_1+0t_1)^2 + (c_2+0t_2)^2 + (c_3 \pm 1t_3)^2 \leq 1, \tag{A.52}$$

hence

$$1 - (t_3 \pm |c_3|)^2 - \left(c_1^2 + c_2^2\right) \geq 0. \tag{A.53}$$

It follows that $f_-(t_3, \vec{c}) < f_+(t_3, \vec{c})$, so that hypothesis H2.2 implies thesis T2.1.

In summary, if $\tau_{B|A}$ has $t_1 t_2 t_3 > 0$ and is positive-semidefinite, then the parameter $\vec{c}$ and $\vec{t}$ satisfy either hypothesis H1 or the hypotheses H2.1, H2.2 and H2.3. In either case, we have shown that $\tau_{B|A}$ must then also satisfy either thesis T1 or the theses T2.1, T2.2 and T2.3, which together imply that $\tau_{B|A}$ is also PPT, and therefore undecidable. Together with the reasoning outlined in the first paragraph, this completes the proof[30]. □

A.4 Complete characterization using observational data given a promise

This section demonstrates that an observational scheme sometimes allows a complete solution of the task of causal discovery. More specifically, we show the following:

Theorem 78. *If one is promised that two qubits are related by a probabilistic mixture of a CE mechanism that realizes a unitary and a CC mechanism that realizes a pure entangled state, then observational probing is sufficient to completely characterize the unitary, the entangled state and their relative weights, up to a sign ambiguity in the general case.*

Proof. We showed in section 3.3.1 that, if the causal relation between two qubits, A and B, is either purely CC or purely CE, so that the inferences one can make from A about B can be represented by a linear one-to-one qubit map $\mathcal{E}_{B|A} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B)$, then observational data is sufficient to completely characterize the inference map $\mathcal{E}_{B|A}$. Notably, the prescription for extracting the Pauli basis coefficients of the conditional $\tau_{B|A}$ from experimental data is the same in both cases. It follows that, if we are promised a probabilistic mixture of purely CC and purely CE relations, the inferences from A to B can still be represented by a single-qubit conditional, which we denote $\tau_{B|A}^{mix}$, and this conditional can be reconstructed from partial tomography data in the same way as described in section 3.3.1. In the case of a probabilistic mixture, the conditional thus obtained can be further decomposed as

$$\tau_{B|A}^{mix} = q\tau_{B|A}^{CC} + (1-q)\,\tau_{B|A}^{CE}, \tag{A.54}$$

where $\tau_{B|A}^{CC}$ and $\tau_{B|A}^{CE}$ are single-qubit quantum conditionals that arise from purely CC and purely CE relations, respectively. In our scenario, we are furthermore promised that $\tau_{B|A}^{CE}$ represents a unitary transformation, and that $\tau_{B|A}^{CC}$ encodes inferences between two qubits that were prepared in a pure, entangled state ρ_{BA} – in short, that both conditionals are extremal. We will show that, one $\tau_{B|A}^{mix}$ has been characterized by partial tomography, its decomposition into a convex combination of an extremal CC conditional and an extremal CE conditional is unique, up to one sign.

We will consider this problem in terms of the geometric representation of conditionals, introduced in lemma 41. As we established in theorem 45, the inference ellipsoid that represents an extremal conditional is the full Bloch sphere, with centre $\vec{c} = \vec{0}$ at the origin and unit radius. The correlation matrix T^{CE} that encodes the effect of $\tau_{B|A}^{CE}$ on the Bloch sphere,

$$\frac{1}{4}\mathrm{Tr}\left(\tau_{B|A}^{CE}\sigma_B^i \otimes \sigma_A^j\right) = T_{ij}^{CE} \quad i,j = 1,2,3, \tag{A.55}$$

is a proper rotation, with singular values ± 1 and $\det T > 0$ (right-handed), whereas the correlation matrix T^{CC} representing the common-cause component is an improper rotation, with singular values ± 1 but $\det T < 0$ (left-handed).

Before turning to a probabilistic mixture of generic extremal CE and CC conditionals, consider how the two are related. Specifically, note that a generic proper rotation of $\mathbb{R}^3$, namely T^{CE}, can

[30] Morgh.

be transformed into a generic improper rotation, T^{CC}, by combining it with two other steps: first reflection through the origin, which we denote by F, followed by rotation about some particular axis $\hat{n}$ by some particular angle. We write the angle as $\pi + \gamma$ (with no expectation that γ be small), so that the rotation can be decomposed into $R_{\hat{n},\pi}$ followed by $R_{\hat{n},\gamma}$. Thus

$$T^{CC} = R_{\hat{n},\gamma} R_{\hat{n},\pi} F T^{CE}. \tag{A.56}$$

The effects of each step of this transformation on the surface of the unit sphere are illustrated in Fig. A.1, where we compare the image of each point (represented by a colour) under T^{CE} alone to the images of the same point under FT^{CE}, $R_{\hat{n},\pi}FT^{CE}$ and $R_{\hat{n},\gamma}R_{\hat{n},\pi}FT^{CE} = T^{CC}$. Once we add the reflection, the image of each point under FT^{CE} is diametrically opposed to its image under T^{CE} alone. Now for the rotations about $\hat{n}$: if the image of a point under T^{CE} lies at $\pm\hat{n}$, then its image under FT^{CE} will lie at $\mp\hat{n}$ and consequently remain there under the rotation; diametrically opposite to the image under T^{CE} alone. Meanwhile, if the image of a point under T^{CE} lies in the plane orthogonal to $\hat{n}$, then if we add reflection followed by a π rotation, we recover the same image – the two coincide. Once we include the final rotation by γ, the images of the unit sphere under T^{CE} and T^{CC} will be separated by an angle γ in the plane orthogonal to $\hat{n}$, while the images under T^{CE} that lie at $\pm\hat{n}$ are diametrically opposed to the images of the same points under T^{CC}.

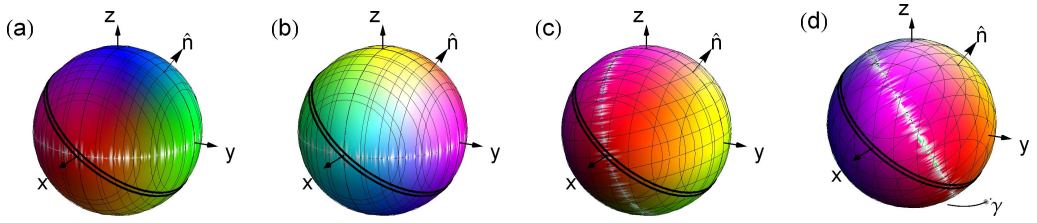

Figure A.1: Relation between a generic proper rotation T^{CE} and a generic improper rotation T^{CC}: one can obtain one from the other by combining with a reflection through the origin (F) followed by a rotation about a specific axis $\hat{n}$ by π, then further by an angle γ. The coloured spheres represent the image of the unit sphere under (a) T^{CE} alone, (b) the reflected FT^{CE}, (c) reflection and π rotation, $R_{\hat{n},\pi}FT^{CE}$, and (d) $R_{\hat{n},\gamma}R_{\hat{n},\pi}FT^{CE} = T^{CC}$. Special relations hold between the images along $\pm\hat{n}$ and in the plane orthogonal to it (highlighted in the figure). Comparing panels (a) and (d), note how the images of a given point (denoted by the same colour) under T^{CE} and T^{CC} are diametrically opposite if they lie along $\hat{n}$, and separated by an angle γ if they lie in the plane orthogonal to $\hat{n}$.

Now consider a probabilistic mixture of the two extremal cases: we will show how the geometric properties of its inference ellipsoid arise from, and consequently reflect, the components T^{CE}, T^{CC} and the weight q. By linearity, it is associated to a correlation matrix

$$T^{mix} \equiv (1-q)\, T^{CE} + q T^{CC}. \tag{A.57}$$

The image of the Bloch sphere under such a combination is shown in Fig. A.2. It must still be an ellipsoid, since T^{mix} is an affine transformation. Furthermore, it inherits the symmetry under rotation about $\hat{n}$. Therefore it has one semi-axis (eigenvector of $T^{mix}\left(T^{mix}\right)^T$) along $\hat{n}$, and a degenerate pair orthogonal to it. The length of the semi-axis (square root of eigenvalue) along $\hat{n}$ is $|1-2q|$, because the images under T^{CE} and T^{CC} along this direction are diametrically opposed. When $q = \frac{1}{2}$, this implies that the ellipsoid reduces to a disk. For $q < \frac{1}{2}$, the contribution from the process dominates, so $\det T^{mix} > 0$, while $\det T^{mix} < 0$ heralds $q > \frac{1}{2}$. The length r of the other two semi-axes, in the plane orthogonal to $\hat{n}$, can be obtained using the geometrical construction in Fig. A.3, which yields the implicit expression

$$\sin^2 \frac{\gamma}{2} = \frac{1-r^2}{4\,(q-q^2)}. \tag{A.58}$$

The images of points under T^{mix} that lie in this plane are rotated from the corresponding images under T^{CE} by an angle γ', in the same direction (same sign) as γ above, and with magnitude given by

$$2r\cos\gamma' = 1^2 + r^2 - \left[2q\sin\frac{\gamma}{2}\right]^2 . \tag{A.59}$$

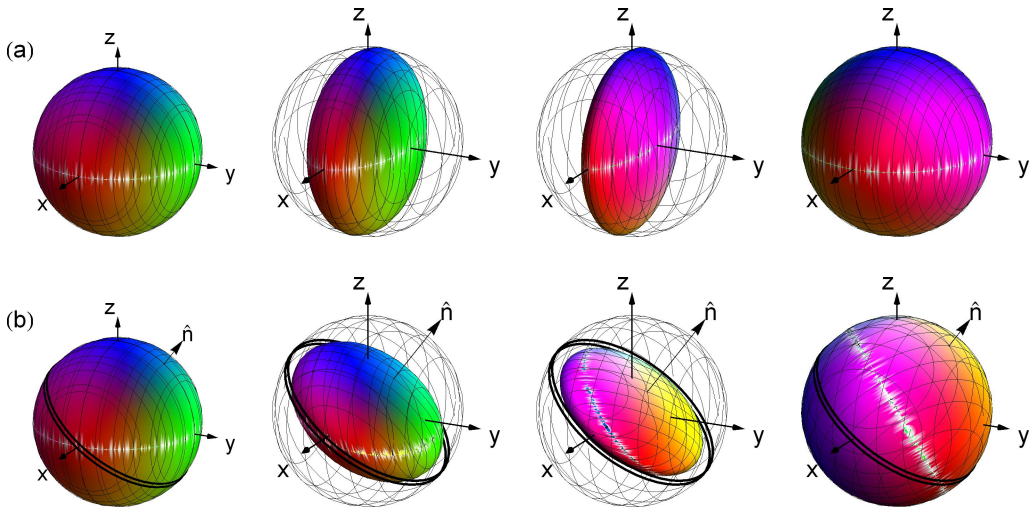

Figure A.2: Bloch sphere representation of a probabilistic mixture of a unitary process and a pure, entangled state. Mixtures of the two extremes (shown for probability of common-cause $q = 0.00,\ 0.25,\ 0.65,\ 1.00$, from left to right) produce ellipsoids that are flattened in the direction $\hat{n}$ (thick arrow) to a height $2\,|1-2q|$, and rotationally symmetric in the plane orthogonal to $\hat{n}$, with radius r. (a) A mixture of the identity channel and the state $|\Phi^+\rangle$, as realized in the experiment described in section 4.2, corresponds to the axis $\hat{n}$ pointing along $\hat{y}$, and a radius in the plane orthogonal to $\hat{n}$ of $r = 1$ throughout the transition. (b) Mixing the identity channel with a generic pure, entangled state produces intermediate ellipsoids with $\hat{n}$ pointing in a generic direction and radius $r \leq 1$.

Given a characterization of an ellipsoid that arose from such a convex combination (specifically the directions and lengths of its axes), it is straightforward to extract the direction $\pm\hat{n}$, the probability q and the angle γ. Note that there remains some ambiguity, because the available data only specify the magnitude of the angle γ and the direction of $\hat{n}$, but not its orientation. The two are related in that a simultaneous sign change of γ and $\hat{n}$ has no net effect. This leaves us free to set $\gamma > 0$ by convention, with the only remaining ambiguity being the orientation of $\hat{n}$. (In the pathological case that all three semi-axes have the same length, one finds that $\gamma = 0$, which implies that the image of T^{CC} is diametrically opposed to that of T^{CE} for all inputs, and there is no need to single out a direction $\hat{n}$. The probability q can still be read off normally.)

Given those parameters, the following steps then allow one to explicitly reconstruct T^{CE} from T^{mix}: (1) scaling by $1/(1-2q)$ in the direction of $\hat{n}$ and $1/r$ in the perpendicular plane, which, as a matrix operation, we denote by $S_{\perp\hat{n},1/r}S_{\hat{n},1/(1-2q)}$; and (2) rotation about $\hat{n}$ by $-\gamma'$, $R_{\hat{n},-\gamma'}$. In all, we have

$$T^{CE} = R_{\hat{n},-\gamma'}S_{\perp\hat{n},1/r}S_{\hat{n},1/(1-2q)}T^{mix}. \tag{A.60}$$

Similarly, the common-cause contribution can be found via

$$T^{CC} = R_{\hat{n},-\gamma+\gamma'}S_{\perp\hat{n},1/r}S_{\hat{n},1/(2q-1)}T^{mix}. \tag{A.61}$$

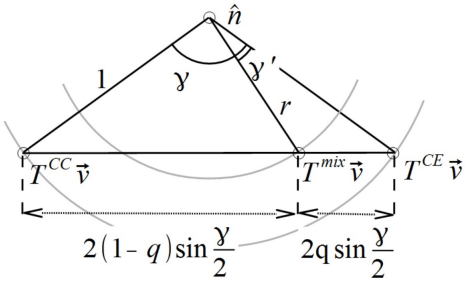

Figure A.3: Geometric construction for characterizing a probabilistic mixture of a unitary process and a pure, maximally entangled state. In the plane orthogonal to $\hat{n}$, the image of a given input Bloch vector $\vec{v}$ under T^{mix} lies on the chord connecting the images under T^{CE} and T^{CC} . Its distance from the centre, which gives the radius r , is related to the angle γ spanned by the chord and the probability p of common cause in the mixture.

Note that the two choices of $\pm\hat{n}$ give rise to different reconstructed T^{CC} and T^{CE}, but they are related by simple rotations about the $\hat{n}$ axis by fixed angles.

The ambiguity is removed when $\gamma = 0$, which is the case in the experiment described in section 4.2, mixing the identity channel and the state $|\Phi^+\rangle$. But even if this ambiguity persists, given a probabilistic mixture of any unitary process and any entangled pure bipartite state, we can uniquely determine the mixing probability q as well as the angle γ and the direction of $\hat{n}$ up to an inversion about the origin. □

A.5 Causal interpretation of separable operators

Theorem. *(59 in the main text) A quantum conditional $\tau_{B|A} \in \mathcal{L}(\mathcal{H}_B \otimes \mathcal{H}_A)$, which can be interpreted as a Jamiołkowski operator representing a trace-preserving inference map from A to B, is separable if it can be explained either by a classical common-cause relation or by a classical cause-effect relation. Moreover, if $\tau_{B|A}$ is separable, then it can be explained both by a classical common-cause relation and by a classical cause-effect relation.*

Proof. ($\Rightarrow$) Assume that there is a classical variable X that screens off A from B; the causal structure being either $A \to X \to B$ or $A \leftarrow X \to B$. Let us show that the resulting conditionals are separable.

In the CE case, both $\tau_{B|X}$ and $\tau_{X|A}$ are CE conditionals between a classical and a quantum variable, so they can be written

$$\tau_{B|X} = \sum_x \rho_B^x \otimes |x\rangle\langle x| \quad \tau_{X|A} = \sum_x |x\rangle\langle x| \otimes \tau_A^x; \tag{A.62}$$

satisfying

$$\begin{cases} T_X \tau_{B|X} \geq 0 \;\Rightarrow\; \rho_B^x \geq 0 \;\forall x \\ T_A \tau_{X|A} \geq 0 \;\Rightarrow\; \tau_A^x \geq 0 \;\forall x \end{cases} \tag{A.63}$$

and

$$\begin{cases} \mathrm{Tr}_B \tau_{B|X} = \mathbb{I}_X \;\Rightarrow\; \mathrm{Tr}_B \rho_B^x = 1 \;\forall x \\ \mathrm{Tr}_X \tau_{X|A} = \mathbb{I}_A \;\Rightarrow\; \sum_x \tau_A^x = \mathbb{I}_A. \end{cases} \tag{A.64}$$

Consequently,

$$\tau_{B|A} = \mathrm{Tr}_X \tau_{B|X} \tau_{X|A} = \sum_x \rho_B^x \otimes \tau_A^x, \tag{A.65}$$

the conditional can be written as a sum of products of terms, each of which is positive-semidefinite, and satisfying the normalization requirement for conditionals, $\mathrm{Tr}_B \tau_{B|A} = \sum_x \tau_A^x = \mathbb{I}_A$. It is therefore a separable quantum conditional.

If A and B are related by a classical common cause, their joint state can be written

$$\rho_{AB} = \mathrm{Tr}_X \tau_{A|X} \tau_{B|X} \rho_X, \tag{A.66}$$

with $\tau_{A|X} = \sum_x \rho_A^x \otimes |x\rangle\langle x|$ satisfying $\rho_A^x \geq 0$ and $\mathrm{Tr}_A \rho_A^x = 1$, and similarly for $\tau_{B|X}$. (Note how the conditional $\tau_{AB|X}$ factorizes, since they are independent variables which merely share a causal parent.) It follows that

$$\rho_{AB} = \sum_x \rho_A^x \otimes \rho_B^x, \tag{A.67}$$

with each factor positive-semidefinite and normalized, forming again a separable state. The separability of the conditional

$$\tau_{B|A} = \rho_A^{-1/2} \rho_{AB} \rho_A^{-1/2} \tag{A.68}$$

follows.

($\Leftarrow$) We will show that any separable conditional $\tau_{B|A}$ admits causal explanations as specified above by explicitly constructing suitable causal models. The hypotheses guarantee that the conditional takes the form

$$\tau_{B|A} = \sum_x \rho_B^x \otimes \tau_A^x, \tag{A.69}$$

with terms that are positive-semidefinite, $\rho_B^x, \tau_A^x \geq 0$, and satisfy the normalization requirement for conditionals,

$$Tr_B \tau_{B|A} = \sum_x \left(Tr_B \rho_B^x\right) \tau_A^x = \mathbb{I}_A. \tag{A.70}$$

We can choose the normalization constants of the ρ_B^x and τ_A^x such that

$$\mathrm{Tr}_B \rho_B^x = 1 \;\Rightarrow\; \sum_x \tau_A^x = \mathbb{I}_A. \tag{A.71}$$

Then one can define

$$\tau_{B|X} = \sum_x \rho_B^x \otimes |x\rangle\langle x| \quad \tau_{X|A} = \sum_x |x\rangle\langle x| \otimes \tau_A^x, \tag{A.72}$$

both of which are PPT and normalized to $\mathrm{Tr}_B \tau_{B|X} = \mathbb{I}_X$, $\mathrm{Tr}_X \tau_{X|A} = \mathbb{I}_A$, and therefore CE conditionals. They characterize a causal model with a CE relation between A and B that is screened off by the classical X, $A \to X \to B$, which reproduces the original conditional $\tau_{B|A}$.

In order to construct a causal model that gives rise to the given $\tau_{B|A}$ and in which A and B are related by a classical common cause X, $A \leftarrow X \to B$, we take $\tau_{B|X}$ to be the same constructed above, but modify the half that involves A. Since a CC model is naturally specified by a conditional of the form $\tau_{A|X}$, we take the conditional $\tau_{X|A}$ constructed above,

$$\tau_{X|A} = \sum_x |x\rangle\langle x| \otimes \tau_A^x, \tag{A.73}$$

and use Bayesian inversion. This requires us to also specify a marginal ρ_A, but since our only goal is to construct some causal model that is compatible with the $\tau_{X|A}$ that we derived from $\tau_{B|A}$, we can take ρ_A to be any positive-semidefinite, trace-one operator. From ρ_A and $\tau_{X|A}$ we compute

$$P(x) = \mathrm{Tr}\left[\rho_A^{\frac{1}{2}} \left(|x\rangle\langle x| \otimes \tau_A^x\right) \rho_A^{\frac{1}{2}}\right] = Tr\left[\tau_A^x \rho_A\right], \tag{A.74}$$

and

$$\tau_{A|X} = \sum_x \frac{1}{P(x)} \rho_A^{\frac{1}{2}} \left(|x\rangle\langle x| \otimes \tau_A^x \right) \rho_A^{\frac{1}{2}} \equiv \sum_x |x\rangle\langle x| \otimes \tilde{\rho}_A^x, \tag{A.75}$$

where we define

$$\tilde{\rho}_A^x = \frac{1}{Tr\left(\tau_A^x \rho_A\right)} \rho_A^{\frac{1}{2}} \tau_A^x \rho_A^{\frac{1}{2}}. \tag{A.76}$$

The elements $\tilde{\rho}_A^x$ are by construction positive-semidefinite and trace-one, viz. valid quantum states, making $\tau_{A|X}$ a valid CE conditional. We have thereby shown that there exist conditionals $\tau_{B|X}$, $\tau_{A|X}$ and a classical prior $P(x)$ that parametrize a model with the causal structure $A \leftarrow X \rightarrow B$ and explain the given conditional $\tau_{B|A}$. □

A.6 Bounds on eigenvalues of partial transposes

Lemma 79. *Consider a convex combination of Hermitian operators,*

$$M = (p) M_1 + (1-p) M_2. \tag{A.77}$$

The eigenvalues of M are bounded by the convex combinations of the eigenvalues of the two terms:

$$\begin{cases} \lambda_{min}(M) \geq p\lambda_{min}(M_1) + (1-p)\lambda_{min}(M_2), \\ \lambda_{max}(M) \leq p\lambda_{max}(M_1) + (1-p)\lambda_{max}(M_2). \end{cases} \tag{A.78}$$

Proof. Eq. (3) in [96], noting that scalar factors multiplying a matrix simply multiply all its eigenvalues. □

Lemma 80. *Given a Hermitian operator ρ_{AB} on a product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, the eigenvalues of its partial transpose with respect to one system*[31] *are bounded by*

$$\mathrm{Tr}\rho_{AB} \geq \lambda\left(T_B \rho_{AB}\right) \geq -\frac{1}{2}\mathrm{Tr}\rho_{AB}. \tag{A.79}$$

Proof. following Rana [97]: Let p_i denote the eigenvalues of ρ_{AB}. Since ρ_{AB} is Hermitian, the p_i are real, but we stress that they need not be positive, since we do not demand positivity of ρ_{AB}. There is a constraint $\sum_i p_i = \mathrm{Tr}\rho_{AB}$ (with the right-hand side usually equal to some integer, depending on convention). Furthermore, the Hermiticity of ρ_{AB} ensures that its eigenvectors $|\psi_i\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can be chosen orthonormal. Subject to these constraints, we can decompose

$$\rho_{AB} = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{A.80}$$

The following holds for each $|\psi_i\rangle$, but for brevity we omit the index i: since $|\psi\rangle$ is a pure state on $\mathcal{H}_A \otimes \mathcal{H}_B$, it admits a Schmidt decomposition,

$$|\psi\rangle = \sum_j c_j |a_j b_j\rangle. \tag{A.81}$$

$\{|a_j\rangle\}$ and $\{|b_j\rangle\}$ can be chosen to be orthonormal bases of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. The only constraint is that they must be eigenbases of the reduced states $\rho_A = Tr_B |\psi\rangle\langle\psi|$ and $\rho_B = Tr_A |\psi\rangle\langle\psi|$, but this leaves the freedom to choose the phases of the basis elements, thereby making the coefficients

[31] The division into factor spaces depends on with respect to which degrees of freedom we want to take the partial transpose; it is not related to the variables used in the rest of the document.

$c_j \equiv \langle a_j b_j|\psi\rangle$ real and positive without loss of generality. Their absolute values are determined by the eigenvalues of ρ_A and ρ_B, and must satisfy

$$\sum_j c_j^2 = 1; \;\; c_j \geq 0. \tag{A.82}$$

The partial transpose of a pure state $|\psi\rangle$ can be written in terms of its Schmidt form (recalling that we made $c_j \in \mathbb{R}$) as

$$T_B|\psi\rangle\langle\psi| = \sum_{jk} c_j c_k |a_j b_k\rangle\langle a_k b_j|. \tag{A.83}$$

Rana points out that it is easy to show that the eigenvectors of this matrix are $|a_j b_j\rangle$ and $|a_j b_k\rangle \pm |a_k b_j\rangle$, with eigenvalues c_j^2 and $\pm c_j c_k$. Since we chose $c_j \geq 0$, only the eigenvalues with an explicit minus sign can be negative. Their maximal absolute value is 1/2, due to the constraint $\sum_j c_j^2 = 1$. It follows that $\lambda_{min}\left(T_B|\psi_i\rangle\langle\psi_i|\right) \geq -\frac{1}{2}$ for each term in the eigendecomposition of ρ_{AB}. Similarly, the largest eigenvalue is bounded by $\lambda_{max}\left(T_B|\psi_i\rangle\langle\psi_i|\right) \leq 1$, which is saturated if one $c_j = 1$ and all others are zero.
The eigenvalues of the partial transpose of ρ_{AB}, which we denote $\lambda_i\left(T_B\rho_{AB}\right)$, can now be bounded using lemma 79, recalling that $T_B\rho_{AB}$ as well as all $T_B|\psi_i\rangle\langle\psi_i|$ are Hermitian. Their smallest eigenvalues satisfy the inequality

$$\lambda_{min}\left(T_B\rho_{AB}\right) \geq \sum_i p_i \lambda_{min}\left(T_B|\psi_i\rangle\langle\psi_i|\right) \geq -\frac{1}{2}\mathrm{Tr}\rho_{AB}, \tag{A.84}$$

and similarly

$$\lambda_{max}\left(T_B\rho_{AB}\right) \leq \mathrm{Tr}\rho_{AB}. \tag{A.85}$$

□

Lemma 81. *Given a Hermitian operator ρ_{AB} on a product Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, the eigenvalues of its partial transpose with respect to one system saturate the lower bound,*

$$\lambda_{min}\left(T_B\rho_{AB}\right) = -\frac{1}{2}\mathrm{Tr}\rho_{AB}, \tag{A.86}$$

if and only if ρ_{AB} is pure, with Schmidt rank two and Schmidt coefficients $\left\{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, ...\right\}$ (additional zeros depending on the dimension d_A).

Proof. Following the proof of the (non-strict) inequality (lemma 80), one can establish conditions for each of the inequalities in the derivation to be saturated, thereby ensuring that the final inequality is also saturated: (a) All[32] eigenvectors $|\psi_i\rangle$ of ρ_{AB} must saturate $\lambda_{min}\left(T_B|\psi_i\rangle\langle\psi_i|\right) \geq -\frac{h_A}{2}$, which occurs if and only if the Schmidt coefficients of the eigenvector $|\psi_i\rangle$ are $\left\{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, ...\right\}$ (extra zeros for dimensions $d_A > 2$). (b) The smallest eigenvalues, $\lambda_{min}\left(T_B|\psi_i\rangle\langle\psi_i|\right) = -\frac{h_A}{2}$, must be associated with the same eigenvector for all i. Note that the explicit form of a partially transposed matrix depends on the basis in which the partial transpose is taken, and therefore so do its eigenvectors (even though its spectrum does not). We therefore demand that the eigenvectors associated with the smallest eigenvalues of $T_B|\psi_i\rangle\langle\psi_i|$ be the same for all i when the partial transpose is taken in the same basis.

The next step is to prove that ρ_{AB} cannot have two or more distinct eigenvectors with non-zero probability. To prove this by contradiction, let us assume that two such vectors exist, denoting

[32] It is understood implicitly that these conditions must hold only for those eigenvectors associated with non-zero probabilities p_i.

them $|\psi_0\rangle$ and $|\psi_1\rangle$. Let us write the Schmidt bases of $|\psi_0\rangle$ on both $\mathcal{H}_A$ and $\mathcal{H}_B$ as $\{|0\rangle, |1\rangle, ...\}$. (Although we use the same notation for the bases of $\mathcal{H}_A$ and $\mathcal{H}_B$ that put $|\psi_0\rangle$ in the Schmidt form, there need not be any particular relation between the two bases.) The Schmidt bases of $|\psi_1\rangle$ on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, may be obtained from the previous ones by unitary transformations: there exist unitaries U, V such that

$$\begin{cases} \{|u_0\rangle, |u_1\rangle, |u_2\rangle, ...\} \equiv U \{|0\rangle, |1\rangle, |2\rangle, ...\} \\ \{|v_0\rangle, |v_1\rangle, |v_2\rangle, ...\} \equiv V \{|0\rangle, |1\rangle, |2\rangle, ...\} \end{cases} \tag{A.87}$$

give the appropriate bases for writing, by condition (a),

$$\begin{cases} |\psi_0\rangle = \frac{1}{\sqrt{2}} \left(|00\rangle + |11\rangle\right) \\ |\psi_1\rangle = \frac{1}{\sqrt{2}} \left(|u_0 v_0\rangle + |u_1 v_1\rangle\right). \end{cases} \tag{A.88}$$

Since the partial transpose must be taken with respect to the same basis on both $|\psi_i\rangle\langle\psi_i|$ in order for us to compare them, let us express $|\psi_1\rangle$ in the $\{|0\rangle, |1\rangle, |2\rangle, ...\}$-bases of $\mathcal{H}_A$ and $\mathcal{H}_B$. In terms of the components of the unitaries U, V,

$$\begin{aligned} |\psi_1\rangle &= \frac{1}{\sqrt{2}} \left(\sum_{i=0}^{h_A - 1} |i\rangle_A U_{i0} \right) \left(\sum_{j=0}^{h_A - 1} |j\rangle_B V_{0j}^T \right) + \frac{1}{\sqrt{2}} \left(\sum_{i=0}^{h_A - 1} |i\rangle_A U_{i1} \right) \left(\sum_{j=0}^{h_A - 1} |j\rangle_B V_{1j}^T \right) \\ &= \frac{1}{\sqrt{2}} \sum_{j=0}^{h_A - 1} \left[\sum_{i=0}^{h_A - 1} |i\rangle \left(U_{i0} V_{0j}^T + U_{i1} V_{1j}^T \right) \right]_A |j\rangle_B \\ &= \frac{1}{\sqrt{2}} \sum_{j=0}^{h_A - 1} |w_j\rangle_A |j\rangle_B, \end{aligned} \tag{A.89}$$

where we introduce an effective transformation on A alone,

$$|w_j\rangle \equiv \sum_{i=0}^{h_A - 1} |i\rangle \left(U_{i0} V_{0j}^T + U_{i1} V_{1j}^T \right) = U \Pi_{01} V^T |j\rangle. \tag{A.90}$$

Note that this is not a unitary, because of the projection onto the subspace $\{|0\rangle, |1\rangle\}$ between V^T and U. The $|w_j\rangle$ need not be orthogonal (we only have $\langle w_j | w_k \rangle = \langle j | V^* \Pi_{01} V^T | k \rangle$, with $|j\rangle, |k\rangle$ orthonormal) and may have norm less than unity (though not greater, since $\langle j | V^* \Pi_{01} V^T | j \rangle \leq \langle j | V^* V^T | j \rangle = 1$).

In terms of this new basis, the partial transpose $T_B |\psi_1\rangle\langle\psi_1|$ with respect to the $\{|0\rangle, |1\rangle, |2\rangle, ...\}$-basis of B takes a simple form:

$$T_B |\psi_1\rangle\langle\psi_1| = \frac{1}{2} \sum_{jk} |w_j k\rangle\langle w_k j|. \tag{A.91}$$

The eigenvector of $T_B |\psi_0\rangle\langle\psi_0|$ associated with the smallest eigenvalue is simply (see proof of lemma 80)

$$|m\rangle = \frac{1}{\sqrt{2}} \left(|01\rangle - |10\rangle\right). \tag{A.92}$$

We demand that this also be an eigenvector of $T_B |\psi_1\rangle\langle\psi_1|$ eigenvalue $-\frac{1}{2}$, that is,

$$\left[\sum_{jk} |w_j k\rangle\langle w_k j| \right] \left(|01\rangle - |10\rangle\right) = \sum_k \left(|w_1 k\rangle\langle w_k | 0\rangle - |w_0 k\rangle\langle w_k | 1\rangle \right) = -\left(|01\rangle - |10\rangle\right). \tag{A.93}$$

Projecting onto $\langle 01|$ and $\langle 10|$, we find the necessary conditions

$$\begin{cases} \langle 0|w_1\rangle\langle w_1|0\rangle - \langle 0|w_0\rangle\langle w_1|1\rangle = -1 \\ \langle 1|w_1\rangle\langle w_0|0\rangle - \langle 1|w_0\rangle\langle w_0|1\rangle = +1. \end{cases} \tag{A.94}$$

On the other hand, we demand that the two eigenvectors of ρ_{AB}, $|\psi_0\rangle$ and $|\psi_1\rangle$, be orthogonal:

$$\langle\psi_0|\psi_1\rangle = \frac{1}{2}\left(\langle 0|w_0\rangle + \langle 1|w_1\rangle\right) = 0, \tag{A.95}$$

that is,

$$\langle 1|w_1\rangle = -\langle 0|w_0\rangle. \tag{A.96}$$

Let us simplify the expressions by introducing the following notation:

$$\begin{cases} |w_0\rangle = \alpha|0\rangle + \beta|1\rangle \\ |w_1\rangle = \gamma|0\rangle + \delta|1\rangle. \end{cases} \tag{A.97}$$

The constraints on the $\langle j|w_k\rangle$ become

$$\begin{cases} \delta = -\alpha \\ \gamma\gamma^* - \alpha\delta^* = -1 \\ \delta\alpha^* - \beta\beta^* = +1, \end{cases} \tag{A.98}$$

which is impossible to satisfy. Therefore, no second eigenvector of ρ_{AB} satisfying all the requirements can exist, and ρ_{AB} must be pure, of the form

$$\rho_{AB} = \frac{1}{2}\left(|00\rangle + |11\rangle\right)\left(\langle 00| + \langle 11|\right) \tag{A.99}$$

in some choice of local bases. □

Theorem. *(68 in the main text) The witnesses defined in eq.(5.25) are lower-bounded by*

$$u_{ce}, u_{cc} \geq -\frac{1}{h_B}, \tag{A.100}$$

where h_B is the dimension of the Hilbert space of variable B. These lower bounds are unique in the following sense: if either of them is saturated, then so is the other, and the inference map is uniquely determined to be

$$\tau_{B|CD} = \frac{1}{h_B}\mathbb{I}_B \otimes \mathbb{I}_{CD}, \tag{A.101}$$

whereby C and D provide no information about B at all.

Proof. Let us first prove the general lower bound on u_{ce}, using h_B, h_C and h to denote the dimensions of the respective Hilbert spaces. The law of total probability,

$$\mathrm{Tr}_B\tau_{B|CD} = \mathbb{I}_{CD}, \tag{A.102}$$

imposes the normalization of the conditional

$$\mathrm{Tr}\left(\tau_{B|CD}\right) = \mathrm{Tr}\mathbb{I}_{CD} = h_C h_D. \tag{A.103}$$

(Note that the trace is independent of the basis, hence we can evaluate it on a product basis, such that partial transposition with respect to C and/or D does not affect the diagonal elements, and

consequently the trace. Thus $\mathrm{Tr}\left(T_{CD}\tau_{B|CD}\right) = h_C h_D$ as well.) On the other hand, the trace is can be related to the eigenvalues $\lambda_{min} \leq \lambda_i \leq \lambda_{max}$ of $\tau_{B|CD}$ by

$$\mathrm{Tr}\left(\tau_{B|CD}\right) = \sum_{i=1}^{h_B h_C h_D} \lambda_i\left(\tau_{B|CD}\right) \geq \left(h_B h_C h_D\right) \lambda_{min}\left(\tau_{B|CD}\right). \tag{A.104}$$

Consequently

$$\lambda_{min}\left(\tau_{B|CD}\right) \leq \frac{1}{h_B} \quad \Leftrightarrow \quad u_{ce} \geq -\frac{1}{h_B}, \tag{A.105}$$

and analogously for $T_{CD}\tau_{B|CD}$ and u_{cc}.

Now assume that u_{ce} saturates this lower bound, which implies that $\lambda_{min}\left(\tau_{B|CD}\right) = \frac{1}{h_B}$. This implies an upper bound for the largest eigenvalue:

$$h_C h_D \geq \left(h_B h_C h_D - 1\right) \cdot \frac{1}{h_B} + \lambda_{max}\left(\tau_{B|CD}\right) \quad \Rightarrow \quad \frac{1}{h_B} \geq \lambda_{max}\left(\tau_{B|CD}\right). \tag{A.106}$$

That is, all eigenvalues λ_i must be between

$$\frac{1}{h_B} \leq \lambda_{min} \leq \lambda_i \leq \lambda_{max} \leq \frac{1}{h_B} \;\Rightarrow\; \lambda_i\left(\tau_{B|CD}\right) = \frac{1}{h_B}\ \forall i. \tag{A.107}$$

It follows that the conditional is proportional to the identity,

$$\tau_{B|CD} = \frac{1}{h_B}\mathbb{I}_{BCD}, \tag{A.108}$$

which represents a "channel" that always outputs a maximally mixed state, destroying all information. As a consequence, $u_{cc} = -\frac{1}{h_B}$ as well. The converse, starting with $u_{cc} = -\frac{1}{h_B}$, follows analogously. □

Theorem. *(69 in the main text) If $\tau_{B|CD}$ represents a non-trivial probabilistic mixture of CC and CE relations (excluding purely CC or CE relations), then the pair (u_{ce}, u_{cc}) lies in the half-plane delimited by*

$$u_{ce} + u_{cc} < \frac{h_A}{2}. \tag{A.109}$$

By the contrapositive, larger values of the sum herald a physical mixture.

Non-strict inequality.

Proof. We begin by proving the non-strict inequality

$$u_{ce} + u_{cc} \leq \frac{h_A}{2}. \tag{A.110}$$

Based on this proof, we will then rule out equality.

For a probabilistic mixture,

$$\tau_{B|CD} = (p)\,\tau_{B|C} \otimes \mathbb{I}_D + (1-p)\,\tau_{B|D} \otimes \mathbb{I}_C, \tag{A.111}$$

we can use lemma 79 to write

$$-\lambda_{min}\left(\tau_{B|CD}\right) \leq -p\lambda_{min}\left(\tau_{B|C}\right) - (1-p)\,\lambda_{min}\left(\tau_{B|D}\right); \tag{A.112}$$

noting that the spectrum of a tensor product $M \otimes \mathbb{I}$ is identical to that of M, up to multiplicity. Furthermore, the purely CC Jamiołkowski operator $\tau_{B|C}$ is positive-semidefinite, so that

$$-\lambda_{min}\left(\tau_{B|C}\right) \leq 0. \tag{A.113}$$

This allows us to modify the above bound to a weaker, but simpler form,

$$-\lambda_{min}\left(\tau_{B|CD}\right) \leq -\left(1-p\right)\lambda_{min}\left(\tau_{B|D}\right). \tag{A.114}$$

For the remaining term, it holds that (a) $T_D\tau_{B|D}$ is Hermitian, and (b) $\mathrm{Tr}\left(T_D\tau_{B|D}\right) = \mathrm{Tr}\mathbb{I}_D = \dim\left(\mathcal{H}_D\right) = \dim\left(\mathcal{H}_A\right) \equiv h_A$, so lemma 80 guarantees that

$$-\lambda_{min}\left(\tau_{B|D}\right) \leq \frac{h_A}{2}. \tag{A.115}$$

Thus the parameter that witnesses a CE influence is upper-bounded by

$$-\lambda_{min}\left(\tau_{B|CD}\right) = u_{ce} \leq \left(1-p\right)\frac{h_A}{2}, \tag{A.116}$$

and a similar derivation for $T_{CD}\tau_{B|CD}$ leads to

$$-\lambda_{min}\left(T_{CD}\tau_{B|CD}\right) = u_{cc} \leq p\frac{h_A}{2}. \tag{A.117}$$

Combining the two,

$$u_{ce} + u_{cc} \leq \frac{h_A}{2}. \tag{A.118}$$

□

Saturating the bound.

Proof. We will now show that the inequality holds strictly, i.e. the bound can not be saturated. To that end, we list necessary (and sometimes sufficient) conditions for saturating all the inequalities involved in the derivation above, and show that they lead to a contradiction.

In order to have $u_{ce} + u_{cc} = \frac{h_A}{2}$, it is necessary and sufficient that both $u_{cc} = \frac{h_A}{2}p$ and $u_{ce} = \frac{h_A}{2}\left(1-p\right)$. The conditions for those two are of the same form, in terms of $T_{CD}\tau_{B|CD}$ and $\tau_{B|CD}$, respectively. Let us consider the conditions for $u_{ce} = \frac{h_A}{2}\left(1-p\right)$ explicitly:

1. In order to saturate

$$-\lambda_{min}\left(\tau_{B|CD}\right) \leq -p\lambda_{min}\left(\tau_{B|C}\otimes\mathbb{I}_D\right) - \left(1-p\right)\lambda_{min}\left(\tau_{B|D}\otimes\mathbb{I}_C\right), \tag{A.119}$$

   it is necessary and sufficient that both summands, $\tau_{B|C}\otimes\mathbb{I}_D$ and $\tau_{B|D}\otimes\mathbb{I}_C$, share (at least) one eigenvector, and that it be associated with the smallest eigenvalue of each[33].

2. The smallest eigenvalue of the CC term $\tau_{B|C}$ must not be negative, but equal to zero, otherwise the inequality is no longer saturated when the term is dropped.

3. The CE term, on the other hand, must saturate

$$-\lambda_{min}\left(\tau_{B|D}\right) \leq \frac{1}{2}\dim\left(D\right) = \frac{h_A}{2}. \tag{A.120}$$

Condition 3 is addressed by lemma 81 (identifying the elements A, B and ρ_{AB} in the lemma with B, D and $T_D\tau_{B|D}$): it is necessary and sufficient that $T_D\tau_{B|D}$ be pure, with Schmidt coefficients $\left\{\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, ...\right\}$. By a similar reasoning, the conditions that ensure $u_{cc} = \frac{h_A}{2}p$, which are cast in

[33]This necessary condition is included for completeness; it is not used in the proof. One may be able to derive stronger statements by including it.

terms of $T_{CD}\tau_{B|CD}$, imply that $\tau_{B|C}$ must also be a pure state, with the same Schmidt coefficients. We may have to choose a different basis from the one that puts $T_D\tau_{B|D}$ in a simple form, but one can find local bases that give $\tau_{B|C} = |\Phi^+\rangle\langle\Phi^+|$. Then it is easy to show that the eigenvalues of $\tau_{B|C}$ are ± 1, which contradicts condition 2 above.

Consequently, there are no conditionals $\tau_{B|C}$, $\tau_{B|D}$ such that a probabilistic mixture of them saturates the bound $u_{ce} + u_{cc} \leq \frac{h_A}{2}$ at any point, except at the extremes $p = 0, 1$. (In those cases, $-p\lambda_{min}\left(\tau_{B|C}\right) \leq 0$ is saturated thanks to the factor $p = 0$, so that condition 2 becomes unnecessary, and analogously for $p = 1$.) □

A.7 Proof that $W = 0$ for probabilistic mixtures

This section shows that the witness W introduced in section 5.3.3 is zero for any probabilistic mixture. Specifically, we show the following:

Theorem 82. *Consider a probabilistic mixture of CC and CE relations between two qubits, with Choi state*

$$\tau^{prob}_{CBD} = q\rho_{CB} \otimes \frac{1}{2}\mathbb{I}_D + (1-q)\rho_C \otimes \tau_{BD}, \tag{A.121}$$

where $\rho_C = Tr_B\rho_{CB}$. The witness is defined in terms of measurements of fixed Pauli observables on C and B and equiprobable preparations of eigenstates of a fixed Pauli observable on D, which generate statistics

$$P(cdb) = \frac{2}{2}\mathrm{Tr}\left[T_D\left(\tau_{CBD}\right)\Pi^{s,c}_C \otimes \Pi^{u,b}_B \otimes \Pi^{t,d}_D\right], \tag{A.122}$$

as

$$W = 8\sum_{b=\pm 1} b\left[P(++b)P(--b) - P(+-b)P(-+b)\right]. \tag{A.123}$$

For this probabilistic mixture, the witness is zero.

Proof. We begin by noting several mathematical properties of probabilistic mixtures that will be useful in the subsequent proof. The first term represents the common-cause scenario, wherein we prepare a bipartite state ρ_{CB} and trace out D; hence the marginal on D of the Choi state is the completely mixed state. The state ρ_{CB} is obtained from the initial state ρ_{CE} by a completely positive, trace-preserving (CPTP) map acting on E, hence the marginal on C is unchanged: $\rho_C = \mathrm{Tr}_E\rho_{CE}$. The second term corresponds to a cause-effect scenario, in which case the marginal state on C is simply the marginal of the initial bipartite state ρ_{CE}. Meanwhile, $\tau_{B|D}$ is the Choi state corresponding to a CPTP map from D to B, hence its marginal on D is again the completely mixed state. In summary, the marginals of the two terms on C and D, respectively, are equal:

$$\mathrm{Tr}_B\rho_{CB} = \mathrm{Tr}_E\rho_{CE} = \rho_C, \tag{A.124}$$

$$\mathrm{Tr}_B\tau_{BD} = \frac{1}{2}\mathbb{I}_D. \tag{A.125}$$

It follows from these equalities that C and D become independent if we ignore B:

$$\mathrm{Tr}_B\left[\tau^{prob}_{CBD}\right] = \rho_C \otimes \frac{1}{2}\mathbb{I}_D. \tag{A.126}$$

The experimental statistics inherit these properties: letting $\mu_D(d) \equiv \frac{1}{2}\ \forall d = \pm 1$ denote the uniform probability distribution, we have

$$P(cdb) = qP_{CB}(cb)\mu_D(d) + (1-q)P_C(c)P_{BD}(bd). \tag{A.127}$$

The marginal distributions over c and d in both terms are identical:

$$\sum_b P_{CB}(cb) = P_C(c), \tag{A.128}$$

$$\sum_b P_{BD}(bd) = \mu_D(d), \tag{A.129}$$

and, if we ignore b, then c and d become independent:

$$\sum_b P(cdb) = [q + (1-q)]P_C(c)\mu_D(d). \tag{A.130}$$

Now we can show that W is zero for any probabilistic mixture of common-cause and cause-effect relations. Recall that, since b only takes two values, the marginal independence (A.130)

$$\sum_b P(cdb) = P_C(c)\mu_D(d) = \frac{1}{2}P_C(c) \tag{A.131}$$

implies that

$$P(cd,-) = \frac{P_C(c)}{2} - P(cd,+). \tag{A.132}$$

This allows us to rewrite the $b=-1$ term in eq. (A.123) as

$$\begin{aligned}
&P(++,-)P(--,-) - P(+-,-)P(-+,-) \\
&= -\frac{P_C(+)}{2}P(--,+) - \frac{P_C(-)}{2}P(++,+) \\
&+ \frac{P_C(+)}{2}P(-+,+) + \frac{P_C(-)}{2}P(+-,+) \\
&+ [P(++,+)P(--,+) - P(+-,+)P(-+,+)],
\end{aligned} \tag{A.133}$$

hence the witness reduces to

$$\begin{aligned}
W &= 4[+P_C(-)P(++,+) - P_C(-)P(+-,+) \\
&\quad - P_C(+)P(-+,+) + P_C(+)P(--,+)] \qquad \text{(A.134)} \\
&= 4\sum_{cd} cd\,[1 - P_C(c)]\,P(cd,+) \qquad \text{(A.135)}
\end{aligned}$$

Our core hypothesis implies that $P(cd,+)$ is a convex combination of two terms, each a product of distributions over c and d. Substituting eq. (A.127) and distributing the sums,

$$\begin{aligned}
W = 4q &\left[\sum_c c\,[1 - P_C(c)]\,P_{CB}(c,+)\right]\left[\sum_d d\mu_D(d)\right] \\
+ 4(1-q) &\left[\sum_c c\,[1 - P_C(c)]\,P_C(c)\right]\left[\sum_d P_{BD}(d,+)\right].
\end{aligned} \tag{A.136}$$

In the first term, we have the average over $d = \pm 1$ under the uniform distribution, which is zero. In the second term, the sum over c gives $P_C(+)P_C(-) - P_C(-)P_C(+) = 0$. Thus

$$W = 0 \tag{A.137}$$

for any probabilistic mixture of the form (A.121). □

A.8 Tight upper bound on the induced mutual information for a probabilistic mixture

This section provides a rigorous proof of the upper bound (5.75) on the mutual information between two causal parents, D and E, induced by post-selection on their common effect, B, if B depends probabilistically either on D or on E. The conditional distribution $P(DE|B)$ is derived in the main text as eq. (5.69); this appendix focuses solely on finding the maximal mutual information for a distribution of this form, which is achieved by explicit construction.

Theorem 83. *If two classical variables D and E of equal cardinality, h, follow the distribution*

$$P(DE) = p\mu(D) P_E(E) + (1-p) P_D(D) \mu(E) \tag{A.138}$$

with μ denoting the uniform distribution and generic distributions $P_E(E)$ and $P_D(D)$, then their mutual information satisfies the tight upper bound

$$I^{prob}(D:E) \leq \left(1+\frac{1}{h}\right) + \log_2 h - \left(1+\frac{1}{h}\right)\log_2(h+1). \tag{A.139}$$

Proof. In general, the parameters $P_D(D|B)$, $P_E(E|B)$ and p_b in (5.69) depend on the value of B on which one post-selects. However, since the following derivation relies only on the abstract form of $P(DE|B)$, the explicit dependence is suppressed for brevity, reducing the expression to

$$P(DE) = p\mu(D) P_E(E) + (1-p) P_D(D) \mu(E). \tag{A.140}$$

We will first show that the mutual information is maximized when $P_D(D)$ and $P_E(E)$ are deterministic, in the sense that they produce one particular value with certainty. To this end, consider the mutual information of two variables as a functional of two arguments: the marginal distribution over one variable, which in our case is

$$P(E) = pP_E(E) + (1-p)\mu(E), \tag{A.141}$$

and the conditional over the other given the first,

$$P(D|E) = \frac{pP_E(E)}{pP_E(E)+(1-p)\mu(E)}\mu(D) + \frac{(1-p)\mu(E)}{pP_E(E)+(1-p)\mu(E)}P_D(D). \tag{A.142}$$

Considered as a functional, the mutual information is convex downward with respect to the second argument (see e.g. [98], theorem 2.7.4); that is, for any

$$P^\lambda(D|E) \equiv \lambda P^0(D|E) + (1-\lambda) P^1(D|E) \ , \ \ \lambda \in [0,1] \tag{A.143}$$

and fixed marginal $P(E)$, it holds that

$$I(D:E)\left[P(E), P^\lambda(D|E)\right] \leq \lambda I(D:E)\left[P(E), P^0(D|E)\right] + (1-\lambda) I(D:E)\left[P(E), P^1(D|E)\right]. \tag{A.144}$$

In particular, if we fix $P(E)$, and therefore the fractions in the expression for $P(D|E)$ above, but take a convex combination

$$P_D^\lambda(D) = \lambda P_D^0(D) + (1-\lambda) P_D^1(D), \tag{A.145}$$

then the resulting $P(D|E)$ will be a convex combination with weight λ as well, and the upper bound on the mutual information follows. Thus, for fixed $P_E(E)$ and p, the largest mutual information is achieved when the distribution $P_D(D)$ is extremal, i.e. produces one value with certainty. By symmetry, $P_E(E)$ must also be extremal. Since the problem is invariant under relabelling the values

of D and E, we can assume without loss of generality that P_D and P_E each produce the first outcome with certainty.

Thus, the maximal mutual information between D and E for a distribution constrained to the form (A.140) is achieved by the distributions

$$P_D(D) = \{1, 0, 0, ...\} = P_E(E) \tag{A.146}$$

(with a varying number of zeros depending on h_D and h), that is,

$$P(DE) = \begin{pmatrix} \frac{p}{h_D} + \frac{1-p}{h_E} & \frac{1-p}{h_E} & \frac{1-p}{h_E} & ... \\ \frac{p}{h_D} & 0 & 0 & ... \\ \frac{p}{h_D} & 0 & 0 & ... \\ \vdots & \vdots & \vdots & \end{pmatrix}, \tag{A.147}$$

where each row corresponds to a value of D and each column to a value of E. Under the simplifying assumption that D and E range over an equal number of values, $h_D = h_E = h$, symmetry suggests that the mutual information is maximal when $p = \frac{1}{2}$. Indeed, it is straightforward to calculate the mutual information of the above distribution explicitly and, if $h_D = h_E = h$, show that it is maximal for $p = \frac{1}{2}$, in which case (assuming the entropy is calculated base 2)

$$I_{prob}^{max}(D : E) \equiv \left(1 + \frac{1}{h}\right) + \log_2 h - \left(1 + \frac{1}{h}\right) \log_2 (h + 1). \tag{A.148}$$

This is a tight upper bound for the mutual information achievable within the given constraints. □