The Topology and Geometry of Causality

Nicola Pinzani
St Hugh's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2022

[. . .] were they concealing no hidden thought, and was it simply visual fatigue that made me see them double in time as one sometimes sees double in space? I could not tell....

Marcel Proust, *In the Shadow of Young Girls in Flower*

Acknowledgements

Oxford, from the countless hours spent in the Bodleian to the flowers of Magdalen college, you calmed the turbulence of my thoughts and what a journey through tumultuous times! And thank you, Edamame, for your miso; your tiny room with a handful of tables was always just the right size. You provided excellent guidance allowing me to fathom the depth of my conscience in the soup, lingering for divination.

Trieste, the promenade, the sea, the scorching sun and the seagulls. The familiar streets and the fragrances of the Adriatic coast were the evanescent essences that collected coherence and beauty in these nine long years. A clarifying sense of belonging to what I've lost.

I'm grateful to my family. My overly-protective (and perfect) Italian mother. My sister, to whom I could have been a better brother. My father fought and lost his battle during my time in Oxford. He would have been so proud of me, and I miss him immensely. Zio Paolo for always being a source of joy, knowledge, pragmatism, and music.

Thank you, Hlèr; I loved sharing the office with you and disagreeing about causality. Thank you, Quantum Group. Aleks K, Nick O, Augustin, James, Aleks vdW, John, Lia, Isaac, Subhayan, Marietta, Fatimah, Nick H, Lucas, Nihil, Tein, David, Sivert, Cole, Aleks C and many others! Cheers to the Quantinuum pioneers, Konstantinos, Carys, Giovanni, Alexis, Robin, Razin, Richie et al. Thank you for sharing the rooftop and the moonlit lives of your neighbours.

Antine, Giuseppe, Alexandra, what happened to Soho? After all, I have always been just a country boy. Bello perdersi nella nostalgia a sole due fermate di treno. Thank you, Duino; you are still silently alive in empty Negroni glasses crystallised in a Viennese attic. Emanuele, grazie per tutti gli scorci discreti e brillanti di un continuo coerente.

Hvala članom ekipe Hudbni Boyz. Kar smo, to bomo in oproščeno mi bo, da še preveč sem bil odsoten.

Bob Coecke, the alpha and omega, has been the most significant source of inspiration and headaches (in the literal sense of the word). Genius or Vandal? You have taught me much—more than you can possibly imagine.

Thank you, Jonathan Barrett for your sharpness, coherence, inspirational rigour, and kindness. Your lobster lunches have been an invaluable compass to navigate the sea of foundations.

To Stefano Gogioso, friend and the best scientific guidance I could have ever hoped for. It has been an honour being your student. I am sure we will both cherish and let the substance of our endless conversations bloom. You taught me everything, and there is still so much to come.

Abstract

Quantum theory is manifestly in tension with the classical notion of causality. How do we recover causal reasoning in the quantum regime? In this dissertation, we propose a framework where such causal idiosyncrasies are identified as obstructions to the existence of global sections for presheaves of causal data. We do so by extending the Abramsky-Brandenburger framework for non-locality and contextuality [6] to situations where measurement contexts are allowed to be signalling. This results in a theory-independent phenomenology of causality, which can be used to reason about causal structure in any theory exhibiting contextuality.

In the first part of this dissertation, we study the specific phenomenology of coherent control of quantum channels, giving rigorous operational meaning to the superposition of causal order. We pursue a bottom-up approach—alternative to the process matrix formalism—by investigating how indefiniteness of causality emerges from specific characteristics of operational theories. This provides the recipe for building processes with indefinite causality, which are then causally analysed using tools described in the second part of the thesis.

The second, more substantial part of this dissertation is devoted to building the sheaf-theoretic framework unifying non-locality, contextuality and indefinite causality. We provide a combinatorial description of the operational assumptions underlying definite and indefinite causal order, and characterise the emergent topologies of classical contexts. We explain how to associate causal data to such topologies and detail the relationship between the covers for a topological space and varying degrees of classicality. We develop a complementary geometric understanding of the space of empirical models for this presheaf, and show how it can be used to perform theory-independent causal analysis of empirical data. We conclude by providing novel examples of such causal analysis, showcasing the existence of the phenomenon of contextual causality. Importantly, our examples demonstrate that such phenomenon can be witnessed in quantum theory, as long as coherently control of causal order is allowed for quantum processes.

Contents

Chapter 1

Introduction

1.1 Prolegomenon

In providing a classical causal explanation for a Bell experiment, the appeal to superluminal causality is unavoidable; no classical causal description fits the operational assumptions underlying the realisation of such protocols. Where does this irreconcilability lie? Studies about the application of classical causal algorithms to quantum theory, an endeavour attempted by Spekkens in [144] and generalised to a broader class of contextuality scenarios in the works of Cavalcanti and Pearl [35, 105], are aimed at a precise explanation of how our classical understanding of causality fails to capture the meaning of Bell's correlations. They show that any causal explanation must contradict the 'no fine-tuning assumption', the principle that the statistical independence between variables should not be explained by the fine-tuning of the causal parameters. For this reason, it seems that quantum theory allows for what we would call—in light of the principles of classical causal modelling—either 'superluminal causes' which are fine-tuned to avoid any possibility of faster than light communication, or the introduction of additional causes entailing superdeterminism.

What structural property of quantum theory makes these fine-tuned correlations possible, if not ubiquitous? How do we characterise this departure from classicality at a purely empirical level? Many foundational research programs have been developed to settle these questions with causality and construct its phenomenology in the broader context imposed by quantum theory. To motivate the kind of interplay one would expect between causality and quantum correlations, we can start by understanding the consensus about the role of probabilities in the 'classical' study of causality.

A comprehensive account of how causal mechanisms are related to uncertainty can be found in Judea Pearl's seminal book [104] where, in the very first sentences, it is made clear that his framework relies on an essential conceptual distinction: '[c]ausality connotes lawlike necessity, whereas probabilities denote exceptionality, doubt, and lack of regularity' [104]. Causality is seen as the embodiment of fundamentally deterministic mechanisms which relate the properties of multipartite systems in an intricate net of causes and effects, in fundamental opposition with probabilities understood as a direct consequence of obfuscating the ontic ‘regularity’. For Pearl, probabilistic uncertainty is described antithetically to the functional relationship between causes and effects even though, in many instances, the causal mechanism breaks through the barrier of uncertainty. Pearl mainly provides two reasons to justify the introduction of an intrinsically probabilistic treatment of causality: the finite granularity of language unavoidably fails to capture the precision needed to reveal the complete underlying functional mechanism, and natural language is unavoidably plagued with uncertainties.

Both justifications for the need for probabilities become controversial in the presence of a fundamental theory of nature, so we should not be too surprised to learn about the failure of causal discovery in framing quantum interactions. The emergence of probabilistic behaviour cannot result from linguistic imprecision or ignorance about objective facts but witnesses a more fundamental descriptive limitation. This intrinsic probabilistic behaviour replaces functions as the machinery of causal mechanisms. It should not be seen as a theoretical limitation but, on the contrary, as what enables the understanding of new types of regularities that ‘exclude analysis on classical lines’[82] and become manifest when one appeals to the unification of complementary perspectives.

This understanding of the nature of probabilities and, therefore, of the failure of classical approaches to causality is strengthened by numerous proofs of the contextual features of the theory, corroborating the implausibility of recovering a fundamental causal description. As Niels Bohr puts it, we are here concerned with ‘new uniformities which cannot be framed into the frame of ordinary causal description’ [26], the exploration of ‘harmonies which cannot be comprehended in the pictorial conceptions adapted to the account of more limited fields of physical experience’[82]: harmonies not amenable to standard causal understanding but governed by precise rules and regularities appearing in a new domain of applicability. Bohr uses these words only two years after the proposal of EPR and decades before Bell or Kochen-Specker proved their seminal theorems but is already hinting at the impression that there is something truly essential in the role played by contextuality. This work is permeated by the belief that the failure to analyse quantum processes with the tools of causal discovery may reveal what Bohr would have described as the unavoidability of synthesising (and not visualising) quantum phenomena by ‘combined use of contrasting pictures’[28].

Given that a completely causal account of quantum mechanics turns out to be so conflicting with the basic principles of the theory, what can be offered as a replacement for classical causal reasoning? If we can easily recognise some idiosyncrasies, finding the conceptual framework that allows a complete diagnosis is more complicated. We need to broaden the conceptual toolbox; one possibility is to find a direct generalisation of classical causal models, replace the deterministic functional relations between classical relata with a quantum generalisation thereof and hope that the classical limit reflects our current understanding of causality [86, 46, 17]. Another perspective, which is the one we endorse in this work, is to consider quantum processes to be a mathematical synthesis of observations, inextricably interconnected with the classical contexts determined by the experimental parameters. The critical point to notice is that, although quantum contextuality rules out the possibility of a simple generalised explanation that works for all contexts, it does not rule out a collection of functional explanations, each one working for a single context. The impossibility of coalescing the classical perspectives showcases the failure of a global functional—and thus causal in the classical sense—description. This second alternative, which is the generalisation that Bohr vehemently proposes in [27], can be characterised by the appropriation of the notion of 'causality' by the more general 'complementarity': causes and effect can be connected across contexts in a way which requires to transcend deterministic functional connections.

> In this novel situation, even the old question of an ultimate determinacy of natural phenomena has lost its conceptional basis, and it is against this background that the viewpoint of complementarity presents itself as a rational generalisation of the very ideal of causality. [27]

The two directions for generalising causal reasoning are radically different, but this is not to say that they cannot coexist; they use different language and have different levels of dependency on the structure of the theory. While an understanding of a quantum notion of causality is conservative insofar as it hopes to preserve certain elements characterising classical causal reasoning while dealing with explicitly quantum processes, our perspective continues to look at a protocol with external 'classical' eyes relaxing the desideratum of underlying functional explanations. The apparent compatibility, however, hides a lot of philosophical nuances. A successful quantum generalisation of causality would be a powerful argument for anyone who understands unitary processes as ontic, as pertaining to the 'real stuff' that connects causes with effects; our approach can be seen as leaning towards a more agnostic perspective in which unitary transformations and quantum states are at the very least symbolic devices unifying classical perspectives. Any specific philosophical commitment from our side is far from being a necessity; the type of causal discovery described in this work can uncontroversially be used for the practical study of protocols independently of any ontological commitment.

The synthesis of empirical data into the theory of Hilbert spaces and unitary transformations presupposes the existence of some more fundamental rules of coordination. The classical notion of causality, undressed by her all-encompassing aura, will always be necessary for quantum theory as it underlies the classical description of experimental arrangements. A classical, and therefore causal, understanding will only constitute one of the possible complementary descriptions; the theory exhibits its fundamentally non-classical or 'fine tuned' correlations, precisely coalescing them into a single process parametrised by the classical settings identifying the contexts. Bohr's suggestion to supersede causality always recognised a privileged role for classical structures. Complementarity is only a ‘wider frame’ that ‘directly expresses [. . .] the account of fundamental properties of matter presupposed in the classical physical description, but outside its scope’ [27]. The description of the measurement devices ‘must always be based on spacetime pictures’, distinctly from the object under investigation, for which ‘observable predictions can in general only be derived by the non-visualizable mechanism’ [27] and a causal account is unavoidable as is unavoidable the description of the parameters of measurement devices:

> It is essential to note that, in any well-defined application of quantum mechanics, it is necessary to specify the whole experimental arrangement and that, in particular, the possibility of disposing of the parameters defining the quantum mechanical problem just corresponds to our freedom of constructing and handling the measuring apparatus, which in turn means the freedom to choose between different complementary type of phenomena to study. [27]

How does this fit in our narrative? In a nutshell, we aim to create a mathematical framework to understand spacetime processes where complementary choices of parameter settings are allowed to causally influence distributions at other events. The ‘new uniformities’ exposed by the quantum theory are to be found in the interplay between the alternative ‘classical snapshots’ specified by the joint local choices of settings, described as a table of conditional probabilities, which we refer to as *empirical models*. Even though compatible with a relativistic notion of no-signalling from the future, a collection of assignments for a family of contexts may nevertheless show a degree of global incoherence —witnessing the impossibility of an underlying causal and classical description— which is nevertheless realisable by quantum instruments.

We propose a general language that can be used to understand contextual phenomena under complex causal assumptions and, at the same time, understand the consequences of quantum causality from the perspective of contextuality. The reconciliation of ‘causality’ and ‘contextuality’ allows us to understand what type of causal inferences can be abstracted from observable distributions of outcomes. The causal relata will be the data which describe a complete classical context, and the quantum mechanical ‘regularities of nature’ are to be captured within the probabilistic description of the outputs conditioned by a given context. In some instances, the table of conditional probabilities will be explainable by classical causal mechanisms; we will call such protocols ‘local’ or ‘classical’. In other, the regularities of nature will express themselves by exhibiting contextuality or nonlocality, and the connection between the events may require other principles of explanation.

Methodologically, the work in this thesis can be seen as a dialogue between the principles of compositionality and decompositionality, the first referring to the possibility of *synthesis*: deriving behaviours from some constitutive elements. The second one of a more *analytical* flavour: understanding structural properties intrinsic in the data. Categorical probabilistic theories (CPTs) [66] will provide the glue between these two notions. A categorical probabilistic theory (CPT) is a mathematical description of processes that makes explicit the classical data used to describe some interaction. Given a CPT, we can construct processes with various causal structures by composing these constitutive elements sequentially and in parallel. When all the inputs and the outputs of a diagram are solely constituted of 'classical systems', we obtain the probability of outcomes for some measurements conditioned on the classical inputs. The decompositional part will be studied through the lenses of the sheaf-theoretic approach to contextuality: a framework developed by Abramsky and Brandenburger in [6] which fits into the broader research program of applying categorical methods to the study of quantum foundations. The sheaf-theoretic approach has been used very successfully to understand the meaning of local and global assignment of classical data that would explain no-signalling protocols; we will show that this use of the sheaf-theoretic machinery is a special case of a more general theory which includes the possibility of signalling.

1.2 Summary of this work

Chapter 1 introduces the theoretical background relevant to the rest of the thesis. We introduce categorical probabilistic theories by accompanying the reader through the main conceptual ideas underlying categorical quantum mechanics. After that, we also briefly and comprehensively introduce the sheaf-theoretic approach to contextuality and non-locality, which is essential for understanding the sheaf theoretically flavoured Chapter 4 and Chapter 5. We give a combinatorial account of causal orders, which will be greatly generalised in Chapter 3. The rest of the chapter is devoted to an extensive review of the literature that has inspired this dissertation.

Chapter 2 is mainly based on the paper 'Giving Operational Meaning to the Superposition of Causal Orders' [107] and other unpublished notes. In this chapter, we discuss the notion of coherent control of casual orders and the problems arising from general descriptions of the quantum control of arbitrary families of channels. We show that —in contrast with the control for arbitrary channels— controlling casual orders is a well-defined notion and provide a procedure to construct circuits between arbitrary laboratories expressing 'indefinite causality'. Chapter 2 aims to assert that a theory of processes where the causal order is coherently or incoherently controlled can be constructed from the standard circuital description of quantum processes and that this construction exhibits a degree of canonicity not usually found in the control of arbitrary families of channels. In particular, we use this framework to construct examples and protocols that are then analysed from a causal perspective in Chapter 6, without relying on the computationally inefficient and unnecessarily general description given by process matrices.

Chapter 3 is based on 'The Combinatorics of Causality' [63] and introduces the main combinatorial object describing causal assumptions, *spaces of input histories*. They provide a generalisation of standard causal orders explicitly accounting for causal interventions at each node. These spaces form a hierarchy of causal assumptions which generalises the lattice of causal orders introducing the possibility for the inputs to have an explicit influence on the causal structure. Once the general theory describing the spaces is established, we study their properties and show that not all spaces of histories correspond to a causally complete scenario, i.e. scenarios where there is no ambiguity about the causal unwrapping of the histories. In a causally complete space, every history represents a sequence of timelike choices and is univocally associated with a compatible causal ordering of its underlying events; this fails with more general causally incomplete spaces. We characterise *non-tight* spaces of histories, cases in which there is an overabundance of possible causal unwrapping, and which describe the conjunction of different assumptions about the causal order between events. We conclude the section by characterising the classes of causally complete spaces on two and three events with binary inputs.

Chapter 4 is based on 'The Topology of Causality' [65] and is devoted to using the sheaf-theoretic language to describe the compatible empirical models. We equip every space of causal histories with a suitable topology describing the hierarchy of contexts induced by the mutual interaction of timelike histories. We provide a description of empirical models as a bundle of compatible data. We explain the importance of the open covers of these topological spaces and explicitly discuss the importance of different covers associated with a space, among which are the 'classical cover', 'the standard cover' and the 'solipsistic cover'. We discuss how the standard notion of locality and contextuality are phrased in our language and provide a general definition of empirical models obtained by fixing a space of input histories and an open cover for its topology.

In Chapter 5, we explore the geometric picture that emerges by assigning compatible families of distributions to open covers and introduce the notion of a *causaltope*. This chapter is based on 'The Geometry of Causality' [64]. We describe causaltopes as polytopes for conditional probability distributions constrained by equations synthesised from the orders between the contexts. The connection between the description of the causaltopes and the empirical models defined in Chapter 4 is made rigorous by constructing a convex linear isomorphism between the causaltope defined for a given space and cover and the space of empirical models on that cover. We define causal separability and inseparability over the causal completions of a given space, providing a more fine-grained notion of what is represented by causal separability in the literature about causal inequalities.

In Chapter 6, based on the last part of [64], we provide several examples of standard empirical models, computing the causal fractions supported by sub-spaces of interest. Through a sequence of novel examples based on quantum switches and entangled states, we prove the existence of 'contextual causality', connecting non-locality—the impossibility of classically explaining outputs at events—with causal inseparability—the impossibility of explaining causal structure between events.

1.3 Background

1.3.1 Categorical Probabilistic Theories

1.3.1.1 Symmetric Monoidal Categories

Category theory can be used to provide a mathematically rigorous description of *process theories*, the backbone of the operational approaches to physics. The formalisation of general theories of processes is based on the monoidal paradigm, according to which *processes can be composed sequentially and in parallel*, respectively representing timelike and spacelike connections. This section aims to introduce categories to a practicing quantum foundationalist by following a slightly different narrative, one that explains *symmetric monoidal categories* by focusing on how they interlace with relativistic causality. We want to convey the idea that general categories can be used to axiomatise the algebra of processes under various spatio-temporal assumptions: from a 'Newtonian spacetime' with totally ordered events to the algebra of processes living in a Minkowski spacetime. For detailed explanations of the relationship between causality and monoidal categories, we invite the reader to consult [80, 44, 43].

We start by describing the ingredients that characterise a process theory C, starting with systems: entities that evoke the properties of a portion of the universe that we care about and are used to coalesce the possible manipulations and observations that can be made about them. We fix a set of symbols obj C [1] which identify the general types of systems inhabiting our theory. The elements of this set constitute the *objects* of our process theory. The importance of an explicit description of objects is relative, fundamental to the process-theoretical or categorical approach is the understanding that the relevant properties should be instead inferred from the structure of the possible interactions and transformations between systems. So, what are the essential ingredients that allow us to speak about processes with a spatio-temporal connotation in the first place? First, we assume that there exists for every object a canonical evolution which identifies their geneidentity, operationally embodying the properties which define the object, which can be (at least in principle) considered stable in 'time'. We call these special transformations *identity transformations*, they are inextricably interconnected with the notion of the object itself. A world where objects can be witnessed only by their potential immutability in time would not be particularly exciting: to allow for the possibility of change, we say that given objects $A, B \in$ obj C there exists a set of transformations, of *morphisms* which we denote by $C(A, B)$.

Given two morphism $f \in C(A, B)$ and $g \in C(B, C)$ the theory must provide a way to compose the two evolutions; we define an abstract function $- \circ - : C(B, C) \times C(A, B) \to C(A, C)$ called the *composition of morphisms* which describes the concatenation of processes. We require this concatenation to be associative so that $f \circ (g \circ h) = (f \circ g) \circ h$. We know that the sets $C(A, A)$

[1]the word 'symbol' tries to convey the objects' abstract nature freeing them from any ontological bias

cannot be empty as it contains at least the canonical identity transformation. This map interacts trivially with any other morphism, in particular we say that if $f \in C(A, A)$ we have that $id_A \circ f = f$ and $f \circ id_A = f$.

The most general theory for which we are only guaranteed the associativity of composition is formalised by the notion of a *category* [53], first proposed by Samuel Eilenberg and Sounders Mac Lane to formalise the notion of 'natural equivalence' of mathematical structures.

Definition 1.1 (Category). *A* category $\mathcal{A}$ *consists of:*

1. *a collection of objects* $obj\mathcal{A}$*;*

2. *for each* $A, B \in obj\mathcal{A}$*, a collection* $\mathcal{A}(A, B)$ *of* arrows *or* morphisms *from* A *to* B*;*

3. *for each* $A, B, C \in obj\mathcal{A}$*, a function*

$$\mathcal{A}(B, C) \times \mathcal{A}(A, B) \to \mathcal{A}(A, C) \tag{1.1}$$

$$(g, f) \mapsto g \circ f \tag{1.2}$$

   called composition*;*

4. *for each* $A \in obj\mathcal{A}$*, an element* 1_A *of* $\mathcal{A}(A, A)$*, called the* identity *on* A*,*

satisfying the following axioms:

1. associativity*: for each* $f \in \mathcal{A}(A, B), g \in \mathcal{A}(B, C)$ *and* $h \in \mathcal{A}(C, D)$ *we have*

$$(h \circ g) \circ f = h \circ (g \circ f); \tag{1.3}$$

2. identity laws*: for each* $f \in \mathcal{A}(A, B)$*, we have* $f \circ 1_A = f = 1_B \circ f$*.*

Example 1.2 (Examples of common categories). *Categories are very general; with no additional structure, they can look like anything resembling physical processes. Here are some commonly found examples:*

- *Let* $\mathcal{S}$ *be a set of sets, we can consider the category* $\mathrm{Set}_{\mathcal{S}}$ *for which objects are given by the elements of* $\mathcal{S}$ *and the morphisms* $\mathrm{Set}_{\mathcal{S}}(U, V)$ *are given by the functions* $U \to V$*.*

- *We can construct the category* $\mathrm{Rel}_{\mathcal{S}}$ *with the same objects as* $\mathrm{Set}_{\mathcal{S}}$ *but where the morphisms* $\mathrm{Rel}_{\mathcal{S}}(U, V)$ *are given by relations. Given* $R : U \to V, T : V \to W$ *we can define the composition* $T \circ R$ *such that:*

$$x(T \circ R)y \iff \exists z \in V \text{ s.t. } xTz \text{ and } zRr$$

- *The category* fHilb *of finite dimensional Hilbert spaces where morphisms* $\mathrm{fHilb}(H, K)$ *are given by linear maps of the type* $H \to K$.

- *For a commutative semiring* R*, the category* $R\text{-Mat}$ *has the positive integers* $m, n, \ldots$ *as objects and the morphisms* $R\text{-Mat}(m, n)$ *are given by the* $m \times n$ R*-valued matrices. The composition is given by matrix multiplication.*

- *Let* V *be a poset, the* V *is a category where the objects are given by the element of* V *and* $V(\omega, \omega') = \{(\omega, \omega')\}$ *where* (ω, ω') *is the ordered pair for which* $\omega \leq \omega'$ *otherwise* $V(\omega, \omega') = \emptyset$*. The existence of a (unique) function composing the morphisms is guaranteed the transitivity of the order relation.*

The class of all (small) categories forms a category itself, the morphisms between categories are transformations mapping objects to objects and morphisms to morphisms that preserve the essential structure of a category provided in Definition 1.1, we call such morphisms *functors*.

Definition 1.3. *A functor* F *from a category* $\mathcal{A}$ *to a category* $\mathcal{B}$ *consists of*

1. *a mapping*
$$obj\mathcal{A} \to obj\mathcal{B}$$
the image of $A \in \mathcal{A}$ *will be denoted* FA*.*

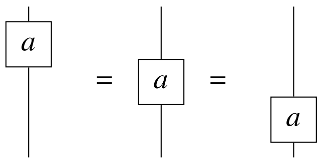
2. *for every pair of objects* A, A' *of* $\mathcal{A}$*, a mapping*
$$\mathcal{A}(A, A') \to \mathcal{B}(FA, FA)$$

subject to the following axioms:

1. *For every pair of morphisms* $f \in \mathcal{A}(A, A'), g \in \mathcal{A}(A', A'')$
$$F(g \circ f) = F(g) \circ F(f);$$

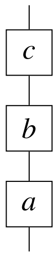
2. *for every object* $A \in \mathcal{A}$
$$F(1_A) = 1_{FA}$$

For the purpose of this thesis, we will use the terms *morphisms* and *processes* interchangeably. Definition 1.1 only requires a well-defined sequential composition and the existence of a special identity morphism. We can graphically depict any category by using boxes to denote morphisms and wires for the identities (with the direction of composition flowing from the bottom to the top of the page). The most general graphical algebra of boxes for a category is not particularly interesting as we are only guaranteed to be able to compose processes sequentially. For example, we can pre-compose and post-compose a process with the identity:



(1.4)

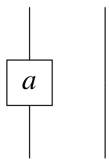
or compose three processes together:



(1.5)

We see that associativity and the properties of identities are absorbed by the topology of the diagrams. At this level of generality, there is no meaning to the notion of parallel interactions of systems. The axioms of a category, when interpreted as a theory of processes, would represent a 'universe' where we can identify a univocal notion of 'time foliation', and for which the possibility of carving out subsystems is fundamentally limited. It is holistic and totally time-ordered. To say that a mathematical structure faithfully represents the possibility of juxtaposing parallel processes, we would at least need to adopt the machinery of a pre-monoidal category that we are about to describe.

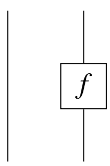
Any notion of disjointness between systems will require that morphisms acting on one of the two parts can cohabit with the identity of the other 'part' so that there is a meaning to describing processes graphically as follows, passing from a one-dimensional to a two-dimensional denotation of processes:



(1.6)

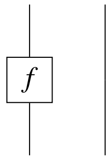
Even though the ability to freely construct such morphisms may seem intuitive, introducing them formally in the framework of categories comes with some subtleties. First we associate to every ordered pair $(A, B) \in \text{obj}\mathcal{A} \times \text{obj}\mathcal{A}$ an object called $A \otimes B$, the whole which includes both parts. We are then required to describe additional morphisms obtained by the actions of the functors $A \rtimes -$ and $- \rtimes A$ on a general $f \in C(A, B)$, which corresponds to the juxtaposition of the identity:

$$A' \rtimes f : A' \otimes A \longrightarrow A' \otimes B \tag{1.7}$$



(1.8)

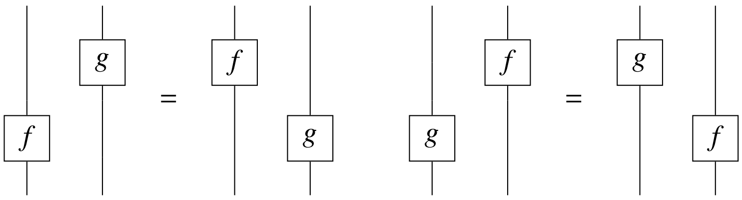
$$f \rtimes A : A \otimes A' \to B \otimes A' \tag{1.9}$$



(1.10)

In this minimal setting, any morphism can be 'trivially extended' to encompass parts of the environment. The *functoriality* requirement guarantees that $(f \circ g) \rtimes A = (f \rtimes A) \circ (g \rtimes A)$, so that juxtaposing the identity wire is always compatible with composition. A category equipped with the two endofunctors (a functor mapping a category to itself) $f \rtimes A$ and $A \rtimes f$ for any object A is a *binoidal category* [114]. In a binoidal category, therefore, there is still no general meaning to the construction of simultaneous evolution from the description of two individual independent transformations. However, there might exist individual processes for which 'simultaneity' becomes meaningful:

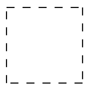
Definition 1.4 (Central morphism)**.** *In a binoidal category $\mathcal{A}$, a morphism $f : A \to B$ is central when for every $g : A' \to B'$ one has that the following diagrams are equal*



(1.11)

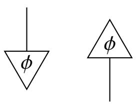
In this case, we denote the equal morphisms on the left by $f \otimes g$ and the equal morphisms on the right $g \otimes f$ and we recover a notion of parallel composition of f and g. To define a *premonoidal category* one only needs to add the following ingredients to a binoidal category:

1. an empty diagram, i.e. a special object I representing the environment, with identity $1_I : I \to I$ graphically denoted as the empty diagram

   (1.12)

   Morphisms from and into the empty diagram are called *states* $I \to A$ and effects $A \to I$ respectively and are usually denoted as

   (1.13)

2. for each triple of objects (A, B, C), a central isomorphism $\alpha_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$ which has the role of unambiguously define products of type $A \otimes B \otimes C$ without worrying about bracketing.

3. for each object A, two central isomorphisms $\lambda_A : A \otimes I \to A$ and $\rho_A : I \otimes A \to A$.

Additionally, we require the following pentagonal diagram to commute:

$$\begin{array}{ccccc} x \otimes (y \otimes (z \otimes w)) & \xrightarrow{\alpha_{x,y,z\otimes w}} & (x \otimes y) \otimes (z \otimes w) & \xrightarrow{\alpha_{x\otimes y,z,w}} & ((x \otimes y) \otimes z) \otimes w \\ {\scriptstyle 1_a \otimes \alpha_{y,z,w}} \downarrow & & & & \uparrow {\scriptstyle \alpha_{a,b,c} \otimes 1_d} \\ x \otimes ((y \otimes z) \otimes w) & & \xrightarrow{\alpha_{x,y\otimes z,w}} & & (x \otimes (y \otimes z)) \otimes w \end{array} \tag{1.14}$$

and the so-called triangle law specifying that α interacts as expected with λ and ρ.

$$\begin{array}{ccc} (x \otimes 1) \otimes y & \xrightarrow{\alpha_{x,1,y}} & x \otimes (1 \otimes y) \\ & {\scriptstyle \rho_x \otimes 1_y} \searrow \quad \swarrow {\scriptstyle \pi_B} & \\ & x \otimes y & \end{array} \tag{1.15}$$

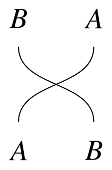
Those properties can be taken to describe the fundamental axioms of a process theory. The pentagon and the triangle equations make sure that the structural morphisms introducing the notion of an empty diagram/trivial system and the introduction of 'spacelike' associativity (similarly to the timelike counterpart, which is essential in the definition of a category) are well behaved. This means that we can interpret morphisms in a premonoidal category as foliated diagrams where α, ρ and λ and the associated coherency laws are absorbed by the topology of the diagram.

Observation 1.4. *The central morphisms $Z(C)$ of a premonoidal category C form a category.*

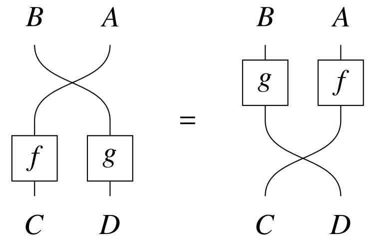
Definition 1.5 (Monoidal Category)**.** *We say that C is monoidal precisely when C coincides with its centre $Z(C)$.*

A monoidal category is a premonoidal category where boxes are always allowed to slide past each other in the sense of Proposition 1.4.

In monoidal categories, there is no guarantee that $A \otimes B$ and $B \otimes A$ are equal or even isomorphic. There is, therefore, a spacial asymmetry in the definition of joint systems which, for many applications, results redundant. For this reason, SMCs or symmetric monoidal categories are equipped with an additional natural isomorphism denoted



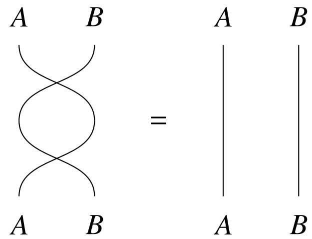
naturality can be described graphically by allowing the boxes to slide through the braid:



Definition 1.6 (Symmetric Monoidal Category)**.** *A symmetric monoidal category is a monoidal category equipped with a natural family of isomorphisms* $B_{x,y} : x \otimes y \to y \otimes x$ *such that*

$$B_{x,y} = B^{-1}_{y,x}$$

so that we can disentangle the double application of the braid:



Typically the categorical operational approaches to quantum processes make use of symmetric monoidal categories: systems are associated with Hilbert spaces, and non-interacting tensor factors evolve independently. A notable exception is the use of premonoidal categories by Richard Blute and Marc Comeau [24] with the aim of abstracting categorical quantum mechanics a la Abramsky and Coecke [7] to include relativistic effects in the style of Algebraic Quantum Field Theory (AQFT). In the field theoretic description, the algebras of operators associated with U and V will commute only if U and V are spacelike separated regions. More recently, and also in the context of modelling spacetime, premonoidal categories have been used by Hefford and Kissinger in [73] to construct a category of spacetime regions and future-directed causal curves.

The treatment of physical processes using quantum circuits or SMCs is based on the assumption that we can always carve out different copies of a physical system which evolve independently. For a study of how an operational theory (a partial SMC) arises by carving out subsystems from global symmetries, we direct the reader to [60].

Example 1.7 (Symmetric monoidal categories)**.** *Symmetric monoidal categories are a very ubiquitous and general object. We now provide a couple of paradigmatic examples which are used in the context of categorical quantum mechanics:*

1. *The category* Set *is a monoidal category where the monoidal product is given by the cartesian product of sets* $A \times B$ *and the tensor unit is given by the singleton set. The parallel composition*

of morphisms $f : A \to B$, $g : C \to D$ is defined as:

$$(f \times g)(x, y) = (f(x), g(x))$$

where $(x, y) \in A \times C$ and $(f(x), g(x)) \in B \times D$. The associator $\alpha_{A,B,C}$ is given by the function

$$(a, (b, c)) \in A \times (B \times C) \mapsto ((a, b), c) \in (A \times B) \times C$$

The left and right unitors λ_A and ρ_A are given by:

$$\rho_A \left[(a, *) \in A \otimes \{*\}\right] \mapsto a \in A$$

and

$$\lambda_A \left[(*, a) \in \{*\} \otimes A\right] \mapsto a \in A$$

2. *The category R-Mat is a monoidal category, where objects are given by R^n, and morphisms $R\text{-Mat}(R^n, R^m)$ are given by $n \times m$ matrices with entries in R. The product of matrices gives the sequential composition of morphisms. The tensor product is given by the Kronecker product $\otimes$. For $R = \mathbb{C}$ we recover a category equivalent to* fHilb*, the category of finite dimensional Hilbert spaces. The tensor unit is given by underlying semiring R*

3. *Finite probability distributions and stochastic maps form a category used to model non-deterministic processes. The objects are given by finite sets, and* $\text{fStoch}(U, V)$ *are given by matrices $U \times V$ values in $\mathbb{R}^+$ such that each column sums to* 1*. The usual Kronecker product of matrices gives the tensor product. A state $p : \{*\} \to A$ is a probability distribution over A.*

1.3.1.2 Axiomatising Quantum-classical Interaction

The spirit of categorical quantum mechanics is to show that much of the logic inherent in the interaction of quantum processes can be recast as abstract structural properties related to their spacelike and timelike composition. The focus is usually on categorical axiomatisation rather than specific models. This notwithstanding, several categories appear in the study of categorical quantum theory; some reflect different treatments of classical systems. The category fHilb of finite dimensional Hilbert spaces and linear maps (introduced in the context of finite dimensional quantum theory in the seminal [7, 8]) is prototypical but does not possess a typing structure which is expressive enough to deal explicitly with classical systems. To do so, one must pass to a category where the objects are mixed states and evolutions given by general quantum channels. Selinger developed a formal categorical bridge between a theory describing the evolution of pure systems and the introduction of mixtures by introducing the abstract CPM construction [129].

Applying the CPM construction to the category of Hilbert spaces gives us CPM[fHilb], where the objects are Hilbert spaces and the morphisms are given by completely positive maps of the form $f : A^* \otimes A \to B^* \otimes B$. A step towards introducing classicality, but an explicit introduction of classical systems is still missing. Classical systems can be recovered through the $\mathrm{CP}^*[]$-construction described by Coecke, Heunen and Kissinger in [42]. Applying the $\mathrm{CP}^*[]$-construction to fHilb gives $\mathrm{CP}^*[\mathrm{fHilb}]$, the category of C^*-algebra and completely positive maps. Both CPM[fHilb] and fStoch embed fully and faithfully in $\mathrm{CP}^*[\mathrm{fHilb}]$.

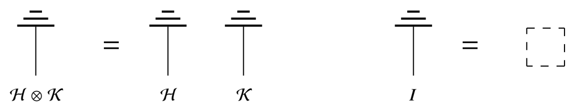
In the CP^*-construction, the focus is on describing quantum observable and classical systems in an elegant algebraic way, but an explicit axiomatisation of the quantum-classical interface is still missing. Referring to this interface is notably important in formulating *operational probabilistic theories* (OPT), introduced by [38, 47]. In OPTs, the probabilistic structure is explicitly axiomatised, but classical systems are *not* treated compositionally as part of the theory. The formulation of probabilistic theories by Gogioso and Scandolo [66] merges the compositional treatment of classical systems with the explicit axiomatisation of probabilistic structure from OPTs.

Definition 1.8. *A* ***probabilistic theory*** *is a (strict) symmetric monoidal category (SMC) C which satisfies the following requirements:*

- *there is a full sub-SMC of C, denoted by C_K, which is equivalent to the SMC* $\mathbb{R}^+$-Mat *modelling* ***classical theory*** *(itself a probabilistic theory);*

- *the SMC C is enriched in commutative monoids, and the induced enrichment on C_K coincides with the one given by the linear structure of* $\mathbb{R}^+$-Mat*;*

- *the SMC C comes with an environment structure, i.e. with a family of effects $\top_A : A \to I$ which satisfy the following requirements:*

  (1.16)

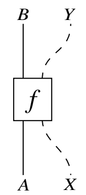
  The environment structure induced on C_K coincides with the one given by marginalisation in $\mathbb{R}^+$-Mat*. Marginalisation is here understood as the linear map sending a row vector to the sum of its coordinate components, i.e the column vector with unit entries:*

$$\begin{pmatrix} a_0 & a_1 & \dots & a_n \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=0}^{n} a_i$$

The requirements above imply that processes $A \to B$ in C have the structure of a convex cone*, i.e. they are $\mathbb{R}^+$-modules. The effects $\top_A : A \to I$ are known as* ***discarding maps****. The systems in the full sub-category C_K are known as* ***classical systems*** *and the processes between them as* ***classical processes****.*

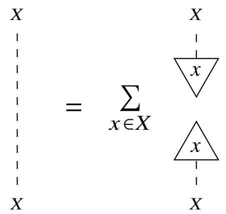
The probabilistic theory most relevant to this work is *quantum theory*, defined by taking completely positive maps together with classical theory and introducing the quantum-classical interface by linearity from families of quantum processes (cf. resolution of the classical identity below).

From a diagrammatic perspective, dashed wires are used to denote systems which are guaranteed to be classical, while solid wires denote generic systems. It is convenient to assume the following general form for processes in probabilistic theories, with distinguished classical input and output systems (finite sets X and Y respectively):



(1.17)

The linear structure of $\mathbb{R}^+$ -Mat can be used to perform resolutions of the classical identity explicitly:



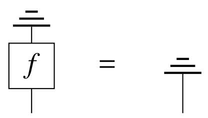
(1.18)

The resolution of the classical identity can be used to equate the following two perspectives, establishing a direct link to the OPT formalism and the empirical models described in [6]:

- processes $F : A \otimes X \to B \otimes Y$, with X and Y classical systems (i.e. finite sets);
- families $F(y|x) : A \to B$ of processes indexed by the classical values $x \in X$ and $y \in Y$.

The discarding maps axiomatise the notion of marginalisation (aka *partial trace* in the context of quantum systems). They can also be used to define a sub-SMC of *normalised* processes:

Definition 1.9. *A process $f : A \to B$ is said to be* ***normalised*** *if it satisfies the following equation:*

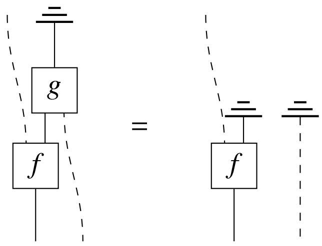


(1.19)

A process $f : A \to B$ is said to be ***sub-normalised*** *if there exists some $g : A \to B$ such that $f + g$ is normalised (in which case g is also sub-normalised).*

In particular, the normalised states on a classical system X are the probability distributions on X, the normalised processes $X \to Y$ are the Y-by-X stochastic matrices, and discarding on a classical output of a classical process is the same as marginalisation.

The axiomatisation of causality as ‘no-signalling from the future’ [47] is embodied in probabilistic theories by the following observation about normalised processes:

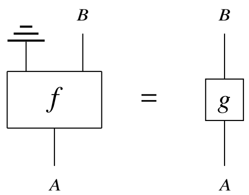


(1.20)

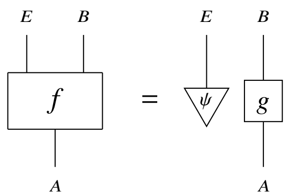
In the above, we see that the classical outcome of a test f cannot be influenced by a controlled process g in its future (i.e. it is independent of the classical input used to control g).

The framework also adopts a notion of *purity*, defined as the lack of non-trivial interaction with a discarded environment.

Definition 1.10. *A process $g : A \to B$ is said to be* ***pure*** *if whenever we can find a system E and a process $f : A \to E \otimes B$ such that the following equality holds:*



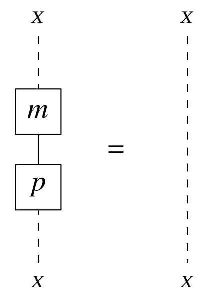
Then there exists a normalised state $\psi : I \to E$, dependent on f, such that the following equality also holds:



Note that neither g nor f are required to be normalised as part of this definition.

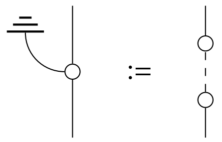
In *quantum theory*, a process is pure and normalised if and only if it is an isometry. Finally, we recall the definition of a sharp preparation-observation pair, capturing the idea of perfect encoding/decoding of classical information in arbitrary systems.

Definition 1.11. *A* ***sharp preparation-observation pair*** *(SPO pair) is a pair (p, m) of a* ***preparation*** *process $p : X \to H$ and an* ***observation*** *process $m : H \to X$ on some classical system X and some* *arbitrary system H, such that the following equation holds:*



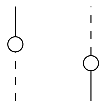
(1.21)

If (p, m) is a sharp preparation-observation pair, we can construct its associated decoherence map:



(1.22)

where the symbols:



(1.23)

are respectively used to denote the preparation and measurement process.

1.3.2 Introduction to the Sheaf Theoretic approach

1.3.2.1 Theoretical Minimum

The sheaf theoretic approach describes contextuality as the impossibility of finding a global assignment of values that accounts for observed conditional distributions. Putting aside (for now) the mathematical formalism that lies at its core, the sheaf theoretic formalism can be intuitively introduced by using bundle diagrams [4] which on their own give a particularly intuitive and insightful account of the logical structure of contextuality. In this section, we briefly introduce the framework developed in [6] starting by building an intuition using bundle diagrams and then introducing the relevant rigorous and general description. Understanding the formalism will be important to grasp the content of this dissertation. The initial aim of the theory developed in [6] was to build a conceptual bridge between causality and contextuality, and we will show that this bridge can be extended to a complete unification with causality.

The analysis of the importance of measurement contexts has been ignited by the celebrated EPR article [54], which claimed that it is possible to measure two complementary observables on an entangled pair of particles and, by doing so, violate the theoretical impossibility of assigning definite values to conjugate observables. The reply by Bohr exposed, for the first time, the importance of measurements contexts. The EPR result does not witness the incompleteness of the newly born theory but reveals a much deeper aspect of its connection with the observable reality. Specifically, Bohr believed that ‘the procedure of measurements has an essential influence on the conditions on which the very definition of the physical quantities’ [25] so that we can assign elements of ‘physical reality’ only conditioned on the global contexts in which they emerge. Believing that an observable can have a well-defined value for contexts in which its measurement excluded a priori leads to incongruence. It is believing that both a value for the momentum and the position of a particle is allowed to exist independently of the context of their realisation that exposes this fallacy. The EPR article is not yet exposing non-locality or contextuality, which strengthens the logical dependence on contexts and will require decades of maturity to be appropriately understood, but it already showcases one of its fundamental ingredients.

Consider the stereotypical quantum scenario where Alice and Bob are two spacelike separated agents, each one with a binary choice of local measurements. The measurements $\{a_0, a_1\}$ are assigned to Alice, and similarly $\{b_0, b_1\}$ is the set of choices for Bob. The measurements are dichotomic so that we can—without loss of generality—associate to each the set of outcome values $\{0, 1\}$. The agents perform several rounds of some protocol before meeting up to collect the global measurement statistics. In each round, one of the four possible contexts $\{(a_0, b_0), (a_0, b_1), (a_1, b_0), (a_1, b_1)\}$ is jointly selected by both parties.

Suppose that the agents find their outcomes to be perfectly anti-correlated when both of them choose to perform the measurement 0 and perfectly correlated otherwise, as in Figure 1.1 (p.19).

The first question that we might reasonably ask is whether this empirical behaviour is compatible with a notion of relativistic causality. Fortunately, it is easy to see, by marginalising the outcomes of one of the two agents, that the table of conditional probabilities in Figure 1.1 (p.19) does not allow any type of superluminal signalling. In other words, spacelike separated agents cannot infer what global context has been selected from local observations.

AB	00	01	10	11
a_0b_0	0	1/2	1/2	0
a_0b_1	1/2	0	0	1/2
a_1b_0	1/2	0	0	1/2
a_1b_1	1/2	0	0	1/2

Figure 1.1: We can represent the observed empirical correlations using a table of conditional probabilities. The rows are indexed by the possible contexts, and the columns by the joint outcome assignments.

Despite its apparent innocuousness and the adherence to the no-signalling principles, the support of the correlations in Figure 1.1 (p.19) hides an interesting topological structure which makes its non-classicality glaring. We start by representing the local measurements as points connected with edges (or hyperedges when the size of a measurement context is greater than two) when they lie in the same context, i.e. when they can be considered jointly measurable (See Figure 1.2 (p.19)).

We refer to this description of measurements and the associated contexts as the *base* of a bundle diagram. Each measurement $i \in M$ is supplemented with a set O_i describing the possible measurement outcomes, this is represented by assigning to each element of the base a *stalk of points* as shown in Figure 1.3 (p.20).

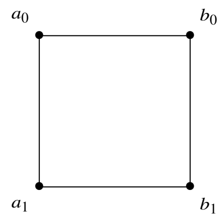


Figure 1.2: Four measurements a_0, a_1, b_0, b_1 are distributed into contexts denoted by the connecting edges.

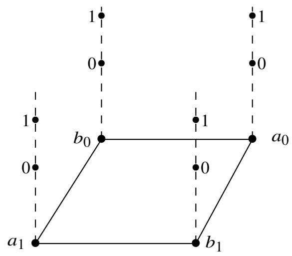


Figure 1.3: The base of the bundle diagram represents the measurements grouped by contexts, the dotted lines denote the stalks, i.e. the sets of possible local outcomes.

The stalks describe the local outcomes. A joint outcome for each context is an edge connecting two local values, one for each measurement in the context. This is graphically denoted by an edge (or again a hyperedge for larger contexts) as in Figure 1.4 (p.20)).

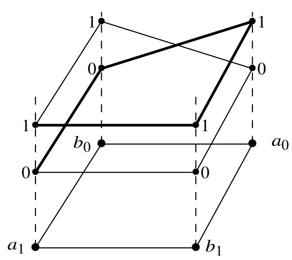


Figure 1.4: Possibilistic structure of the empirical correlations described in Figure 1.1

Starting from the correlations in Figure 1.4 (p.20) and reasoning 'across the contexts', with the underlying assumption that there always exist hidden values for unperformed counterfactual measurements, leads to a contradiction. If Alice observes 1 after having performed the measurement

a_0, we can draw, based on the knowledge of the possibilistic structure of the conditional probability table, the following chain of logical implications:

1. $a_1 \mapsto 1$ implies that $b_1 \mapsto 1$
2. $b_1 \mapsto 1$ implies that $a_1 \mapsto 1$
3. $a_1 \mapsto 1$ implies that $b_0 \mapsto 0$
4. $b_0 \mapsto 0$ implies that $a_1 \mapsto 0$

These take the form of the path shown in Figure 1.4 (p.20): starting with a given outcome value for a measurement in a context and traversing the unperformed measurements, we deduce a contradictory valued for a performed measurement. Notice that in this particular scenario, an analogous argument can be drawn by starting from the outcome of any other measurement and choice of context.

A *deterministic hidden variable* is a function which represents a value assignment to *all* measurements simultaneously and independently on the context. A mechanism that—conditioned on the past information common to both measurement sites—assigns a definite joint value to all the 4 possible measurements. Deterministic hidden variables would look look like surfaces in our bundle diagram, connecting a single value for every measurement (see Figure 1.5 (p.21)). A table of conditional probabilities has a realisation in terms of deterministic hidden variable theories exactly when it can be reproduced by taking a probabilistic mixture of these deterministic assignments. Here probabilistic mixtures allow for the possibility that the determination of the value is correlated to some latent variable.

We note that this description of hidden variables assumes that the assignment cannot depend on the local choice performed to select the context, the deterministic outcome assigned by the hidden mechanisms is statistically independent of the choice of measurement context.

From the bundle diagram in Figure 1.4 (p.20), it is clear that no joint outcome can occur as a deterministic assignment of values to all measurements. This particular no-signalling protocol is known in the literature as the Popescu-Rohrilch box (PR box) [112] and exhibits the strongest possible version of non-locality.

To give a simple example of correlations which are *local*—classically realisable—and for which a hidden variable model exists, we can consider the case where the outcomes are always perfectly correlated with 00 and 11 occurring with the same probability.

This gives rise to the bundle diagram presented in Figure 1.7 (p.22), which has a simple description in terms of deterministic hidden variables by assigning to the measurements (a_0, a_1, b_0, b_1) the outcomes $(1, 1, 1, 1)$ or $(0, 0, 0, 0)$ with equal probability.

We have described the way non-locality is intuitively treated in Abramsky and Brandenburger's framework. To provide a more formal characterisation, we introduce the language of *presheaves* and

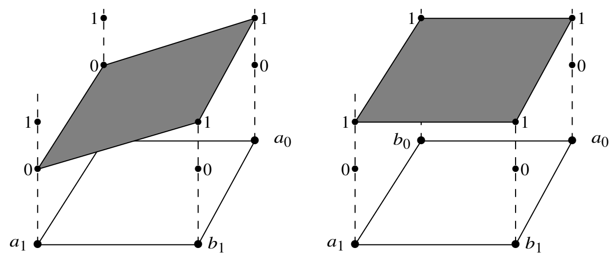


Figure 1.5: Examples of deterministic assignments of values for all measurements. The picture on the left represents the assignment $(a_0, b_0, a_1, b_1) \mapsto (0, 0, 1, 1)$ and the picture on the right $(a_0, b_0, a_1, b_1) \mapsto (1, 1, 1, 1)$

AB	00	01	10	11
a_0b_0	1/2	0	0	1/2
a_0b_1	1/2	0	0	1/2
a_1b_0	1/2	0	0	1/2
a_1b_1	1/2	0	0	1/2

Figure 1.6: The measurements are always correlated independently of the choice of context.

sheaves, of which the usefulness lies in particular in its mathematical generality, a feature that will be exploited in this thesis to include the discussion about causal structures. The tension between local and global assignments of values is a recurring phenomenon, and this toolbox provides the perfect mathematical grammar to describe it.

The first fundamental ingredient is the notion of a *presheaf*, a (contravariant) set-valued functor: concrete data assigned to the objects of some category C. Although a presheaf can be described over arbitrary categories, we are here interested in their application to topological spaces:

Definition 1.12 (Presheaves)**.** *Let X be a topological space and let $\mathcal{T}(X) \subseteq \mathcal{P}(X)$ be its collection of open sets, which we also refer to as* contexts. *A (set-valued)* presheaf *P on X is an association of:*

- *a set $P(U)$ to each $U \in \mathcal{T}(X)$, specifying the possible values for* contextual data *on U;*
- *a* restriction *$P(U, V) : P(U) \to P(V)$ for each open set U and each open subset $V \subseteq U$, restricting contextual data on U to corresponding contextual data on V.*

The restrictions are required to satisfy the following conditions:

1. *$P(U, U) = id_{P(U)}$, i.e. the trivial restriction from U to U is the identity on $P(U)$;*
2. *$P(V, W) \circ P(U, V) = P(U, W)$, i.e. restrictions are stable under function composition.*

In more categorical terms, a presheaf is a functor

$$F : \mathcal{T}(X) \to \text{Set}^{\text{op}} \tag{1.24}$$

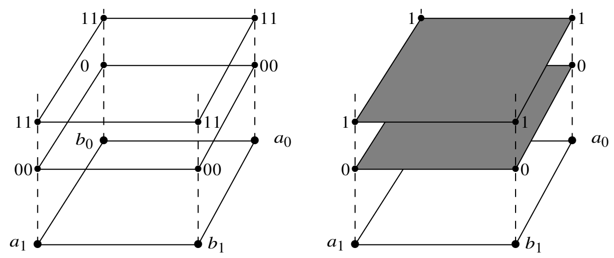


Figure 1.7: Bundle diagram for the conditional distribution in Figure. On the right, we have a choice of hidden variables realising the perfect correlation.

where the partial order $(\mathcal{T}(X), \subseteq)$ is seen as a posetal category of opens associated with the topological space X. When the presheaf is clear from the context, we will denote the restriction of contextual data $a \in P(U)$ to some open subsets $V \subseteq U$ as follows:

$$a|_V := P(U, V)(a)$$

The two restriction conditions can then be rewritten as follows, for all $a \in P(U)$

$$\text{(i)} \quad a|_U = a \qquad\qquad \text{(ii)} \quad (a|_V)|_W = a|_W \tag{1.25}$$

We denote a choice of some element assigned to the open sets as follows:

1. given an open set $U \in \mathcal{O}(X)$ we call the elements of $F(U)$ the *sections* at U
2. the elements of $F(X)$ are the *global sections* of the presheaf.

A compatible family for a family of objects $\{C_i\}_{i \in I}$ is then a choice of sections which agrees on the 'intersections':

Definition 1.13 (Compatible Family)**.** *Let X be a topological space and let P be a presheaf on X. Let $\mathcal{U} \subseteq \mathcal{T}(X)$ be a family of open sets in X and let $a = (a_U)_{U \in \mathcal{U}}$ be a family specifying contextual data on the open sets in $\mathcal{U}$. We say that a is a* compatible family *(for P) is for every $U, U' \in \mathcal{U}$ we have:*

$$(a_U)|_{U \cap U'} = (a_{U'})|_{U \cap U'} \tag{1.26}$$

We say that P is a separated *presheaf if every compatible family has at most one gluing. At most one element $e \in \bigvee_{U \in \mathcal{U}} U$ such that $e|_U = a_U$. We say that P is a* sheaf *is every compatible family has exactly one gluing.*

A presheaf can be thought of as an assignment of local data to the open set of a topological space. We are not a priori requiring the data to be globally consistent, i.e. that pairwise compatible local sections can always be glued in a consistent way. When consistent gluing is always well defined and unique, the presheaf forms a *sheaf*.

The definition of presheaves would not be complete if we did not explain what it means for different assignments of data to be isomorphic, i.e. what it means to assign data in an essentially equivalent way:

Definition 1.14. *Let X be a topological space and let $\mathcal{T}(X) \subseteq \mathcal{P}(X)$ be its collection of open sets. We day that two presheaves P and P' on X are* naturally isomorphic*, written $P \simeq P'$, if there is a family $\phi = (\phi_U)_{U \in \mathcal{T}}$ of bijections $\phi_U : P(U) \to P'(U)$ such that for all inclusions $V \leq U$ of open sets we have:*

$$\phi_V \circ P(U,V) = P'(U,V) \circ \phi_U \tag{1.27}$$

The commutation condition can be written more succinctly as follows, for all $a \in P(U)$:

$$\phi_V(a|_V) = \phi_U(a)|_V \tag{1.28}$$

If we wish to specify a specific natural isomorphism *ϕ, we can also write $\phi : P \simeq P'$.*

An implicit assumption in Abramsky and Brandenburger is to endow the set of measurements X with the discrete topology. In general, what makes the sheaf theoretic machinery work is the fact that in an arbitrary topological space X, the set of opens $\mathcal{T}(X)$ has the structure of a 'locale', which for the specific case of [6] is the locale $\mathcal{P}(X)$ of all subsets of X ordered by inclusion. The general definition of sheaves (although not the most general) can be given by replacing the opens of a topological space with one of these abstract order-theoretic structures:

Definition 1.15. *A* locale *is a partially ordered set $(\mathcal{L}, \subseteq)$ satisfying the following properties:*

1. *$(\mathcal{L}, \leq)$ has all finite meets, i.e. for all finite $F \subseteq \mathcal{L}$ there is a $\bigwedge F \in \mathcal{L}$ such that:*
   - *$f \geq \bigwedge F$ for all $f \in F$*
   - *$f \geq g$ for all $f \in F$ then $\bigwedge F \geq g$*
2. *$(\mathcal{L}, \leq)$ has all joins, i.r. for all $F \subseteq \mathcal{L}$ there is a $\bigvee F \in \mathcal{L}$ such that:*
   - *$f \leq \bigvee F$ for all $f \in F$*
   - *if $f \leq g$ for all $f \in F$ then $\bigvee F \leq g$*

satisfying the infinite distributive law:

$$x \vee \left(\bigvee_i y_i \right) = \bigvee_i (x \vee y_i)$$

For a topological space, the axioms for the topology $\mathcal{T}(X) \subseteq \mathcal{P}(X)$ are equivalent to asking that $(\mathcal{T}(X, \subseteq)$ is a locale. A locale describes the minimal abstract axiomatisation of ordered objects that can be glued together arbitrarily and of which we can take finitely many intersections; we do not require them to arise explicitly as open sets of some topological space.

Going back to empirical data, the fundamental structure is the notion of a *measurement scenario*, a tuple $\Sigma = (X, \mathcal{M}, \{O_x\}_{x\in X})$ composed of a set of *measurements*, a measurement cover $\mathcal{M}$, and a set of outcomes $\underline{O} = \{O_x\}_{x\in X}$ parametrised by the inputs.

A measurement cover $\mathcal{M}$ is given by a family of subsets of X satisfying:

- $\bigcup \mathcal{M} = X$.
- $\mathcal{M}$ is an *anti-chain*, i.e. for $U, V \in \mathcal{M}$, $U \subseteq V$ implies that $U = V$.

Here is where the implicit choice of topology on the set X becomes relevant. The notion of a measurement cover that can be found in [6] is the particular case of the more general notion of *open cover* for a general topological space $\mathcal{T}(X)$.

Definition 1.16. *Let X be a topological space and $\mathcal{T}(X) \subseteq \mathcal{P}(X)$ be its topology. An* open cover*, or simply* cover*, for X is an antichain in the partial order $\mathcal{T}(X)$, i.e. a collection $C \subseteq \mathcal{T}(X)$ of open sets which are incomparable:*

$$\forall U, V \in C. V \leq U \Rightarrow V = U$$

and such that:

$$\bigvee_{U\in C} U = X$$

If C and C' are covers on X, we say that C' is finer than C, written $C' \prec C$, if the following holds:

$$C' \prec C \Leftrightarrow \forall V \in C'. \exists U \in C. \ s.t. \ V \subseteq U$$

Equivalently, we sat that C is coarser *than C'. Note that $\prec$ is a partial order on covers for X, known as the* refinement order.

Given two covers C' and C such that $C' \prec C$, a compatible family for the finer cover automatically induces a compatible family on the coarser. In particular global assignments for $\mathcal{M} = \{X\}$ restrict to compatible families for arbitrary measurement covers.

Observation 1.16. *Let X be a topological space and let P be a presheaf on X. Let C be a cover for X and $a = (a_U)_{U\in C}$ be a compatible family over C. If $C' \prec C$ is a finer cover for X, then the following is a compatible family over C', known as the* restriction *of a to C':*

$$a|_{C'} := (a_{U_V}|_V)_{V\in C'}$$

where $U_V \in C$ is any open such that $V \subseteq U_V$.

The presheaf, which is relevant for the study of contextuality and nonlocality, can be defined in two stages. First, to each subset U of X, we associate a set of sections $\mathcal{E}(U) = O^U$ representing the possible joint outcomes for the set of events U. The map $\mathcal{E}$ is a functor $\mathcal{E} : \mathcal{P}(X) \to \mathrm{Set}$, and therefore a *presheaf*: for any $U, U' \in X$ such that $U \subseteq U'$ we can associate to it a restriction map $\mathcal{E}(U \subseteq U') = \text{res}_U^{U'}$

$$\text{res}_U^{U'} : \mathcal{E}(U') \to \mathcal{E}(U) :: s \to s|U \tag{1.29}$$

Locally compatible assignment of outcomes can always univocally be glued into a coherent whole, making $\mathcal{E}$ a sheaf.

In a purely deterministic theory, it is unreasonable to expect any tension between *local* and *global* assignments of data; the presence of fundamental indeterminism gives the theory the necessary flexibility to exhibit such discrepancy. Compatible partial functions can always be glued together univocally. [2]

Definition 1.17 (R-distribution [6]). *Let R be a* commutative semiring, *an R-distribution on X is a function $d : X \to R$ which has finite support such that*

$$\sum_{x \in X} d(x) = 1 \tag{1.30}$$

We can construct a functor $\mathcal{D}_R : \mathcal{P}(X)^{\text{op}} \to \text{Set}$ assigning to a subset U the set $\mathcal{D}_R(U)$ of R-distribution on U.

Definition 1.18. *The* distribution monad $\mathcal{D}_R$ *is the following mapping on sets and functions:*

- *If X is a set, $\mathcal{D}(X)_R$ is the set of probability distributions over X with finite support:*

$$\mathcal{D}(X) := \left\{ d : X \to R \;\middle|\; \sum_{x \in X} d(x) = 1, \textit{supp}(d) \textit{ is finite} \right\} \tag{1.31}$$

  where the support of a distribution is the set of points over which it is non-zero:

$$\textit{supp}(d) := \{ x \in X \mid d(x) \neq 0 \} \tag{1.32}$$

- *If $f : X \to Y$ is a function between sets, $\mathcal{D}(f)_R$ is the function $\mathcal{D}(X)_R \to \mathcal{D}(Y)_R$ defined as the linear extension of f to probability distributions with finite support:*

$$\mathcal{D}(f) := d \mapsto \sum_{x \in X} d(x) \delta_{f(x)} \tag{1.33}$$

  where $\delta_y \in \mathcal{D}(Y)$ is the delta distribution at y:

$$\delta_y := y' \mapsto \begin{cases} 1 \textit{ if } y' = y \\ 0 \textit{ otherwise} \end{cases} \tag{1.34}$$

[2] This will cease to be true for an entire class of causal assumptions that we introduce in Chapter 3 and Chapter 4 in which even the deterministic causal data will at times require a type of global compatibility which makes the presheaf of 'joint outcomes' separable but not a sheaf.

For $f : U \to V$, $\mathcal{D}_R(f)$ is defined as

$$\mathcal{D}_R(f) : \mathcal{D}_R(U) \to \mathcal{D}_R(V) \tag{1.35}$$

$$d \mapsto [y \mapsto \sum_{f(x)=y} d(x)] \tag{1.36}$$

Remark 1.19. *The term 'monad' comes from category theory, where it defines a functor with specific additional structure. We do not need this additional structure in our work, but we have preserved the name for compatibility with other sheaf-theoretic work.*

The distribution monad is an endofunctor of the type Set $\to$ Set, the composition $\mathcal{D}\,(D)_R\,\mathcal{E}$ describes a presheaf $\mathcal{T}(X) \to$ Set. The functor is not a sheaf, and the failure to extend local compatible families into global sections constitutes the essence of contextuality. An empirical model is then an assignment of empirical behaviours to each set in the measurement cover: a compatible family over $\mathcal{M}$:

Definition 1.20 (Empirical model)**.** *An empirical model for the measurement scenario $(X, \mathcal{M}, O)$ is a compatible family for $\mathcal{D}\,(D)_R\,\mathcal{E} : \mathcal{P}(X) \to$ Set over the cover $\mathcal{M}$.*

In non-locality scenarios, spacelike separability induces a particular measurement cover. The set of measurements is given by:

$$X = \coprod_{s \in S} I_s \tag{1.37}$$

where S represents the set of spacelike separated sites and I_s is the set of local measurements available at $s \in S$. The measurement cover by sets identifying a single choice of measurement for each site (See for example [88]):

$$\mathcal{M} = \prod_{s \in S} I_s \tag{1.38}$$

The compatibility with respect to the aforementioned cover $\mathcal{M}$ encompasses the usual no-signalling requirement:

Observation 1.20. *Suppose that $(e_M)_{M \in \mathcal{M}}$ is a compatible family for the cover for non-locality $\mathcal{M}$ and fix two arbitrary disjoint subsets $S_1, S_2 \subseteq S$. Compatibility implies that the choice of measurement at S_2 cannot influence the distribution of outcomes at S_1. Take $M, M' \in \mathcal{M}$ such that, $M|_{S \setminus S_2} = M'|_{S \setminus S_2}$. Set $C := M|_{S \setminus S_2}$. Compatibility implies that for all $t \in \mathcal{E}(C)$:*

$$e_M|_C\,(t) = e'_M|_C\,(t) \implies \sum_{\substack{t' \in \mathcal{E}(M) \\ t'|_C = t}} e_M(t') = \sum_{\substack{t' \in \mathcal{E}(M') \\ t'|_C = t}} e_{M'}(t') \tag{1.39}$$

no choice performed at S_2 can influence the 'local' outcomes at S_1. The compatible families for this scenario coincide with conditional distributions satisfying the multipartite no-signalling principle.

The table of Figure 1.1 (p.19) describes a compatible family for the scenario where $S = \{A, B\}$ and $I_A = \{a_0, a_1\}$, $I_B = \{b_0, b_1\}$. This compatible family does not arise by marginalising a global section.

In the Abramsky and Brandenburger approach, the measurement contexts are allowed to be arbitrary subsets of X, and the choice of the discrete topology guarantees that all the possible restrictions have to be 'operationally accessible': the distributions defined on $U \in \mathcal{M}$ need to have the property that for every $V \subseteq U$ the marginalisation from the bigger to the smaller set is always well defined. Generalising the framework to arbitrary causal structures requires us to transcend this limitation by letting us play with the underlying topology imposed on a set of measurements X. We will see in Chapter 4 that the choice of a particular topological space can be used to reflect a signalling structure on the measurements.

The global sections can be thought of as context-independent assignments of outcomes to every measurement. The functor $\mathcal{D}_R\mathcal{E}(X)$ describes the R-distributions of this local assignment. We call such assignments *deterministic hidden variable assignment (DHV)*. Artur Fine, in his work 'Hidden variables, joint probability, and the bell inequalities' [55] proves the following theorem in the context of nonlocality scenarios:

Theorem 1.21 (Fine's theorem [55])**.** *There exists a factorizable stochastic hidden-variable model for a correlation experiment if and only if there exists a deterministic hidden-variables model for the experiment.*

One of the salient aspects of the mathematical generality achieved by the sheaf theoretic approach is that it can be used to extend the equivalence of DHV assignments and factorisability beyond non-locality. We know that DHVs are just global sections $e \in \mathcal{D}_{\mathbb{R}}\mathcal{E}(X)$. Factorisable hidden variable models (such as the one implied by Bell's local realism) can be recast in the sheaf theoretic language as follows:

Definition 1.22 (Hidden variable model [6])**.** *Given a measurement cover $\mathcal{M}$ a stochastic hidden variable model is given by a set Λ and a distribution $h_\Lambda \in \mathcal{D}_R(\Lambda)$. Each value of the hidden variable describes a compatible family $\{h_C^\lambda\}_{C \in \mathcal{M}}$.* [3]

A hidden variable model realises an empirical model e if the conditional distribution e can be obtained by averaging over the value of the hidden variable:

$$e_C(s) = \sum_{\lambda \in \Lambda} h_C^\lambda(s) h_\Lambda(\lambda) \tag{1.40}$$

This definition may seem strictly more general then DHVs, however, Abramsky and Brandenburger show that finding a realisation in terms of stochastic hidden variables is, in general, equivalent to e

[3]Compatibility here is equivalent to the assumption that the knowledge of λ cannot be used to infer any information about the 'global' context given some 'local' measurement outcomes.

having a global section. When restricting to the case where the set of measurements is of the form of Equation 1.37, and the set of contexts is given by Equation 1.38, Abramsky-Brandenburger's theorem restricts to Fine's.

The sheaf theoretic framework can be used to recast and unify several paramount results in the study of non-locality and contextuality. Abramsky and Brandenburger show that substituting the semiring of the positive reals with the ring of reals, thus admitting a quasi-probabilistic description of the outcomes assigned to each context, is enough to equiparate no-signalling empirical model with global sections: the linear spaces generated by the non-contextual models and the no-signalling models coincide.

The sheaf theoretic analysis can be used to go beyond a merely qualitative understanding of the phenomenon of contextuality and provides a valuable characterisation of a variety of different contextual behaviours. Abramsky and Brandeburger distinguish a hierarchy of contextuality: *strong contextuality*, the strongest form in which no joint outcome in a context can be extended by a global section (e.g. PR boxes), *possibilistic non-locality*, i.e. models which have no global section when the probabilities of the empirical model are substituted by the values of the boolean semiring representing possibilities (e.g. see Hardy [71]), and the usual *probabilistic non-locality* exemplified by the Bell theorem.

In a paper by Abramsky, Barbosa and Mansfield [5], the authors also introduce a measure which can be used to quantify the degree of non-classicality of a given empirical model:

Definition 1.23 (Non-contextual Fraction [5])**.** *Given two empirical model e and e' on a measurement scenario $S = (X, \mathcal{M}, \underline{O})$ and $\lambda \in [0, 1]$ the convex sum $\lambda e + (1 - \lambda)e'$ forms a well defined empirical model over S. For an empirical model e, we can find*

$$e = \lambda e^{NC} + (1 - \lambda)e^{C}$$

such that e^{NC} is non-contextual and e^{C} is contextual. The maximal value of λ in such decompositions id the non-contextual fraction *of e.*

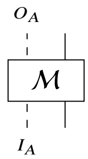
In Chapter 6 of this dissertation, we will directly compare these quantitative measures with the output of our causal analysis, the causal fractions. We will showcase the correspondence between contextuality and indefinite causality when we endow quantum theory with the possibility of controlling the causal order of the application of the instruments.

1.3.2.2 Extending the framework?

The existing framework for contextuality and non-locality is limited to spacelike separated protocols and Kochen-Specker type of contextuality arguments. Abramsky and Brandenburger implicitly assume that any subset of a maximal measurement context can be considered a measurement context for some operational restriction of the protocol, i.e. by marginalising some event's output or by restricting the possible choices of inputs. This is true when considering the usual Bell scenario, i.e. when the contexts are given by $\{a_0, b_0\}$, $\{a_1, b_0\}$, $\{a_0, b_1\}$ and $\{a_1, b_1\}$, and any subset of arbitrary measurements induces a valid probability distribution, either by marginalising some of the outcomes or by conditioning on particular sets of inputs.

This property cannot be upheld in the presence of signalling, an observation that originally motivated our extension of the sheaf-theoretic approach. Suppose that we are given classically controlled quantum instruments: $\{\mathcal{M}_i^o\}_{(i,o)\in I_A\times O_A}$ and $\{\mathcal{N}_i^o\}_{(i,o)\in I_B\times O_B}$.

Definition 1.24. *A* classically controlled quantum instrument *is a set of CPTPS $\mathcal{S}$ with a given Kraus decomposition $\{\mathcal{S}_o^i\}_{o\in O_A}$ for every choice of input I_A. Such an instrument can be described in the framework of categorical probabilistic theories as a box parametrised by a classical input and a classical output wire:*



(1.41)

every quantum instrument, when applied to some initial state and when the final system is discarded, induces an empirical model with rows I_A and columns indexed by O_A.

We can imagine applying them sequentially and in parallel, as shown in Figure 1.8. Both ways of

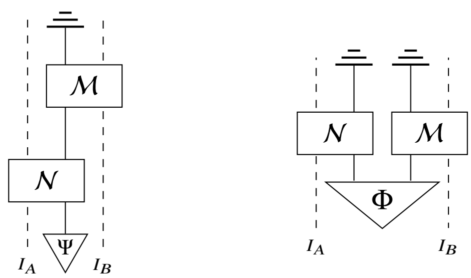


Figure 1.8: Sequential (on the left) and parallel application of quantum instruments (on the right).

using the 'same' resources, the same channels $\mathcal{M}$ and $\mathcal{N}$, induce a different joint instrument with input values $I_A \times I_B$ and output values $O_A \times O_B$. More importantly, composing the instruments sequentially or in parallel is reflected in what sub-contexts can be considered well-defined. For non-locality (right hand side of Figure 1.8 (p.30)), the no-signalling property allows us to identify any subset of $I_A \sqcup I_B$ as a well defined operational context (see Figure 1.9 (p.30)), while in the signalling case, the contexts 'Bob performs the measurement b_0' is not necessarily associated to a well defined distribution of outcomes and it can crucially depend on the choices at Alice's side.

For no-signalling scenarios, the Abramsky-Brandenburger framework endows the set of measurements X with the discrete topology $\mathcal{P}(X)$, indicating that every subset should be thought of as a

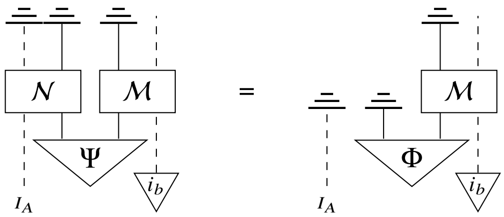


Figure 1.9: The outcome distribution associated with the context $i_b \in I_A \sqcup I_B$ is well-defined since independent on the choice of Alice, as witnessed by the disconnected diagram on the right

well defined context. To give causal semantics to a broad set of possible causal structures, we need to update the notion of measurement from being a point of a discrete finite topological space to a sequence of spacetime events inducing what we call a 'space of input histories'. Nonlocal scenarios, conditional distributions that agree on all intersections, will be thought of as a 'one-dimensional' case of a more general topological and geometrical study of the hierarchy of contexts.

1.3.3 Causal Graphs

The attempt to describe spacetime involving discrete atomic components is as old as special relativity itself. Notably, the first appearance of a structure resembling a causal set can be found in the 1914 book by Alfred Robb [121] 'A Theory of Time and Space'. Robb attempts an axiomatic derivation of special relativity involving only relations of temporal succession of individual events. The relativity of simultaneity is then captured by introducing a different type of time ordering, one which considers the possibility that individual events may be incomparable with respect to the succession of time. This desideratum is embodied in the description of a partial order of events. What is the matter of fact of these connections between events? Robb responds that they represent the possibility of causal dependence:

> In an instant B be distinct from an instant A, then B will be said to be after A, if and only if, it be abstractly possible for a person at the instant A to produce an affect at the instant B [121]

This description in terms of agency and the 'abstract possibility of causation' seemed to Norbert Weiner, reviewing the work in [142] 'utterly pointless' by appealing to the fact that the notion of causality itself is at least as obscure to the one of time succession. In this work, we provide a combinatorial description of a generalisation of causal orders, in which interventions are explicit and fundamental.

The celebrated result by Malament [87] extended the reconstruction of Robb to general relativity. Malament's theorem shows that an isomorphism of causal structures between two different spacetime $(\mathcal{M}, g_{ab})$ and $(\mathcal{M}', g'_{ab})$ which are future- and past-distinguishing can be extended to a smooth conformal isometry, so that an isomorphism of the causal structures preserves the topological, the differential, and the conformal structure, showing that the geometry of spacetime up to a conformal structure is entirely determined by its causal structure.

Chapter 3 will generalise this idea by transcending deterministic events to a more operational space of 'eventualities'. Every 'event' will be endowed with a set of inputs characterising the possible 'input values'. The combinatorial picture that emerges, which can be thought of as a direct operational generalisation of causal sets, allows for the flexibility to go beyond definite causality by describing situations where the input at some event can even affect the causal structure of subsequent events. The picture that emerges is that of a poset where the individual elements describe timelike histories instead of individual events.

Malament's result is the motivation behind many past and current lines of enquiry in causality: examples include the "causal sets" research programme [29], the domain-theoretic investigations of Martin and Panangaden [93, 94], and the functorial approach to quantum field dynamics [67].

The works mentioned above are all concerned with recovering relativistic structures or understanding quantum fields in an approximation of the spacetime continuum. Our efforts are directed towards the needs of quantum information protocols and experiments, where operations are performed locally at a finite set of spacetime events. Furthermore, our approach is independent of the theory underpinning the experiments and the concrete realisation of the local operations involved. Consequently, we will work directly with finite order structures without ever needing Lorentzian geometry to be involved.

Definition 1.25. *A* causal order Ω *is a preorder: a set* $|\Omega|$ *of events—finite, in this work—equipped with a symmetric transitive relation* $\leq$*, which we refer to as the* causal relation. *In cases where multiple cause orders are involved, we might also use the more explicit notation* $\leq_\Omega$*, to indicate that the relation is order-dependent.*

There are four possible ways in which two distinct events $\omega, \xi \in \Omega$ can relate to each other causally:

- ω *causally precedes* ξ if $\omega \leq \xi$ and $\xi \not\leq \omega$, which we write succinctly as $\omega \prec \xi$ (to distinguish it from $\omega < \xi$, meaning instead that $\omega \leq \xi$ and $\omega \neq \xi$)

- ω *causally succeeds* ξ if $\xi \leq \omega$ and $\omega \not\leq \xi$, which we write succinctly as $\omega \succ \xi$ (to distinguish it from $\omega > \xi$, meaning instead that $\omega \geq \xi$ and $\omega \neq \xi$)

- ω and ξ are *causally unrelated* if $\omega \not\leq \xi$ and $\xi \not\leq \omega$

- ω and ξ are in *indefinite causal order* if $\omega \neq \xi$, $\omega \leq \xi$ and $\xi \leq \omega$, which we write succinctly as $\omega \simeq \xi$

We say that a causal order is *definite* when the last case cannot occur, i.e. when $\leq$ is anti-symmetric ($\omega \leq \xi$ and $\omega \geq \xi$ together imply $\omega = \xi$); otherwise, we say that it is *indefinite*. A definite causal order is thus a *partial order*, or *poset*: in this case, $\omega \prec \xi$ is the same as $\omega < \xi$, and $\omega \succ \xi$ is the same as $\omega > \xi$.

Definition 1.26. *We say that two events ω, ξ are* causally related *if they are not causally unrelated, i.e. if at least one of $\omega \leq \xi$ or $\omega \geq \xi$ holds. We also define the* causal past $\omega\downarrow$ *and* causal future $\omega\uparrow$ *of an event $\omega \in \Omega$, as well as its* causal equivalence class $[\omega]_{\simeq}$*:*

$$\omega\downarrow \quad := \{\, \xi \in \Omega \,|\, \xi \leq \omega \,\} \tag{1.42}$$

$$\omega\uparrow \quad := \{\, \xi \in \Omega \,|\, \xi \geq \omega \,\} \tag{1.43}$$

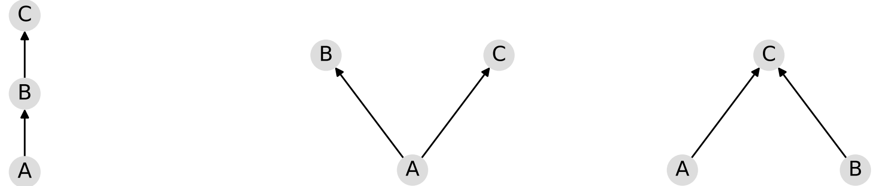
$$[\omega]_{\simeq} \quad := \{\, \xi \in \Omega \,|\, \xi \simeq \omega \,\} = \omega\downarrow \cap\, \omega\uparrow \tag{1.44}$$

Note that the ω always lies in both its own causal future and its own causal past, but also that their intersection can comprise more events (if the order is indefinite).

Differently from Robb we treat causality 'negatively': when ω causally precedes ξ we are not so much interested in the "possibility" of causal influence from ω to ξ (because $\omega \leq \xi$) as we are in the "impossibility" of causal influence from ξ to ω (because $\xi \not\leq \omega$). This generalises the "spatial" no-signalling case, where one is interested in the statements $\omega \not\leq \xi$ and $\xi \not\leq \omega$. Far from being merely an interpretation, such no-signalling approach to causality permeates the entirety of this work. From a topological perspective, it points to the lattice of lowersets $\Lambda(\Omega)$ of a causal order Ω as the correct combinatorial object to consider. Indeed, the inclusion order $U \subseteq V$ of lowersets is equivalent to the following condition.

$$U \subseteq V \;\Leftrightarrow\; \forall \xi \in V \backslash U.\, \forall \omega \in U.\, \xi \not\leq \omega$$

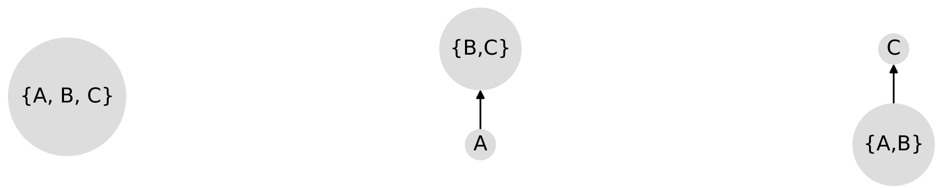
Definite causal orders have an equivalent presentation as directed acyclic graphs (DAGs), known as *Hasse diagrams*: vertices in the graph correspond to events $\omega \in \Omega$, while edges $x \rightarrow y$ correspond to those causally related pairs $\omega \leq \xi$ with no intermediate event (i.e. where there is no $\zeta \in \Omega$ such that $\omega < \zeta < \xi$). For example, below are the Hasse diagrams for three definite causal orders on three events A, B and C.



On the left, A causally precedes B, which in turn causally precedes C: this is an example of a *total* order, one corresponding to a line Hasse diagram. In the middle, A causally precedes both B and C, which are causally unrelated to one another. On the right, C causally succeeds both A and B, which are causally unrelated to one another.

More precisely, there is a bijective correspondence between finite partial orders and finite *intransitive* DAGs—loosely speaking, those without unnecessary edges [108]. The correspondence further generalises to *locally finite* partial orders—where any two elements have finitely many elements in between—and arbitrary intransitive DAGs [67].

The Hasse diagram representation extends to arbitrary causal orders, by making vertices in the graph correspond to causal equivalence classes instead of individual events; in the case of definite orders, the equivalence classes are all singletons, and can be safely identified with the unique event they contain. For example, below are the Hasse diagrams for three indefinite causal orders on three events A, B and C.



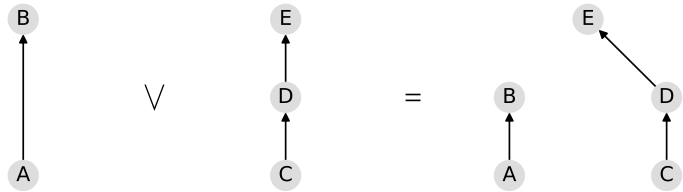
On a given set of events, all orders lie between two extremes: the *discrete order*, where all elements are causally unrelated, and the *indiscrete order*, where all elements lie in the same equivalence class. Additionally, the $n!$ possible *total orders* on n events are often of interest.

Definition 1.27. *For any finite set X of events, we write discrete (X) for the discrete order on the events and indiscrete (X) for the indiscrete order. For any finite sequence $\omega_1, \dots, \omega_n$ of events, we write total $(\omega_1, \dots, \omega_n)$ for the total order on the events which matches the sequence order.*

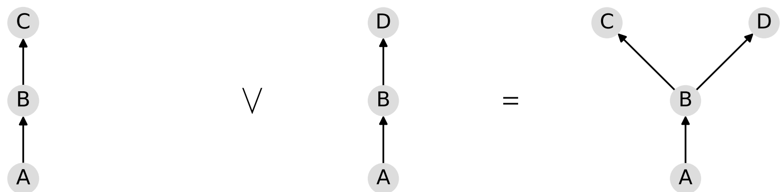
1.3.3.1 Join/union of causal orders

Definition 1.28. *The* join *of a family $(\Omega_j)_{j=1}^n$ of causal orders, denoted by $\bigvee_{j=1}^n \Omega_j$, is the union of their events equipped with the transitive closure of the union of the respective causal relations. Explicitly, two events ω and ξ are related by $\omega \leq \xi$ in the join $\bigvee_{j=1}^n \Omega_j$ iff there is a sequence of events $(\omega_k)_{k=0}^m$ and a sequence of causal orders $(\Omega_{j_k})_{k=1}^m$ such that $\omega_0 = \omega$, $\omega_m = \xi$ and $\omega_{k-1} \leq_{\Omega_{j_k}} \omega_k$ for all $k = 1, \dots, m$.*

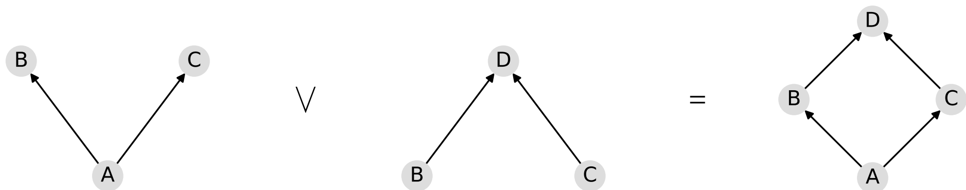
The join operation is commutative (order does not matter), associative (bracketing does not matter) and idempotent (repetition does not matter). When two causal orders are *disjoint*, i.e. when they share no common events, their join represents a scenario where all events from one order are causally unrelated to all events from the other, as in the following example.



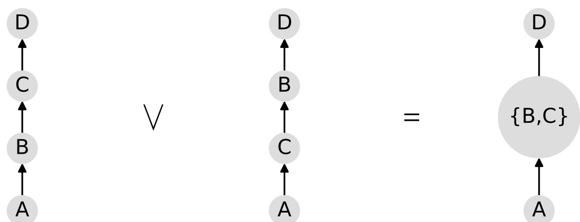
Because of this behaviour, we will refer to disjoint joins as *parallel composition*. When two causal orders have events in common, their join 'glues' them along the common events. Below is an example of two orders sharing an initial totally-ordered segment A $\rightarrow$ B, followed by two distinct events. Their join then has the same initial totally ordered segment, with a fork at B that leads to two causally unrelated events.



Below is a second example, where the two order share a pair of causally unrelated events B and C. Taking their join glues the two orders into a diamond, with causally unrelated events B and C separating the bottom event A from the top event D.



In the two examples above, the common events had an identical mutual causal relation in both others. However, it is generally the case for causal orders involved in a join to impose different causal relations on their common events. In particular, if two events are causally related in different ways in two causal orders, then the same two events will be in indefinite causal order in the join. For example, event B causally precedes C in the first causal order below, while the same event B causally succeeds C in the second causal order: in join (on the right), events B and C are therefore in indefinite causal order.

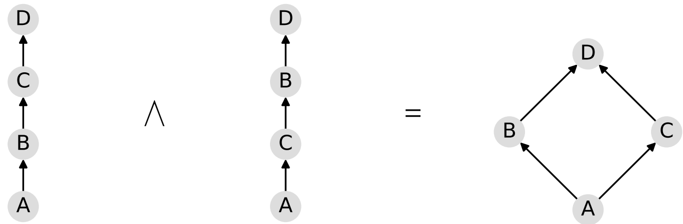


The space above is not a total order, but its Hasse diagram takes the same shape, so we extend our notation slightly and write total (A, $\{B, C\}$, D) to denote it.

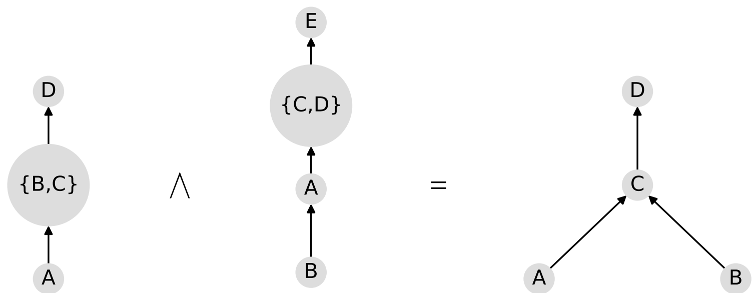
1.3.3.2 Meet/intersection of causal orders

Definition 1.29. *The* meet *of a family* $(\Omega_j)_{j=1}^n$ *of causal orders, denoted by* $\bigwedge_{j=1}^n \Omega_j$*, is the intersection of the events from the individual orders, equipped with the intersection of the respective causal relations. Explicitly, two events* ω *and* ξ *are related by* $\omega \leq \xi$ *in the meet* $\bigwedge_{j=1}^n \Omega_j$ *iff they are related in all orders, i.e. if* $\omega \leq_{\Omega_j} \xi$ *for all* $j = 1, ..., n$.

For example, below is the intersection of two total orders on the same 4 events $\{A, B, C, D\}$: in both orders, we have that event A causally precedes events B and C, which in turn causally precede event D. However, B precedes C in the first order, while it succeeds it in the second, resulting in B and C being causally unrelated in the meet. We will sometimes refer to this as the *diamond order*.



Below is a more complicated example, involving events in indefinite causal order. Events B and C are in indefinite causal order on the left, but B causally precedes C on the right, so B causally precedes C in the meet. Similarly, events C and D are in indefinite causal order on the right, but C causally precedes D on the left, so C causally precedes D in the meet. The situation leading to events A and B being causally unrelated in the meet is analogous to the one from the previous example.



1.3.3.3 Hierarchy of Causal Orders

For the scope of the dissertation is important not only to mention causal orders in isolation but to see them embedded in a hierarchy of possible causal assumptions on a finite set of events. This idea will be generalised in Chapter 3 where it will be shown that a more general notion of causal assumptions can be used to refine the hierarchy.

Causal orders are naturally ordered by inclusion: $\Omega \leq \Xi$ if $|\Omega| \subseteq |\Xi|$ as sets and $\leq_\Omega \subseteq \leq_\Xi$ as relations (i.e. as subsets $\{ (\omega, \omega') \mid \omega \leq_\Omega \omega' \} \subseteq |\Omega|^2$ and $\{ (\xi, \xi') \mid \xi \leq_\Xi \xi' \} \subseteq |\Xi|^2$). The requirement that $\leq_\Omega \subseteq \leq_\Xi$ explicitly means that for all $\omega, \omega' \in \Omega$ the constraint $\omega \not\leq_\Xi \omega'$ in Ξ implies the constraint $\omega' \not\leq_\Omega \omega'$. Put in different words:

- If ω and ω' are causally unrelated in Ξ , then they are causally unrelated in Ω.
- If ω causally precedes ω' in Ξ, then it can either causally precede ω' in Ω or it can be causally unrelated to ω' in Ω; it cannot causally succeed ω' or be in indefinite causal order with it.
- If ω and ω' are in indefinite causal order in Ξ, then their causal relationship in Ω is unconstrained: ω can causally precede ω', causally succeed it, be causally unrelated to it or be in indefinite causal order with it.

From a causal standpoint, $\Omega \leq \Xi$ means that Ω imposes on its own events at least the same causal constraints as Ξ, and possibly more. In particular, if Ξ is definite (no two events in indefinite causal order) then so is Ω; conversely, if Ω is indefinite, then so is Ξ.

Observation 1.29. *Causal orders on a given set of events form a finite lattice, which we refer to as the* hierarchy of causal orders*. The join and meet operations on this lattice are those described in the previous subsections, the indiscrete order is the unique maximum (all events in indefinite causal order, i.e. no causal constraints), while the discrete order is the unique minimum (all events are causally unrelated).*

The hierarchy of causal orders on three events $\{A, B, C\}$ is displayed by Figure 1.10 (p.37), with definite causal order coloured red and indefinite ones coloured blue. The definite causal order always form a lowerset in the hierarchy—if Ξ is definite then all $\Omega \leq \Xi$ are also definite—and the maxima for this lowerset are exactly the total orders. However, total orders do not form a separating set: there are inclusions of definite orders into indefinite ones that do not factor through a total order (cf. Figure 1.10, arrows from red notes in the second layer to blue nodes in the fourth layer).

1.3.3.4 Lattice of Lowersets

When we discuss causality from a theory independent perspective we will be concerned with a certain class of operational scenarios: blackbox devices are operated locally at events in spacetime, determining a probability distribution on their joint outputs conditional to their (freely chosen) joint inputs. In such scenarios, causality constraints essentially state that the output at any subset of events cannot depend on inputs at events which causally succeed them or are causally unrelated to them. Furthermore, the output at any event is only well-defined conditional to inputs for all events in its past: we are not interested in all sub-sets of events of a causal order, but rather in its lowersets.

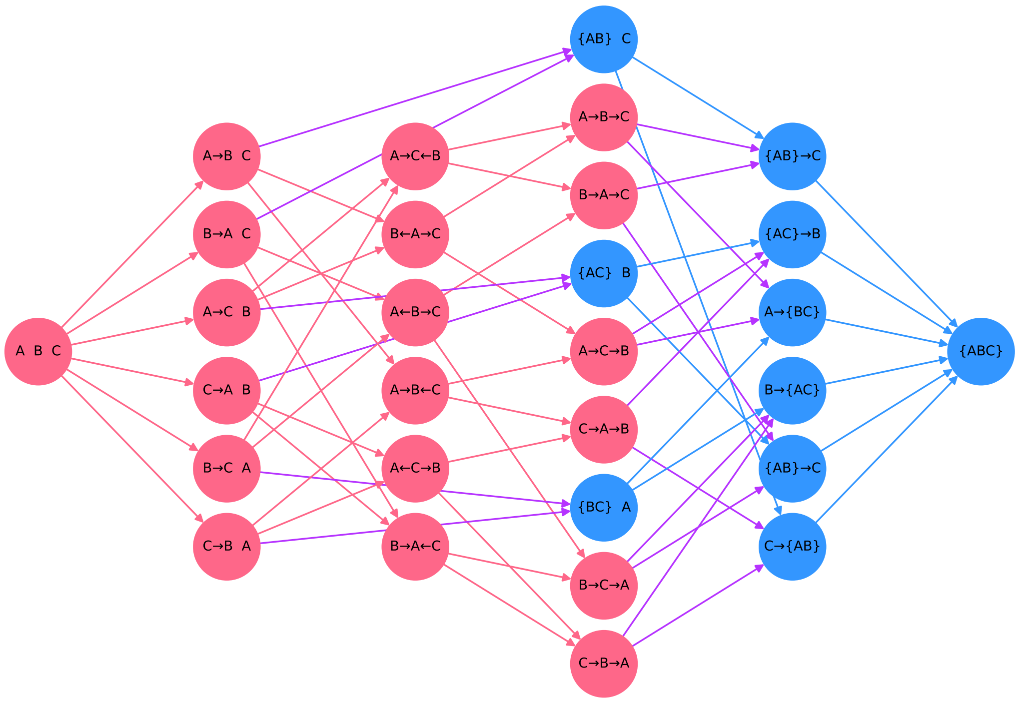


Figure 1.10: Hasse diagram for the hierarchy of causal orders on three events $\{A, B, C\}$, left-to-right in inclusion order. Definite causal orders (left and middle) are coloured red, while indefinite causal orders (middle and right) are coloured blue. Inclusions between definite orders are coloured red, inclusions between indefinite orders are coloured blue, inclusions of a definite order into an indefinite one are coloured violet. In the order labels, a space is used to indicate causal unrelatedness, arrows are used to indicate that the event at the tail causally precedes the event at the head, and braces are used to indicate that the events contained are in indefinite causal order.

The discussion above indicates that the object we seek to understand is not the causal order Ω itself, but rather its *lattice of lowersets* $\Lambda(\Omega)$. This is the subsets of events closed in the past, ordered by inclusion:

$$\Lambda(\Omega) := \{ U \subseteq \Omega \,|\, \forall \omega \in U .\, \omega\!\downarrow\, \subseteq U \} \tag{1.45}$$

In this case, being a lattice means that lowersets are closed under both intersection and union; we always omit the empty set from our Hasse diagrams, for clarity.

Inclusions between lowersets determine the causality constraints for the causal order: if $U, V \in \Lambda(\Omega)$ are such that $U \subseteq V$, then the output at events in U cannot depend on the inputs at events in $V \backslash U$. Consider the total order $A \to B \to C$, and its associated lattice of lowersets: the inclusion $\{A, B\} \subseteq \{A, B, C\}$, for example, tells us that the outputs at events A and B cannot depend on the input at event C; the inclusion $\{A\} \subseteq \{A, B\}$, additionally, tells us that the outputs at event A cannot depend on the input at event B.

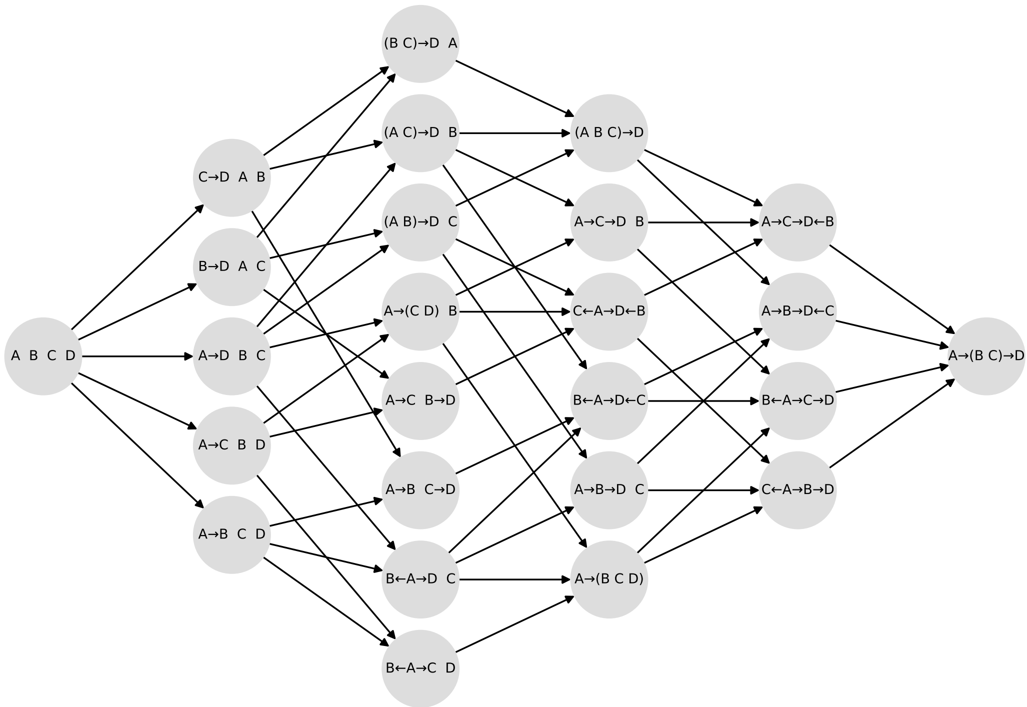
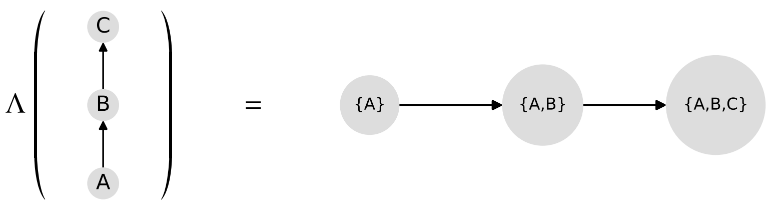
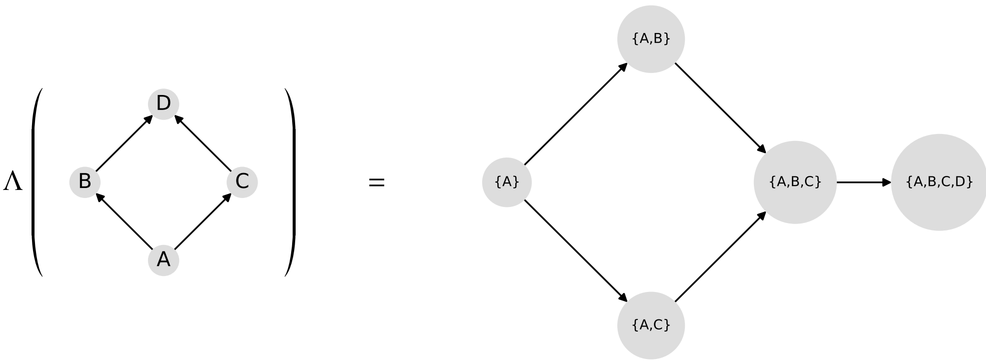


Figure 1.11: Hasse diagram for the hierarchy of sub-orders of the diamond order, left-to-right in inclusion order. All orders are definite, so no colour-coding of nodes and edges is necessary. In the order labels, a space is used to indicate causal unrelatedness, arrows are used to indicate that the event(s) at the tail causally precedes the event(s) at the head, and brackets are used to group multiple causally unrelated events together (for ease of notation). For example, A→(B C)→D on the right indicates that event A precedes events B and C, which are causally unrelated to each other and both precede event D.



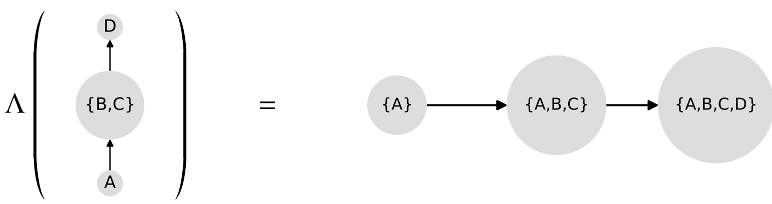
Below is a more complicated example, for the diamond order: the inclusion $\{A, B\} \subseteq \{A, B, C, D\}$, for example, tells us that the outputs at events A and B cannot depend on the input at events C and D.



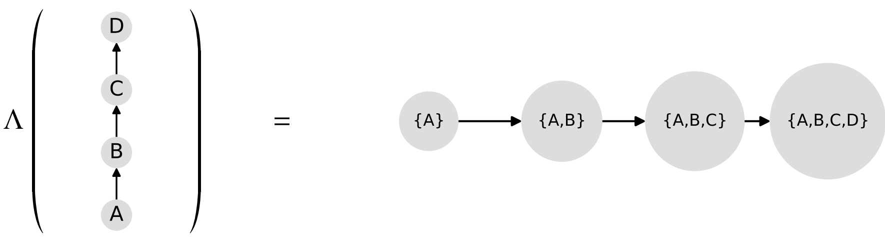
Here, we note for the first time how lowersets are more general than downsets: we have $A\downarrow = \{A\}$, $B\downarrow = \{A, B\}$, $C\downarrow = \{A, C\}$ and $D\downarrow = \{A, B, C, D\}$, but lowerset $\{A, B, C\}$ does not originate from any individual event. Hence, lowersets strictly generalise the notion of causal past from individual events to arbitrary subsets of events:

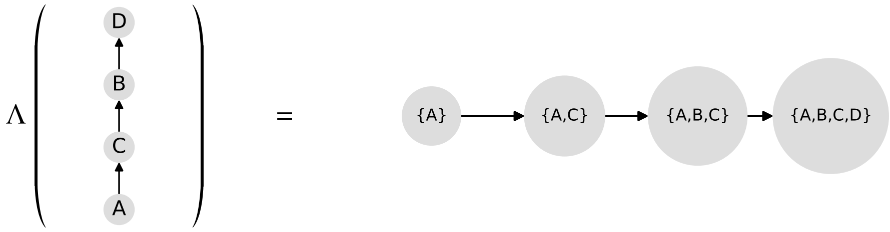
$$\{A, B, C\} = B\downarrow \cup C\downarrow = \{B, C\}\downarrow$$

When the causal order is indefinite, lowersets cannot split causal equivalence classes: either no event from the class is in the lowerset, or all events are. We can see this in the lattice of lowersets for the indefinite causal order $A \to \{B, C\} \to D$, where events $\{B, C\}$ form a causal equivalence class.

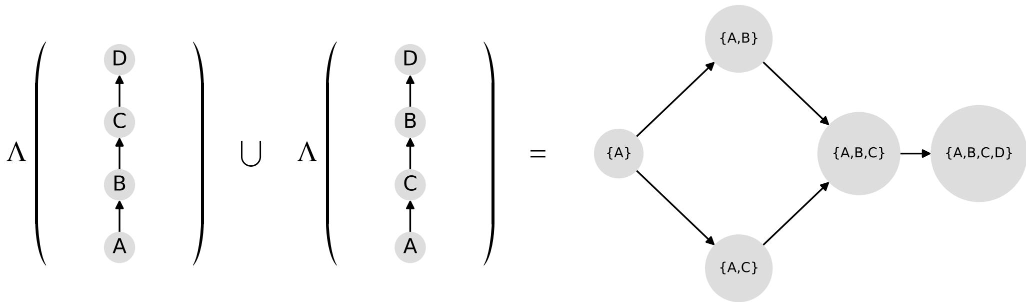


An interesting question arises when we consider the interaction of causal constraints for multiple causal orders. A scenario is explainable by two causal orders Ω and Ω' if it satisfies the causal constraints of both: in terms of lowersets, such constraints correspond to the union $\Lambda(\Omega) \cup \Lambda(\Omega')$ of the lowersets for the individual orders. Consider, for example, the total orders $\Omega = A \to B \to C \to D$ and $\Omega' = A \to C \to B \to D$, together with the associated lowersets.





For a scenario to satisfy both orders, it has to satisfy the constraints derived from $\Lambda\,(\mathrm{A} \to \mathrm{B} \to \mathrm{C} \to \mathrm{D}) \cup \Lambda\,(\mathrm{A} \to \mathrm{C} \to \mathrm{B} \to \mathrm{D})$, depicted below.



We immediately recognise the lowersets as those of the diamond order, which we also know to take the form:

$$\mathrm{diamond}_{\mathsf{ABCD}} = (\mathrm{A} \to \mathrm{B} \to \mathrm{C} \to \mathrm{D}) \wedge (\mathrm{A} \to \mathrm{C} \to \mathrm{B} \to \mathrm{D})$$

So the question arises: is simultaneously satisfying the causal constraints for two (or more) causal orders always the same as satisfying the causal constraints for their meet? To answer it, we first note that the hierarchy of causal orders is (contravariantly) related to the hierarchy formed by the corresponding lattices of lowersets under inclusion.

Proposition 1.29. *For any two causal orders Ω and Ω' such that $|\Omega| = |\Omega'|$, we have:*

$$\Omega \leq \Omega' \;\Leftrightarrow\; \Lambda\,(\Omega) \supseteq \Lambda\,(\Omega') \tag{1.46}$$

Proof. We separately prove the two sides of the equivalence:

1. $\Omega \leq \Omega' \implies \Lambda\,(\Omega) \supseteq \Lambda\,(\Omega')$. Let $U \in \Lambda\,(\Omega')$, then $U \subseteq |\Omega|$. We need to show that $U \in \Lambda\,(\Omega)$ let $u \in U$, then for $v \in \Omega$ such that $u \leq_{\Omega} v$ we have that $u \leq_{\Omega'} v$ and since $U \in \Lambda\,(\Omega')$ we conclude that $u \in U$.

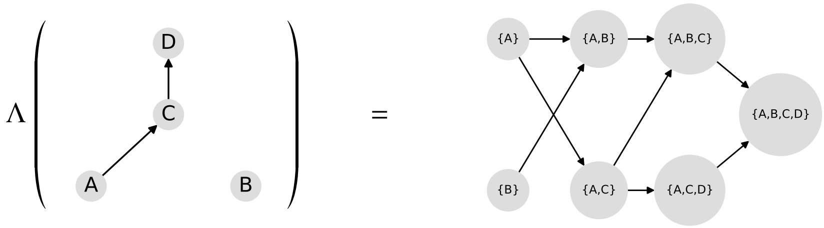
2. Let $\Lambda\,(\Omega) \supseteq \Lambda\,(\Omega')$, and $u \leq_{\Omega} v$. Since $|\Omega| = |\Omega'|$, we also have that $u, v \in \Omega'$. Consider $u\!\downarrow\, \in \Lambda(\Omega')$, then $v \in u\!\downarrow$ since $u\!\downarrow$ is a lowerset in Ω and by assumption $u \leq_{\Omega} v$, therefore $v \leq_{\Omega'} u$.

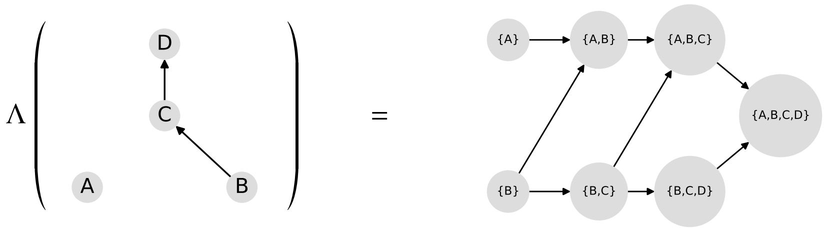
□

Corollary 1.30. *For any two causal orders* Ω *and* Ω' *such that* $|\Omega| = |\Omega'|$, *we have:*

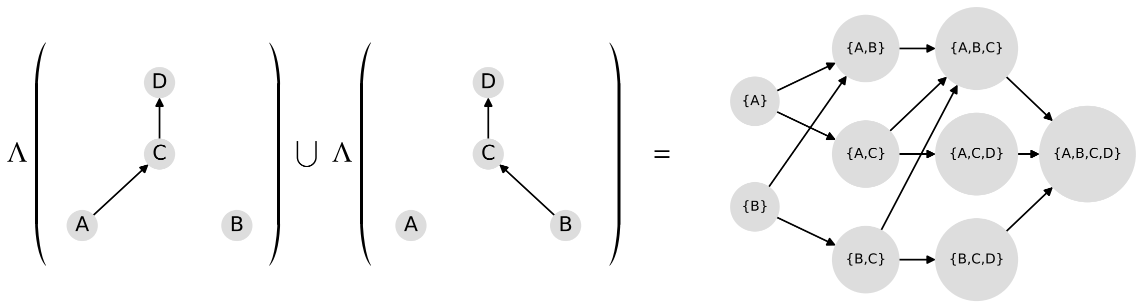
$$\Lambda(\Omega) \cup \Lambda(\Omega') \subseteq \Lambda(\Omega \wedge \Omega') \tag{1.47}$$

Unfortunately, the above inclusion cannot be strengthened to an equality: in general, $\Lambda(\Omega) \cup \Lambda(\Omega')$ is not even a lattice! For a counterexample, we consider the following orders on four events and their associated lattices of lowersets.

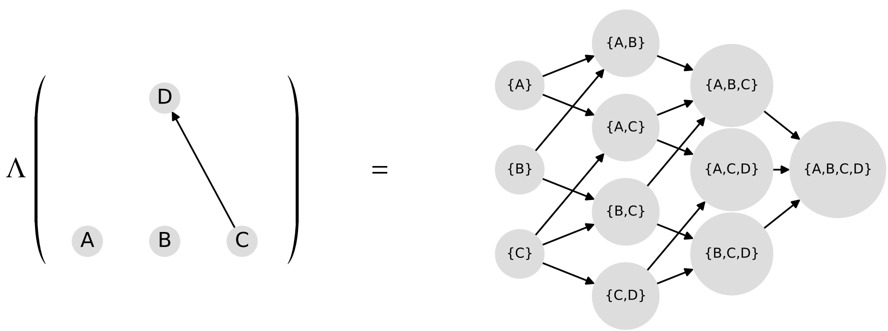




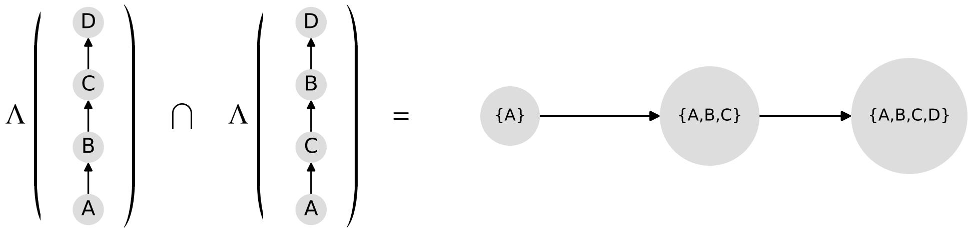
The union of the corresponding lattices of lowersets is the following set, which is evidently not closed under intersection: the intersections $\{C\} = \{A, C\} \cap \{B, C\}$ and $\{C, D\} = \{A, C, D\} \cap \{B, C, D\}$ are both conspicuously missing.



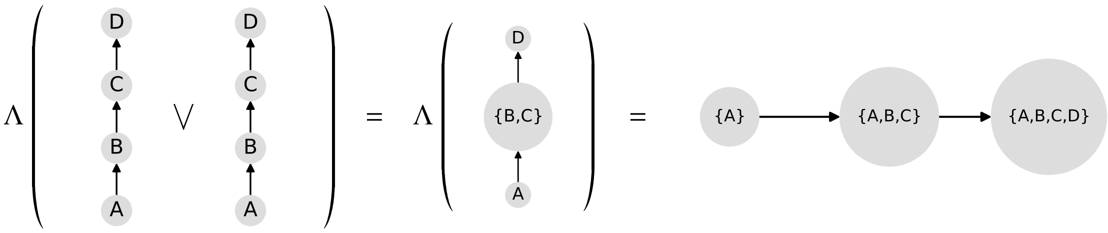
In particular, the collection of lowersets displayed above is not the lattice of lowersets for the meet of the two orders, which (in this case) is obtained by including the two missing lowerset intersections.



Dually to the above, we can ask whether the intersection of the lattices of lowersets for two (or more) causal events is the lattice of lowersets for their join, i.e. whether it holds that $\Lambda(\Omega) \cap \Lambda(\Omega') = \Lambda(\Omega \vee \Omega')$. This is more promising: at the very least, the intersection $\Lambda(\Omega) \cap \Lambda(\Omega')$ is always a lattice! As a motivating example, we go back to the total orders $\Omega = \mathrm{A} \to \mathrm{B} \to \mathrm{C} \to \mathrm{D}$ and $\Omega' = \mathrm{A} \to \mathrm{C} \to \mathrm{B} \to \mathrm{D}$.



Indeed, we immediately recognise the intersection as the lattice of lowersets for the join of the two total orders, where the events B and C are in indefinite causal order:



In this dual scenario, the contravariant relation between the hierarchy of causal orders and the associated lattices of lowersets implies that $\Lambda(\Omega) \cap \Lambda(\Omega') \supseteq \Lambda(\Omega \vee \Omega')$. This time, the inclusion can be strengthened to an equality.

Proposition 1.30. *For any two causal orders* Ω *and* Ω' *we have:*

$$\Lambda(\Omega) \cap \Lambda(\Omega') = \Lambda(\Omega \vee \Omega') \tag{1.48}$$

Proof. We need to show that $\Lambda(\Omega) \cap \Lambda(\Omega') \subseteq \Lambda(\Omega \vee \Omega')$. Take $v \in U$ and u such that $u \leq_{\Omega \vee \Omega'} v$: by definition of the join order, there is a sequence of events $(\omega_k)_{k=0}^m$ such that $\omega_0 = u$, $\omega_m = v$ and for all $k = 1, ..., m$ either $\omega_{k-1} \leq_\Omega \omega_k$ or $\omega_{k-1} \leq_{\Omega'} \omega_k$. We follow the sequence backwards, starting from $\omega_m = \xi$: for each $k = m, ..., 1$, $\omega_k \in U$ implies $\omega_{k-1} \in U$. Hence $u = \omega_0 \in U$, so that U is a lowerset for $\Omega \vee \Omega'$. □

The above result provides operational meaning to joins of causal orders: the causal constraints imposed by the join order are exactly the constraints common to all causal orders involved.

1.4 Literature review

The conflict between classical causality and quantum theory animated the foundational discussions from their earliest stages. Despite persisting in the background of any foundational work, it took centre stage only relatively recently. Ideas about indefinite causal order and dynamical spacetime have shaped the recent work on process matrices; Judea Pearl's systematic account of causal inference has opened the doors for a formal analysis of the clash between quantum experiments and causal intuition; contextuality has exposed the tensions of quantum theory with a classical account of observable facts. In the following subsections, we will review the works that interlace with this study and, to facilitate the reading, we have divided this review into four parts: *no-signalling correlations* and generalisations thereof, the study of *indefinite causality*, *quantum causal models*, and *contextuality*. The purpose of this section is to discuss works which share both similarities and fundamental differences from our narrative. It is not essential to understanding the dissertation's content but exposes the works that were influential in the maturing process of our framework. This thesis aims to provide a cohesive mathematical language that restores conceptual unity among several independent research directions. Due to the inherent variety of topics, we cannot hope to be exhaustive enough, and we content ourselves with weaving a thread before entering the labyrinth.

1.4.1 Signalling and No-signalling correlations

The study of non-local correlations has a long history, starting with John Bell's work and his account of violations of local realism. Despite its seminal character and the importance that it had in shaping quantum foundations, it is noteworthy to mention that (as explained by Pitowsky in [109, 110]) Bell's inequalities can be thought of as a special case of a much earlier George Boole's formulation of 'conditions for possible experience' [30]. Boole's aimed to derive the conditions that logical dependencies impose on relative frequencies. Suppose we are given a set of probabilities $p_1, p_2, \dots, p_n$ all arising as relative frequencies of logically disconnected sets. The only requirements to be imposed is that: $p_i \geq 0$, and $p_i \leq 1$ for $1 \leq i \leq n$. If p_1, p_2, p_3 represent the relative frequencies of occurrence of elements from the sets E_1, E_2 and $E_1 \cap E_2$ then the latter condition implies that the probability p_3 is correlated by the values observed for p_1 and p_2 by an additional requirement, namely:

$$-p_1 - p_2 + p_3 + 1 \geq 0 \tag{1.49}$$

Suppose we are performing an experiment by extracting balls from a box and that the balls can be wooden or plastic, red or black, and find out that that there is a probability of 60% of extracting a red ball, 75% of them are wooden, and 30% are both red and wooden. We therefore have that $p_1 = 0.6$, $p_2 = 0.75$ and $p_3 = 0.3$ violating the inequality of Equation 1.49. If we observe these relative frequencies, we immediately conclude that there is something wrong with some of the logical assumptions of the experiment or about the way this probabilistic behaviour is associated with the underlying objective properties of the samples.

Boole's contribution was to realise that for finite sets of events, with finitely many logical connections, these always induce a finite set of linear constraints. Logical constraints, therefore, form a polytope of compatible probabilistic behaviours. Bell's type inequalities can all be seen as a particular case of Boole's conditions in which the logical relations are derived from the assumption of local causality [109].

The impossibility of superluminal signalling required to establish the compatibility between quantum theory and special relativity has often been abstracted and studied as a theory independent principle, notably in the seminal work by Popescu and Rohrilch [112, 113]. The connection between no-signalling correlations and polytopes originates in [16], where the authors explicitly characterise various sets of no-signalling boxes and pave the way for a more systematic study of such correlations.

Let us denote by $p(a, b|x, y)$ the probability that the outcomes $a \in O_A$ and $b \in O_B$ are observed given the settings $x \in I_A$ and $y \in I_B$. For the bipartite case, the standard no-signalling conditions are given by the following equations:

$$\sum_b p(a, b|x, y) = \sum_b p(a, b|x, y') \quad \forall y, y', x, a \tag{1.50}$$

The probability that A observes an output a is independent of B's local setting. This family of equations can be interpreted as linear constraints together with the additional requirement of positivity:

$$p(a, b|x, y) \geq 0 \quad \forall x, y \tag{1.51}$$

and normalisation:

$$\sum_{a,b} p(a, b|x, y) = 1 \quad \forall x, y \tag{1.52}$$

These constraints bound the *non-signalling polytope*. A table of correlations for the bipartite scenario has 2^4 entries; subtracting the number of independent constraints, we get an 8-dimensional polytope with 24 vertices. In [16] they also characterise a 'local' sub-polytope which is the convex hull of 16 of the total 24 vertices, obtained by the deterministic no-signalling functions correlating joint inputs to joint outputs.

A correlations is said to admit a *deterministic hidden variable (DHV)* if for a (finite) set of hidden values Λ there exist response functions

$$p_A : I_A \times \Lambda \to O_A \tag{1.53}$$

$$p_B : I_B \times \Lambda \to O_B \tag{1.54}$$

and a distribution $p_\Lambda(\lambda)$ on Λ such that the observed statistics can be reproduced by averaging over the possible values of the hidden variable (here the response functions described above are treated as deterministic conditional distributions):

$$p(a,b|x,y) = \sum_{\lambda\in\Lambda} p_\Lambda(\lambda) p_A(a|x,\lambda) p_B(b|y,\lambda) \tag{1.55}$$

The functions p_A and p_B represent an underlying deterministic mechanism through which chosen inputs and hidden variables concur to influence the outcome. We already encountered these hidden variables when reviewing the sheaf theoretic approach.

One may think about a more general hidden variable theory, where the response mechanisms are themselves allowed to be genuinely stochastic under the assumption that they are mutually independent given knowledge of the hidden variable. We call such models *factorisable hidden variable models (FHVs)*. We observe that DHVs are a subset of FHVs where p_A and p_B are deterministic maps. Conversely, we can always dilate a stochastic map $p : I \to O$ to a deterministic $\tilde{p} : I \times \Lambda \to O$ such that there exist a distribution p_Λ for which $p(o|i) = \sum_{a\in\Lambda} \tilde{p}(o|i,a) p_\Lambda(a)$. This means that for any factorisable hidden variable model, we can shift the randomness of the response functions into additional hidden variables and convert them into a DHV. Both notions of hidden variables are equivalent for the no-signalling case (as shown in [6] also for more general contextuality scenarios). This result is often associated with Arthur Fine [55] and is known in its more general formulation as the Fine-Abramsky-Brandenburger theorem [6].

The same work by Barrett et al. [16] also shows that generalising the bipartite no-signalling condition comes with some subtleties. In general tripartite no-signalling is identified as follows:

$$\sum_a p(a,b,c|x,y,z) = \sum_a p(a,b,c|x',y,z) \;\; \forall\, b,c,y,z,x,x' \tag{1.56}$$

However, if we only focus on pairwise signalling, we can imagine situations where no agent alone is allowed to signal to any other party, but the scenario does not form an empirical model compatible with three spacelike separated agents. An example of this is given by the correlations described in Figure 1.15 (p.52). Even though no single party sends direct signals to any other party, the table does not satisfy all the constraints given by Equation 1.56.

If we marginalise the output of C, local choices at C will determine if the measurements of A and B are correlated or anti-correlated. For example, fixing $i_c = 0$ and marginalising the output at C, the probability of obtaining $o_A = 0, o_B = 0$ given $i_A = 0, i_B = 0$ is:

$$\sum_{o_c\in O_C} p(o_c, o_a := 0, o_b := 0 | i_c := 0, i_a := 0, i_b := 0) = 1/2 \tag{1.57}$$

setting $i_C = 1$ instead:

$$\sum_{o_c\in O_C} p(o_c, o_a := 0, o_b := 0 | i_c := 1, i_a := 0, i_b := 0) = 0 \tag{1.58}$$

Even though the protocol is not compatible with a genuine tripartite spacelike separation, the fact that no agent alone can signal is witnessed by constraints of the type:

$$\sum_{a,b} p(a,b,c|x,y,z) = \sum_{a,b} p(a,b,c|x',y,z) \;\; \forall\, c,y,z,x,x' \tag{1.59}$$

When we marginalise the outputs of A and B, the choices at A cannot influence the outcome distribution at C and similarly for any other permutation of agents. We can therefore marginalise any subset comprising of two agents and obtain a well defined probability distribution. This does not mean that the marginalisation of the output of any single agent will leave us with a well defined conditional distribution.

The table in Figure 1.15 (p.52) is not compatible with a causal structure where all three agents are spacelike separated, but it is also not univocally compatible with a specific signalling structure synthesised from a causal graph between three agents. We will see later that several different causal orders can reproduce the correlations exhibited by such an empirical model. In particular, we can reproduce the correlations by assigning conditional stochastic maps to the vertices of the following three different causal graphs (nota bene, we need to allow the possibility of arbitrary shared resources in the causal past of the agents, in direct generalisation of multipartite Bell scenarios):

- A after C, with no-signalling to/from B
- B after C, with no-signalling to/from A
- A and B after C

There is no unique minimal causal graph—we understand minimality as witnessing all the possible no-signalling constraints—that characterises the signalling structure of the empirical model. One would need to appeal to a more fine-grained description of causal structures: fewer restrictions than genuine tripartite no-signalling, and more restrictions than any other causal ordering between the events A, B, C. One of the aims of our investigation is to provide the tools to characterise polytopes of correlations that are compatible with various operational assumptions.

No-signalling constraints describe properties of the conditional distribution in a theory independent way. It is just about the well-definedness of the marginals attributed to independent subsystems. Of different nature is the additional logical requirements characterising local causality. When the number of parties increases, we expect a greater interplay between the possible explanatory causal mechanisms. In [16] Barrett et al. discuss the observation (originally by Svetlichny [136] and generalised in [45]) that for the tripartite case there is already room for various 'degrees' of nonlocality. What they call 'local correlations' satisfy a straightforward symmetric tripartite generalisation of local causality:

$$p(a,b,c|x,y,z) = \sum_{\lambda\in\Lambda} p(\lambda)p(a|x,\lambda)p(b|z,\lambda)p(c|x,\lambda) \tag{1.60}$$

Given the knowledge of the shared variable $\lambda \in \Lambda$, the response distributions factor in three local stochastic maps. The weaker notion of 'two-way locality' describes correlations that may not be fully local but admit a classical explanation concerning some bipartition of the boxes. These are correlations that can be decomposed as follows:

$$\begin{aligned} p(a,b,c|x,y,z) &= p_{12}\textstyle\sum_{\lambda\in\Lambda_{12}} p(\lambda)p(a,b|x,y,\lambda)p(c|z,\lambda) && (1.61)\\ &+p_{13}\textstyle\sum_{\lambda\in\Lambda_{13}} p(\lambda)p(a,c|x,z,\lambda)p(b|z,\lambda) && (1.62)\\ &+p_{23}\textstyle\sum_{\lambda\in\Lambda_{23}} p(\lambda)p(b,c|y,z,\lambda)p(a|x,\lambda) && (1.63)\end{aligned}$$

Two-way local correlations can be considered classical under a 'coarse-graining' of two of the three events. Similarly to how in the sheaf theoretic description, locality is seen as the possibility of finding a global section that accounts for the empirical model; in our work we will show that the hierarchy of multipartite locality is related to the possibility of lifting compatible families onto finer measurement covers.

The set of the quantum realisable correlations (although convex) does not form a polytope [68, 111, 9, 48, 98] and it is of a less straightforward characterisation. Navascues et al. have introduced an infinite hierarchy of conditions which are necessarily satisfied by any quantum correlation in [97] and showed that it is complete, i.e. that every correlation satisfying all conditions has a quantum realisation, in [98]. Our work will not be concerned with identifying quantum realisable correlations, and we leave the possible extension of the work by Navascues et al. to our causal polytopes for future work.

Extensions of multipartite non-locality can be found in [59, 115, 143, 120, 137, 57]. These allow common sources to be distributed only to specific subsets of agents, differing from the generalisation that we want to present in this work. It has been shown in [32, 56, 120, 57] that—with the extra assumption of factorisability of common sources—quantum nonlocality becomes detectable without any reference to freely chosen inputs. For us, free choice of inputs will be an essential ingredient of the notion of an event, a requirement for the definition of empirical models. Moreover, assuming factorisability of common sources may result in a space of compatible correlations (local and non-local), which is no longer convex. Our approach to causal discovery will rely on the spaces being convex polytopes, enabling the crucial employment of convex geometry and linear programming.

An example exhibiting the failure of convexity can be found in the triangle scenario discussed in [120, 57]. Three observers perform a single measurement and are pairwise connected by three independent sources as shown in Figure 1.12 (p.48). The deterministic response R_0, where everyone measures the outcome 0 and the deterministic response R_1, where the outcome for each measurement is 1 are compatible with this bipartite repartition of the common causes. However, the convex mixture of the two possible global outcomes $1/2(R_0 + R_1)$ requires common resources to be shared among all three parties simultaneously and is therefore incompatible with the original assumptions.

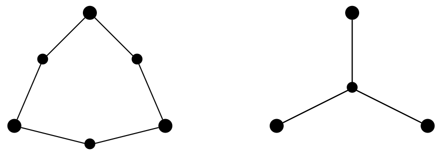


Figure 1.12: The bigger dots are agents, while the smaller ones indicate shared common resources. The picture on the left represents the assumption that three sources are separated, and the picture on the right describes a standard tripartite no-signalling scenario. The pairwise correlated common sources on the left cannot realise the distribution $1/2(R_0 + R_1)$, while the scenario on the right can reproduce any tripartite distribution of outputs if we do not allow the introduction of measurement settings.

There have been proposals for a generalisation of the study of non-locality to arbitrary Causal Bayesian networks, most notably by Henson et al. and Fritz [58, 74]. Fritz in [58] shows that all the relevant assumptions in a Bell scenario can be directly derived from the causal structure shown in Figure 1.13 (p.49). If we write $o[V]$ for the vector or tuple $(o[v])_{v\in V}$ consisting of all the outcomes $o[v]$ for the nodes $v \in V$ of a causal network G, then a necessary condition for a joint distribution $P(o[V])$ to be explainable in terms of hidden variables on a given causal structure is that of being a 'correlation':

Definition 1.31 (Correlation [57]). *$P(o[V])$ is a* correlation *if for any collection of subsets $V_1, \ldots, V_m \subseteq V$ with disjoint causal past, the distribution factors as*

$$P(o[V_1]\ldots o[V_m]) = P(o[V_1])\ldots P(o[V_m]) \tag{1.64}$$

This condition is far from sufficient. In particular, any assignment of deterministic quantum processes (even the ones violating Bell's inequalities) to the nodes will respect the factorisation assumptions of Definition 1.31. In fact, for the particular case of Figure 1.13 (p.49), $P(o[V])$ is a correlations if and only if it satisfies the usual no-signalling and free will equations. This way of treating quantum protocols allows generalising Bell's theorem by avoiding explicit dependencies on free-will assumptions [57].

In our work, the philosophy will be different in quite a fundamental way. Causality is represented as the possibility of signalling between subsets of intervening agents, and we do not want to associate a particular distribution with the settings themselves. Our assumption allows us to explore the correlations induced by more complicated operational assumptions, such as indefinite causality, without worrying about re-deriving tailored free-will conditions for particular causal models (cfr. the discussion of [77]). For us, free will is a global structural assumption describing the class of admissible dynamical hidden mechanisms: the inputs should be chosen freely and are paramount for describing the structure of operational contexts.

Another framework to analyse quantum causal correlation beyond standard no-signalling has recently been described by Guryanova et al. [70, 76, 117]. They consider the restrictions to

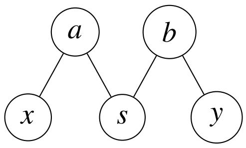


Figure 1.13: Causal structure of the two-party Bell scenario as described in [58]. The dots represent observable random variables, and the links are causal dependencies. In particular, we see that there is no distinction between settings and outcomes. The free will condition, similarly to no-signalling, is encoded in the causal relationship between the variables.

correlations imposed by the 'no-backwards-in-time signalling condition' (NBTS). For Guryanova et al., a closed laboratory is described by an agent receiving a classical random variable only after performing a fixed measurement on his incoming system. The agent can then use measurement results and the classical random variable to apply a correction before sending the system outside the laboratory. In their framework, this no-signalling condition already limits the correlations that are realisable by a single laboratory, i.e. for every external input x and measurement outcome a we must have that $p(a|x) = p(a)$, the probability of outcomes is independent of the input communicated from outside the laboratory.

In the context of our work, it is important to allow correlations between classical variables parametrising an event (the choices of possible interventions) and the classical variables describing the results of chosen measurements. In particular, the convex polytope of correlations associated with causally ordered agents will be different if one uses the NBTS condition proposed in [70] against the usual 'no-signalling from the future' which is ubiquitous in operational approaches to quantum theory (a formulation of this principle can be found for example in [38, 47] or in Section 1.3.1.2 where we present CPTs, i.e. categorical probabilistic theories).

1.4.2 Indefinite causality

The idea of intertwining indefinite causality with quantum information dates back to work by Hardy [72], where he proposes a generalised notion of computation consistent with a dynamical background spacetime. More recently, discussions on indefinite causality were reignited with the work of Chiribella et al. [39, 69] on the *quantum switch*, and the attempts to justify the experimental realisability of indefinite causal orders [69, 122, 22, 50, 124, 116, 123]. It has been shown in [141, 107] that superposing the causal orders of channels overcomes the subtleties about the general notion of 'quantum control' exposed in [40, 3, 99] and can be given a sound operational semantic. Discussions about the feasibility of indefinite causal order triggered from one side the attempt to understand the most general way of composing quantum instruments without breaking the local structure of the theory, and from the other, the realisation that a clear understanding of the correlations compatible with more or less exotic descriptions of causal composition was still missing. This prompted a characterisation of general quantum supermaps that would generalise the sequential and parallel composition of laboratories. Assuming that a theory enhanced with the possibility of exotic causal orders would be locally embodied by the usual quantum theory, Oreshkov, Costa and Bruckner introduced the concept of *process matrices* [101] which developed in an active research area [31, 12, 1, 141]. Process matrices describe the spaces of general quantum 'supermaps', i.e. operations mapping quantum instruments or classically controlled quantum instruments to valid empirical models.

This formalism showcases that beyond the standard paradigm of the sequential and parallel composition of quantum process, other types of supermaps interact well with the linear and probabilistic nature of the theory. While the idea of process matrices is very distant from the external and theory independent treatment of signalling relations that we want to discuss here, the notion of *causal inequalities* which has been developed from the study of process matrices, can be used—by virtue of it theory independent nature—to draw fruitful connections between the two perspectives. In [101], these inequalities are presented as linear bounds which are respected by distributed protocols exhibiting definite causal order or a probabilistic mixture thereof. The causal GYNI (Guess Your Neighbour's input) inequality provided in [101] is derived from the following bipartite game: Bob has the task of communicating a bit to Alice or guessing Alice's bit depending on the value of a classical random variable b'. If we assume that Alice causally precedes Bob or vice versa, or if the order is given by a classical mixture of the two causal orders, the probability of success for this game is always bounded by $3/4$. It can be shown that the framework of process matrices allows for a violation of causal inequalities [101, 20, 21]. The OCB protocol gives the archetypical example of a violation of causal inequalities. In this protocol, two agents freely choose the bits a, b, b' and record the output of their measurement in x, y:

- Alice: always measures the incoming qubit in the z basis, assigning the value $x = 0$ to the outcome 0 and $x = 1$ to the outcome 1. Then prepares a qubit encoding a in the computational basis.

- Bob: for $b' = 0$ he measures in the x basis and, if the outcome of the measurement is $|+\rangle$, encodes b in the z basis of the outgoing qubit as follows: $0 \mapsto |0\rangle$ and $1 \mapsto |1\rangle$. Otherwise, if the outcome of the measurement is $|-\rangle$, b is encoded as: $0 \mapsto |1\rangle$, $1 \mapsto |0\rangle$. The value of y is not relevant to the protocol and can be set arbitrarily. When $b' = 1$, Bob measures the incoming qubit in the z basis and saves the outcome of the measurement in the bit y independently on whether $b = 0$ or $b = 1$.

The receipt above can be turned into well defined classically controlled quantum instruments, which can be inputted into the process matrix given by:

$$W^{A_1A_2B_1B_2} = 1/4\left[1^{A_1A_2B_1B_2} + 1/\sqrt{2}\left(\sigma_z^{A_2}\sigma_z^{B_1} + \sigma_z^{A_1}\sigma_x^{B_1}\sigma_z^{B_2}\right)\right] \tag{1.65}$$

CAB	000	001	010	011	100	101	110	111
000	0	0	1/2	0	0	1/2	0	0
001	1/2	0	0	0	0	0	0	1/2
010	0	1/2	0	0	0	0	1/2	0
011	0	0	1/2	0	0	1/2	0	0
100	0	0	1/2	0	0	1/2	0	0
101	0	1/2	0	0	0	0	1/2	0
110	1/2	0	0	0	0	0	0	1/2
111	0	0	1/2	0	0	1/2	0	0

Figure 1.14: The maximally tripartite causally nonseparable process from [19]

CAB	000	001	010	011	100	101	110	111
000	1/4	0	0	1/4	1/4	0	0	1/4
001	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
010	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
011	1/4	0	0	1/4	0	1/4	1/4	0
100	0	1/4	1/4	0	0	1/4	1/4	0
101	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
110	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
111	1/4	0	0	1/4	0	1/4	1/4	0

Figure 1.15: A tripartite process where single parties do not signal to each other and C encodes a signal in the correlations of the outcomes measured by A and B.

The probability of success in the GYNI game is, in this case, given by $(2+\sqrt{2})/4 > 3/4$ proving the violation of a causal inequality.

Another interesting example can be found in [21]. In this work, Baumeler and Feix classify the set of 'classical' process matrices, i.e. process matrices which can be thought of as stochastic matrices from sets of classical inputs to sets of classical outputs and which are compatible with any classical local operation. Here by local operation we mean a stochastic process $P : A \times I \to O \times X$ where X is a classical output (i.e. the result of a measurement), A is a choice of intervention, I, O are respectively the local incoming and outgoing systems which are then fed into the classical process matrix.

Of particular interest is trying to understand the relationship between the 'internal' description of the exotic causality described by process matrices and the 'external' class of correlations that this formalism induces. It is well-known [13, 102] that causal non-separability, i.e. the impossibility of decomposing the general process matrix as two causally ordered super-operators, is not sufficient to induce a violation of causal inequalities. An example of this is provided by the process matrix representing the quantum switch, which cannot be written as a sum of two causally ordered matrices, and for which no choice of quantum instruments can lead to a violation of causal inequalities.

Most of the early literature on process matrices focuses on the bipartite case. Going beyond this assumption, the notion of definite causality becomes particularly subtle. In the three-partite case, we encounter situations where the causal order can be chosen contextually to some agents' choices. Similarly to signalling, this form of contextuality (not to be distinguished from the contextuality generalising non-locality) would have to be tamed to avoid influences from the future to the past. Oreshkov and Giarmatzi in [102] provide the following inductive description of multipartite definite causality, flexible enough to describe protocols where events can determine the causal orders of subsequent parties[4]:

Definition 1.32 (Multipartite causality [102]). *A conditional distribution $p(\underline{o}|\underline{i})$ is causal if it can be constructed inductively as follows:*

- *For $N = 1$, any valid probability distribution $P(a_1|x_1)$ is causal;*
- *For $N \geq 2$, an N-partite correlation is* causal *if and only if it can be decomposed in the form*

$$P(\underline{a}|\underline{x}) = \sum_{k \in N} q_k P_k(a_k|x_k) P_{k,x_k,a_k}(\underline{a}_{\setminus k}|\underline{x}_{\setminus k}) \tag{1.66}$$

  with $q_k \geq 0$ for each k, $\sum_k q_k = 1$ where (for each k) $P_k(a_k|x_k)$ is a single-party (and hence causal) probability distribution and (for each k,x_k,a_k), $P_k(a_k|x_k)P_{k,x_k,a_k}(\underline{a}_{\setminus k}|\underline{x}_{\setminus k})$ is a causal $(N-1)$-partite correlation.

This notion of definite causality is used by Abbott et al. [1] to generalise causal inequalities to an arbitrary number of parties. Abbott et al. also show that any causal process can be written as a convex sum of deterministic functions satisfying Equation 1.66. They form a polytope in which vertices are given by causal deterministic functions, and its facets define all causal inequalities.

The consequences of this simple observation are far-reaching even though not explicitly discussed in the paper; it suggests that when we allow maximal flexibility in the causal structure explaining the empirical behaviour, the difference between local and non-local scenarios vanishes. Every causal conditional distribution of outcomes satisfying causal inequalities can be described as arising from some underlying functional (classical) behaviour where hidden random variables obfuscate the causal structure. Any non-local behaviour can be explained classically if we allow explanations exhibiting arbitrary causal structures. We will explain that this ceases to be the case for more restrictive but still operationally meaningful causal assumptions, which preserve a crucial distinction between local and non-local processes allowing us to define *contextual causality*.

A focus on different types of subtleties that must be considered when talking about causal correlations is given in [2]: one of the problems associated with device independent verification of genuinely non-causal behaviour lies in the assumption that agents are 'causally localised'. We can

[4]Note that we take this definition in the form provided by [1]

expect such non-casual behaviour to vanish when two or more parties are coalesced, similarly to how certain tripartite correlations could have been considered local for some coarse-graining of the agents. One of the possibilities is, as done by the authors of [1], to tackle this problem by providing definitions for non-causal behaviours that cannot be coarse-grained into causal definiteness.

In this dissertation, we offer a description of 'definite' causality which subsumes the one provided in [102] and where the burden gets shifted from an inductive description of the allowed distributions to the description of topological spaces onto which data can be assigned to construct the relevant polytopes. We will substantially generalise the causal inequalities found in [1] and [2] to show how to derive a convex space of correlations from a plethora of operational assumptions. The observation from [1] that any causal process can be written as a convex sum of deterministic behaviours will then amount to the fact that the set of global sections for space of histories, which are associated with the indiscrete causal order, coincides with polytope obtained taking the convex hull of causaltope exhibiting definite causality (more about this in Chapter 5 and Chapter 6). In general, however, the requirement of satisfying causal inequalities may be too strict; substituting the indiscrete preorder with a different overarching preorder leads to the phenomenon of 'contextual causality', which will be exemplified in Chapter 6.

No 'realisable quantum process (even though there is some ambiguity in defining what a realisable quantum process means) is known for violating the most general causal inequalities. On the contrary, it has been proved in [118] and [141] that quantum processes endowed with the possibility of the quantum control of causal orders cannot violate them and, independently from the observation in [1], that they can always be decomposed as probability distributions on the set of deterministic functions with definite causality. The work by Wechs et al. [141] describes classes of circuits where the order of gates is classically and causally controlled, similarly to what we propose in [107] to allow a broader class of dynamically controlled causal orders.

There is an extensive literature on the study of how process matrices [75, 100, 31, 13, 1, 102, 12, 141] interface with causality. The generality of their definition trades off with the compositional ambitions of operational quantum theory: they must be thought of as describing the entire 'mediating spacetime' surrounding some sets of laboratories. We cannot input into the arguments of a process matrices two laboratories which are signalling without expecting the creation of paradoxical time loops. Therefore, a compositional theory of how general higher-order maps interact with non-trivial causal structures as inputs has to consider the different possible 'signalling types' of multipartite instruments. This is studied in [81, 23]. In [75], Oreshkov and Hoffreumon generalise process matrices to allow their inputs to be not just simple quantum instruments but general quantum combs with arbitrary causal structures.

1.4.3 Quantum and Classical Causal Modelling

A different approach from the attempts to frame quantum causality inside classical outcomes and interventions is based on describing it directly as a quantum mechanical phenomenon. For the classical case, Judea Pearl [104] considers a causal structure on a set of variables V to be described by a directed acyclic graph where each node corresponds to a random variable, and each directed edge witnesses causal dependency. A causal model $M = (D, \Theta_D)$ is a directed acyclic graph D and a set of causal parameters Θ_D given by a function for every $x_i \in V$ such that $x_i = f_i(pa_i, u_i)$, where pa_i denotes the set of causal parents of v_i in D and u_i a local disturbance described by a probability measure $p(u_i)$. Complete knowledge of all the values u_i and the structural equations can be used to deterministically predict the value of a node given values for its ancestors. The assumption that the local noise represented by the u_i's is independently distributed makes the casual model Markovian for D, i.e. the distribution on each variable (conditional on values for its parents) is independent of its non-descendants.

Fine-tuning, or stability as it is known in [104], is the property that the relationship of independence for a given causal explanation is stable with respect to a change of the causal parameters:

Definition 1.33 (Stability [104])**.** *Let $P(M)$ be the distribution arising from the causal model M, and let $I(P)$ denote the set of all conditional independence relationships embodied in P. A causal model $M = (D, \Theta_D)$ generates a stable distribution if and only $P(M)$ contains no extraneous independences - that is, if and only if $I(P((D, \Theta_D))) \subseteq I(P((D, \Theta'_D)))$ for any set of parameters Θ'_D.*

We cannot always recover a stable causal model explaining quantum correlations. Wood and Spekkens have provided a novel characterisation of Bell's theorem in [144] by showing that: choice independence (no super-deterministic correlation between the choices of the agents), local setting dependence (the local measurement output is allowed to be causally dependent on the settings), no-signalling (conditional independence of each local outcome from the distant local setting), and no fine-tuning (or stability) imply bell inequalities. A quantum violation of such inequalities witnesses the impossibility of stable causal discoveries. Cavalcanti extends this result, proving that Bell inequalities can be recovered from a subset of the original assumptions: no-signalling and no-fine-tuning [35, 105]. The result in [35] has been generalised in [105] to an arbitrary number of parties and general contextual scenarios, going beyond the particular structure of contexts imposed by nonlocality.

While works like [35, 144, 105] look at quantum correlations from the perspective of classical causal modelling, there have been several proposal of a quantum generalisation of causal modelling itself [17, 18, 36, 74, 106, 37, 10, 46], including quantum generalisations [17, 18] of Reichenbach's Principle [119]. Barrett, Lorentz and Oreshkov generalise Pearl's framework by going from a fundamentally functional description of causal mechanisms to unitary transformations, taking the causal relata to be 'loci of possible interventions', i.e. pairs of Hilbert spaces representing an incoming and outgoing system. One of the main consequences is a generalisation of the notion of the Causal Markov condition to the quantum domain:

Theorem 1.34 (Factorization of a unitary channel from no-influence conditions [17])**.** *Let $\rho_{B_1 \ldots B_k | A_1 \ldots A_n}$ be the CJ representation of a unitary channel with n input and k output systems. Let $S_i \subseteq \{A_1, \ldots, A_n\}, i = 1, \ldots, k$, be the k subsets of input systems such that there is no influence from the complementary sets to B_i. Then the operator factorizes in the following way*

$$\rho^U_{B_1, \ldots B_k | A_1 \ldots A_n} = \prod_{i=1}^{k} \rho_{B_1 | S_I} \tag{1.67}$$

where the marginal channels commute pairwise, $[\rho_{B_i|S_i}, \rho_{B_j|S_j}] = 0$ for all i, j.

Barrett et al. understand causal relationships as structural properties of unitary transformation. As such, the notion of causal influence is more general than requiring factorisability of a unitary matrix into causally faithful quantum circuits as in [46]. Causal nodes are not understood operationally, with compositional semantics, but are of a more decompositional flavour. While it is possible to show that the absence of a causal relation between A and B (respectively, a tensor factor of the input and output of the unitary operator) can be witnessed by a circuital realisation of the process that makes this particular causal disconnection evident, in general, no circuit faithfully models all the causal constraints simultaneously. Attempts to understand the decomposition of the unitary operators using a causally faithful graphical language can be found in [86, 139].

1.4.4 Contextuality and the sheaf-theoretic approaches

Almost simultaneously to the proof of the inadequacy of local realism, Simon Kochen and Ernst Specker tackled the classical quantum dichotomy without relying on the causal structure of a protocol. They considered the possibility of assigning deterministic values to measurements [132] in a way which adheres to our classical intuition. The failure of such classical assignments initiated the study of *quantum contextuality*. Similarly to non-locality, contextuality is a property exhibited by quantum systems which limits the possible classical explanation of a quantum phenomenon. Finding a clear understanding of how and where quantum formalism diverges from classical reasoning is as multifaceted as the study of the dichotomy between classical and quantum causality: many different notions of contextuality appeared in the literature, some of which will be briefly discussed in this section.

In their seminal paper, Kochen and Specker [132] explain that a reasonable notion of hidden variables would have to entail an embedding from the set of quantum observables to a commutative algebra (such as the space of real-valued functions of phase space). The set of quantum mechanical observables constitutes a partial algebra if we restrict the operation of sum and product to be defined on commutative operators, and they show that its partial subalgebra of idempotents cannot be embedded into a Boolean algebra and, as a consequence, that the desired embedding from the observables to a commutative algebra describing the assignment of a classical evaluation is impossible.

Kochen and Specker's work aims at proving a mathematical statement about quantum theory, and it differs in intentions from Bell's violation of local causality. This is because the theorem relies on assumptions about the algebraic structure of quantum observables, while Bell inequalities depend on general assumptions about the background's causal structure. Kochen and Specker's contextuality tells us that using quantum theory to describe specific empirical observations reveals an intrinsic degree of non-classicality implicit in the theory, keeping open, at least in principle, the possibility that a different theoretical description of the same empirical behaviour might cease to exhibit this contextual behaviour. A violation of Bell's inequality certifies (upholding the assumption of non-disturbance between spacelike separated interactions) the inadequacy of local causality as a descriptive principle of nature, independently of the theory used to describe the correlations.

Some operational assumptions will need to persist in both arguments. However, the cogent question is whether it is possible to develop a more general phenomenology of contextuality, which describes it as a general property of empirical observations under operational assumptions that are as theory-independent as possible.

Contextuality in quantum theory is studied with a plethora of different mathematical tools and representation [134, 6, 34, 51, 134, 135, 127]. It is beneficial for our research to make an initial distinction between two approaches which are often not very clearly differentiated in the literature. Using the terminology introduced by Budroni et al. in [33], contextuality can be understood either from the *effect perspective* or from the *observable perspective*. In the effect perspective, single measurements constitute contexts which are then inhabited by self adjoint positive operators. In the observable perspective, the content of these contexts inextricably depends on the high-level description of the theory, specifically on how the theory describes individual 'effects' updating the state upon the record of a measurement outcome. In the observable perspective, the primary objects are the observables themselves, to which one has to impose additional operational criteria that allow one to discriminate the sets of observables which constitute a context.

Any attempt to experimentally 'verify' contextuality is controversial, not only for the technical difficulties that 'noisy' measurements may entail, but also conceptually [79, 15, 96, 131]. The original aim of contextuality was to exclude a possible classical description of the standard quantum formalism. What happens if one wants to assume as little as possible about the theory explaining an empirical model? What structural assumptions must one impose to be able to speak about a contextual or a non-contextual explanation of raw data?

Kent and Barrett have eloquently expressed an argument questioning the relevance of contextuality in [15], where they suggest that the very notion of empirical verification of contextuality may not stand on substantial conceptual grounds:

> We only wish to note that the class of such alternatives is not merely as general and natural as the class of locally causal theories. So far as the project of verifying the contextuality of Nature is concerned (as opposed to the contextuality of hidden variable interpretations of standard quantum formalism), the question is of rather limited relevance and interest. [15]

In our work, we are not particularly worried about what constitutes a context or whether there is or there is not an experimentally meaningful way of justifying that two measurements lie in the same context. Contexts are merely considered part of the operational description of the experimental protocol, regardless of whether they come from causal assumptions related to no-signalling or more general assumptions of non-disturbance.

The observable and the effect perspective are equivalent for the ur-case, which was of interest to Kochen and Specker. For a three-dimensional Hilbert space, a family of commuting dichotomic observables $\{\Pi_i\}_{i=0,1,2}$ where P_i is the projector associated to the value 1 and $1 - P_i$ the projector associated to the value 0 can be equivalently described as a single observable with effects $\{P_0, P_1, P_2\}$. There is no particular benefit in adhering to a particular view, but this coincidence of perspectives might be an exclusive feature of quantum theory. Quantum nonlocality can be seen as a particular example of contextuality, and it naturally speaks the language of the 'operational perspective': contexts are given by sets of sharp measurements which must be non-disturbing due to the causal relationship between the two agents.

The approach developed by Spekkens in [134], generally known as Spekkens's contextuality, has been developed out of the 'effect perspective' to account for the possibility of noisy measurements and to provide contexts of effects with more solid operational foundations. Spekkens contextuality has been further developed, including several discussions about experimental verifiability in [83, 128, 126, 127, 95]. This redefinition of contextuality, even if it may, at first sight, seem akin to our operational spirit (also considering the extensive use of the notion of process theories [126, 127]) will not be relevant to our work, which will not assume the structure which is implicit in the definition of an operational theory or an ontic model.

Despite the apparent differences, Bell's nonlocality (focusing on correlations between spacelike separated events) and KS contextuality (dealing with ideal measurements that do not disturb the outcome statistics of compatible observables) can be easily reconciled. Abramsky and Brandenburger [6] make this unification explicit. Nonlocality and contextuality are instances of the same mathematical problem: finding a global section to some compatible family of distributions for a given presheaf.

The sheaf theoretic approach to non-locality is not only a robust unifying framework but also provides a general category-theoretical flavoured description of many of the notions ubiquitous in the literature about hidden variables. *Parameter independence*, *λ-independence*, *factorisability* and *determinism* are all cast out in great generality and the same mathematical arena using the language of monads and presheaves. The generality of the sheaf-theoretic machinery guarantees that once we redefine the topological spaces and the presheaves that are relevant to our investigation, the formalism will be automatically suited to generalise many of the results that already shaped the study of non-locality and contextuality and bring them into the domain of definite and indefinite causality.

One of the salient features of the Abramsky-Brandenburger approach is the possibility of discussing contextuality and nonlocality both qualitatively and quantitatively. The contextual fraction for a given empirical model e is defined in [5] as the maximal value of λ in the decomposition:

$$e = \lambda e^{NC} + (1 - \lambda)e' \tag{1.68}$$

where e^{NC} is a non-contextual model and e' is a non-signalling empirical model. It parametrises the distance between the classical polytope and the faces of the non-contextual polytope characterising *maximally contextual* scenarios.

For example, Figure 1(a) of [5] shows the degree of nonlocality exhibited by quantum instruments on the Bell state where the agents have the choice to either measure in the XY plane by an angle ϕ_1 or an angle ϕ_2 from the X axis in the clockwise direction. In the context of our work, these quantitative approaches will prove particularly relevant (in particular, we will reference Figure 1 of [5] again in Chapter 6). Quantifying contextuality is mainly concerned with the interplay between the classical and the contextual polytope. On the other hand, when considering causal scenarios, different polytopes represent different causal assumptions (causaltopes). The quantitative part of our sheaf-theoretic formalism will not only allow us to quantify the 'non-classical' behaviour but also to perform causal discovery. For example, by finding the decomposition of an empirical model for the polytopes associated with causally definite scenarios, we can quantify the degree of casual indefiniteness showcased by empirical data.

Most of the literature in quantum contextuality assumes hidden variables as being static; not much has been said about the possibility of hidden variables being allowed to evolve in time, let alone general causal scenarios. A notable exception is 1985 Leggett and Garg [84, 14, 92] work on macro-realism where, under some rather strong assumption of classicality: *macroscopic realism* and *non-invasive measurability* they construct an inequality aimed at ruling out macro-realist explanations.

The work of Legget-Garg is the first example where sequentiality in time and contextuality are seen as interconnected. Suppose that a quantity Q which can take values $-1, +1$ is measured at three different times t_0, t_1, t_2. In between the measurements, the system evolves according to its natural dynamics. To construct the table of conditional probabilities, we consider three contexts: the cases in which the measurements are performed at times (Q_0, Q_1), (Q_1, Q_2), (Q_0, Q_2) and record the respective outcomes. No-signalling from the future implies that:

$$p(Q_0 = q_i, Q_1 = q_j) = \sum_{q_k \in (1,-1)} p(Q_0 = q_i, Q_1 = q_j, Q_2 = q_k) \tag{1.69}$$

Furthermore, non-disturbing measurability can be mathematically rephrased as:

$$p(Q_0 = q_i, Q_1 = q_j) = \sum_{q_k \in (1,-1)} p(Q_0 = q_i, Q_1 = q_j, Q_2 = q_k) \tag{1.70}$$

$$p(Q_0 = q_i, Q_1 = q_j) = \sum_{q_k \in (1,-1)} p(Q_0 = q_i, Q_1 = q_j, Q_2 = q_k) \tag{1.71}$$

With these inequalities at hand, we can infer the following condition on the expected value of the correlations

$$-1 \leq \langle Q_0 Q_1 \rangle + \langle Q_0 Q_2 \rangle + \langle Q_1 Q_2 \rangle \leq 3 \tag{1.72}$$

Leggett-Garg shows that a violation of less than -1 can be obtained in quantum theory.

There is a bit of a controversy regarding what Leggett-Garg inequalities actually show. Maroney and Timpson [92] argue that despite the formal similarity of the approach to Bell's inequalities, they are methodological very different and cannot be used to rule out macro-realism in a suitable model-independent manner. The class of theories violating the inequality cannot be characterised in a simple way without appealing to Spekkens' operational formalism as done in [92]. The opinion of Timpson and Maroney is that if a methodological parallel between Bell and Leggett-Garg, then "it must be drawn between non-invasive measurability and non-locality". Bacciagaluppi has posed similar criticisms in [14]. Rephrasing Legget-Garg in our framework will support this view and show that a violation of the above inequalities is entirely ascribable to a failure of no-signalling and not to a generalised notion of contextuality as it is often thought.

Another work on this line is Mansfield and Kashefi's [89] proof that contextuality in time is necessary and sufficient to get an advantage for the deterministic computation of non-linear functions. A notion of non-contextuality in time can be recovered from the ontological model framework: a system is contextual when the empirical data obtained by sequential transformations cannot be reproduced by analogous modular transformation on the ontic states. In our work, the relationship between contextuality and causality is different; we never appeal to the existence of an ontic theory that describes the evolution of hidden variables but to mere logical consistency.

The relationship between contextuality and indefinite causality has been previously investigated in [130]. They approach the problem from the ontological framework perspective and show that, in general, it is not possible to construct an ontological theory which is both instrument and process non-contextual that can provide an explanation for all protocols described by process matrices.

From our perspective, however, which differs from Spekken's approach, contextuality is seen as the incompatibility of the empirical model expressed by quantum theory with a possible explanation in terms of classical functions coordinating the inputs with the outputs. As shown in [1], every process which does not violate causal inequalities can be deemed as non-contextual if we allow the classical explanation to exhibit arbitrary definite causality between the events. It follows that, at least for quantum controlled processes, the possibility of arbitrary causal explanation renders such correlations always local.

Chapter 2

Giving operational meaning to indefinite causality

2.1 Introduction

The first appearance of the notion of 'indefinite causality' in the context of quantum information can be traced back to [72]. Whether two events are timelike or spacelike separated depends on the spacetime metric, which is a dynamical variable in the context of GR. In a putative theory of quantum gravity, we expect to be able to promote these degrees of freedom from classical to quantum. In the seminal [72] Hardy attempts to provide a description of what *indefinite causal structure* could refer to:

> Indefinite causal structure is when there is, in general, no matter of fact as to whether the separation between two events is time-like or not. (Lucien Hardy [72])

These words suggest that a description of indefinite causality must be amenable to empirical testing. Matters of facts are extracted analytically from observational data. Suppose indefinite causality is witnessed by the impossibility of describing systems' evolution in time, as Hardy develops in his paper. In that case, this must come at odds with our empirical access to quantum phenomena being unavoidably described as evolving in time. Interactions giving rise to measurements take place in spacetime, and the empirical detectability of indefinite causality will be interconnected with the a priori causal assumptions making measurement well defined in the first place.

So can we even make sense of indefinite causality operationally? We will see in Chapter 5 and Chapter 6 what it means to certify indefinite causality from the bare correlations alone. In this chapter, however, we start with a more straightforward task: we extend a theory by giving semantics to indefinite causality from the bottom up. If this reminds us of the process matrix formalism, we underline that while they dictate a top-down approach by considering all causal correlations that are compatible with local quantum experiments, here we investigate how the concept of indefinite causality can be supplemented to a theory where compositionality and the importance of the monoidal composition of processes are at the forefront.

A discussion about the realisability and the advantages offered by indefinite causality in the context of quantum theory can be found in [39] where they introduce a hypothetical device which can coherently control the order of two arbitrary quantum channels: the *quantum switch*:

Definition 2.1 (Quantum switch [39])**.** *Given two arbitrary quantum channels f and g with Kraus forms $f(\rho) = \sum_i f_i \rho f_i^\dagger$ and $g(\rho) = \sum_j g_j \rho g_j^\dagger$. We say that* switching the causal order *of f and g gives the channel $\mathcal{W}_{f,g}(\sigma)$:*

$$\mathcal{W}_{f,g}(\sigma) = \sum_{i,j} W_{f_i,g_j} \sigma W_{f_i,g_j}^\dagger \tag{2.1}$$

where the Kraus operator W_{f_i,g_j} is given by:

$$W_{f_i,g_j} := |0\rangle\langle 0| \otimes (f_i \circ g_j) + |1\rangle\langle 1| \otimes (g_j \circ f_i) \tag{2.2}$$

Chiribella et al. point out that this definition is independent of the Kraus decomposition of f and g, and that the assignment $\mathcal{W} : f \otimes g \mapsto \mathcal{W}_{f,g}$ is linear and positive, thus being an example of a valid supermap (or process matrix using the language of [101]). The following years saw a proliferation of attempts to pinpoint the information theoretical and communication advantages provided by the ability of superposing channels and their causal order [99, 3, 40, 11, 125, 101] and the possibility of experimentally detecting such superpositions [91, 91, 41]. In particular, there has been some recent speculation about the possibility of realising a genuine quantum switch by coherently controlling the 'spatial degree of freedom' [103]. Unfortunately, the issue of delimiting a tight operational setting in which to interpret the results is not ordinarily viewed as a necessity, sometimes leading to misinterpretation of their physical significance. The goal of this section is first to provide a rigorous standpoint from which to discuss the operational phenomenology of causal superposition and control in quantum theory (without referring to the full generality offered by process matrices) and to provide a recipe to construct empirical models which exhibit exotic causality. These models will be used as proofs of concepts for our investigation in Chapter 6.

Reasoning about operational theories in the context where spacetime itself becomes a dynamical variable carries a number of additional complications. For example, thinking about the setup and outcomes of an experiment presupposes the existence of causally stable surroundings, where the notions of cause and consequence take their familiar form independently of the specific processes being performed. Failing these assumptions, how can we ensure that a mathematical model of quantum theory in the presence of dynamical spacetimes is empirically testable? This is an important question upon which many others stumbled before us. For example, the following reflection can be found in a prominent piece of literature on the application of sheaves and topoi to quantum theory [78, 52]:

"[A]round fifteen years ago, I came to the conclusion that the use of standard quantum theory was fundamentally inconsistent, and I stopped working in quantum gravity proper [...] [W]hat could it mean to 'measure' properties of space or time if the very act of measurement requires a spatiotemporal background within which it is made?" (Chris J. Isham [78])

We want to provide a sound way of constructing and characterising localised processes which do not take place against a fixed causal background but rather against a 'superposition' of causal backgrounds, determined by some "wave-function" over the set of all fixed causal backgrounds compatible with the processes in question. We do so by first defining a general notion of 'control' of processes, accommodating both the classical case—where the choice of process to execute can be captured by some hidden variable—and the coherent case—where it might not be possible to establish which one process was executed without reference to a specific measurement context. Armed with such a notion, we show how to construct superposition of diagrams, giving semantics to the execution of processes and operations against a dynamical causal background.

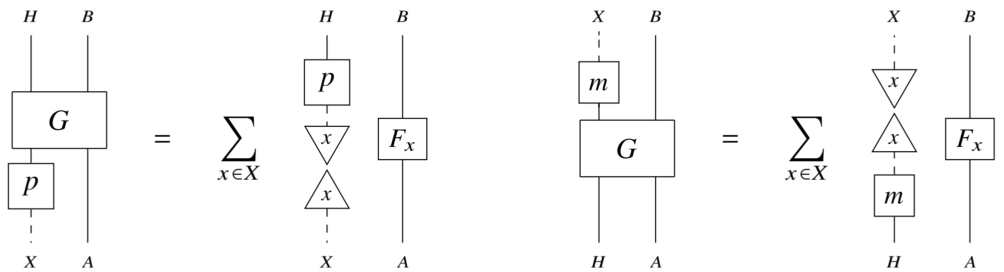
Control of causal orders is a case of a more general notion of control of quantum channels. The recent years have been very prolific for the study of the quantum control of processes [138, 3, 107, 49]. A problem underlined by several works is that the notion of the quantum control of a general family of quantum channels is itself ambiguous. An example of this is the claim made by [3] that *'the output of the interferometric circuit depends [...] on a more detailed description of the implementation of the channels'*. A similar problem has been highlighted by Daniel Oi [99], which unequivocally concludes that *'interferometry can be applied to the case of non-unitary processes to extract information about the underlying physical processes which implement them'*. This has been successively investigated in several foundationally flavoured works [3, 40, 99, 138, 107] and prima facie seems to impose a rupture of the notion of essentially unique purifications: channels acting on a single arm of an interferometer can be physically purified in 'inequivalent ways', giving rise to different observational statistics.

2.2 Controlled processes

Daniel Oi claimed [99] that by using interferometry to control quantum channels, the output interference pattern depends on their particular Kraus decomposition. On the other hand, we have seen that interferometry can hardly be interpreted as implementing the quantum control of a channel but refers to a genuinely different resource. Araujo et al. proved in [11] that there can be no well-defined higher order maps realising the control of channels, arguing that the circuit formalism should be extended to allow for the description of the elementary interferometric setup for quantum control. Similarly, Abbott et al. [3] explain this dependence on the Kraus decomposition as a physical phenomenon standing 'in contrast to the usual paradigm of quantum channel' and hinting at fundamental inequivalence of different purifications that would require the description of channels to be integrated by a *transformation matrix* depending on their physical realisation. We proceed to formally study the notion of control of quantum channels and clarify some of the conceptual fog surrounding it by providing an exact characterisation of the possible notion of 'quantum control'.

We start by spelling out the most general definition of a controlled family of processes. Our definition captures the idea of an agent being able to control the choice of morphisms by suitably encoding and/or decoding classical information about their choice into and/or from a physical system.

Definition 2.2. *Let C be a probabilistic theory, and let $A, B \in \text{obj } C$ be any two systems in the theory. Let $(F_x)_{x \in X}$ be a family of processes $F_x : A \to B$, not necessarily normalised or sub-normalised. A* ***controlled process*** *for the family is a triple (G, p, m) consists of a sharp preparation-observation (SPO) pair $(p : X \to H, m : H \to X)$—where X is a classical system and H is a generic system—together with a process $G : H \otimes A \to H \otimes B$ satisfying the following equations:*

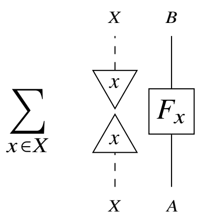


(2.3)

Conversationally, we will also say that (G, p, m) is a 'control of' the family $(F_x)_{x \in X}$.

The system H acts as a physical control system, while the classical system X contains the logical information about the process choice. The SPO pair is used to encode the logical information into the physical system and/or to decode it from the physical system. From the perspective of an actor using the SPO pair to encode/decode the logical information, the controlled process is no different from classical control.

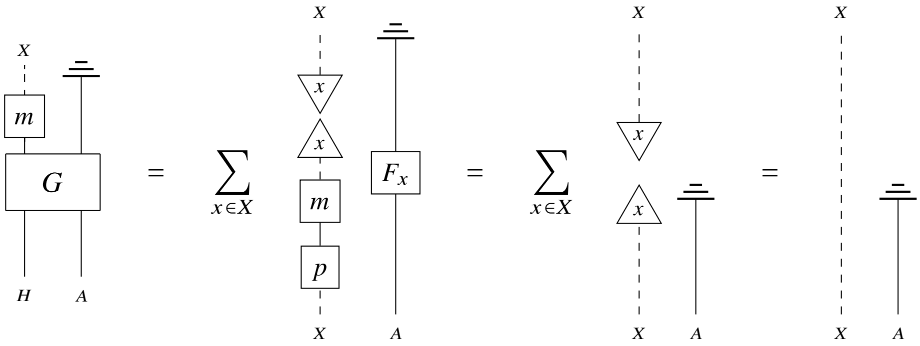
Example 2.3. *The following* ***classically controlled process*** *always exists in every probabilistic theory:*



(2.4)

Conversationally, we will also refer to the above as the *'classical control of' the family $(F_x)_{x \in X}$. Note that if (G, p, m) is any control of $(F_x)_{x \in X}$ then the triple $((m \otimes id_B) \circ G \circ (p \otimes id_A), id_X, id_X)$ is always the classical control of the same $(F_x)_{x \in X}$.*

The definition allows for much more general notions of control, as we shall shortly see, but it also limits the amount of leakage between the input/output systems A, B and the physical control system H. In particular, an agent without access to the output system B cannot, through the SPO pair alone, extract any information about the input state on system A if the maps F_x are normalised:

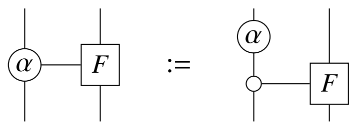


(2.5)

2.3 Coherent Control in Quantum Theory

Coherent control of families of pure maps is well defined up to the choice of a *phase gate* on the control system.

Example 2.4. *If $(F_x)_{x \in X}$ is a family of pure CP maps in quantum theory (e.g. isometries), a generic* ***coherently controlled process*** *for the family uses $\mathbb{C}^X$ as a control system and takes the following form:*

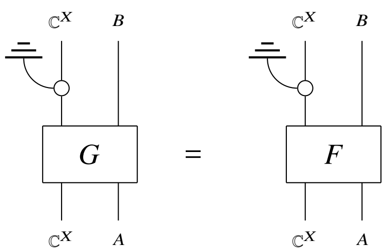


(2.6)

where α denotes an arbitrary phase in the canonical basis $(|x\rangle)_{x \in X}$ for the control system. Conversationally, we will also say that the above is a 'coherent control of' the family $(F_x)_{x \in X}$.

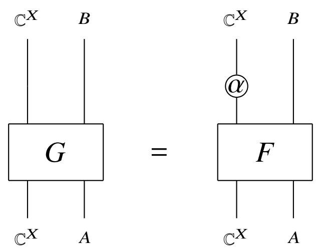
To prove the next proposition, we will need the following lemma, which establishes the uniqueness up to a phase of two bipartite completely positive maps when they differ by local decoherence:

Lemma 2.5. *Suppose that two pure completely positive channels are equal under local decoherence, so that:*



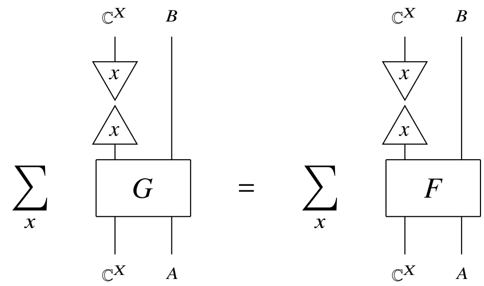
(2.7)

Then there exists a unitary phase $\textcircled{\alpha} : \mathbb{C}^X \to \mathbb{C}^X$ *such that*



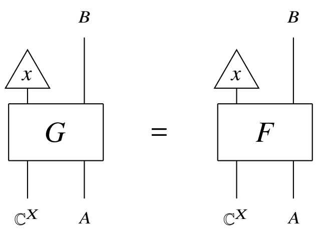
(2.8)

Proof. We start by expressing the decohered channels as a sum in which local projectors are applied to the decohered systems:



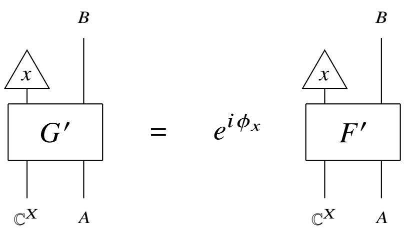
(2.9)

by postselecting by an arbitrary $|x\rangle$ we have that for all $(|x\rangle)_{x \in X}$:



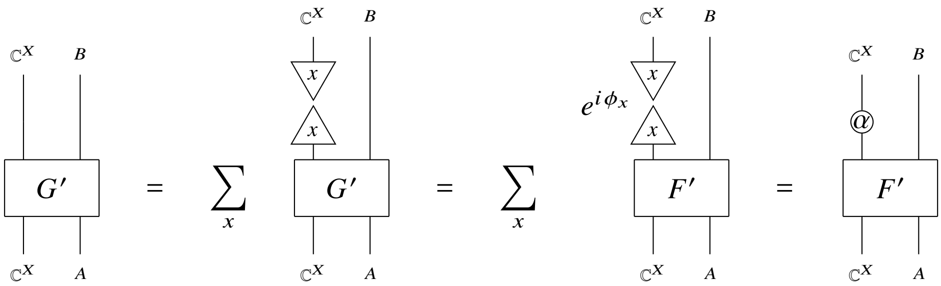
(2.10)

The equivalence between pure morphisms in the category of completely positive maps is witnessed by an equality of the underlying linear maps up to a global phase so that for all $|x\rangle$ in $\{|x\rangle\}_{x \in X}$ we get:



(2.11)

where G' and F' are two Kraus operators of the corresponding channels $G(\rho) = G'^{\dagger}\rho G'$ and $F(\rho) = F'^{\dagger}\rho F'$. Using Equation 2.11 we can show that the Kraus operators for the original pure channels differ by a phase gate:



(2.12)

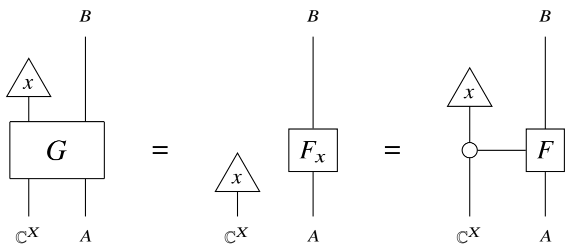
proving Equation 2.8. □

Proposition 2.5. *Let $(F_x)_{x\in X}$ be a family of pure CP maps in quantum theory, not necessarily normalised (i.e. not necessarily trace-preserving). Assume that (G, p, m) is a controlled process for the family with control system $\mathbb{C}^X$, where (p, m) is the SPO for the canonical basis $(|x\rangle)_{x\in X}$ and where G is itself a pure CP map. Then G takes form* (2.6) *for some phase α.*

Proof. In quantum theory, the normalised SPO pairs connecting a classical system X and the quantum system $\mathbb{C}^X$ are exactly the preparations and measurements in some orthonormal basis of $\mathbb{C}^X$. Without loss of generality, assume that the SPO pair is for the canonical basis $(|x\rangle)_{x\in X}$ of the system $\mathbb{C}^X$:

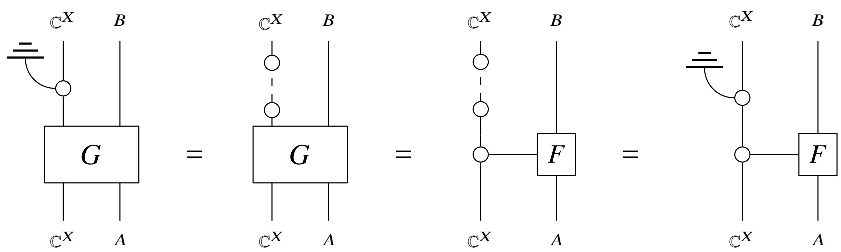
(2.13)

Now assume that (G, p, m) is a controlled process for the family $(F_x)_{x\in X}$. In particular, it satisfies the following equalities for all $x \in X$, where on the right we have used the coherent control of Example 2.4:



(2.14)

Equation 2.14 together with the resolution of classical identity implies that:



(2.15)

By Lemma 2.5 the two maps are equivalent up to a local phase, proving Proposition 2.5. □

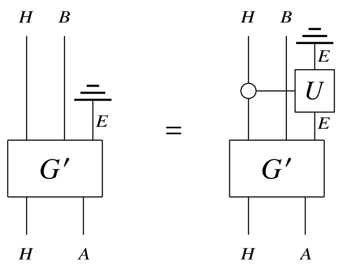
Moving away from pure maps, the question arises what the 'coherent' control for arbitrary CP maps should be. We adopt the following definition within our framework.

Definition 2.6. *Let $(F_x)_{x\in X}$ be a family of CP maps in quantum theory, not necessarily normalised (i.e. not necessarily trace-preserving). A* ***coherently controlled process*** *for the family is a controlled process (G, p, m) for the family such that G is obtained as $\mathrm{Tr}_E(G')$ for some pure CP map $G' : H \otimes A \to H \otimes B \otimes E$. Conversationally, we will also say that the above is a 'coherent control of' the family $(F_x)_{x\in X}$.*

If each map F_x in a family $(F_x)_{x\in X}$ of CP maps comes with a chosen purification $F_x = \mathrm{Tr}_E(\hat{F}_x)$—without loss of generality, using the same environment system E for all purifications—then it is easy to construct a coherent control of the family, by taking (G', p, m) to be the coherent control of the family $(\hat{F}_x)_{x\in X}$ of purifications and then discarding the environment E of G'. However, it would be desirable for such a construction to be a function of the family $(F_x)_{x\in X}$ alone, without dependence on additional information. As shown by Proposition 2.5, this is indeed possible when all CP maps in the family are pure. However, the following no-go result shows this to no longer be the case when generic CP maps are considered.

Proposition 2.6. *Let $(F_x)_{x\in X}$ be a family of CP maps in quantum theory. It is not generally possible to construct a coherent control of the family which is a function of the family $(F_x)_{x\in X}$ alone, i.e. one which is independent of a choice of purification for the CP maps in the family.*

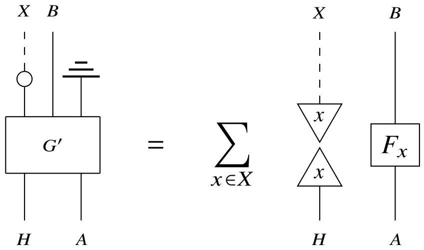
Proposition is the same as the statement that it is not generally possible to construct a coherent control $(\mathrm{Tr}_E(G'), p, m)$ of the family $(F_x)_{x\in X}$ in such a way that the following equation holds for all choices of unitaries $U_x : E \to E$:



(2.16)

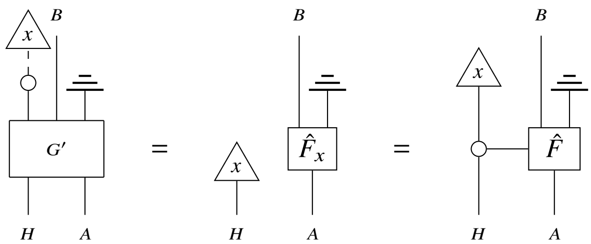
Note that the unitaries $(U_x)_{x\in X}$ correspond to all possible choices of purification for the CP maps $(F_x)_{x\in X}$.

Proof. Consider a coherent control $(\mathrm{Tr}_E(G'), p, m)$ of a family $(F_x)_{x\in X}$ of CP maps, where G' is pure. In particular, by Equation 2.3, the following equality is satisfied:



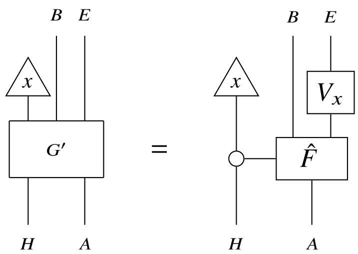
(2.17)

If $\hat{F}_x$ is a chosen purification of F_x for each $x \in X$—without loss of generality all with environment E—then we can create the coherent control of the family $(\hat{F}_x)_{x \in X}$ and obtain the following equality:



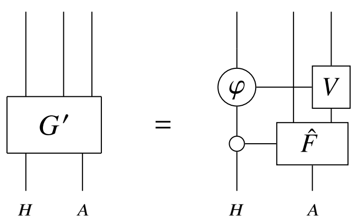
(2.18)

By essential uniqueness of purification, the following equality of pure CP maps must hold for some choice of unitary $V_x : E \to E$, dependent on each specific value of $x \in X$:



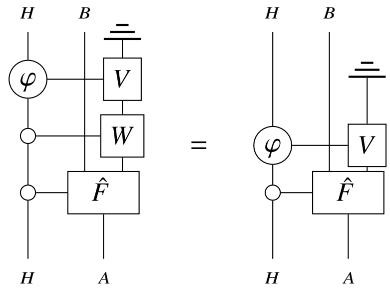
(2.19)

The equality above is equivalently an equality of linear maps up to a global phase φ_x, also dependent on each specific value of $x \in X$. We can therefore put all $x \in X$ together and obtain the following equality of pure CP maps:



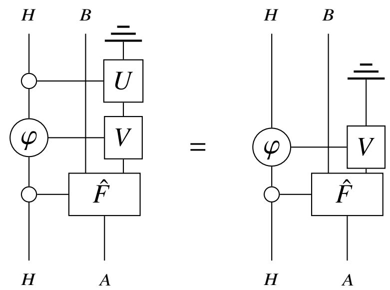
(2.20)

Any alternative choice of purification for each CP maps F_x can be obtained by applying some unitary $W_x : E \to E$ to the environment of our current choice of purification $\hat{F}_x$. If the coherent control is to be invariant under this choice of purification, the following equality must hold for all possible choices of unitaries $(W_x)_{x \in X}$:



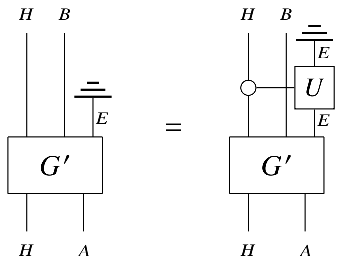
(2.21)

For each $x \in X$, we can define a unitary $U_x := V_x W_x V_x^\dagger$ such that $U_x V_x = V_x W_x$, so that the equality above for all possible choices of unitaries $(W_x)_{x \in X}$ can be equivalently recast as the equality below for all possible choices of unitaries $(U_x)_{x \in X}$:



(2.22)

The above is equivalent to Equation (2.16) in the statement of this Proposition holding for all $(U_x)_{x \in X}$:



(2.23)

For a general such G'—i.e. for a general choice of $(\hat{F}_x)_{x \in X}$—this equation cannot always be made to hold for all $(U_x)_{x \in X}$, leading to the statement of the proposition.

□

The classical control of arbitrary families of CP maps is trivially possible, as shown in Example 2.3. The formulation of Proposition 2.3 shows that, on the other hand, the question of coherently controlling families of CP maps is much more sophisticated, leading to some confusion in the literature about its feasibility.

Oi in [99] interprets the failure to construct such a coherent control independently of the choice of purification (aka choice of Kraus operators) as a sign that an interferometric realisation of such coherent control would extract information about the underlying physical implementation of the CP maps themselves. We have seen that this statement can be easily misinterpreted: the CP maps involved in the experiment are already the 'physical' ones—defined on the direct sum of the vacuum sector and the 1-particle sector—and the results of the experiment are independent of the choice of purification for them. This is obvious since the experiment itself can be easily written as a circuit. What the results of the experiment actually depend on is the choice of purification for the 'logical' CP maps involved, those restricted to the 1-particle sector. This is due to the specific design of the experiment: the implementation of the 'physical' CP maps is such that they react to the vacuum state on their input by emitting a non-vacuum state $|e\rangle$ on the environment E. This is not physically unreasonable, e.g. if the environment system comprises some static massive particle which is made to interact with the photons passing in the interferometric setup. However, this dependence on the choice of purification for the 'logical' CP maps goes away as soon as we allow the environment 'rest' state $|e\rangle$ to be transformed covariantly with the choice of purification, i.e. if we set $|e\rangle \mapsto U|e\rangle$ whenever we change the purification by applying a unitary $U : E \to E$ to the environment.

It is this last observation which helps us frame the discussion by [99] within the context of Proposition 2.3: given the two CP maps, acting on the vacuum and non-vacuum sectors respectively, it is very much possible to find a coherent control which is invariant under the application of the same unitary U to the environment of both purifications. This is always the case: if all $U_x : E \to E$ are chosen to be equal to some fixed U, then Equation (2.16) always holds (because U is trace-preserving). What is found to be impossible in the discussion by [99] is to choose such coherent control in a way which is invariant under application of U to the environment of the non-vacuum sector map and of the identity to the environment of the vacuum sector map. This issue indeed generalises and formed the inspiration for our proof of Proposition 2.3.

2.4 Definite and Indefinite Causal Scenarios

When operational scenarios with definite causal order are depicted diagrammatically in the context of probabilistic theories, it is easy to conflate the boxes in the diagrams with processes happening locally at events (i.e. points in spacetime), and the wires in the diagrams with the information flow establishing the causal relationships between said events. It has been previously argued [108] that this practice—though natural and notationally pleasant—is not mathematically well-founded, as there need not be a canonical way to decide how a process should be decomposed into a diagram compatibly with a given definite causal structure. As a consequence, two ingredients are needed when talking about such operational scenarios:

(i) a causal graph, representing the events in the scenario and their definite causal order;

(ii) a map assigning each event in the scenario to the process happening at that event.

The mathematical structure introduced in [108] as the substrate for such operational scenarios is that of *framed causal graphs*. For reasons which will become clear later on, we generalise the original definition to include the possibility of multiple edges between the same pairs of events. Furthermore, we include some additional information about the classical interface of the local processes/experiments, in the form of finite sets of input values that can be used to control them and output values for their outcomes.

Definition 2.7. *A **framed multigraph** is an directed multigraph Γ [1] equipped with the following data:*

- *a sub-set* $\mathrm{in}(\Gamma) \subseteq \mathrm{nodes}(\Gamma)$ *of the nodes of* Γ*—the **input nodes**—such that each* $x \in \mathrm{in}(\Gamma)$ *has zero incoming edges and a single outgoing edge;*
- *a sub-set* $\mathrm{out}(\Gamma) \subseteq \mathrm{nodes}(\Gamma)$ *of the nodes of* Γ*—the **output nodes**—such that each* $x \in \mathrm{out}(\Gamma)$ *has zero outgoing edges and a single incoming edge;*
- *a **framing** for* Γ*, which consists of the following:*
  - *a total order on* $\mathrm{in}(\Gamma)$*;*
  - *a total order on* $\mathrm{out}(\Gamma)$*;*
  - *for each node* $x \in \mathrm{nodes}(\Gamma)$*, a total order on the edges outgoing from* x*;*
  - *for each node* $x \in \mathrm{nodes}(\Gamma)$*, a total order on the edges incoming to* x*;*

We refer to nodes in $\mathrm{in}(\Gamma)$ *or in* $\mathrm{out}(\Gamma)$ *as **boundary nodes** and to all other nodes in* Γ *as **internal nodes**. An **acyclic framed multigraph** is a framed multigraph which is acyclic (and in particular has no loops).*

Remark 2.8. *The input and output nodes of a framed multigraph are designed to behave as 'half-edges': when two framed multigraphs* Γ *and* Γ' *are composed sequentially, the outputs of* Γ *and the inputs of* Γ' *are joined and disappear, each pair of corresponding output/input resulting in a single edge of the composite framed multigraph* $\Gamma' \circ \Gamma$*. (We do not use such composition here.)*

Definition 2.9. *A **definite causal scenario** is a triple* $\Theta = (\Gamma, \underline{I}, \underline{O})$ *of an acyclic framed multigraph* Γ *with:*

- *a finite set* I_ω *of **classical inputs** for each* $\omega \in \mathrm{ev}(\Theta)$*, i.e. the values available locally to control the process at the event;*
- *a finite set* O_ω *of **classical outputs** for each* $\omega \in \mathrm{ev}(\Theta)$*, i.e. the values that the process at the event can return locally as its outcome.*

[1] A directed multigraph Γ consists of a set $\mathrm{nodes}(\Gamma)$, a set of $\mathrm{edges}(\Gamma)$ and a pair of functions $\mathrm{tail} : \mathrm{edges}(\Gamma) \to \mathrm{nodes}(\Gamma)$ and $\mathrm{head} : \mathrm{edges}(\Gamma) \to \mathrm{nodes}(\Gamma)$ specifying the tail and head of each edge respectively.

*In the above, we have defined the **events** in the scenario as the set* $\mathrm{ev}(\Theta) := \mathrm{nodes}(\Gamma) \setminus (\mathrm{in}(\Gamma) \sqcup \mathrm{out}(\Gamma))$ *of internal nodes for* Γ.

Remark 2.10. *Compared to the original [108], we have restricted our attention to* chronology respecting *scenarios, i.e. those corresponding to acyclic framed multigraphs. However, one could easily extend Definition 2.9 to one for* chronology violating *scenarios, by allowing the framed multigraph to be cyclic and/or to have loops.*

We now define exactly what it means to 'draw a diagram over' one such definite causal scenario, with semantics valid in any probabilistic theory.

Definition 2.11. *Let* $\Theta = (\Gamma, \underline{I}, \underline{O})$ *be a definite causal scenario and let* C *be a probabilistic theory. A **diagram over** Θ **in** C is a pair of functions* $\mathrm{sys} : \mathrm{edges}(\Gamma) \to \mathrm{obj}\, C$ *and* $\mathrm{proc} : \mathrm{ev}(\Gamma) \to \mathrm{mor}(C)$, *associating each* $e \in \mathrm{edges}(\Gamma)$ *to a system* $\mathrm{sys}(e)$ *in* C *and each* $\omega \in \mathrm{ev}(\Gamma)$ *to a process* $\mathrm{proc}(\omega)$ *in* C *with the following type:*

$$\mathrm{proc}(\omega) : I_\omega \otimes \bigotimes_{e \in \mathrm{in}(\omega)} \mathrm{sys}(e) \longrightarrow O_\omega \otimes \bigotimes_{e' \in \mathrm{out}(\omega)} \mathrm{sys}(e') \tag{2.24}$$

Above we denoted by $\mathrm{in}(\omega) := \{ e \in \mathrm{edges}(\Gamma) \mid \mathrm{head}(e) = \omega \}$ *the edges of* Γ *coming into* ω *and we similarly denoted by* $\mathrm{out}(\omega) := \{ e \in \mathrm{edges}(\Gamma) \mid \mathrm{tail}(e) = \omega \}$ *the edges of* Γ *going out of* ω.

Even though it specifies concrete processes in a probabilistic theory, the definition of diagram Δ above is still partly abstract, as it does not explicitly state how the processes fit together. What gives it fully concrete semantics is the following definition of the overall process $[\![\Delta]\!]$ *associated* to the diagram Δ. See Figure 2.1 (p.73) for an exemplification.

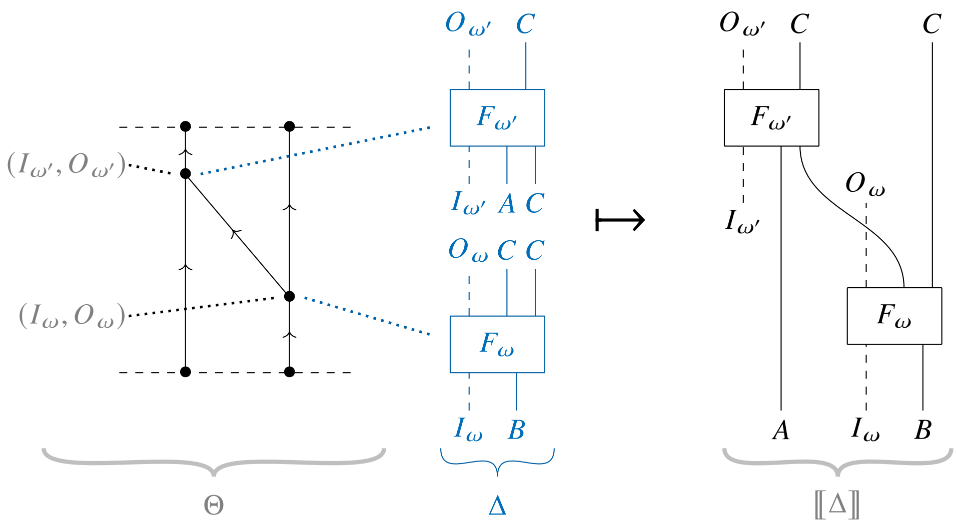


Figure 2.1: Graphical exemplification of the association of the process $[\![\Delta]\!]$ to a diagram Δ over a definite causal scenario Θ.

Definition 2.12. *Let $\Theta = (\Gamma, \underline{I}, \underline{O})$ be a definite causal scenario, $\mathcal{C}$ be a probabilistic theory and $\Delta = (\mathrm{sys}, \mathrm{proc})$ be a diagram over Γ in $\mathcal{C}$. The* ***process associated with*** Δ *is the unique process $[\![\Delta]\!]$ in $\mathcal{C}$ obtained by joining the outputs and inputs of the processes* $\mathrm{proc}(\omega)$ *in the diagram Δ according to the directed multigraph Γ, resulting in a process with the following overall type:*

$$[\![\Delta]\!] : \left(\left(\bigotimes_{\omega \in \mathrm{ev}(\Gamma)} I_\omega\right) \otimes \left(\bigotimes_{e \in \mathrm{in}(\Gamma)} \mathrm{sys}(e)\right)\right) \longrightarrow \left(\left(\bigotimes_{\omega \in \mathrm{ev}(\Gamma)} O_\omega\right) \otimes \left(\bigotimes_{e' \in \mathrm{out}(\Gamma)} \mathrm{sys}(e')\right)\right) \quad (2.25)$$

Each classical input system I_ω of $[\![\Delta]\!]$ is wired to the classical input system I_ω of the process $\mathrm{proc}(\omega)$. *Similarly, each classical output system O_ω of $[\![\Delta]\!]$ is wired to the classical output system O_ω of* $\mathrm{proc}(\omega)$.

Definite causal scenarios and diagrams over them are perfectly adequate when it comes to discussion of operational scenarios over definite causal orders, but they don't have the necessary flexibility to accommodate operational scenarios where causality is indefinite. The main contribution of this work will now be to define diagrams over *indefinite causal scenarios*, giving them semantics in probabilistic theories using controlled processes. As a special case, we will be able to describe the idea of superposition of causal orders in quantum theory.

Firstly, we define the purely operational canvas against which the indefinite causal scenario takes place. This includes all the black-box information locally available to the actors in our scenario but does not include any information about causal order nor any information about the specific implementation of the local processes.

Definition 2.13. *An* ***indefinite causal scenario*** Φ *is specified by the following data:*

- *the set Ω of* ***events*** *at which the processes (operations, experiments, etc.) take place;*
- *a set Ξ of* ***system labels****, used to abstractly indicate which physical systems are guaranteed to be the same across different experiments;*
- *a finite set I_ω of* ***classical inputs*** *for each $\omega \in \Omega$, i.e. the values available locally to control the process at the event;*
- *a finite set O_ω of* ***classical outputs*** *for each $\omega \in \Omega$, i.e. the values that the process at the event can return locally as its outcome;*
- *a finite sequence Σ^{in}_ω of elements of Ξ for each $\omega \in \Omega$, the system labels for the physical systems coming into the process from outside the event;*
- *a finite sequence Σ^{out}_ω of elements of Ξ for each $\omega \in \Omega$, the system labels for the physical systems coming out of the process and leaving the event;*

- *a finite sequence* Π^{in} *of elements of* Ξ*, the system labels for the physical systems coming in from outside the region where the scenario is taking place;*
- *a finite sequence* Π^{out} *of elements of* Ξ*, the system labels for the physical systems going out from the region where the scenario is taking place.*

Formally, the scenario is the tuple $\Phi = (\Omega, \Xi, \underline{I}, \underline{O}, \underline{\Sigma}^{in}, \underline{\Sigma}^{out}, \Pi^{in}, \Pi^{out})$*, where the underlined letters indicate* Ω*-indexed families (e.g.* $\underline{I} := \omega \mapsto I_\omega$*).*

Remark 2.14. *Even though the definitions involved are formally different, the definite causal scenarios of Definition 2.9 arise naturally as a special case of the indefinite causal scenarios from Definition 2.13. Indeed, consider an indefinite causal scenario* $\Phi = (\Omega, \Xi, \underline{I}, \underline{O}, \underline{\Sigma}^{in}, \underline{\Sigma}^{out}, \Pi^{in}, \Pi^{out})$ *and assume that each symbol* $e \in \Xi$ *appears exactly twice as follows:*

- *it appears once either in an output set* $\Sigma^{out}_{\mathrm{tail}(e)}$ *for some* $\mathrm{tail}(e) \in \Omega$ *or otherwise at some place* $\mathrm{tail}(e) \in \{1, ..., |\Pi^{in}|\}$ *in the sequence* Π^{in}*;*
- *it appears once either in an input set* $\Sigma^{in}_{\mathrm{head}(e)}$ *for some* $\mathrm{head}(e) \in \Omega$ *or otherwise at some place* $\mathrm{head}(e) \in \{1, ..., |\Pi^{out}|\}$ *in the sequence* Π^{out}*.*

This defines a framed multigraph Γ *with* $\mathrm{nodes}(\Gamma) := \Omega \sqcup \{1, ..., |\Pi^{in}|\} \sqcup \{1, ..., |\Pi^{out}|\}$ *and* $\mathrm{edges}(\Gamma) := \Xi$*. The internal nodes of* Γ *are the events in* Ω*, yielding a definite causal scenario* $(\Gamma, \underline{I}, \underline{O})$*. Conversely, each definite causal scenario* $(\Gamma, \underline{I}, \underline{O})$ *can be turned into an indefinite causal scenario by taking* $\Xi := \mathrm{edges}(\Gamma)$ *and defining the sequences* Σ^{in}_ω*,* Σ^{out}_ω*,* Π^{in} *and* Π^{out} *from* $\mathrm{in}(\omega)$*,* $\mathrm{out}(\omega)$*,* $\mathrm{in}(\Gamma)$ *and* $\mathrm{out}(\Gamma)$ *respectively.*

Secondly, we define the set of definite causal scenarios which are *compatible* with a given indefinite causal scenario: they correspond exactly to all possible ways of joining the output and input physical systems into a multigraph in such a way as to respect the system labels for the indefinite causal scenario. Each indefinite causal scenario gives rise to a set of compatible definite causal scenarios, each definite causal scenario equipped with a labelling associating each edge of the multigraph to the corresponding system label from the indefinite causal scenario. See Figure 2.2 (p.76) for an exemplification.

Definition 2.15. *Let* $\Phi = (\Omega, \Xi, \underline{I}, \underline{O}, \underline{\Sigma}^{in}, \underline{\Sigma}^{out}, \Pi^{in}, \Pi^{out})$ *be an indefinite causal scenario. A definite causal scenario* $\Theta = (\Gamma, \underline{I}', \underline{O}')$ *is* ***compatible*** *with* Φ *if the following conditions hold:*

(i) the events of Θ *(internal nodes of* Γ*) are the events of the scenario, i.e. we have* $\mathrm{ev}(\Theta) = \Omega$*;*

(ii) we have $\underline{I}' = \underline{I}$ *and* $\underline{O}' = \underline{O}$*;*

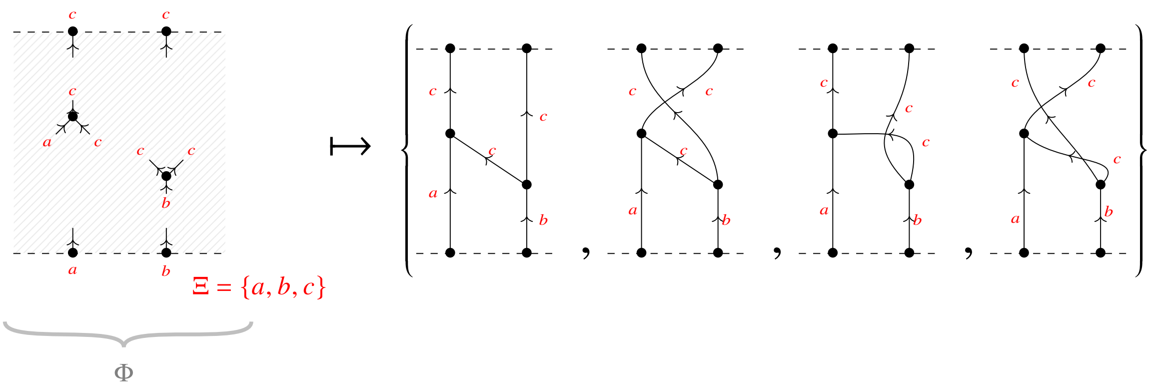


Figure 2.2: Graphical exemplification of the association of definite causal scenarios to an indefinite causal scenario Φ.

(iii) *for each* $\omega \in \Omega$, $\mathrm{in}(\omega)$ *and* Σ^{in}_{ω} *have the same number of elements; write* $\Sigma^{in}_{\omega}(e)$ *for the system label in the totally ordered set* Σ^{in}_{ω} *at the same position as edge* e *in the totally ordered set* $\mathrm{in}(\omega)$*;*

(iv) *for each* $\omega \in \Omega$, $\mathrm{out}(\omega)$ *and* Σ^{out}_{ω} *have the same number of elements; write* $\Sigma^{out}_{\omega}(e)$ *for the system label in the totally ordered set* Σ^{out}_{ω} *at the same position as edge* e *in the totally ordered set* $\mathrm{out}(\omega)$*;*

(v) *the input nodes* $\mathrm{in}(\Gamma)$ *and* Π^{in} *have the same number of elements; write* $\Pi^{in}(e)$ *for the system label in the totally ordered set* Π^{in} *at the same position as* $\mathrm{tail}(e)$ *in the totally ordered set* $\mathrm{in}(\Gamma)$*;*

(vi) *the output nodes* $\mathrm{out}(\Gamma)$ *and* Π^{out} *have the same number of elements; write* $\Pi^{out}(e)$ *for the system label in the totally ordered set* Π^{out} *at the same position as* $\mathrm{head}(e)$ *in the totally ordered set* $\mathrm{out}(\Gamma)$*;*

(vii) *for each edge* $e \in \mathrm{edges}(\Gamma)$*, the system label at its tail and at its head coincide, i.e. we have* $\Sigma^{out}_{\mathrm{tail}(e)}(e) = \Sigma^{in}_{\mathrm{head}(e)}(e)$*; by convention, we set* $\Sigma^{out}_{\mathrm{tail}(e)}(e) := \Pi^{in}(e)$ *when* $\mathrm{tail}(e)$ *is an input node and* $\Sigma^{in}_{\mathrm{head}(e)}(e) := \Pi^{out}(e)$ *when* $\mathrm{head}(e)$ *is an output node (both can be true at the same time).*

A definite causal scenario Θ *which is compatible with the indefinite causal scenario* Φ *comes equipped with an edge labelling* $\mathrm{syslabel}_{\Phi,\Theta} : \mathrm{edges}(\Gamma) \to \Xi$*, sending each edge* $e \in \mathrm{edges}(\Gamma)$ *to the system label* $\mathrm{syslabel}_{\Phi,\Theta}(e) \in \Xi$ *which the indefinite causal scenario associates to the endpoints of the edge. We write* $\mathrm{DCaus}(\Phi)$ *for the set of definite causal scenarios compatible with* Φ.

Finally, we are in a position to define what it means to draw a diagram over an *indefinite* causal scenario Φ. This is a generalisation of the abstract notion of drawing a diagram over a definite causal scenario from Definition 2.11: processes are associated with the events, but now taking additional care that the induced diagrams over all *definite* scenarios compatible with Φ are well-defined. See Figure 2.3 (p.77) for an exemplification.

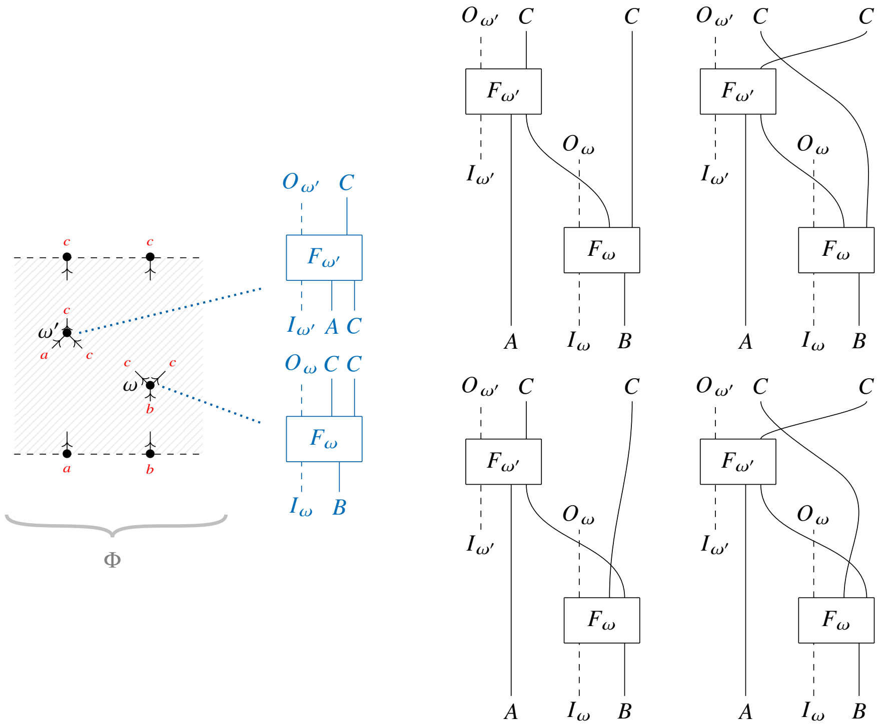


Figure 2.3: Graphical exemplification of how a diagram over an indefinite causal scenario Φ gives rise to diagrams over the compatible definite causal scenarios (cf. Figure.

Definition 2.16. *Let* $\Phi = (\Omega, \Xi, \underline{I}, \underline{O}, \underline{\Sigma}^{in}, \underline{\Sigma}^{out}, \Pi^{in}, \Pi^{out})$ *be an indefinite causal scenario and let* $\mathcal{C}$ *be a probabilistic theory. A* ***diagram over*** Φ ***in*** $\mathcal{C}$ *is a pair of functions* $\mathrm{sys} : \Xi \to \mathrm{obj}\,\mathcal{C}$ *and* $\mathrm{proc} : \Omega \to \mathrm{mor}\,(\mathcal{C})$*, associating each system label* ξ *to a system* $\mathrm{sys}\,(\xi)$ *in* $\mathcal{C}$ *and each event* $\omega \in \Omega$ *to a process* $\mathrm{proc}\,(\omega)$ *in* $\mathcal{C}$ *with the following type:*

$$\mathrm{proc}\,(\omega) : I_\omega \otimes \bigotimes_{\xi \in \Sigma^{in}_\omega} \mathrm{sys}\,(\xi) \longrightarrow O_\omega \otimes \bigotimes_{\xi' \in \Sigma^{out}_\omega} \mathrm{sys}\,(\xi') \tag{2.26}$$

If Δ *is such a diagram and* $\Theta = (\Gamma, \underline{I}, \underline{O})$ *is a definite causal scenario compatible with* Φ*, then the* ***induced diagram*** $\Delta|_\Theta = (\mathrm{sys}|_\Theta, \mathrm{proc}|_\Theta)$ *over* Θ *is defined by setting* $\mathrm{sys}|_\Theta : \mathrm{edges}\,(\Gamma) \to \mathrm{obj}\,\mathcal{C}$ *to be* $\mathrm{sys}|_\Theta\,(e) := \mathrm{sys}\left(\mathrm{syslabel}_{\Phi,\Theta}\,(e)\right)$ *and setting* $\mathrm{proc}|_\Theta : \mathrm{ev}\,(\Theta) \to \mathrm{mor}\,(\mathcal{C})$ *to be* $\mathrm{proc}|_\Theta\,(\omega) := \mathrm{proc}\,(\omega)$.

The semantics for a diagram over an indefinite causal scenario Φ will no longer be given by a single process with a definite causal order—as was the case for the semantics of diagrams over definite scenarios—but rather a controlled process for the family of all diagrams over all definite scenarios compatible with Φ.

Definition 2.17. *Let Φ be an indefinite causal scenario, C be a probabilistic theory and $\Delta = (\mathrm{sys}, \mathrm{proc})$ be a diagram over Φ in C. A* ***controlled process associated with*** *Δ is a controlled process (G, p, m) in C associated with the family $(\llbracket \Delta|_\Theta \rrbracket)_{\Theta \in \mathrm{DCaus}(\Phi)}$ of induced diagrams over all definite causal scenarios Θ compatible with Φ.*

It is worth noting that the semantics for diagrams over indefinite causal scenarios result in a *controlled* process (G, p, m): the control system is left open on the side, allowing preparations and observations to be used to control the causal order in all possible ways. Regardless of the specific controlled process and regardless of the specific probabilistic theory, the following will always work.

If we pre-compose the controlled process G with p on the control system we are able to classically control the causal order. Specifically, feeding a specific value $\Theta \in \mathrm{DCaus}(\Phi)$ as input to p results in the diagram $\Delta|_\Theta$ for the definite causal scenario Θ. More generally, feeding a a probability distribution $\sum_{\Theta \in \mathrm{DCaus}(\Phi)} \mathbb{P}(\Theta)\delta_\Theta$ as input to p results in a convex mixture of the diagrams associated with the definite scenarios, each diagram $\Delta|_\Theta$ happening with probability $\mathbb{P}(\Theta)$.

If we pre-compose the controlled process G with a normalised state ρ on the control system and we post-compose it with the observation m on the control system, we obtain again a convex mixture of the diagrams associated with the definite scenarios, with probability distribution given by the classical state $m \circ \rho$. If instead of the observation m we use the discarding map on the control system, we obtain the same convex mixture, but now without being able of extracting information about the causal order from the classical outcome of the observation m. [2]

The last observation, that discarding the control system always yields a convex mixture of causal orders, will play an important role in the next section, when we construct superpositions of causal orders in quantum theory. Indeed, the observation implies that we cannot obtain superposition of causal orders by preparing the control system in a superposition and then discarding it after we are done. This is because the act of discarding is an *epistemic* one: it simply means that information about the system is not locally available, not that it can never be recovered.

2.5 Superposition of Causal Orders

In this Section, we show how quantum superposition of causal orders can be modelled within our framework. We start by looking at the *quantum switch* and then proceed to generalise our construction to arbitrary indefinite causal scenarios. We conclude by showing that superpositions of causal orders can be constructed purely as a function of the quantum instruments operated at the events, with no dependence on a choice of purification for the CP maps. This is somewhat surprising—in light of Proposition 2.3—and it is a consequence of the fact that each discarded environment refers to the

[2] This is because applying the discarding map on the control system is the same as first applying the observation m and then discarding its classical outcome, resulting in a mixture.

same local CP map in all branches of the superposition (something which is not true in general coherent control of CP maps, e.g. in the circumstances considered by [99]).

In the n-partite generalisation of the quantum switch, n parties operate their quantum instruments sequentially on the same physical system, resulting in a superposition of $n!$ causal orders (corresponding to all possible permutations of the parties). It is now straightforward to model such a scenario within our framework.

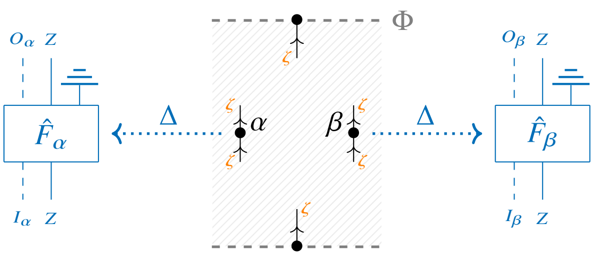
Definition 2.18. *An **n-partite switch** is an indefinite causal scenario satisfying the following conditions:*

- *there are n events, corresponding to the n parties;*
- *the set Ξ has a single element, as the same physical system is operated upon by all parties;*
- *the sequences Σ_ω^{in}, Σ_ω^{out}, Π^{in} and Π^{out} each have a single element, forcing the parties to operate on the same physical system one after the other.*

The classical inputs I_ω and classical outputs O_ω are free to choose.

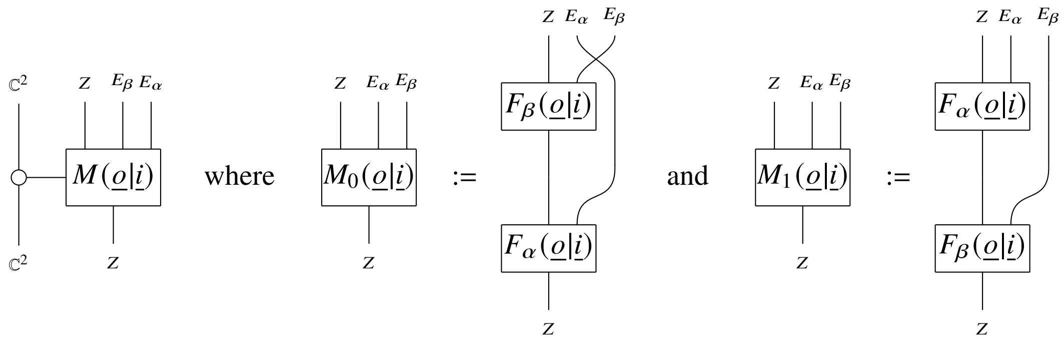
If Φ is an n-partite switch, the semantics of a diagram Δ over Φ in a probabilistic theory is given by a controlled process where each choice of permutation $\sigma \in S_n$ for the parties $\{1, ..., n\}$ results in party $\sigma(1)$ acting first, followed by party $\sigma(2)$, followed by all other parties in order until party $\sigma(n)$.

For the sake of simplicity, we will now restrict our attention to the bipartite ($n = 2$) case. The two parties are our beloved Alice and Bob, the corresponding events are called α and β, the physical system will be some quantum system Z and the quantum instruments operated by Alice and Bob will be $F_\alpha : I_\alpha \otimes Z \to O_\alpha \otimes Z$ and $F_\beta : I_\beta \otimes Z \to O_\beta \otimes Z$ respectively. Each quantum instrument $F_\omega : I_\omega \otimes Z \to O_\omega \otimes Z$ (for $\omega \in \{\alpha, \beta\}$) is defined by a family of CP maps $F_\omega(o_\omega|i_\omega) : Z \to Z$ indexed by each possible classical input value $i_\omega \in I_\omega$ and classical output value $o_\omega \in O_\omega$, subject to the normalisation requirement that $\sum_{o_\omega \in O_\omega} F_\omega(o_\omega|i_\omega)$ be a CPTP map for each choice of classical input $i_\omega \in I_\omega$. Let $\hat{F}_\omega(o_\omega|i_\omega) : Z \to Z \otimes E_\omega$ be a family of purifications for the CP maps, chosen (without loss of generality) to all have the same environment E_ω. This results in the following scenario Φ and diagram Δ:



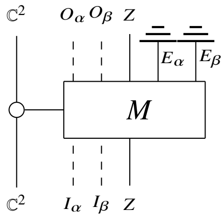
(2.27)

We can construct a controlled process associated with Δ by using our definition of coherently controlled processes from Section 2.3. Specifically, for each fixed $\underline{i} = (i_\alpha, i_\beta) \in \underline{I}$ and $\underline{o} = (o_\alpha, o_\beta) \in \underline{O}$ we can define the following pure controlled processes from the purifications:



(2.28)

If we discard the two environments E_α and E_β and we reintroduce the classical inputs and outputs, we obtain the controlled process for Δ:



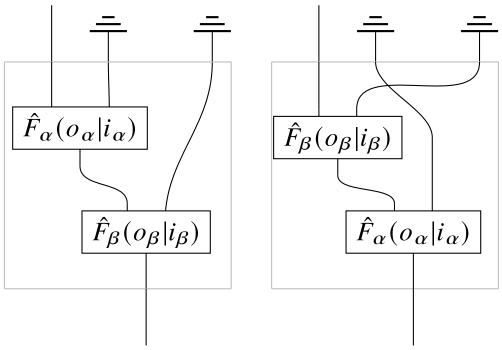
(2.29)

As mentioned in the previous Section, the controlled process above is very general: amongst other things, we can plug any qubit state into the control systems and perform any measurement on it afterwards. In order to obtain a true superposition of the two causal orders, we use some phase state $|\varphi\rangle := \frac{1}{\sqrt{2}}(|0\rangle + e^{i\varphi}|1\rangle)$ and measure in the Pauli X basis. The two measurement outcomes $\langle\pm|$ then correspond to families of processes—indexed by the classical inputs $\underline{i} \in \underline{I}$ and classical outputs $\underline{o} \in \underline{O}$—which see a superposition with phase of the two causal orders. In traditional notation, the processes can be written as follows:

$$\frac{1}{4}\,\mathrm{Tr}_{E_\alpha \otimes E_\beta}\Big[\mathrm{dbl}\Big[\Big(\hat{F}_\alpha(o_\alpha|i_\alpha) \otimes id_{E_\beta}\Big) \circ \hat{F}_\beta(o_\beta|i_\beta)\Big] \pm$$
$$e^{i\varphi}\,\mathrm{dbl}\Big[\Big(id_Z \otimes \sigma_{E_\beta, E_\alpha}\Big) \circ \Big(\hat{F}_\beta(o_\beta|i_\beta) \otimes id_{E_\alpha}\Big) \circ \hat{F}_\alpha(o_\alpha|i_\alpha)\Big]\Big]$$

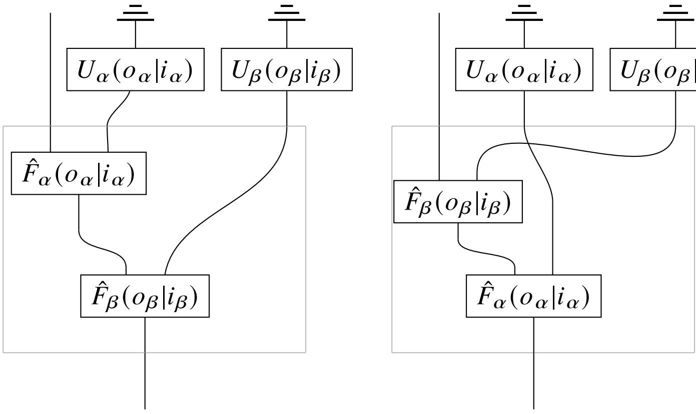
where we introduced the short-hand $\mathrm{dbl}[U]$ for the CP map $\mathrm{dbl}[U] := UU^\dagger$ corresponding to a linear map U and we have freely confused the pure CP maps $\hat{F}_\omega(o_\omega|i_\omega)$ with the corresponding linear maps.

Proposition 2.3 tells us that, in general, a controlled process such as (2.29) will depend on our choices of purification $\hat{F}_\omega$. However, this turns out not to be the case for the switch. Indeed, we can pull the two environments to the boundary of the scenario, keeping them throughout the superposition of causal orders and only discarding them afterwards:



(2.30)

Any alternative choice of purification can be obtained by applying a suitable unitary to each environment. However, the same map $\hat{F}_\omega$ appears at the bottom of the environment E_ω in all branches of the superposition, and hence the same unitary appears applied to the environment. This means that we can pull the unitaries themselves out of the scenario:



(2.31)

The unitaries will then be cancelled by the discarding maps, leading us to conclude that the controlled process we constructed for the switch was actually independent of our choices of purification $\hat{F}_\omega$ and is, therefore, a function of the original quantum instruments F_ω. This argument is not unique to the switch, but instead generalises to coherent control for all diagrams over indefinite causal scenarios in quantum theory, as dictated by our final result below. As a consequence, our framework can be used to give well-defined semantics to superposition of causal orders in quantum theory.

Definition 2.19. *Let* $\Phi = (\Omega, \Xi, \underline{I}, \underline{O}, \underline{\Sigma}^{in}, \underline{\Sigma}^{out}, \Pi^{in}, \Pi^{out})$ *be an indefinite causal scenario and* Δ *be a diagram over* Φ *in quantum theory. The* ***purification*** *of* Φ *is the indefinite causal scenario obtained from* Φ *by adding fresh symbols* ε_ω *to* Ξ *and* Π^{out} *for all* $\omega \in \Omega$ *(in some chosen order). A* ***purification of*** Δ ***with environments*** $(E_\omega)_{\omega\in\Omega}$ *is the diagram over the purification of* Φ *obtained by considering purifications of all CP maps in* Δ*, with fixed environment* E_ω *for each event* $\omega \in \Omega$. [3]

[3]The purifications considered here are for each fixed choice of classical input and output values, i.e. they are indexed families of purifications $\hat{F}_\omega(o_\omega|i_\omega)$ for the CP maps $F_\omega(o_\omega|i_\omega)$ corresponding to each given choice of $i_\omega \in I_\omega$ and

The purification of a diagram—and the necessary corresponding 'purification' of the underlying indefinite causal scenario—generalise the process seen in (2.30) and (2.31) above, where the environment wires were 'pulled to the boundary' of the diagram. We use purifications of a diagram Δ to define its coherent control.

Definition 2.20. *Let Φ be an indefinite causal scenario and Δ be a diagram over Φ in quantum theory. Let $P(\varphi) := \sum_{\Theta\in \mathrm{DCaus}(\Phi)} e^{i\varphi_\Theta}|\Theta\rangle\langle\Theta|$ be a phase gate for the computational basis of $\mathbb{C}^{\mathrm{DCaus}(\Phi)}$. The* ***coherent control of*** *Δ* ***with phase*** *$P(\varphi)$ is defined to be the coherent control—with phase $P(\varphi)$ and control system $\mathbb{C}^{\mathrm{DCaus}(\Phi)}$—of the following family of processes, where Δ^{pure} is a purification of Δ with environments $(E_\omega)_{\omega\in\Omega}$ and we have defined the global environment $E := \bigotimes_{\omega\in\Omega} E_\omega$:*

$$\Big(\mathrm{Tr}_E[\![\, \Delta^{pure}|_\Theta]\!]\Big)_{\Theta\in\mathrm{DCaus}(\Phi)} \tag{2.32}$$

Note that $\mathrm{Tr}_E[\![\, \Delta^{pure}|_\Theta]\!]$ *is a diagram over Θ for all $\Theta \in$* DCaus (Φ).

Proposition 2.20. *Let Φ be an indefinite causal scenario and Δ be a diagram over Φ in quantum theory. For each possible choice of phase gate $P(\varphi)$ for the computational basis of $\mathbb{C}^{\mathrm{DCaus}(\Phi)}$, the coherent control of Δ with phase $P(\varphi)$ is well-defined independently of the choice of purification for the processes in Δ.*

Having shown that the coherent control with phase for diagrams over indefinite causal scenarios is a well-defined concept, we conclude by providing the following definition for sake of clarity and future use.

Definition 2.21. *Let Φ be an indefinite causal scenario and Δ be a diagram over Φ in quantum theory. By a* ***superposition of causal orders for*** *Δ* ***with phase*** *$P(\varphi)$ we mean the quantum instrument obtained by:*

(i) considering the coherent control of Δ with phase $P(\varphi)$;

(ii) pre-composing the control system with the uniform superposition state $|+\rangle$ below:

$$|+\rangle := \frac{1}{\sqrt{|\,\mathrm{DCaus}\,(\Phi)\,|}} \sum_{\Theta\in\mathrm{DCaus}(\Phi)} |\Theta\rangle \tag{2.33}$$

(iii) post-composing the control system with some measurement which is unbiased with respect to the computational basis (e.g. one in the Fourier basis for some group structure on DCaus (Φ)*).*

Note that the choice of fixed input state $|+\rangle$ was made to avoid redundancy in the relative phase between the different causal orders, which is already controlled by $P(\varphi)$.

$o_\omega \in O_\omega$.

In this chapter, we discussed a way to define the control of causal order between laboratories or classically controlled quantum instruments. Here the explicitly compositional part of the dissertation ends. We constructed a way to synthesise a table of conditional distributions of probabilities in a way which pinpoints the type of information and labelling that is necessary to construct the superposition of arbitrary causal orders unambiguously. In the following chapters, we observe an inherently decompositional and theory independent approach. If it is possible, as we have discussed at length in this chapter, to define a notion of superposition of causal orders it is also paramount to understand what are the traces left by this procedure on the tables of conditional probability distributions describing the empirically observed correlations in the presence of indefinite causality.

The procedure obtained above will be used at the end of the thesis in Chapter 6 to provide examples of the type of causal analysis entailed by our sheaf-theoretic perspective. This can be done without making explicit use of process matrices, which describe a broader class of possible 'causal connections', one of which the semantics has not been fully understood [141, 12].

The following chapters will therefore try to answer the questions: 1) Given a table of probabilities correlating local interventions with local output, what is the shape of the circuits compatible with such a table? 2) How 'non-local' must a circuit with a given causal structure be to account for these correlations?

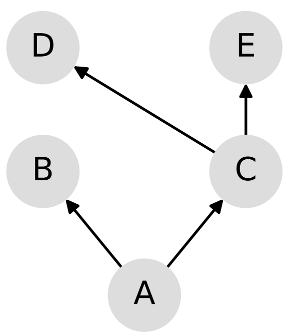
Chapter 3

The topology of causality I: Spaces of Histories

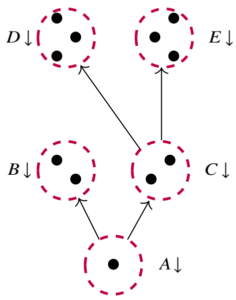
3.1 Motivation

If the causal order between events is assumed to be static and the event deterministic, the causal description given by a poset can be understood as consistently specifying for each element a unique history of 'past events'. There is a simple way to understand each element of some causal order identified with the set of events leading to their realisation. The structure of the causal order is then reflected in the containment structure between these sets. Such a perspective is employed by Markopolou in [90] to provide a topos theoretical account of causal sets; the 'internal viewpoint' of an event p in a causal set is associated with the events composing its past, providing an algebraic view to code the causal information of a discrete spacetime.

Formally this means that for a set of events Ω, there is a full and faithful injection of the causal order into $\Lambda(\Omega)$ which associates to each element of Ω the set corresponding to its downset $\omega\downarrow$. For example, the causal order Ω given by:



can be embedded into the lattice of lowersets of Ω as follows (see also Figure 3.1):



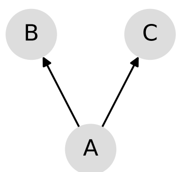
Our arena will not be a static net of 'deterministic' events; we fundamentally depart from this perspective: in our case, the set of 'possible' operational contexts can vary due to choices made at earlier events.

To give a conceptual glimpse of what is the direction that characterises our generalisation, we note that for 'deterministic spacetimes', we can always describe subsets of events as partial functions with a singleton codomain:

$$\mathrm{PFun}\left(\underline{Y}\right) := \bigcup_{U \subseteq |\Omega|} \prod_{x \in U} \{*\}$$

in particular, a downset $\omega\downarrow$ can be associated to the unique partial function for which $\mathrm{dom}\left(\omega\downarrow\right) = \omega\downarrow$. The space of general partial functions—seen as a partial order—is isomorphic to the lattice of all possible subsets of Ω. This characterisation is a mathematical triviality, but a generalisation of this way to describe subsets of events turns out to be particularly insightful. We aim to obfuscate the deterministic nature of events by assigning finite sets of interventions. The strategy is straightforward: causal histories will assume the structure of a more general class of partial functions, one that considers general finite sets of possible intervention at each node, in direct generalisation of the deterministic case represented by the singleton $\{*\}$. From this elementary extrapolation, a rich mathematical structure interpreting statical and dynamical causality emerges naturally. In this chapter we formalise the *spaces of input histories* (the generalised corresponding of Ω) and *spaces of extended input histories* (generalising $\Lambda(\Omega)$).

A causal constraint is reflected in the impossibility for causally disconnected events to share signals. In the absence of causal constraint, it is possible for the output at each event to depend arbitrarily on all inputs. Consequently, the only well-defined conditional distribution is one on all joint outputs for all events. A more refined causal structure guarantees that more conditional distributions become well defined. For example, from the causal graph given by:



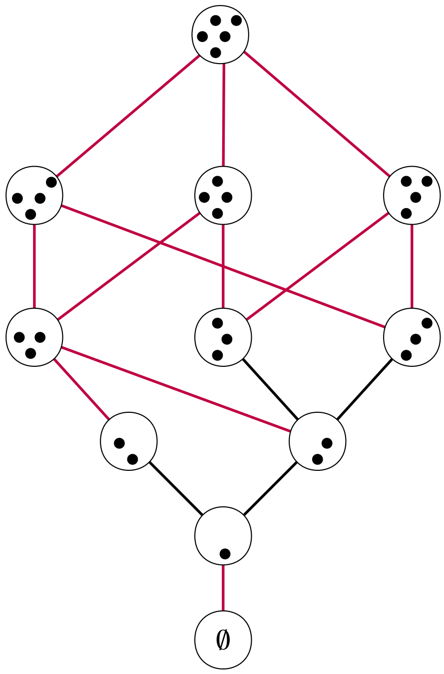


Figure 3.1: The deterministic history embed fully and faithfully in the poset of lowersets of Ω.

we can extract the following causal constraints between subsets of vertices: C and B cannot be mutually signalling, B cannot signal to A, C cannot signal to A, and the joint inputs of B and C cannot signal to A either. Notice that it is the *absence of a directed arrow* which establishes a causal constraint. Additionally to the global conditional distribution on $\{A, B, C\}$, we would in this case also get distributions on the following lowersets of Ω: $\{A\}$, $\{A, B\}$, $\{A, C\}$.

The lowersets of a given poset form a complete lattice under inclusion, which is the locale of a finite topology on the elements of Ω. The intuition we are about to develop in this and the subsequent chapters is that causal distributions are data assigned to the 'open sets' of a topology associated with Ω. A big part of our investigation will consist of generalising this observation to construct more abstract topologies than the ones describing definite causal orders. This part of the thesis, which we describe as the 'topological study of causality', will be divided into two chapters, of which the following constitutes the first part. For now, topologies will not play an explicit role; we will treat the spaces from a more combinatorial perspective in what could be described as a 'combinatorial study of causality'. However, keeping the topological perspective in mind as the story progresses will be particularly helpful.

Describing contexts as open sets of some topological space can be seen as a direct mathematical generalisation of the work in [6]. In [6], protocols are not thought to have any timelike component, and the local choices of measurements always constitute a complete history. Arbitrary subsets of these points, forming possible contexts in the original sheaf-theoretic framework, can therefore be seen as collections of timelike histories trivialised by the absence of any explicit dependence on past settings. Understanding the points of the space as timelike histories of possible interventions is, therefore, relatively straightforward but only a posteriori. The form formalised in this chapter results from intense shifts of perspectives. Of course, there is no arguing that we could define the same lattice describing the relationship between the possible operational contexts differently. What matters is only the order theoretic structure represented by the frame of open sets describing what can be obtained by marginalising more extensive contexts.

In an early attempt showcased in [61], we take a different perspective by trying to characterise this hierarchy by starting from general lowersets of settings. If it does entail the same frame of contexts for scenarios with definite causality, this approach's expressive power is limited and it does not straightforwardly generalise to indefinite causality. Moreover, the history-based perspective, providing the elementary building blocks of each context, allows for—as will be discussed in depth in Chapter 4—a particularly elegant description of the causal functions ascribable to each particular context.

3.2 Spaces of input histories

3.2.1 Partial functions and operational scenarios

This work is concerned with the study of the causal structure of a particular type of protocols, where the events are characterised by the local operation of some black box device. At each run of the protocol, inputs are chosen for each device and each black-box responds with an element from a finite set of possible outputs. The black boxes act locally, i.e. no information about the rest of the protocol is explicitly used in the operation, and the dependence between inputs and outputs will be entirely mediated by the causal structure alone. We partition our study of the causality of such protocols and experiments into two distinct concerns: the *operational scenario*, defining the set of events together with their local inputs and outputs, and the *empirical model*, assigning probabilities to joint outputs conditional to joint inputs. The operational scenario is the canvas upon which empirical models are specified, defining their combinatorial interface without constraining the concrete behaviour of the devices. The definition of an operational scenario will be different from the 'definite causal scenarios' defined in Chapter 2, Section 2.4; here, we do not explicitly require the a priori specification of a causal structure between the events.

Definition 3.1 (Operational scenario)**.** *An* operational scenario $(E, \underline{I}, \underline{O})$ *specifies a finite non-empty set E of* events*, a finite non-empty set* I_ω *of* inputs*, and a finite non-empty set* O_ω *of* outputs *for each event* $\omega \in E$*; we write* $\underline{I} = (I_\omega)_{\omega \in E}$ *and* $\underline{O} = (O_\omega)_{\omega \in E}$*. The set of* joint inputs *is defined by:*

$$\prod_{\omega \in E} I_\omega = \{ (i_\omega)_{\omega \in E} \mid i_\omega \in I_\omega \}$$

Similarly, the set of joint outputs *is defined by:*

$$\prod_{\omega \in E} O_\omega = \{ (o_\omega)_{\omega \in E} \,|\, i_\omega \in O_\omega \}$$

Intuitively we want the notion of an input history to characterise the possible 'timelike' sequences of events induced by causal relations. We start by formally defining the set of all 'partial functions' for an operational scenario:

Definition 3.2 (Partial functions). *Given a family* $\underline{Y} = (Y_x)_{x \in X}$ *of sets, the* partial functions $PFun\left(\underline{Y}\right)$ *on* $\underline{Y}$ *are defined to be all possible functions* f *having subsets* $D \subseteq X$ *as their domain* $\mathrm{dom}\,(f) := D$ *and such that* $f(x) \in Y_x$ *for all* $x \in D$*.*

$$PFun\left(\underline{Y}\right) := \bigcup_{D \subseteq X} \prod_{x \in D} Y_x \tag{3.1}$$

Partial functions are partially ordered by restriction:

$$f \leq g \quad \overset{def}{\Leftrightarrow} \quad \mathrm{dom}\,(f) \subseteq \mathrm{dom}\,(g) \;\; \textit{and} \;\; g|_{\mathrm{dom}(f)} = f \tag{3.2}$$

When $Y_x = Y$ for all $x \in X$ we recover the usual notion of partial function $X \to Y$ for all $D \subseteq X$. There exists a domain function dom : PFun $\left(\underline{Y}\right) \to \mathcal{P}(X)$ which is order preserving satisfying:

$$f \leq g \Rightarrow \mathrm{dom}\,(f) \subseteq \mathrm{dom}\,(g) \tag{3.3}$$

Partial functions ordered by restriction form a semilattice, with the empty function $\emptyset$ as minimum and meets given by:

$$\mathrm{dom}\,(f \wedge g) = \{x \in \mathrm{dom}\,(f) \cap \mathrm{dom}\,(g) \,|\, f(x) = g(x)\} \tag{3.4}$$

$$f \wedge g = f|_{\mathrm{dom}(f \wedge g)} \tag{3.5}$$

It is always true that $\mathrm{dom}\,(f \vee g) \subseteq \mathrm{dom}\,(f) \vee \mathrm{dom}\,(g)$ If the domain function preserves the meets of two partial functions we say that they are *compatible*.

Definition 3.3 (Compatible functions). *We say that* f *and* g *are* compatible *when:*

$$\mathrm{dom}\,(f \wedge g) = \mathrm{dom}\,(f) \cap \mathrm{dom}\,(g) \tag{3.6}$$

We say that $\mathcal{F} \subseteq PFun\left(\underline{Y}\right)$ *is a compatible set of functions if* f *and* g *are pairwise compatible for every* $f, g \in \mathcal{F}$*.*

Definition 3.4 (Joins of partial functions). *Let* $\mathcal{F}$ *be a compatible set of partial functions. The join of* $\mathcal{F}$ *is given by*

$$\begin{aligned} \mathrm{dom}\,(\bigvee \mathcal{F}) &= \bigcup_{f \in \mathcal{F}} \mathrm{dom}\,(f) \\ \bigvee \mathcal{F} &= x \mapsto f(x) \textit{ for any } f \textit{ such that } x \in \mathrm{dom}\,(f) \end{aligned} \tag{3.7}$$

The requirement of compatibility allows the definition of the join of two or more compatible histories by 'stitching' them about their common values. Well definedness is guaranteed by compatibility. The maxima of the space PFun $(\underline{Y})$ ordered by restriction are the total functions defined on the entirety of dom $(\underline{Y})$.

3.2.2 Spaces of input histories

3.2.2.1 Space of input histories for causal orders

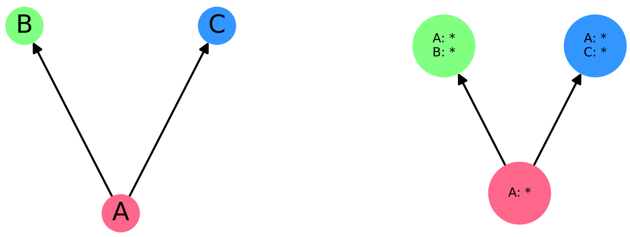
We associate a casual order to an operational scenario by selecting a set of partial functions which denotes the compatible timelike sequences of events entailed by the order:

Definition 3.5 (Input histories for a causal order)**.** *The* input histories *for a given choice of order* Ω *and inputs* $\underline{I} = (I_\omega)_{\omega\in\Omega}$ *are defined to be the partial functions:*

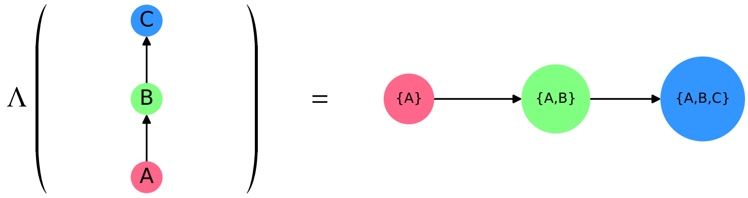
$$Hist\left(\Omega, \underline{I}\right) := \bigcup_{\xi\in\Omega} \prod_{\omega\in\xi\downarrow} I_\omega \subseteq PFun\left(\underline{I}\right) \tag{3.8}$$

We refer to the partially ordered set Hist $\left(\Omega, \underline{I}\right)$*–where the order is inherited from the restriction order of partial functions PFun* $\left(\underline{I}\right)$*–as a* space of input histories induced by the causal order Ω.

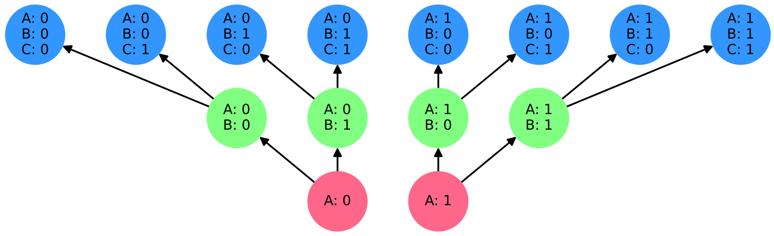
Example 3.6. *Let* Ω *be a poset and consider* $\underline{I} = \prod_{\omega\in\Omega}\{*\}$*. A single classical choice input is associated with every event. The input histories for this causal order are in bijective correspondence with its downsets* $\{\omega\downarrow | \omega \in \Omega\}$*. For the preorder given by discrete* $(A) \vee$ *indiscrete* (B, C)*(on the left), the poset of causal histories (on the right) recovers the original causal order:*



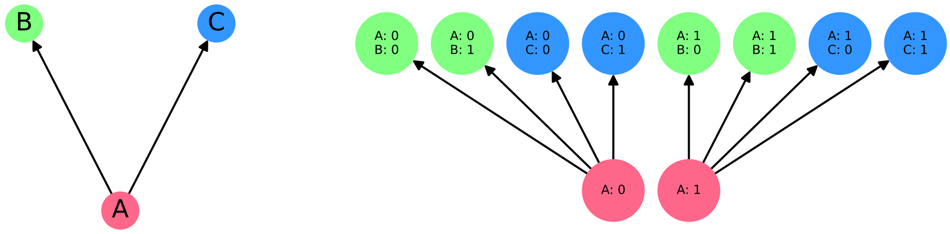
Example 3.7. *Let* Ω *be the total order on* 3 *events with binary inputs* $I_\omega = \{0, 1\}$*, and consider its associated lattice of lowersets* $\Lambda(\Omega)$



Because the space is a total order, the downsets coincide with the lowersets, and the associated space of input histories consists of all the possible assignments of joint inputs to the subsets of events $\{A\}, \{A, B\}, \{A, B, C\}$*. The Hasse diagram of the space of histories associated with this total order is given by:*

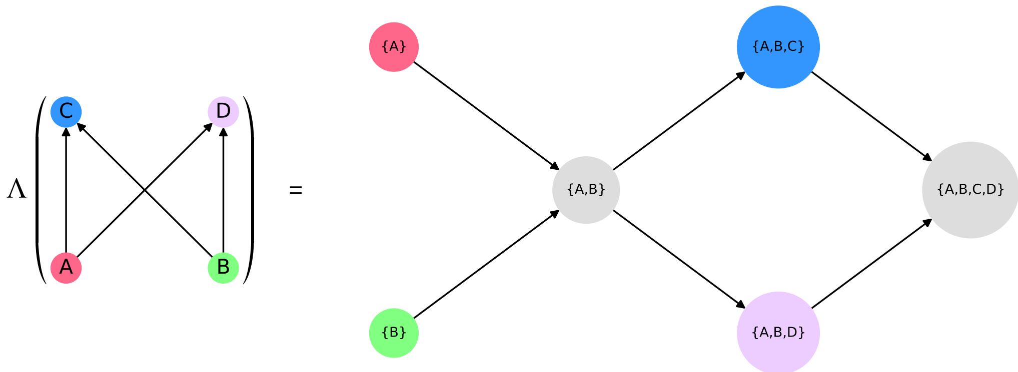


Example 3.8. *Let Ω by the poset discrete* $(A) \vee$ *indiscrete* (B, C) *and assume dichotomic interventions. The poset on the right gives the associated space of histories.*

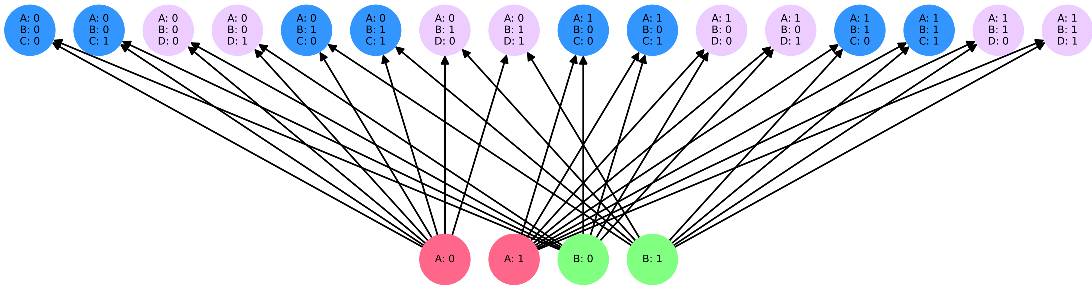


In this and the previous example, we have colour-coded the histories to highlight the associated downsets. For the case above $\{A{:}0\}$ *,*$\{A{:}1\}$ *are associated to the downset* $A{\downarrow} = \{A\}$*, the histories* $\{B{:}0,\ A{:}0\}$ *,*$\{B{:}1,\ A{:}0\}$ *,*$\{B{:}0,\ A{:}1\}$ *,*$\{B{:}1,\ A{:}1\}$ *to the downset* $B{\downarrow} = \{A, B\}$ *and* $\{C{:}0,\ A{:}0\}$ *,*$\{C{:}1,\ A{:}0\}$ *,*$\{C{:}0,\ A{:}1\}$ *,*$\{C{:}1,\ A{:}1\}$ *are associated to* $C{\downarrow} = \{A, C\}$*.*

The spaces of input histories are not generally closed under meets or joins. Consider the following "M"-shaped causal order on 4 events. We observe that both the intersection $\{\mathrm{A}, \mathrm{B}\}$ and the union $\{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}\}$ of the causal pasts $\mathrm{C}{\downarrow} = \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$ and $\mathrm{D}{\downarrow} = \{\mathrm{A}, \mathrm{B}, \mathrm{D}\}$ are not causal pasts of events themselves.



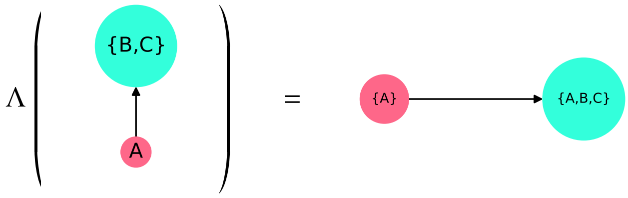
The associated space of input histories doesn't feature any meets $f \wedge g$ or joins $f \vee g$ for compatible histories f, g with domain $\mathrm{C}{\downarrow}$ and $\mathrm{D}{\downarrow}$ respectively (we remind the reader that the meets and joins being referred to are those in $\mathrm{PFun}\left(\underline{I}\right)$).



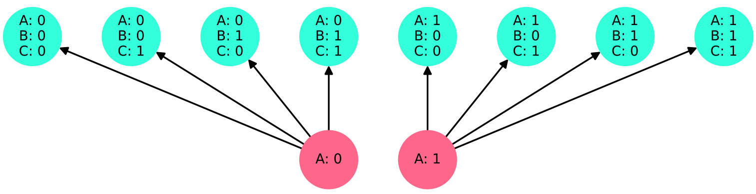
The space above is also an example where the maxima of the space of input histories differ from those of PFun $(\underline{I})$: $\{A, B, C, D\}$ is not the causal past of an event, so the total functions in PFun $(\underline{I})$ are not input histories.

Definition 3.5 does not exclusively refer to situations where a causal poset gives the order between events of some operational scenario; we can apply the same recipe to preorders:

Example 3.9. *Consider the following indefinite causal order* Ω *on* 3 *events, and its associated lattice of lowersets* $\Lambda(\Omega)$.

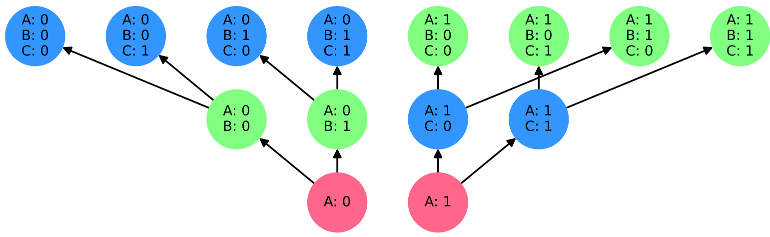


Because events B and C are in indefinite causal order, they have the same causal past, and no causal history separates them. For the associated space of histories, we have 2^3 *histories of size* 3 *and* 2^1 *histories of size* 1.



This space of input histories does not entail any causal separability between the events B and C. We will revisit this specific issue later when talking about 'causal completeness', but it already prompts the question: can we extend our spaces of input histories to capture more specific situations where the causal order between events depends on choices performed in the past?

For example, we might want to consider a '3-party causal switch', in which an event A controls the order of events B and C, e.g. by setting $B \to C$ when the input is 0, and $C \to B$ when the input is 1. In this case, the output at B is fully determined by the inputs at events A and B when the input at A is 0, but C also becomes relevant when the input at A is 1. Considering this observation—and the analogous one about the output at C— we obtain our putative space of input histories.



The space above describes all the possible 'time-like histories' of events that can occur when the causal order is controlled by A. It is a subset of PFun $(\underline{I})$ but does not arise from Hist $(\Omega, \underline{I})$ for any causal order Ω: the order between B and C is indefinite overall, but the input histories are now able to discriminate between the events based on the input assigned at the event A (colour coding of input histories reflects this fact).

3.2.2.2 Abstract spaces of input histories

We will see that spaces of histories associated with causal orders can be generalised to encompass a variety of other operational assumptions. How do we axiomatically characterise these sets of histories? We first notice that the spaces obtained as Hist $(\Omega, \underline{I})$ satisfy two special properties: they are $\vee$-prime (read 'join prime'), and they satisfy the 'free-choice condition'.

Join primality guarantees that each history is not decomposable in simpler histories and that the space is composed of causally 'atomic' constituents. The free-choice condition guarantees that there are enough input histories to reconstruct all possible global behaviours characterising a scenario. We start by formalising join-primality:

Definition 3.10 (Join-prime subsets)**.** *A subset $\Theta \subseteq PFun\left(\underline{I}\right)$ is said to be* $\vee$-prime *(read 'join-prime') if no $h \in \Theta$ can be written as the compatible join $h = \bigvee \mathcal{F}$ of a subset $\mathcal{F} \subseteq \Theta$ such that $h \notin \mathcal{F}$:*

$$\left(\mathcal{F} \subseteq \Theta \textit{ compatible and } \bigvee \mathcal{F} \in \Theta\right) \Rightarrow \bigvee \mathcal{F} \in \mathcal{F}$$

Dually, a subset $W \subseteq PFun\left(\underline{I}\right)$ is said to be $\vee$-closed (read 'join-closed') if for every pair of compatible $h, k \in W$ the join $h \vee k$ is itself in W. This implies that, more generally:

$$\mathcal{F} \subseteq \Theta \textit{ compatible } \Rightarrow \bigvee \mathcal{F} \in \Theta$$

Proposition 3.10. *For any causal order Ω, Hist $\left(\Omega, \underline{I}\right) \subseteq PFun\left(\underline{I}\right)$ is always a $\vee$-prime subset of PFun $\left(\underline{I}\right)$.*

Proof. Let $\mathcal{F}$ be a set of compatible histories. For all $f \in \mathcal{F}$ we have that dom $(f) = \omega_f\downarrow$. Suppose that $\bigvee \mathcal{F} \in$ Hist $\left(\Omega, \underline{I}\right)$ then by definition there exists $e \in \Omega$ such that $e\downarrow = \bigcup_{f \in \mathcal{F}} \omega_f\downarrow$. This can only be true when $e\downarrow = \omega_{f'}\downarrow$ for some $f' \in \mathcal{F}$. Since $\bigvee \mathcal{F}$ and f' are compatible they must agree on all elements of their domain and we conclude that $f' = \bigvee \mathcal{F}$ and $\bigvee \mathcal{F} \in \mathcal{F}$. □

Equipped with this notion, we describe general spaces of input histories:

Definition 3.11 (Spaces of input histories)**.** *A* space of input histories *is a finite set* Θ *of partial functions which is* $\vee$*-prime. To every space of input histories we can associate an* event set E^{Θ} *and a family of* input sets $\underline{I^{\Theta}} = (I^{\Theta}_{\omega})_{\omega \in E^{\Theta}}$ *as follows:*

$$\begin{array}{lcl} E^{\Theta} & := & \bigcup_{h \in \Theta} \mathrm{dom}\,(h) \\ I^{\Theta}_{\omega} & := & \{\, h_{\omega} \mid h \in \Theta, \omega \in \mathrm{dom}\,(h) \,\} \end{array} \tag{3.9}$$

We have $\Theta \subseteq PFun\left(\underline{I^{\Theta}}\right)$ *and the space* Θ *is equipped with the partial order inherited from* $PFun\left(\underline{I^{\Theta}}\right)$.

We also define the extended version of the spaces by taking their join closure in $\mathrm{PFun}\left(\underline{I}\right)$:

Definition 3.12 (Spaces of extended input histories)**.** *The* space of extended input histories $Ext\,(\Theta)$ *associated to* Θ *is defined to be its* $\vee$*-closure:*

$$Ext\,(\Theta) := \left\{ \bigvee \mathcal{F} \,\middle|\, \mathcal{F} \subseteq \Theta \textit{ compatible} \right\} \tag{3.10}$$

We have $Ext\,(\Theta) \subseteq PFun\left(\underline{I^{\Theta}}\right)$ *and the space* $Ext\,(\Theta)$ *is equipped with the partial order inherited from* $PFun\left(\underline{I^{\Theta}}\right)$.

The duality between spaces of input histories and extended spaces (their join closure in $\mathrm{PFun}\left(\underline{I}\right)$) will play an important role in this work. We can now appreciate the importance of $\vee$-primality: it gives a 'normal form' that establishes the one-to-one equivalence between spaces of input histories Θ and their join closure $\mathrm{Ext}\,(\Theta)$. From an arbitrary subset of histories $W \subseteq \mathrm{PFun}\left(\underline{I}\right)$, we can always obtain a space of input histories by considering its $\vee$-prime elements.

Definition 3.13 (Join prime elements)**.** *Given any subset* $W \subseteq PFun\left(\underline{I}\right)$ *the* $\vee$*-prime elements associated to* W *are given by*

$$Prime\,(W) := \left\{ w \in W \,\middle|\, \forall \mathcal{F} \subseteq W \textit{ compatible}.\, w = \bigvee \mathcal{F} \Rightarrow w \in \mathcal{F} \right\} \tag{3.11}$$

In particular, we can recover the space of input histories from its correspondent space of extended histories:

$$\mathrm{Prime}\,(\mathrm{Ext}\,(\Theta)) = \Theta \tag{3.12}$$

Conversely, any $\vee$-closed subset $W \subseteq \mathrm{PFun}\left(\underline{I}\right)$ is the extension of its prime elements

$$\mathrm{Ext}\,(\mathrm{Prime}\,(W)) = W \tag{3.13}$$

Proposition 3.13. *The extended input histories for a given choice of order* Ω *and inputs* $\underline{I} = (I_{\omega})_{\omega \in \Omega}$ *are given by:*

$$Ext\left(Hist\left(\Omega, \underline{I}\right)\right) \;=\; \bigcup_{U \in \Lambda(\Omega)} \prod_{\omega \in U} I_{\omega} \;\;\subseteq\;\; PFun\left(\underline{I}\right) \tag{3.14}$$

Proof. Let $h \in \text{Ext}\left(\text{Hist}\left(\Omega, \underline{I}\right)\right)$, by definition there exists a finite compatible set of histories $\mathcal{F} \subseteq \Theta$ such that $h = \bigvee \mathcal{F}$. For every $f \in \mathcal{F}$, we have $\omega_f \in \Omega$ such that $\text{dom}\,(f) = \omega_f \downarrow$ by the definition of $\text{Hist}\left(\Omega, \underline{I}\right)$. Compatibility imposes that:

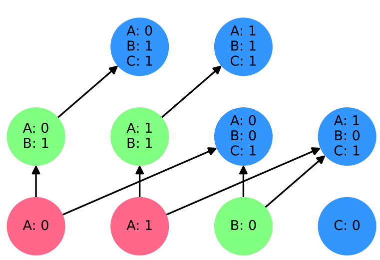
$$\text{dom}\,(h) = \text{dom}\left(\bigvee \mathcal{F}\right) = \bigcup \text{dom}\,(\mathcal{F})$$

Let $\omega \in \bigcup \text{dom}\,(\mathcal{F})$ and $\omega' \leq \omega$, there exists $f \in \mathcal{F}$ such that $\omega \in \text{dom}\,(f)$. Since $\text{dom}\,(f) = \omega_f \downarrow$ for some $\omega_f \geq \omega$, we conclude that $\omega' \in \bigcup \text{dom}\,(\mathcal{F})$ making $\text{dom}\,(h)$ a lowerset. □

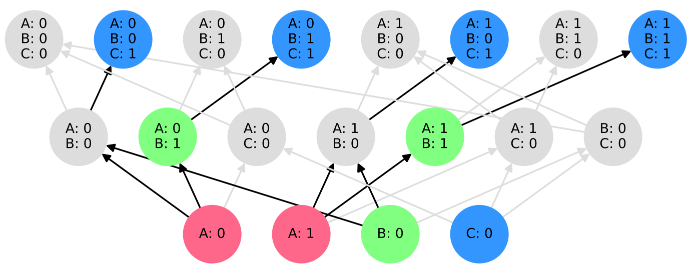
Intuitively extended input histories provide an abstract description of families of joint inputs that are allowed to affect the outcomes of an entire subset of spacelike events. In contrast, input histories alone describe the joint inputs affecting a single event. This picture will become clear only when we attach causal data to the histories.

This abstract characterisation of what it means to assign histories of events to an operational scenario not only goes beyond the deterministic description of events but also gives, as mentioned earlier, the additional flexibility needed to describe situations where the order itself is indefinite. We provide three examples of spaces and their extension: in the first one, the events can ‘choose’ to causally disconnect themselves from other events, and the second one is the by now familiar case where the past is allowed to influence the order of subsequent events. The causal meaning of the third one is more esoteric and will become more apparent when discussing non-tight spaces:

Example 3.14. *Consider the space of histories Θ (on the left) and its corresponding space of extended input histories $Ext\,(\Theta)$. The grey-coloured extended input histories on the right are those which are not input histories (i.e. they arise by join). We can interpret this space, which does not occur as a space of histories for a fixed preorder, as a refinement of the total order total (A, B, C) in which choices at B can cause a causal disconnection from A. Similarly, a choice for C can causally disconnect it from the past events A and B.*

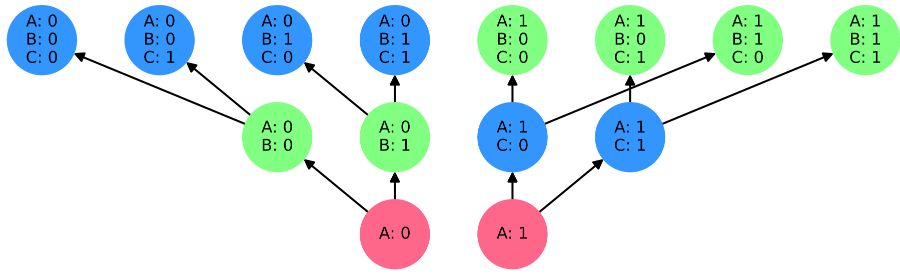


Θ

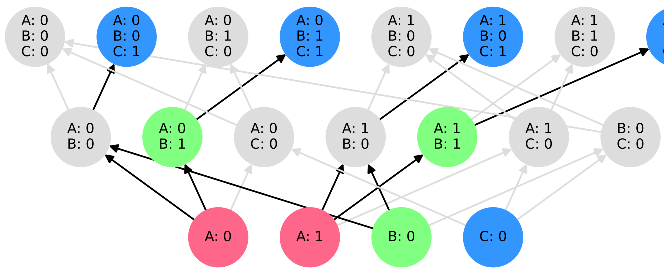


Ext (Θ)

Example 3.15. *Recall that the order for the quantum switch is given by:*



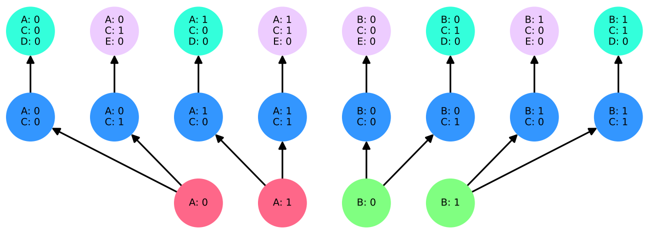
Θ



Ext (Θ)

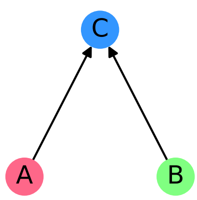
The interpretation for this space of histories is straightforward $\{A:0,\ B:0\}$ *,* $\{A:0,\ B:1\}$ *,* $\{A:1,\ C:0\}$ *,* $\{A:1,\ C:1\}$ *witness the possibility of A to control the causal order of the events B,C. The values at A are always independent of any other input, and this is witnessed by the histories* $\{A:0\}$ *,* $\{A:1\}$ *.*

Example 3.16. *Spaces of histories arising from preorders and switches do not exhaust the expressive power of* $\vee$*-prime spaces. Consider the following example:*

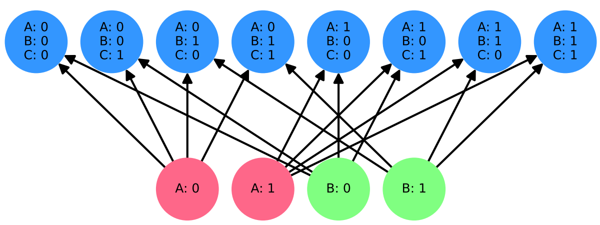


Θ

This is an example of a non-tight space of histories. We will formally study this class of spaces later on, but we can attempt a preliminary analysis. Both A and B are independent of any other choice performed by C or D: their inputs alone constitute valid input histories. If we focus on the event C, it seems that there is a dependence on both inputs A and B. However, it is not the type of dependence induced by space associated with the following order Ω*:*



Ω



Hist $(\Omega, \underline{I})$

In particular there are no histories $\{A\!:\!a,\ B\!:\!b,\ C\!:\!c\}$ *. It seems, therefore, that C can be both thought of as being in the causal future of A and in the causal future of B but not of the two events simultaneously. Such spaces of input histories arise as meets of spaces with a more straightforward causal description, and they are not without operational significance. We conclude our preliminary analysis by observing that the the choice of C will condition A to be followed by either E or D as witnessed by the histories* $\{A\!:\!a,\ C\!:\!1,\ E\!:\!e\}$ *or* $\{A\!:\!a,\ C\!:\!0,\ D\!:\!d\}$ *and the same happens to the event B:* $\{B\!:\!b,\ C\!:\!1,\ D\!:\!d\}$ *,* $\{B\!:\!b,\ C\!:\!0,\ E\!:\!e\}$ *.*

Note that the graphical depiction of the Hasse diagram for $\mathrm{Ext}(\Theta)$ can be used to illustrate the $\vee$-primality and the $\vee$-closure condition. Consider the space of histories in Example 3.14: for $\vee$-primality observe that no input histories (the coloured nodes) can be obtained as the minimal common successor of two other histories. With regards to the $\vee$-closure condition, note how all compatible input histories have some common successor in the graph: for some of them—such as $\{\mathrm{A}\!:\!1\}$ and $\{\mathrm{B}\!:\!0\}$ —this is an immediate common successor, namely $\{\mathrm{A}\!:\!1,\ \mathrm{B}\!:\!0\}$; for others—such as $\{\mathrm{B}\!:\!0\}$ and $\{\mathrm{C}\!:\!0\}$ —this is a common successor further up the graph, e.g. $\{\mathrm{A}\!:\!1,\ \mathrm{B}\!:\!0,\ \mathrm{C}\!:\!0\}$. Extended input histories without a join are always incompatible ones, such as $\{\mathrm{A}\!:\!0,\ \mathrm{C}\!:\!0\}$ and $\{\mathrm{A}\!:\!1,\ \mathrm{C}\!:\!0\}$ (differing in value on a common event A).

A crucial ingredient, common to all the examples shown above, is hidden in the important assumption that the inputs at each event can be chosen independently. This desideratum is embodied in the following condition:

Definition 3.17 (Free-choice condition)**.** *A space of input histories is said to satisfy the* free-choice condition *if:*

$$\max \mathit{Ext}(\Theta) = \prod_{\omega \in E^{\Theta}} I^{\Theta}_{\omega} \tag{3.15}$$

In spaces satisfying the free-choice condition, we refer to the histories in $\prod_{\omega\in E^{\Theta}} I^{\Theta}_{\omega}$ *as the* maximal extended input histories.

Example 3.18. *The set of partial functions* $\Theta = \{\{A\!:\!0\}, \{A\!:\!1\}, \{A\!:\!0,\ B\!:\!1\}\}$ *satisfies* $\vee$*-primality but doesn't satisfy the free-choice condition. A timelike history exists where measurement is performed at B only if Alice's input is* 0*, a clear violation of the requirement that the measurement choices are locally independent. Also, the space* $\{\emptyset\}$*, containing only the empty function, does not satisfy free choice for any non-empty* $\underline{I}$*, but it trivially satisfies* $\vee$*-primality.*

Proposition 3.18. *The space of input histories* $\mathit{Hist}(\Omega, \underline{I})$ *constructed from causal orders always satisfy the free-choice condition.*

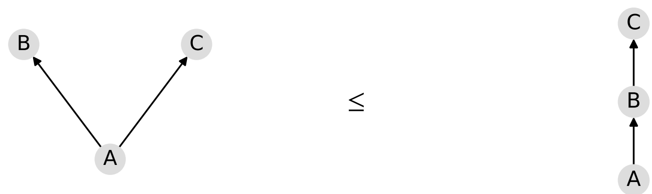
Proof. Let $h \in \prod_{\omega\in\Omega}$ be a maximal input history. Let $\mathcal{H}$ be the subset of $\mathrm{Hist}(\Omega, \underline{I})$ defined by $\mathcal{H} = \{h|_{\omega\downarrow} | \omega \in \mathrm{dom}(h)\}$, then $\mathcal{H}$ is composed of compatible histories as they all arise as restriction of h, and we have that $h = \bigvee \mathcal{H}$. □

3.3 Hierarchy of Spaces of Input Histories

We have seen in Chapter 1, Section 1.3.3.3 that for causal order hierarchies can be defined unproblematically: $\Omega \leq \Xi$ is $|\Omega| \subseteq |\Xi|$ as sets and $\leq_\Omega \subseteq \leq_\Xi$ as relations. Generalising the type of causal explanations to general spaces of input histories, we lose the possibility of describing abstract causality relations with binary edges. How can we endow spaces of histories with a hierarchy which naturally extends what we know about causal orders?

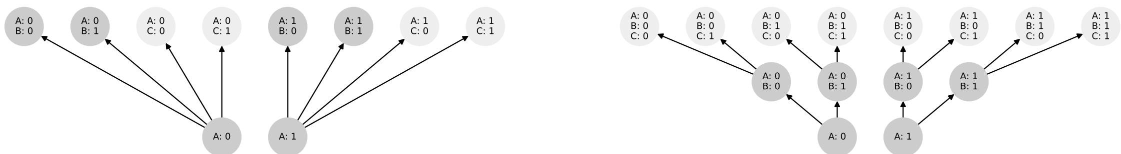
We appeal to the extended version of the spaces, where more causal constraints are always identified with the presence of more extended histories. This is particularly easy to appreciate by keeping in mind the standard no-signalling case where the events are all spacelike separated. The space of input histories associated with the scenario endowed with a discrete causal order consists of all the elements of the form $\{\omega : i\}$ where $\omega \in E$ and $i \in I_\omega$. These elementary histories generate the entire space of possible partial functions $\bigcup_{U \in \mathcal{P}(X)} \prod_{\omega \in U} I_\omega$ by iteratively constructing all joins. The no-signalling case is extremal; it entails situations in which all subsets of choices of inputs can be considered operationally meaningful, i.e. they can always be expressed as unions of compatible timelike choices for some events.

Consider the following 'causal fork' Ω and total order Ξ on 3 events. We know that in the hierarchy of preorders $\Omega \leq \Xi$:



Ω Ξ

The corresponding spaces of input histories $\text{Hist}\left(\Omega, \underline{I}\right)$ and $\text{Hist}\left(\Xi, \underline{I}\right)$ are not related by inclusion in either direction. To make this fact evident, no colour coding is used for the input histories in these diagrams: instead, the common input histories have been highlighted with a darker colour.



$\text{Hist}\left(\Omega, \underline{I}\right)$ $\text{Hist}\left(\Xi, \underline{I}\right)$

If the direct comparison fails for spaces of input histories, the order between the spaces is witnessed by the associated $\text{Ext}\left(\Theta\right)$ and is compatible with the hierarchy of casual orders:

Proposition 3.18. *For any two causal orders Ω and Γ, we have:*

$$\Omega \leq \Gamma \quad \Leftrightarrow \quad Ext\left(Hist\left(\Omega, \underline{I}\right)\right) \supseteq Ext\left(Hist\left(\Gamma, \underline{I}\right)\right) \tag{3.16}$$

Proof. Suppose that $\Omega \leq \Gamma$, by Proposition 1.29 $\Lambda(\Gamma) \subseteq \Lambda(\Omega)$. Let $h \in \mathrm{Ext}\left(\mathrm{Hist}\left(\Gamma, \underline{I}\right)\right)$ then $\mathrm{dom}\,(h) \in \Lambda(\Gamma) \subseteq \Lambda(\Omega)$ hence $h \in \prod_{\omega \in \mathrm{dom}(h)} I_\omega \subseteq \mathrm{Ext}\left(\mathrm{Hist}\left(\Omega, \underline{I}\right)\right)$ by Proposition 3.13. Conversely, assume that $\mathrm{ExtHist}\left(\Omega, \underline{I}\right) \supseteq \mathrm{ExtHist}\left(\Lambda, \underline{I}\right)$ let $V \in \Lambda(\Gamma)$ and consider $h \in \prod_{\omega \in V} I_\omega$. By assumption $h \in \mathrm{Ext}\left(\mathrm{Hist}\left(\Omega, \underline{I}\right)\right) = \bigcup_{U \in \Lambda(\Omega)} \prod_{\omega \in U} I_\omega$, from which we infer that $\mathrm{dom}\,(h) \in \Lambda(\Omega)$, and therefore $\Lambda(\Gamma) \subseteq \Lambda(\Omega)$ implying $\Omega \leq \Gamma$. □

We proceed to define general spaces on input histories:

Definition 3.19 (Partial order on spaces of input histories)**.** *We define the following partial order on spaces of input histories:*

$$\Theta' \leq \Theta \iff Ext\,(\Theta') \supseteq Ext\,(\Theta) \tag{3.17}$$

We say that Θ' *is a* causal refinement *of* Θ*, or that* Θ *is a* causal coarsening *of* Θ'.

A causal coarsening for a space will entail fewer causal constraints, a refinement more causal constraints and therefore more extended histories. It may seem inconsistent that the definition of the partial order does not explicitly mention the underlying set of events. Recall from Definition 3.11 that these can be thought of as implicit in the definition of spaces of input histories.

Observation 3.19. *Recall the definition of the events associated with a space of input histories* E^Θ *from Definition 3.11. If* θ *is a space of input histories and* $\Theta \leq \Theta'$*, then for all* $\omega \in E^\Theta$ *we have that:*

$$E^{\Theta'} \subseteq E^\Theta \tag{3.18}$$

$$I^{\Theta'} \subseteq I^\Theta \tag{3.19}$$

Proof. From Definition 3.11 the event sets of Θ and Θ' are respectively given by

$$E^\Theta = \bigcup_{h \in \mathrm{Ext}(\Theta)} \mathrm{dom}\, h \qquad \text{and} \qquad E^{\Theta'} \bigcup_{h \in \mathrm{Ext}(\Theta')} \mathrm{dom}\, h$$

if $\Theta \leq \Theta'$ then $\mathrm{Ext}\,(\Theta') \subseteq \mathrm{Ext}\,(\Theta)$ from which follows that $E^{\Theta'} \subseteq E^\Theta$. For the input sets $I^{\Theta'}_\omega$ and I^Θ_ω, where ω is any $\omega \in E^\Theta$, we have:

$$\begin{aligned} I^\Theta_\omega &= \{h_\omega | h \in \mathrm{Ext}\,(\Theta)\,, \omega \in \mathrm{dom}\,(h)\} \\ &\supseteq \{h_\omega | h \in \mathrm{Ext}\,(\Theta')\,, \omega \in \mathrm{dom}\,(h)\} = I^{\Theta'}_\omega \end{aligned}$$

□

Causal orders for a given set of events form a lattice, the indiscrete order is the maximum (minimal set of causal restrictions), and the discrete causal order is the minimum (maximal set of causal restrictions). The complete lattice structure of the hierarchy of preorders extends to general spaces of input histories:

Proposition 3.19. *Input histories—with no restriction on the underlying events and input values—form an infinite lattice Spaces under the partial order of Definition 3.18. The join and meets take the following form*

$$\Theta \vee \Theta' = \mathit{Prime}\,(\mathit{Ext}\,(\Theta) \cap \mathit{Ext}\,(\Theta')) \tag{3.20}$$

$$\Theta \wedge \Theta' = \mathit{Prime}\,(\mathit{Ext}\,(\Theta) \cup \mathit{Ext}\,(\Theta')) \tag{3.21}$$

We can think of the join $\vee$ *in Spaces as the* closest common coarsening *and of the meet* $\wedge$ *as the* closest common refinement *of two spaces of histories.*

Proof. For all spaces we have that $\text{Prime}\,(\text{Ext}\,(\Theta)) = \Theta$, because of $\vee$-primality. For subsets of histories W which are $\vee$-closed, Ext is a retraction of Prime so that $\text{Ext}\,(\text{Prime}\,(W)) = W$.

- For the join. We have that $\text{Ext}\,(\Theta) \cap \text{Ext}\,(\Theta') \subseteq \text{Ext}\,(\Theta')$ and $\text{Ext}\,(\Theta) \cap \text{Ext}\,(\Theta') \subseteq \text{Ext}\,(\Theta)$. The spaces Θ and Θ' are $\vee$-prime and therefore:

  $$\Theta \leq \text{Prime}\,(\text{Ext}\,(\Theta) \cap \text{Ext}\,(\Theta'))$$
  $$\Theta' \leq \text{Prime}\,(\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta'))$$

  Consider an arbitrary Θ'' such that $\Theta'' \geq \Theta, \Theta'$. By definition $\text{Ext}\,(\Theta'') \subseteq \text{Ext}\,(\Theta) \cap \text{Ext}\,(\Theta')$. Recall that for all $\vee$-prime Θ, $\text{Prime}\,(\text{Ext}\,(\Theta)) = \Theta$, so that

  $$\text{Prime}\,(\text{Ext}\,(\Theta'')) = \Theta'' \leq \text{Prime}\,(\text{Ext}\,(\Theta) \cap \text{Ext}\,(\Theta'))$$

- For the meet. We have that $\text{Ext}\,(\Theta), \text{Ext}\,(\Theta') \subseteq \text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$ and the $\vee$-closure of $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$ is the smallest $\vee$-closed superset, so that

  $$\text{Prime}\,(\vee\text{-closure of } \text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')) \leq \Theta, \Theta'$$

  Let $\Theta'' \leq \Theta, \Theta'$ then $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta') \subseteq \text{Ext}\,(\Theta'')$. Since $\text{Ext}\,(\Theta'')$ is $\vee$-closed, then the $\vee$-closure of $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$ must also be contained by $\text{Ext}\,(\Theta'')$.

  $$\Theta'' = \text{Prime}\,(\text{Ext}\,(\Theta'')) \geq \text{Prime}\,(\vee\text{-closure of } \text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta'))$$

  Note that the elements that are in the $\vee$-closure of $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$ which are not in $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$ cannot be $\vee$-prime, therefore:

  $$\Theta'' \geq \text{Prime}\,(\vee\text{-closure of } \text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')) = \text{Prime}\,(\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta'))$$

□

Proposition 3.19. *Spaces of input histories Θ such that $E^{\Theta} \subseteq E$ and $\underline{I}^{\Theta} \subseteq \underline{I}$ for a fixed set of 'events' $\underline{I} := (I_\omega)_{\omega \in E}$ form a finite upperset of Spaces.*

$$Spaces\left(\underline{I}\right) \hookrightarrow Spaces$$

Proof. The maximum of Spaces $\left(\underline{I}\right)$ is the empty space with a unique empty history $\{\emptyset\}$, which is trivially a $\vee$-prime set of histories for any $\underline{I}$. For the minimum, observe that

$$\text{Ext}\left(\text{Hist}\left(\text{discrete}\left(E\right), \underline{I}\right)\right) = \text{Ext}\left(\prod_{\omega \in E} I_\omega\right) = \bigcup_{U \in \mathcal{P}(E)} \prod_{\omega \in E} I_\omega = \text{PFun}\left(\underline{I}\right)$$

so that $\text{Ext}\left(\Theta\right) \subseteq \text{Ext}\left(\text{Hist}\left(\text{discrete}\left(E\right), \underline{I}\right)\right)$ for all $\Theta \subseteq \text{PFun}\left(\underline{I}\right)$. Correspondingly, for all $\Theta \in \text{Spaces}\left(\underline{I}\right)$ we get $\text{Hist}\left(\text{discrete}\left(E\right), \underline{I}\right) \leq \Theta$. Observe that for $\Theta \geq \text{Hist}\left(\text{discrete}\left(E\right), \underline{I}\right)$, we have that $\text{Ext}\left(\Theta\right) \subseteq \text{PFun}\left(\underline{I}\right)$ so in particular $\Theta \subseteq \text{PFun}\left(\underline{I}\right)$ and $\Theta \in \text{Spaces}\left(\underline{I}\right)$. The set is therefore upward closed with a minimum and a maximum, i.e. a full-sublattice of Spaces. □

Proposition 3.19. *Denote by $Spaces_{FC}\left(\underline{I}\right)$ the set of spaces satisfying the free-choice condition. The set $Spaces_{FC}\left(\underline{I}\right)$ forms a lowerset in $Spaces\left(\underline{I}\right)$ so that together with Proposition 3.19 we have the following chain of full inclusions (see Figure 3.3 (p.100)):*

$$Spaces_{FC}\left(\underline{I}\right) \hookrightarrow Spaces\left(\underline{I}\right) \hookrightarrow Spaces$$

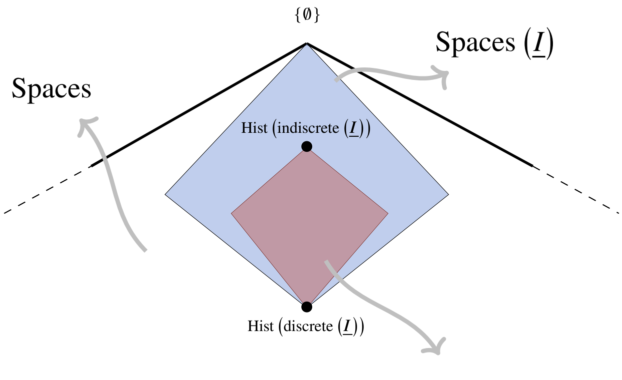


Figure 3.2: The empty space $\{\emptyset\}$ is the maximum of the hierarchy, but it does not satisfy the free-choice condition for a non-empty $\underline{I}$. For a given $\underline{I}$ the set Spaces $\left(\underline{I}\right)$ is given by the upperset covering $\text{Hist}\left(\text{indiscrete}\left(E\right), \underline{I}\right)$ —highlighted in blue. Introducing the free-choice condition singles out all the spaces between $\text{Hist}\left(\text{discrete}\left(E\right), \underline{I}\right)$ and $\text{Hist}\left(\text{indiscrete}\left(E\right), \underline{I}\right)$.

Proof. Recall that

$$\text{Hist}\left(\text{indiscrete}\left(E\right), \underline{I}\right) = \text{Ext}\left(\text{Hist}\left(\text{indiscrete}\left(E\right), \underline{I}\right)\right) = \prod_{\omega \in E} I_\omega$$

so that for every space Θ satisfying the free choice condition we get $\prod_{\omega\in E} I_\omega \subseteq \mathrm{Ext}(\Omega)$ implying $\Theta \leq \mathrm{Hist}\left(\mathrm{discrete}(E), \underline{I}\right)$. □

It follows from Proposition 3.18 that for every Ω with events inputs $\underline{I}$ we have that $\mathrm{Hist}\left(\Omega, \underline{I}\right) \in \mathrm{Spaces}_{\mathrm{FC}}\left(\underline{I}\right)$ since in particular $\mathrm{Hist}\left(\mathrm{discrete}\left(\underline{I}\right)\right) \in \mathrm{Spaces}_{\mathrm{FC}}\left(\underline{I}\right)$. We show that the operation of extracting the space of input histories from a causal order commutes with the join operation. $\mathrm{Hist}\left(\Omega, \underline{I}\right)$ therefore embeds the hierarchy of causal orders (studied in Chapter 1) in a sup-semilattice of $\mathrm{Spaces}\left(\underline{I}\right)$.

Proposition 3.19. *For any given $\underline{I} = (I_\omega)_{\omega\in E}$ the function*

$$\Omega \mapsto \mathit{Hist}\left(\Omega, \underline{I}\right)$$

commutes with the join operation

$$\mathit{Hist}\left(\Omega, \underline{I}\right) \vee \mathit{Hist}\left(\Omega', \underline{I}\right) = \mathit{Hist}\left(\Omega \vee \Omega', \underline{I}\right)$$

Proof. Recall that the join is defined as:

$$\mathrm{Hist}\left(\Omega, \underline{I}\right) \vee \mathrm{Hist}\left(\Omega', \underline{I}\right) := \mathrm{Prime}\left(\mathrm{Ext}\left(\mathrm{Hist}\left(\Omega, \underline{I}\right)\right) \cap \mathrm{Ext}\left(\mathrm{Hist}\left(\Omega', \underline{I}\right)\right)\right)$$

The extended histories are obtained by the lowersets $\Lambda(\Omega)$ and $\Lambda(\Omega')$. For arbitrary preorders the following equality holds $\Lambda(\Theta \cap \Theta') = \Lambda(\Theta) \cap \Lambda(\Theta')$ from which we conclude

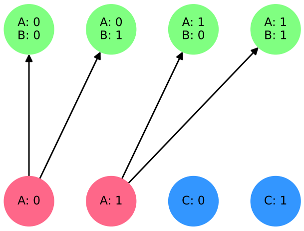
$$\mathrm{Ext}\left(\mathrm{Hist}\left(\Omega, \underline{I}\right)\right) \cap \mathrm{Ext}\left(\mathrm{Hist}\left(\Omega', \underline{I}\right)\right) = \mathrm{Ext}\left(\mathrm{Hist}\left(\Omega \cap \Omega', \underline{I}\right)\right)$$

□

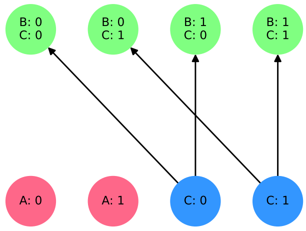
Proposition 3.19 shows that the joins for spaces obtained from preorders are compatible with the hierarchy of causal orders described in Chapter 1. When considering meets, the situation changes as we need to consider union of extended histories. For example, consider $\Omega = \mathrm{total}(A, B) \vee \mathrm{discrete}(C)$ and $\Omega' = \mathrm{discrete}(A) \vee \mathrm{total}(C, B)$:

Ω Ω'

The spaces of histories for $\mathrm{Hist}(\Omega, \{0, 1\})$ and $\mathrm{Hist}(\Omega', \{0, 1\})$ are given by:

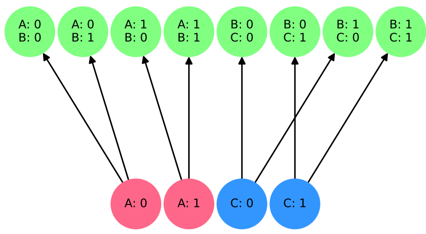


$\text{Hist}(\Omega, \{0,1\})$

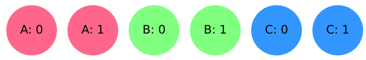


$\text{Hist}(\Omega', \{0,1\})$

The meet in the hierarchy of causal orders is the discrete space $\Omega \wedge \Omega' = \text{discrete}(A, B, C)$. To find the meet of the spaces of histories, we need to consider the $\vee$-prime elements of the union of corresponding extended input histories which is given by the poset on the left hand side (the right hand side represents the histories for the discrete space):



$\text{Hist}(\Omega, \{0,1\}) \wedge \text{Hist}(\Omega', \{0,1\})$



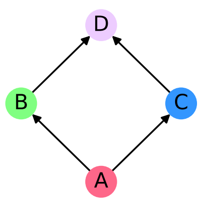
$\text{Hist}(\Omega \wedge \Omega', \{0,1\})$

Satisfying the causal constraints for both Ω and Ω' is different from satisfying tri-partite no-signalling constraints. This observation is witnessed at the level of generality imposed by input histories. We will further describe the space $\text{Hist}(\Omega, \{0,1\}) \wedge \text{Hist}(\Omega', \{0,1\})$, which does not arise from a causal order and is a non-tight space when talking about the hierarchy of causally complete spaces on three events Section 3.3.1.3.

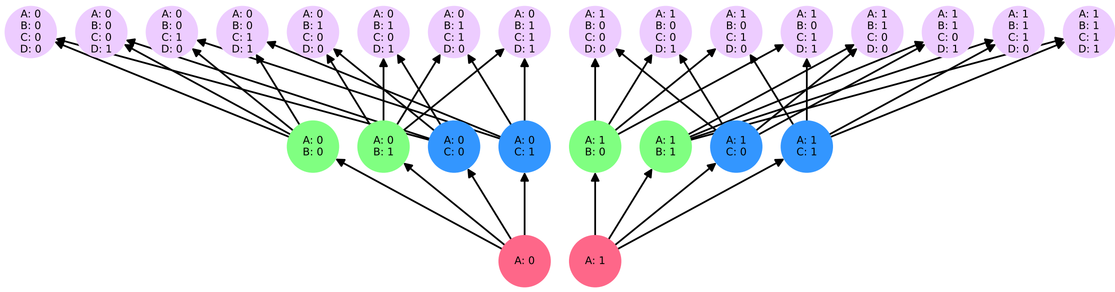
3.3.1 Causally Complete Spaces

In our operational interpretation, input histories are the data upon which the output values at individual events are allowed to depend. When the causal order is given, it is always clear which histories refer to which outputs: the output at event ω is determined by the set of input histories with domain $\text{dom}(h) = \omega\downarrow$ concincing with its causal past and each one describing a different configurations of settings in the causal past of ω. In the more general setting of spaces of input histories, where a definite causal order is not explicit in its definition, is there a way to determine to which event is each history referring to?

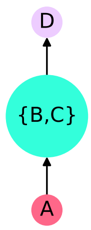
Understanding the phenomenon in $\text{Hist}(\Omega, \underline{I})$ will, once more, inform a suitable generalisation. Consider the example of the causal diamond Ω:



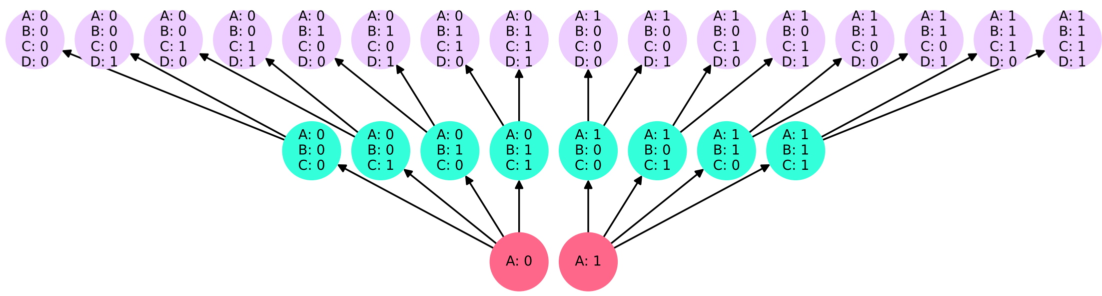
Looking at the space of input histories for a causal poset we observe that an association between input histories and events can be made from the order of histories alone without explicitely mentioning Ω. Indeed, if h is a history with $\mathrm{dom}\,(h) = \omega\downarrow$, then we can look at all input histories $k < h$ strictly below it and recover ω as the only event in $\mathrm{dom}\,(h) \setminus \bigcup_{k<h} \mathrm{dom}\,(k)$: this is the only event not covered by the domains of the histories strictly below h. Iterating this procedure we can colour-code input histories according to the events they refer to as follows



If two or more events are in indefinite causal order, they will togheter index some of the histories in the space. Indeed, consider the following indefinite version of the diamond order above: the space total $(\mathrm{A}, \{\mathrm{B}, \mathrm{C}\}, \mathrm{D})$, where the events B and C are in indefinite causal order:



Because B and C cannot be distinguished by input histories, the histories in the middle layer are referring to two events instead of one:



The operational interpretation of histories referring to multiple events is challenging: in a naive sense, the output value is to be produced ‘simultaneously’. This is problematic because indefinite causal order should not trivialise to causal collapse: under our operational interpretation, distinct events should retain their independent local nature. It should not, for example, be possible to perform the 'swap' function $(b, c) \mapsto (c, b)$ on two events B and C in indefinite causal order: the devices would have to wait for both inputs to be given before producing their outputs, with the effect of delocalising the events.

However, there is an alternative way to look at the presence of multiple terminal events as a form of 'causal incompleteness'. Rather than interpreting such spaces as allowing event de-localisation, we think of them as not providing sufficient information for causal inference to be performed. We will not use causally incomplete spaces directly in our framework: we will focus our efforts on 'causally complete' spaces, and study the incomplete ones through the lens of all possible 'causal completions' they admit.

We formalise the notion described above by introducing 'tip events': set of events assigned to single histories which tell us how to think of them as a generalised notion of a complete 'operational past'. The tip for a single history is not a property of the history itself, i.e thought of simply as a partial function, but rather depends on how the histories are intelaced in forming a space of input histories.

Definition 3.20 (Tip events)**.** *Let Θ be a space of input histories. Given an extended input history $h \in Ext(\Theta)$, we define the* tip events *of h in Θ as the events which are in the domain of h but not in the domain of any history strictly below it:*

$$\begin{aligned} tips_{\Theta}(h) &:= \mathrm{dom}(h) \setminus \bigcup_{k<h} \mathrm{dom}(k) \\ &= \{ \omega \in \mathrm{dom}(h) \,|\, \forall k < h.\, \omega \notin \mathrm{dom}(k) \} \end{aligned} \tag{3.22}$$

Proposition 3.20. *Every input history $h \in \Theta$ has at least one tip event. Every extended input history $h \in Ext(\Theta)$ which is not an input history—i.e. one such that $h \notin \Theta$—has no tip events.*

Proof. Let $h \in \Theta$ be an input history. If $\mathrm{tips}_{\Theta}(h) = \emptyset$ then for every $\omega \in \mathrm{dom}(h)$ there exists h_ω such that $\omega \in \mathrm{dom}(h_\omega)$ and $h_\omega \leq h$. Clearly $\{h_\omega\}_{\omega \in \mathrm{dom}(h)}$ is a set of compatible histories since for all $\omega \in \Omega$ we have that $\mathrm{h}_\omega \leq h$ and $\mathrm{dom}\left(\bigvee_{\omega \in \mathrm{dom}(h)} h_\omega\right) = \mathrm{dom}(h)$. Since $\bigvee_{\omega \in h} h_\omega$ is compatible with h we have that $\bigvee(h_\omega) = h$ contradicting join-primality.

Now let $h \in \mathrm{Ext}(\Theta)$ be an extended input history such that $h \notin \Theta$. Then $h = \bigvee_{k<h} k$ implies that $\mathrm{tips}_{\Theta}(h) = \mathrm{dom}(h) \setminus \bigcup_{k<h} \mathrm{dom}(k) = \emptyset$. □

The definition of causal completeness does not explicitly refer to the free-choice condition. For the rest of the work, all the relevant causally complete spaces will satisfy free will. We, therefore, decided to directly include this property in the definition of *causally complete spaces*.

Definition 3.21 (Causally complete spaces)**.** *Let Θ be a space of input histories satisfying the free-choice condition. We say that Θ is* causally complete *if all input histories $h \in \Theta$ have exactly one tip event and that it is* causally incomplete *otherwise. If Θ is causally complete and $h \in \Theta$, we define the* tip event *of h in Θ to be the unique event in $tips_{\Theta}(h)$:*

$$\Theta \text{ causally complete } \Leftrightarrow \forall h \in \Theta.\, tips_{\Theta}(h) = \{tip_{\Theta}(h)\} \tag{3.23}$$

In the following proposition, we show that when restricting to the causal histories of the type $\text{Hist}\left(\Omega, \underline{I}\right)$, causal completeness characterises the spaces arising from causal posets.

Proposition 3.21. *A space of input histories $\Theta = Hist\left(\Omega, \underline{I}\right)$ induced by a causal order Ω is causally complete if and only if the causal order Ω is causally definite.*

Proof. For all $\omega \in \Omega$ and all $h \in \Theta$ with $\text{dom}\,(h) = \omega\!\downarrow$, we must have:

$$\text{tips}_{\Theta}\,(h) := \text{dom}\,(h) \setminus \bigcup_{k<h} \text{dom}\,(k) = \omega\!\downarrow \setminus \bigcup_{\omega'<\omega} \omega'\!\downarrow = [\omega]_{\simeq}$$

Hence, Θ is causally complete if and only if all causal equivalence classes $[\omega]_{\simeq}$ have size 1. This is precisely the characterisation of causal definiteness for Θ. □

Observation 3.21. *The 'minimal' extended input histories $k \in Ext\,(\Theta)$ are those without sub-histories, i.e. those such:*

$$\forall k' \in Ext\,(\Theta)\,.k' \leq k \Rightarrow k' = k$$

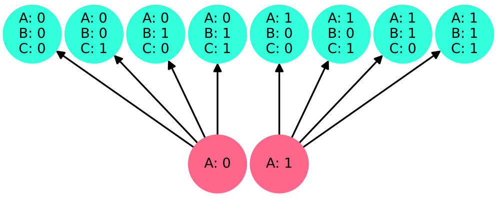
Such k are necessarily $\vee$-prime, so we refer to them as the minimal input histories. *For a minimal input history $h \in \Theta$, we always have $tips_{\Theta}\,(h) = \text{dom}\,(h)$. If Θ is causally complete, this forces any minimal input history h to have $|\,\text{dom}\,(h)\,| = 1$.*

Definition 3.22 (Causal completion)**.** *Let Θ be a space of input histories satisfying the free-choice condition. The* causal completions *of Θ are the closest refinements of Θ which are causally complete, i.e. the maxima of the set of causally complete spaces which are causal refinements of Θ:*

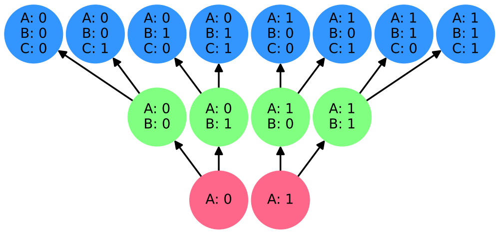
$$CausCompl\,(\Theta) := \max\,\{\,\Theta' \leq \Theta\,|\,\Theta' \textit{ causally complete}\,\} \tag{3.24}$$

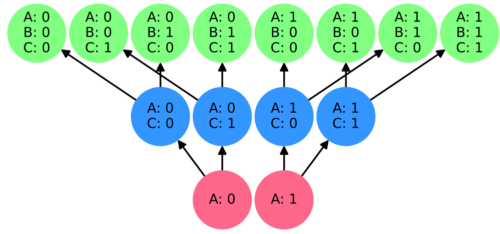
Since the discrete space $Hist\left(discrete\left(E^{\Theta}\right), \underline{I}^{\Theta}\right)$ is always causally complete, the set of causal completions of Θ is never empty. If Θ is itself causally complete, then $CausCompl\,(\Theta) = \{\Theta\}$.

As an example of causal completion, we refer back to the indefinite causal order total (A, {B, C}). The associated space of input histories is causally incomplete because B and C always appear together as tip events (coloured aquamarine, at the top).

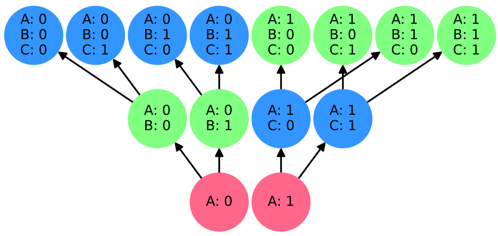


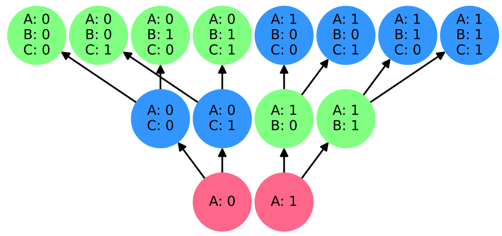
There are four possible causal completions for this space. Two of the causal completions are obtained by imposing a fixed order on events B and C: either B causally precedes C (left below) or B causally succeeds C (right below).





The remaining two causal completions are obtained by imposing an order on events B and C that depends on the input at event A: either B causally precedes C when the input at A is 0 and causally succeeds C when the input at A is 1 (left below), or B causally succeeds C when the input at A is 0 and causally precedes C when the input at A is 1 (right below).





3.3.1.1 Hierarchy of Causally Complete Spaces

Causally complete spaces do not admit any casual ambiguity. Each causally complete space constitutes a causal explanation, and, as is the case for any other space of histories, some can be thought of as being more constrained than others. In this subsection, we study the properties of the hierarchy of spaces when restricting them to causally complete spaces. This investigation transcends the mere mathematical exercise; the hierarchy of causally complete space will constitute the arena for the type of causal discovery described in Chapter 5.

Proposition 3.22. *Let Θ and Θ' be causally complete spaces satisfying $\underline{I}^{\Theta} = \underline{I}$. In general, $\Theta \vee \Theta'$ is not a causally complete space.*

Proof. By Proposition 3.21 we have that considering $\text{Hist}\left(\Omega, \underline{I}\right)$ and $\text{Hist}\left(\Omega', \underline{I}\right)$ for $\Omega = \text{total}\,(A, B)$ and $\Omega' = \text{total}\,(B, A)$ are causally complete spaces of input histories. By Proposition 3.19 we know that $\text{Hist}\left(\Omega, \underline{I}\right) \vee \text{Hist}\left(\Omega', \underline{I}\right) = \text{Hist}\left(\Omega \vee \Omega', \underline{I}\right)$. However, $\Omega \vee \Omega'$ is an indefinite causal order, and Proposition 3.21 shows that the space cannot be causally complete. □

Proposition 3.22. *Causally complete spaces Θ satisfying $\underline{I}^{\Theta} = \underline{I}$ form a subset $\mathit{CCSpaces}\left(\underline{I}\right) \subseteq \mathit{Spaces}_{FC}\left(\underline{I}\right)$ which is closed under meet. We refer to the $\wedge$-semilattice $\mathit{CCSpaces}\left(\underline{I}\right)$ as the* hierarchy of causally complete spaces *for $\underline{I}$.*

Proof. Regarding closure under meet, consider two causally complete spaces $\Theta, \Theta' \in \text{Spaces}_{\text{FC}}\left(\underline{I}\right)$. By Proposition 3.19, the meet $\Theta \wedge \Theta'$ is obtained by taking the $\vee$-prime elements in $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$: this means that every input history $h \in \Theta \wedge \Theta'$ (i.e. a $\vee$-prime element in $\text{Ext}\,(\Theta) \cup \text{Ext}\,(\Theta')$) is either an input history in $h \in \Theta$ (i.e. a $\vee$-prime element in $\mathrm{Ext}(\Theta)$) or an input history $h \in \Theta'$ (i.e. a $\vee$-prime element in $\mathrm{Ext}(\Theta')$). Without loss of generality, assume $h \in \Theta$. We have:

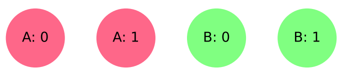
$$\begin{array}{rcl} \mathrm{tips}_{\Theta\wedge\Theta'}(h) & = & \mathrm{dom}(h) \setminus \bigcup_{k\in\Theta\wedge\Theta' \text{ s.t. } k<h} \mathrm{dom}(k) \\ & \subseteq & \mathrm{dom}(h) \setminus \bigcup_{k\in\Theta \text{ s.t. } k<h} \mathrm{dom}(k) \ = \mathrm{tips}_{\Theta}(h) \end{array}$$

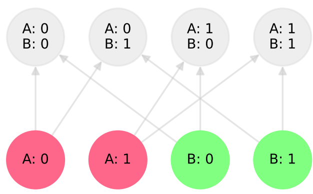
Because Θ is causally complete, $\mathrm{tips}_{\Theta}(h)$ is a singleton, which forces $\mathrm{tips}_{\Theta\wedge\Theta'}(h)$ to also be a singleton (by Proposition 3.20, $h \in \Theta \wedge \Theta'$ has at least one tip event in $\Theta \wedge \Theta'$). □

3.3.1.2 Hierarchy of causally complete spaces on 2 events

As our simplest non-trivial example, we look at the hierarchy of causally complete spaces $\mathrm{CCSpaces}\left((\{0,1\})_{\omega\in\{A,B\}}\right)$ on 2 events A and B with binary inputs $\{0,1\}$. This hierarchy contains 7 causally complete spaces of input histories, ordered in 3 layers. For additional ease of understanding, each space of input histories we examine is displayed together with the associated space of extended input histories: this way, it is easy to check whether a given space refines another.

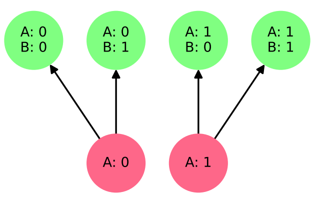
At the bottom of the hierarchy $\mathrm{CCSpaces}\left((\{0,1\})_{\omega\in\{A,B\}}\right)$ is the discrete space, induced by the discrete order on two events. This space has 4 histories: because the two events are causally unrelated, the input histories $\{A:0\}$ and $\{A:1\}$ determine the output on event A, while the input histories $\{B:0\}$ and $\{B:1\}$ determine the output on event B.

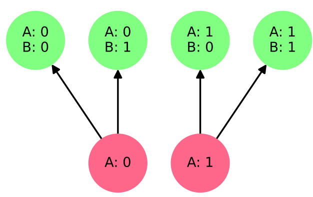




Θ $\mathrm{Ext}(\Theta)$

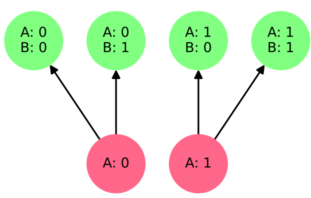
At the top of the hierarchy are the 2 spaces induced by the two possible total orders on two events. Below is the space corresponding to the total order $A \to B$. This space has 6 histories: the input histories $\{A:0\}$ and $\{A:1\}$ determine the output on event A, while the remaining four histories are needed to determine the output on event B, because the latter causally succeeds A.

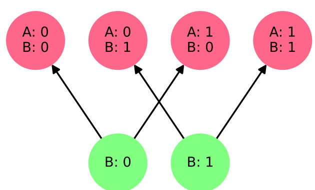




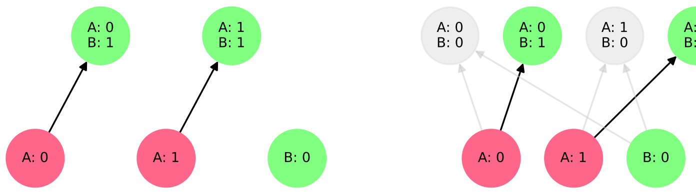
Θ $\mathrm{Ext}(\Theta)$

The two spaces induced by total orders are related by event permutation symmetry $S(\{A,B\})$.



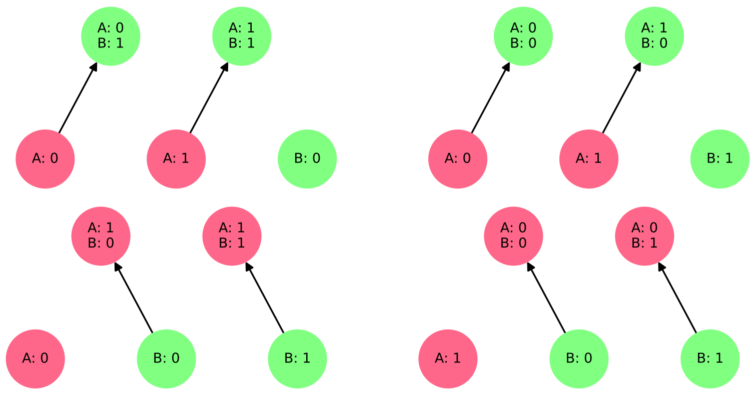


The middle layer of the hierarchy contains 4 spaces, each of them a coarsening of the discrete space and a refinement of one of the two total order spaces. Below is one of the four spaces. This space is a refinement of the space for the total order $\mathrm{A} \to \mathrm{B}$: by looking at the space of extended input histories, we note that the input history $\{\mathrm{B}:0\}$ has been added, with tip event B. This means that the output at B does not depend on the input at A when the input at B is 0: choosing 0 at B causally disconnects B from A. When the input at B is 1, the output at B can still depend on the input at A, as demonstrated by the two input histories $\{\mathrm{A}:0,\ \mathrm{B}:1\}$ and $\{\mathrm{A}:1,\ \mathrm{B}:1\}$ with tip event B.



Θ Ext (Θ)

The four spaces in the middle layer are related by event-input permutation symmetry $S(\{\mathrm{A},\mathrm{B}\}) \times S(I_{\mathrm{A}}) \times S(I_{\mathrm{B}})$: that is, by independently permuting the event set $\{\mathrm{A},\mathrm{B}\}$ and each of the input value sets I_ω (in fact, permuting one of the input sets is enough in this case).



Event-input permutation symmetry is extremely helpful when classifying spaces: because the event and input labels are arbitrary, permutations do not contribute to our general understanding of causality. For a general $\underline{I} = (I_\omega)_{\omega \in E}$, it corresponds to the following group, where $S(X)$ is the group of permutations on a set X:

$$S(E) \times \prod_{\omega \in E} S(I_\omega)$$

Permutation symmetry is broken once an empirical model is specified because conditional probability distributions are not, in general, invariant under its action. In those cases where empirical models retain some symmetry, the latter can be used to reduce the computational burden for causal decomposition. Figure 3.3 (p.109) shows the action of permutation symmetry on a causally complete space on 3 events with binary inputs: the symmetry group does not act freely on this particular equivalence class (which features 24 spaces), but it does on other equivalence classes (27 in total, e.g. equivalence class 30).

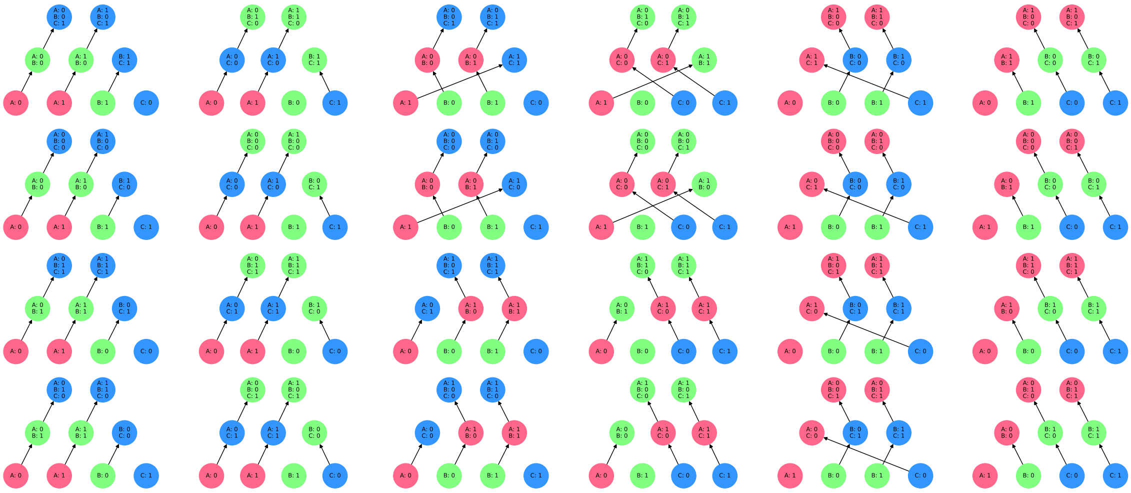


Figure 3.3: All 24 permutations of a causally complete space on 3 events with binary inputs. Specifically, these are the contents of equivalence class 28 in the hierarchy, as depicted in Figure 3.4. Each row is a coset for the action of event permutation symmetry $S(\{A, B, C\})$, which acts freely (on this equivalence class).Each column is a coset for the action of input permutation symmetry $\prod_{\omega \in \{A,B,C\}} S(I_\omega)$, which doesn't act freely (on this equivalence class).

3.3.1.3 Hierarchy of causally complete spaces on 3 events

Having completed our exposition of the hierarchy of spaces on two events with binary inputs, we now move to the hierarchy $\text{CCSpaces}\left((\{0,1\})_{\omega \in \{\text{A,B,C}\}}\right)$ on three events. This hierarchy has 2644 spaces, forming 102 equivalence classes under event-value permutation symmetry. Explaining the algorithm used to enumerate spaces and equivalence classes is outside the scope of this dissertation. An in-depth discussion and detailed anatomy of these spaces can be found in [62].

While the full hierarchy is too complex to display, Figure 3.4 (p.110) depicts the corresponding hierarchy of 102 equivalence classes: in this condensed graph, an edge $i \to j$ indicates that every space in equivalence class i is a closest refinement of some space of equivalence class j. To get a reasonably orderly 3D view of the complete hierarchy, one could imagine stacking all spaces in each equivalence class vertically: edges between spaces in equivalence classes i and j would line up, and their 2D vertical projections would form the edges seen in Figure 3.4 (p.49).

At the bottom of the hierarchy, we find the discrete space, induced by the discrete order discrete (A, B, C), sitting alone in equivalence class 0. This space has 6 histories—one for each event and input choice at that event—all unrelated: this is the no-signalling scenario, where the output at each event depends only on the input at that event. The corresponding space of extended input histories contains all 26 binary-valued partial functions on the 3 events: histories supported by more than one event are not $\vee$-prime in this space.

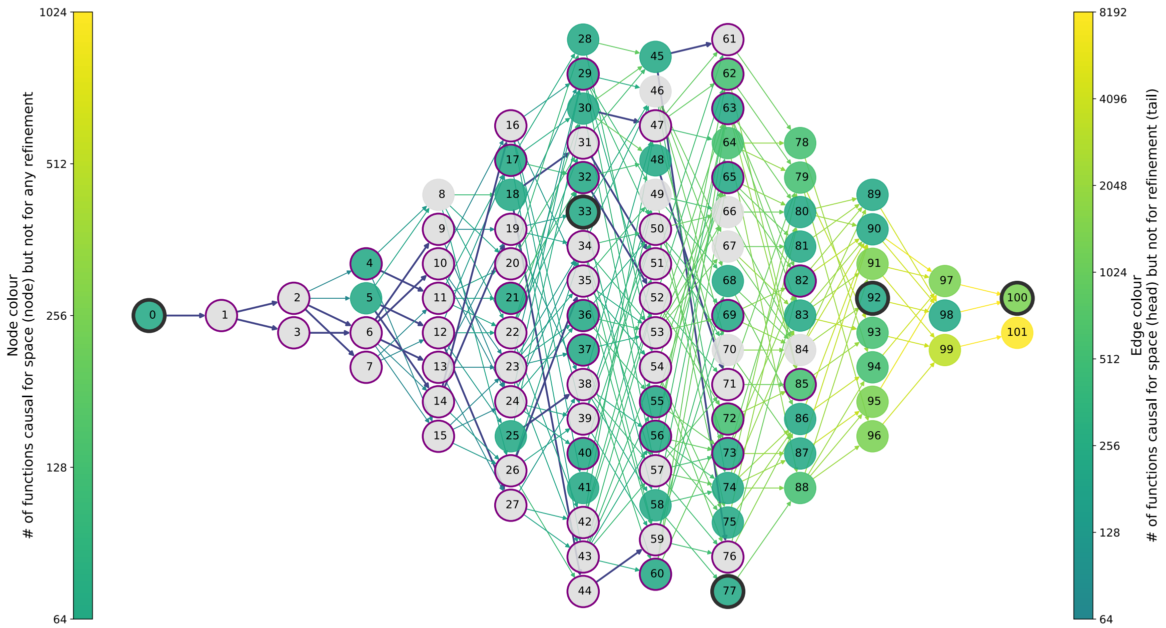
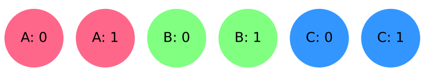
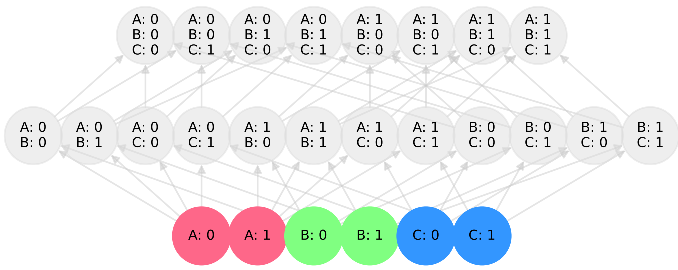


Figure 3.4: The hierarchy of causally complete spaces on 3 events with binary inputs, grouped into 102 equivalence classes under event-input permutation symmetry. An edge $i \rightarrow j$ indicates that some space in eq. class i is a closest refinement for some space in eq. class i. Node colour indicates the number of causal functions for a space which are not causal for any of its subspaces, while edge colour indicates the number of causal functions for the head space that are not causal for the tail space. Grey nodes (e.g. eq. class 1) indicate spaces where all causal functions are also causal for some subspace, while thicker dark blue edges (e.g. edge $0 \rightarrow 1$) indicate that all causal functions for the head space are also causal for the tail space. Causal functions will be defined and extensively discussed in the next chapter. Thin purple borders for nodes indicate eq. classes of non-tight spaces (e.g. eq. class 1). Thick black borders for nodes indicate the eq. classes of spaces induced by causal orders.

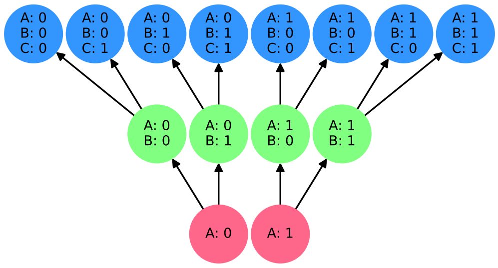


Θ_0

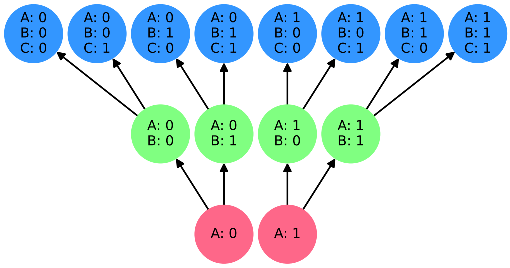


$\text{Ext}(\Theta_0)$

At the top of the hierarchy, we find two equivalence classes of spaces, labelled 100 and 101. Equivalence class 100 contains the 6 spaces induced by total order: below is the space induced by order total (A, B, C). This space has 14 histories, covering all possible combinations of inputs for event A (determining the output at event A), for events $\{A, B\}$ (determining the output at event B) and for events $\{A, B, C\}$ (determining the output at event C). This space coincides with its own space of extended input histories.



Θ_{100}

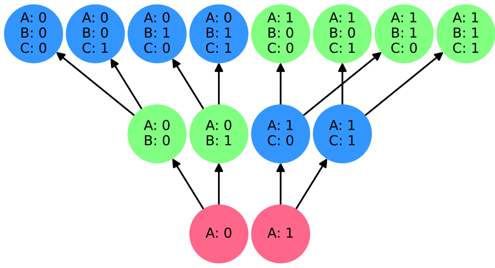


$\text{Ext}\,(\Theta_{100})$

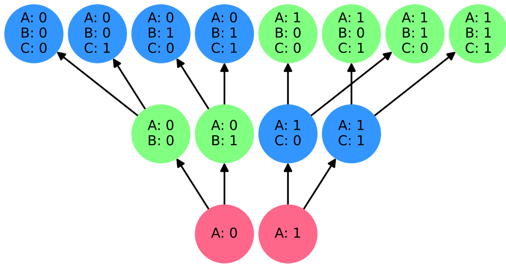
Equivalence class 101 contains the 6 spaces for a 3-party causal switch: below is the space where the input of A determines the total order between B and C, with input 0 at A setting B $\rightarrow$ C and input 1 at A setting C $\rightarrow$ B. This space has 14 histories, covering:

- all inputs for event A, determining the output at A and the total order between B and C
- all inputs for event B when A has input 0, determining the output at B
- all inputs for events $\{B, C\}$ when A has input 0, determining the output at C
- all inputs for event C when A has input 1, determining the output at C
- all inputs for events $\{C, B\}$ when A has input 1, determining the output at B

This space coincides with its own space of extended input histories.



Θ_{101}



$\text{Ext}\,(\Theta_{101})$

The spaces in equivalence class 101 are examples of causally complete spaces not admitting a fixed definite causal order: they are not refinements of $\text{Hist}\,(\Omega, \{0, 1\})$ for any definite causal order Ω on A, B and C. There are 13 equivalence classes consisting of spaces that don't admit a fixed definite causal order, highlighted in Figure 3.5 (p.112).

A thick black border marks the 5 equivalence classes of spaces induced by total orders in Figure 3.4 (p.110). We have already seen equivalence class 0 (for the discrete order) and equivalence class 100 (for total orders): we now look at the remaining three. Equivalence class 92 contains the 3 spaces induced by wedge orders: below is the space induced by order $\text{total}\,(A, C) \vee \text{total}\,(B, C)$. This space has 12 histories, covering all possible combinations of inputs for event A (determining the output at event A), for event B (determining the output at event B), and for events $\{A, B, C\}$ (determining the output at event C). The extended input histories supported by $\{A, B\}$ are not $\vee$-prime in this space.

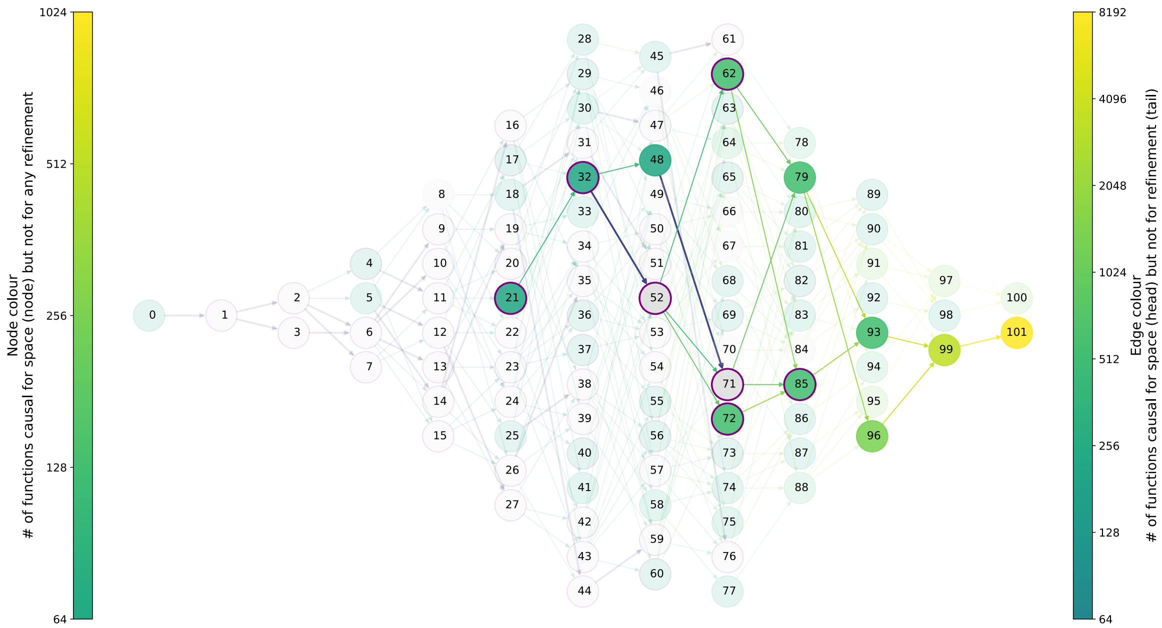
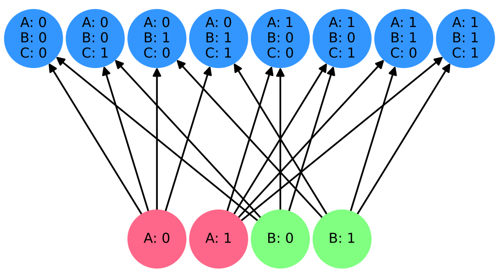
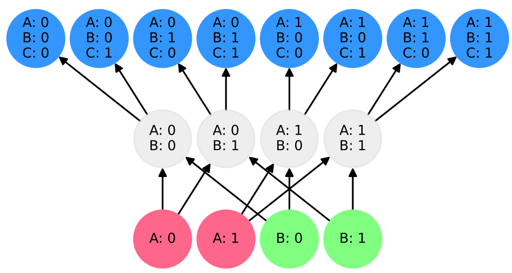


Figure 3.5: The 13 equivalence classes not admitting a fixed definite causal order, highlighted within the hierarchy of causally complete spaces on 3 events with binary inputs. See Figure 3.4 for a discussion of colours and markings.

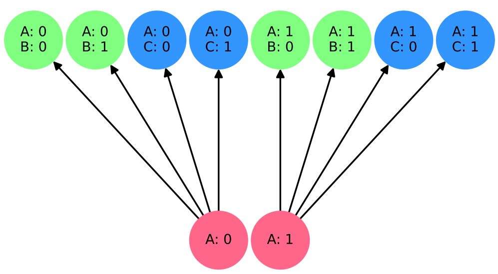


Θ_{92}

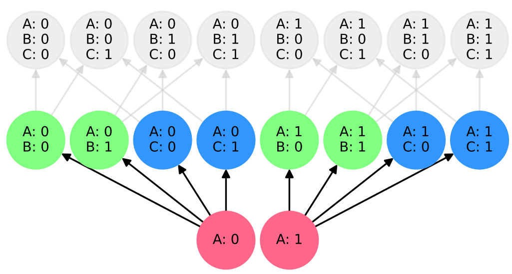


$\mathrm{Ext}(\Theta_{92})$

Equivalence class 77 contains the 3 spaces induced by fork orders: below is the space induced by order total (A, B) ∨ total (A, C). This space has 10 histories, covering all possible combinations of inputs for event A (determining the output at event A), for events {A, B} (determining the output at event B), and for events {A, C} (determining the output at event C). The extended input histories supported by all three events are not ∨-prime in this space.

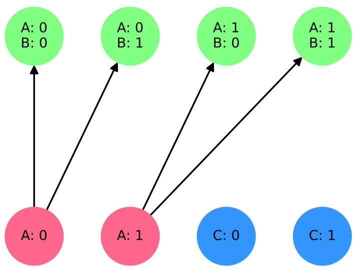


Θ_{77}

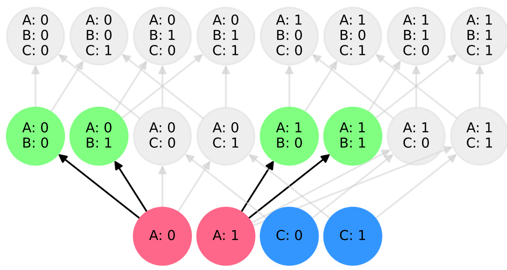


$\mathrm{Ext}(\Theta_{77})$

Equivalence class 33 contains the 6 spaces induced by the disjoint join of a total order on two events with a discrete third event: below is the space induced by order total $(\mathrm{A}, \mathrm{B}) \vee$ discrete (C). This space has 8 histories, covering all possible combinations of inputs for event A (determining the output at event A), for events $\{\mathrm{A}, \mathrm{B}\}$ (determining the output at event B), and for event C (determining the output at event C). The extended input histories supported by either $\{\mathrm{A}, \mathrm{C}\}$ or by all three events are not $\vee$-prime in this space.



Θ_{33}



$\mathrm{Ext}(\Theta_{33})$

Spaces not induced by causal orders can all be understood as introducing input-dependent causal constraints. We already saw this in the 3-party causal switch space Θ_{101}: it refines the (non causally complete) order-induced space Hist (total $(\mathrm{A}, \{\mathrm{B}, \mathrm{C}\}), \{0, 1\})$, by introducing causal constrains on $\{\mathrm{B}, \mathrm{C}\}$ which depend on the input at event A. The spaces in equivalence class 101 might be the iconic example of this mechanism, but all 97 equivalence classes of non-order-induced spaces can be understood this way: we take some order-induced coarsening and study the additional input-dependent causal constraints.

In the most general case of this procedure, we consider a space $\Theta \in \mathrm{Spaces}(\underline{I})$ and a causal order Ω such that $\Theta < \mathrm{Hist}(\Omega, \underline{I})$, i.e. such that:

$$\mathrm{Ext}(\Theta) \supseteq \mathrm{Hist}(\Omega, \underline{I})$$

In particular, the above implies that $\Theta \in \mathrm{Spaces}_{\mathrm{FC}}(\underline{I})$. The extended input histories in $\mathrm{Ext}(\Theta) \setminus \mathrm{Hist}(\Omega, \underline{I})$ correspond to causal constraints that Θ imposes additionally to $\mathrm{Hist}(\Omega, \underline{I})$: if there is a unique minimal choice for Ω (cf. equivalence class 98, discussed below), then the additional constraints are truly input-dependent; if there are multiple minimal choices for Ω (cf. equivalence class 3, discussed below), then the additional constraints might instead be those of a different causal order, independent of any input values. We can restrict our attention to the input histories in $h \in \Theta \cap (\mathrm{Ext}(\Theta) \setminus \mathrm{Hist}(\Omega, \underline{I}))$, because all additional extended input histories arise as their join. For each such input history h, we consider the set K_h of minimal extended input histories from the order-induced space which lie above h:

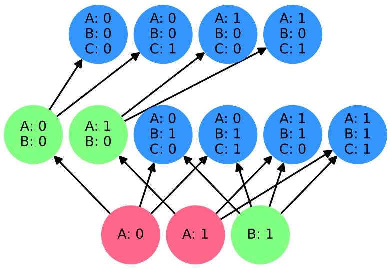
$$K_h := \min\left(h\uparrow \cap \mathrm{ExtHist}(\Omega, \underline{I})\right) \subseteq \mathrm{Ext}(\Theta)$$

We then consider the set E_h of all events which are in the domain of some $k \in K_h$ but not of h:

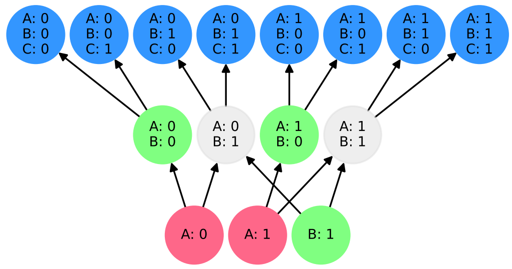
$$E_h := \bigcup_{k \in K_h} \mathrm{dom}\,(k) \setminus \mathrm{dom}\,(h)$$

The additional constraint imposed by h can then be understood as follows: when the events in $\mathrm{dom}\,(h)$ have inputs specified by h, the outputs at the tip events $\mathrm{tips}_\Theta\,(h)$ are independent of the inputs at events in E_h.

As the simplest example of input-dependent causal constraints, we consider space Θ_{98} below, a representative from equivalence class 98 which is a closest refinement of $\mathrm{Hist}\,(\mathrm{total}\,(\mathrm{A}, \mathrm{B}, \mathrm{C})\,, \{0, 1\})$. The only additional history, in this case, is $\{\mathrm{B}:1\}$, imposing the following constraint: when the input at B is 1, the output at B is independent of the input at event A.

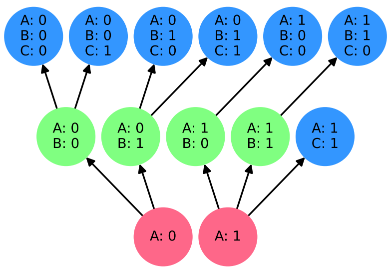


Θ_{98}

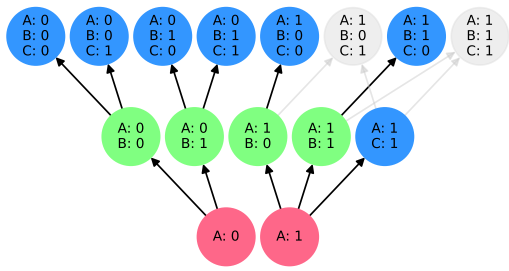


$\mathrm{Ext}\,(\Theta_{98})$

Another simple example is given by Θ_{97} below, a representative from equivalence class 97 which is also a closest refinement of $\mathrm{Hist}\,(\mathrm{total}\,(\mathrm{A}, \mathrm{B}, \mathrm{C})\,, \{0, 1\})$. The only additional history, in this case, is $\{\mathrm{C}:1\}$, imposing the following constraint: when the input at C is 1, the output at C is independent of the input at event B (but not necessarily of the input at event A).

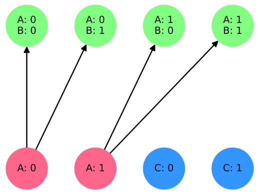


Θ_{97}

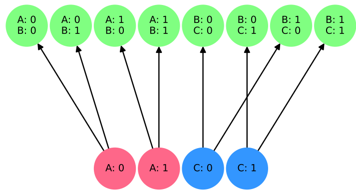


$\mathrm{Ext}\,(\Theta_{97})$

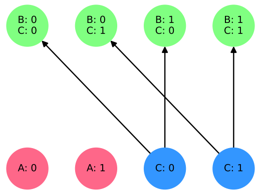
Both examples above are clear cases of input-dependent causal constraints. However, we mentioned that additional causal constraints need not be truly input dependent, as witnessed by our previous example on the meet of order-induced spaces for causal orders $\Omega = \mathrm{total}\,(\mathrm{A}, \mathrm{B}) \vee \mathrm{discrete}\,(\mathrm{C})$ and $\Omega' = \mathrm{discrete}\,(A) \vee \mathrm{total}\,(\mathrm{C}, \mathrm{B})$.



$\Theta_{33} = \text{Hist}(\Omega, \{0,1\})$



$\Theta_{3} = \text{Hist}(\Omega', \{0,1\}) \wedge \text{Hist}(\Omega', \{0,1\})$

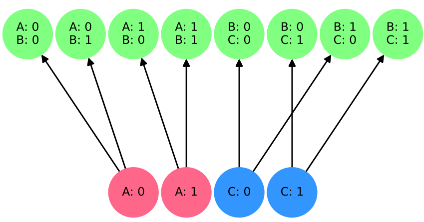


$\text{Hist}(\Omega', \{0,1\})$

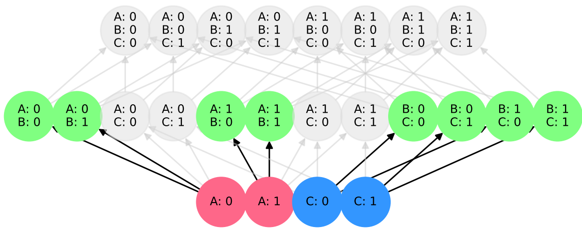
Indeed, the spaces in equivalence class 3 are exactly the meets of 3 pairs of spaces from equivalence class 33 (the other 15 non-trivial meets of pairs in equivalence class 33 all yield the discrete space in equivalence class 0). For space Θ_3, specifically, we get the following additional constraints:

- as a coarsening of order-induced space $\Theta_{33} = \text{Hist}(\text{total}(A,B) \vee \text{discrete}(C), \{0,1\})$, the additional constraints come from the 4 histories with domain $\{B,C\}$: they state that the output on B is independent of the input on A for all possible choices of inputs on $\{B,C\}$.
- as a coarsening of order-induced space $\text{Hist}(\text{discrete}(A) \vee \text{total}(C,B), \{0,1\})$, the additional constraints come from the 4 histories with domain $\{A,B\}$: they state that the output on B is independent of the input on C for all possible choices of inputs on $\{A,B\}$.

Because the additional constraints appear for all possible choices of inputs on their common support, they are not truly input-dependent in this case.



Θ_3



$\text{Ext}(\Theta_3)$

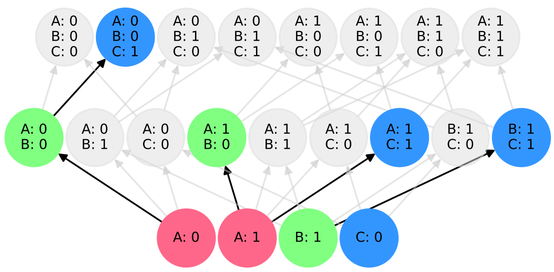
The description of the constraints for space Θ_3 is a bit confusing: one would indeed be forgiven for thinking that these constraints should be equivalent to the no-signalling ones, generated by the discrete space. And, in a sense, they are: the spaces in equivalence class 3 have exactly the same causal functions as the discrete space (as do the spaces in equivalence classes 1, 2, 6, 7, 9, 10 and 13). Furthermore, we will see in Chapter 6 that the causaltope for space Θ_3 (as well as Θ_1) coincides with the no-signalling polytope—the causaltope of the discrete space Θ_0—when the 'non-locality cover' is considered. However, this does not mean that the spaces Θ_3 and Θ_0 are causally equivalent: they are for non-locality purposes, but the former admits strictly more contextual empirical models than the latter, for certain other choices of cover.

3.4 On non-tight spaces

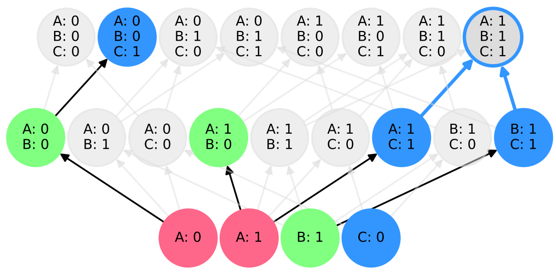
Space Θ_3 is also an example of a 'non-tight' space, where multiple causal orders constrain the events in some histories. Lack of tightness is a peculiar pathology: it implies a form of contextuality where deterministic functions defined compatibly on certain subsets of input histories cannot always be glued together into functions defined on all histories. Put it in more technical terms, we will see later that the pre-sheaf of causal functions on a non-tight space of input histories is not necessarily a sheaf.

Definition 3.23. *Let Θ be a space of input histories. We say that Θ is* tight *if for every (maximal) extended input history $k \in Ext(\Theta)$ and every event $\omega \in \mathrm{dom}(k)$ there is a unique input history $h \in \Theta$ such that $h \leq k$ and $\omega \in tips_\Theta(h)$. We say that Θ is* non-tight *otherwise.*

Non-tight spaces are indicated in Figure 3.4 (p.110) by a thin violet border, and they constitute the majority of examples: out of 102 equivalence classes, 58 are 'non-tight' and 44 are 'tight'. To understand what lack of tightness means concretely, let us consider space Θ_{17} below. In the input histories below extended input history $\{A{:}1,\ B{:}1,\ C{:}2\}$ (circled in blue), the event C appears as a tip event in two separate histories, namely $\{A{:}1,\ C{:}1\}$ and $\{B{:}1,\ C{:}1\}$; edges from the latter input histories to the former extended input histories have also been highlighted blue, for clarity.



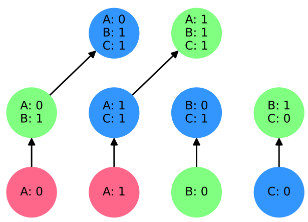
Ext (Θ_{17})



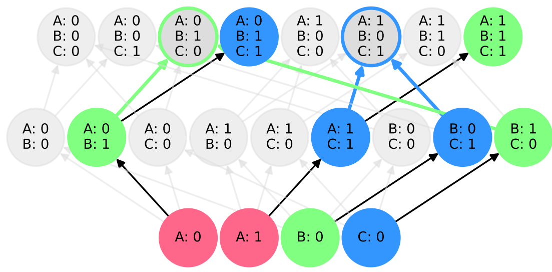
Ext (Θ_{17}) with highlights

The effect of these multiple appearances of C as a tip event is that causal functions on space Θ_{17} must yield identical output values at event C for both input histories $\{A{:}1,\ C{:}1\}$ and $\{B{:}1,\ C{:}1\}$, which would have otherwise been unrelated. Put in other words, in history $\{A{:}1,\ B{:}1,\ C{:}2\}$ the output at event C must satisfy the constraints of two different causal orders: total (A, C, B) (from $\{A{:}1\} \to \{A{:}1,\ C{:}1\} \to \{A{:}1,\ B{:}1,\ C{:}1\}$) and total (B, C, A) (from $\{B{:}1\} \to \{B{:}1,\ C{:}1\} \to \{A{:}1,\ B{:}1,\ C{:}1\}$).

A further example of non-tight space is given by space Θ_{21}, which does not admit a fixed definite causal order: B causally precedes C when the input at B is 0 or the input at A is 1, while it causally succeeds C when the input at C is 0 or the input at A is 1. In this space, there are two extended input histories with 'tip event conflicts' below them: the extended input history $\{A{:}0,\ B{:}1,\ C{:}0\}$ (circled in green) sees B appearing as tip event in the two input histories $\{A{:}0,\ B{:}1\}$ and $\{B{:}1,\ C{:}0\}$ below it, while the extended input history $\{A\colon 1,\ B\colon 0,\ C\colon 1\}$ (circled in blue) sees C appearing as tip event in the two input histories $\{A\colon 1,\ C\colon 1\}$ and $\{B\colon 0,\ C\colon 1\}$ below it.

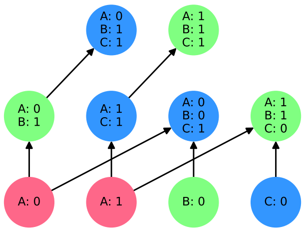


Θ_{21}

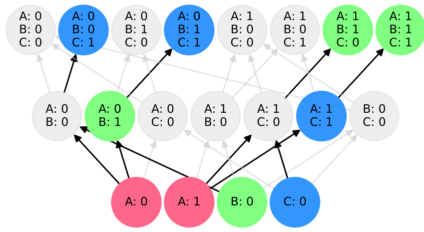


$\mathrm{Ext}\left(\Theta_{21}\right)$ with highlights

The (unique) closest causal coarsening of Θ_{21} which is tight is the space in equivalence class 48 obtained by removing the 'conflicting' input histories $\{B\colon 1,\ C\colon 0\}$ (for event B) and $\{B\colon 0,\ C\colon 1\}$ (for event C). The space is displayed below as Θ_{48}, and it also does not admit a fixed definite causal order.

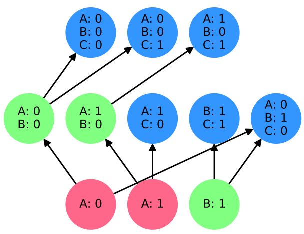


$\Theta_{48,1}$

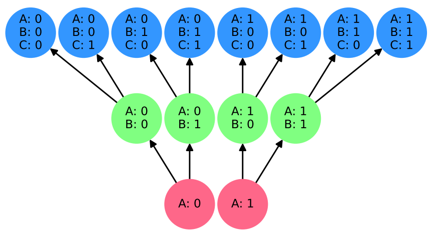


$\mathrm{Ext}\left(\Theta_{48,1}\right)$

A more complicated example of tight space—imposing multiple input-dependent causal constraints—is given Θ_{80}, a representative of equivalence class 80 and causal refinement of $\mathrm{Hist}\left(\mathrm{total}\left(A, B, C\right), \{0, 1\}\right)$.



Θ_{80}

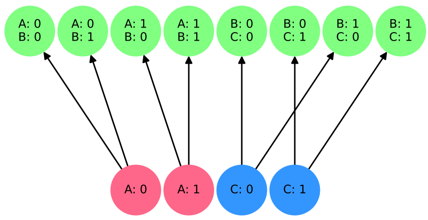


$\mathrm{Hist}\left(\mathrm{total}\left(A, B, C\right), \{0, 1\}\right)$

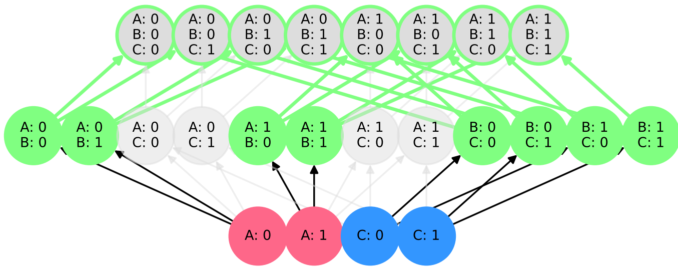
In addition to the causal constraints associated with the total order $\mathrm{total}\left(A, B, C\right)$, space Θ_{80} imposes the following input-dependent causal constraints:

- From the additional history $\{B\colon 1\}$, with tip event B, we get that the input at B is independent of the input at A when the input at B is 1.
- From the additional history $\{A\colon 1,\ C\colon 0\}$, with tip event C, we get that the output at C is independent of the input at B when the input at A is 1 and the input at C is 0.
- From the additional history $\{B\colon 1,\ C\colon 1\}$, with tip event C, we get that the output at C is independent of the input at A when the input at B is 1 and the input at C is 1.

We also show that non-tight spaces arise inevitably when meets of causally complete spaces are considered, even in the simplest case of order-induced spaces (with at least 3 events). Indeed, the (causally complete) space Θ_3 on 3 events which originally sparked our investigation gives an example of such a non-tight meet of order-induced (causally complete) spaces.



Θ_3



$\mathrm{Ext}\left(\Theta_3\right)$ with highlights

Proposition 3.23. *Let Ω be a causal order and let $\underline{I} = (I_\omega)_{\omega\in\Omega}$ be a family of non-empty input sets. The space of input histories $\Theta := \mathit{Hist}\left(\Omega, \underline{I}\right)$ is tight.*

Proof. Let $k \in \mathrm{ExtHist}\left(\Omega, \underline{I}\right)$ and $\omega \in \mathrm{dom}\left(k\right)$. The input history $h := k|_{\omega\downarrow} \in \Theta$ is the unique input history $h \leq k$ with $\omega \in \mathrm{tips}_\Theta\left(h\right)$. As a consequence, Θ is tight. □

Proposition 3.23. *Let Ω and Ω' be two causal orders on the same set of events $E := |\Omega| = |\Omega'|$. Let $\Theta := \mathit{Hist}\left(\Omega, \underline{I}\right)$ and $\Theta' := \mathit{Hist}\left(\Omega', \underline{I}\right)$ be the spaces of input histories induced by the two causal orders for the same family of input sets $\underline{I}$. The meet $\Theta \wedge \Theta'$ is tight if and only if for all $\omega \in E$ we have $\omega\downarrow_\Omega \subseteq \omega\downarrow_{\Omega'}$ or $\omega\downarrow_{\Omega'} \subseteq \omega\downarrow_\Omega$.*

Proof. For every input history $h \leq k$ in the meet $\Theta \wedge \Theta'$ we must have that $h \in \Theta$ or $h \in \Theta'$, because the extended input histories in $\mathrm{Ext}\left(\Theta \wedge \Theta'\right)$ arise the the compatible joins of input histories in the set $\Theta \cup \Theta'$; for the same reason, we must also have that $\mathrm{tips}_{\Theta\wedge\Theta'}\left(h\right) \subseteq \mathrm{tips}_\Theta\left(h\right)$ and $\mathrm{tips}_{\Theta\wedge\Theta'}\left(h\right) \subseteq \mathrm{tips}_{\Theta'}\left(h\right)$. Let $k \in \prod_{\omega\in E} I_\omega$, which is a maximal extended input history for Θ, Θ' and $\Theta \wedge \Theta'$.

In one direction, assume that $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}\left(h\right)$ and $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}\left(h\right)$ for two distinct input histories $h, h' \leq k$: then h and h' cannot be both in Θ or both in Θ', because the two spaces are tight, and without loss of generality we can assume that $h \in \Theta$ and $h' \in \Theta'$. Since $h = k|_{\mathrm{dom}(h)}$ and $h' = k|_{\mathrm{dom}(h')}$, we must have $\mathrm{dom}\left(h\right) \neq \mathrm{dom}\left(h'\right)$; since h and h' both have ω as a tip event, we must furthermore have $\mathrm{dom}\left(h\right) \not\subseteq \mathrm{dom}\left(h'\right)$ and $\mathrm{dom}\left(h'\right) \not\subseteq \mathrm{dom}\left(h\right)$. Because $\mathrm{dom}\left(h\right) = \omega\downarrow_\Omega$ and $\mathrm{dom}\left(h'\right) = \omega\downarrow_{\Omega'}$, we conclude that $\omega\downarrow_\Omega \not\subseteq \omega\downarrow_{\Omega'}$ and $\omega\downarrow_\Omega \not\subseteq \omega\downarrow_{\Omega'}$.

In the other direction, assume that $\omega\downarrow_\Omega \not\subseteq \omega\downarrow_{\Omega'}$ and $\omega\downarrow_\Omega \not\subseteq \omega\downarrow_{\Omega'}$ for some $\omega \in E$. Let h be any input history $h \in \Theta \wedge \Theta'$ such that $h \leq k|_{\omega\downarrow_\Omega}$ and $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}\left(h\right)$: one must exist, because $\omega \in \mathrm{dom}\left(k|_{\omega\downarrow_\Omega}\right)$; analogously let h' be any input history $h' \in \Theta \wedge \Theta'$ such that $h' \leq k|_{\omega\downarrow_{\Omega'}}$ and $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}\left(h'\right)$. If it were the case that $h \in \Theta'$, then $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}\left(h\right) \subseteq \mathrm{tips}_{\Theta'}\left(h\right)$ would imply that $\mathrm{dom}(h) = \omega\downarrow_{\Omega'}$: this would contradict the definition of $h \leq k|_{\omega\downarrow_\Omega}$, and hence we must have $h \in \Theta$; analogously, we must have $h' \in \Theta'$. We conclude that there exist distinct $h, h' \leq k$ such that $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}(h)$ and $\omega \in \mathrm{tips}_{\Theta\wedge\Theta'}(h')$, making $\Theta \wedge \Theta'$ non-tight. □

3.5 Conclusions

In this chapter, we have explored the combinatorial properties of spaces of input histories. We have provided an axiomatic description of the spaces of input histories and described the properties of the associated hierarchy. We explicitly characterised causally complete spaces on 2 and 3 events and briefly discussed non-tightness. This chapter is based on the preprint [63]; however, we must warn the reader that to make the narrative more fluent, we decided to omit reference to the part of the paper dealing with the compositional properties of the spaces of input histories. Unfortunately, this did not allow us to provide an in-depth description of a significant result proved in 'The Combinatorics of Causality'[63], which we briefly mention in these conclusions.

If we have a look at the hierarchy for three events represented in Figure 3.4 (p.110), we see that the top spaces of the hierarchy are given by the equivalence classes for totally ordered events or by the switch spaces on three events. Respectively, Θ_{100} and Θ_{101}. Does this property generalise to arbitrary events? First, we need to generalise the notion of switch spaces:

Definition 3.24. *Let E be a set of events and $\underline{I} = (I_\omega)_{\omega\in E}$ be a family of non-empty input sets. The* causal switch spaces *$CSwitchSpaces\left(\underline{I}\right)$ are defined as follows. If $E = \emptyset$, then $CSwitchSpaces\left(\underline{I}\right) = \emptyset$. Otherwise, for each $\omega_1 \in E$ we can consider:*

$$\begin{aligned} \underline{I}\big|_{\{\omega_1\}} &= (I_\omega)_{\omega\in\{\omega_1\}} \\ \underline{I}\big|_{E\setminus\{\omega_1\}} &= (I_\omega)_{\omega\in E\setminus\{\omega_1\}} \end{aligned}$$

Then the set $CSwitchSpaces\left(\underline{I}\right)$ is defined inductively as follows:

$$\bigcup_{\omega_1\in E}\left\{ Hist\left(\{\omega_1\}, \underline{I}\big|_{\{\omega_1\}}\right) \rightsquigarrow \underline{\Theta} \,\middle|\, \underline{\Theta} \in CSwitchSpaces\left(\underline{I}\big|_{E\setminus\{\omega_1\}}\right)^{I_{\omega_1}} \right\} \tag{3.25}$$

Where the operation $\Theta \rightsquigarrow \Theta'$ is defined in [63] and is a way to construct the timelike composition of spaces of input histories. With this general definition of the class of spaces given by $\mathrm{CCSpaces}\left(\underline{I}\right)$, we can prove that the maximal causally complete spaces are all causal switch spaces.

Theorem 3.25. *Let E be a set of events and $\underline{I} = (I_\omega)_{\omega\in E}$ be a family of non-empty input sets. The maxima of $CCSpaces\left(\underline{I}\right)$ are exactly the causal switch spaces $CSwitchSpaces\left(\underline{I}\right)$.*

The theorem above will be particularly relevant in Chapter 6. To prove that an empirical model is causally definite, i.e. entirely supported in the hierarchy of causally complete spaces, it will be enough to check the decomposition for the switch spaces $\mathrm{CCSpaces}\left(\underline{I}\right)$, significantly simplifying causal discovery.

Chapter 4

The topology of causality II: Causal Data

4.1 Causal functions for tight causally complete spaces

In the previous chapter, we provided a combinatorial description of spaces of input histories and analysed the order theoretic properties of the emerging hierarchy of spaces. Here we invoke sheaf theory to describe a way to track causal data assigned to the spaces of histories.

Describing the space of input histories as a topological space can be understood as passing from a description of the poset of timelike histories, to the explicit description of the hierarchy of contexts induced by an operational scenario.

Our approach differs from the classical causal modelling perspective in that instead of requiring a global inter-contextual explanation, we admit explanations which are valid for each element of a family of contexts provided that a notion of compatibility is retained. We therefore need to explain what constitutes a contextual assignment of causal data.

A choice of open cover corresponds to the possible families of contexts over which probability distributions are simultaneously definable. The hierarchy formed by open covers under refinement corresponds to all possible kinds and degrees of contextuality, including:

- The ‘standard cover’, accommodating generic causal distributions on joint outputs conditional to the maximal extended input histories. It models settings where it is, at the very least, possible to define conditional distributions when all events are taken together.
- The ‘classical cover’ is the coarsest cover, lying at the top of the hierarchy. It models settings admitting a deterministic causal hidden variable explanation.
- The ‘solipsistic cover’ is the finest cover, lying at the bottom of the hierarchy. It models settings more restrictive than those modelled by the standard cover, where it might only be possible to define distributions over the events in the past of some event. That is, the solipsistic cover accommodates all causal distributions on joint outputs conditional to the maximal input histories.

We will see in particular see that a conditional distributions for the standard cover, which is the usual empirical description of a protocol, can be considered classical when it arises by restricting data definable on the global classical cover. Before we embark to formalise this intuition we need to explain what it means for a causal function to be compatible with a given causal assumption.

Causal functions are thought of as deterministic assignment of joint outputs to joint inputs embodying some causal constraints. We, therefore, start with a very general definition in this direction:

Definition 4.1 (joint IO functions)**.** *A* joint input-output function *for an operational scenario* $(E, \underline{I}, \underline{O})$ *is any function mapping joint inputs to joint outputs:*

$$F : \prod_{\omega \in E} I_\omega \to \prod_{\omega \in E} O_\omega$$

For every choice $k \in \prod_{\omega \in E} I_\omega$ *of joint inputs and every event* $\omega \in E$*, we refer to the component* $F(k)_\omega \in O_\omega$ *as the output of* F *at* ω*.*

We now need to classify general input-output functions with respect to some externally imposed causal constraints. When is a function causal for the space Θ? In general, the most intuitive answer would be the following:

Definition 4.2 (Causal joint IO functions)**.** *Let* $F : \prod_{\omega \in E} I_\omega \to \prod_{\omega \in E} O_\omega$ *be a joint IO function for an operational scenario* $(E, \underline{I}, \underline{O})$ *and let* Θ *be a space of histories such that* $\underline{I}^\Theta = \underline{I}$*. We say that* F *is* causal *for* Θ *if for all input histories* $h \in \Theta$*, the outputs* $F(k)|_{tips_\Theta(h)}$ *at the tips of* h *are the same for all joint inputs* $k \in \prod_{\omega \in E} I_\omega$ *such that* $h \leq k$*.*

Definition 4.2 makes it easy to check whether a function is causal for a given Θ, but it is not given in a form which is amenable to our discussion. Instead of providing the conditions that a causal joint IO function has to satisfy to be considered valid, we aim to fully characterise the data needed to define such causal functions.

For the special case of tight causally complete spaces, this is an easy endeavour, which provides an equivalent definition to the aforementioned characterisation of *causal functions*:

Definition 4.3 (Causal functions for tight causally complete spaces)**.** *Let* Θ *be a tight causally complete space and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty set of outputs. The* causal functions *CausFun* $(\Theta, \underline{O})$ *for space* Θ *and outputs* $\underline{O}$ *are the functions mapping each history in* Θ *to the output value for its tip event:*

$$\mathit{CausFun}\left(\Theta, \underline{O}\right) := \prod_{h \in \Theta} O_{tip_\Theta(h)}$$

For the case in which O_ω *is the same for all* $\omega \in E$ *then the set of causal functions takes the simplified form:*

$$\mathit{CausFun}\,(\Theta, O) = \Theta \to O \tag{4.1}$$

In such cases we use the shorthand CausFun (Θ, O) *to denote CausFun* $\big(\Theta, (O)_{\Theta \in E^\Theta}\big)$.

We see that input histories are key to understanding the structure of causal functions: the joint IO functions which are causal for a tight and causally complete space Θ are in exact correspondence with functions mapping input histories to the outputs at their tip events. The two characterisations are however not so straightforwardly equivalent, in particular standard IO functions are defined on the joint inputs $\prod_{\omega \in E^\Theta} I_\omega$, representing the maximal extended histories of a space (provided that space satisfies the free-choice condition), while CausFun (Θ, O) are valued on all the prime histories. In order to relate the perspectives offered by Definition 4.3 and Definition 4.2 it is useful to describe general 'extended causal functions', defined on all extended input histories.

Definition 4.4 (Extended functions)**.** *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_\omega \in E^\Theta$ *be a family of non-empty output sets. The* extended functions *on* Θ *(with output* $\underline{O}$*) are the elements of the following set:*

$$ExtFun\left(\Theta, \underline{O}\right) := \prod_{k \in Ext(\Theta)} \prod_{\omega \in \mathrm{dom}(k)} O_\omega$$

Given a causal function for a tight and causally complete space Θ in the form prescribed by Definition 4.3 we can turn it into an *extended causal functions* by gluing the output values of f over compatible input histories:

Definition 4.5 (Extended causal functions)**.** *Let* Θ *be a tight causally complete space and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty set of outputs. For each causal function* $f \in CausFun\left(\Theta, \underline{O}\right)$*, the extended causal function* $Ext\,(f) \in ExtFun\left(\Theta, \underline{O}\right)$ *is defined as follows:*

$$Ext\,(f)\,(k) := (f(h_{k,\omega}))_{\omega \in \mathrm{dom}(k)} \text{ for all } k \in Ext\,(\Theta) \tag{4.2}$$

where $h_{k,\omega}$ *is the unique input history* $h \in \Theta$ *such that* $h \leq k$ *and* $tip_\Theta\,(h) = \omega$*. Recall that such a history is guaranteed to be unique by the tightness of the space.*

Definition 4.6. *We say that* $\hat{F} \in ExtFun\left(\Theta, \underline{O}\right)$ *is causal if it is an extended causal function, i.e if it takes the form*

$$\hat{F} = Ext\,(f)$$

for same $f \in CausFun\left(\Theta, \underline{O}\right)$*. The set of extended causal functions is denoted by ExtCausFun* $\left(\Theta, \underline{O}\right)$.

Alternatively, one can see causal functions for a space as being the extended functions that satisfy some compatibility requirement with respect to the underlying space of histories.

Definition 4.7 (Consistency and gluing condition)**.** *Let* Θ *be a space of input histories and let* $\hat{F} \in ExtFun\left(\Theta, \underline{O}\right)$ *be an extended function.*

1. *We say that* $\hat{F}$ *satisfies the* consistency condition *is* $\hat{F}(k') \leq \hat{F}(k)$ *for all* $k, k' \in Ext\,(\Theta)$ *such that* $k' \leq k$.

2. *We say that* $\hat{F}$ *satisfies the* gluing condition *if it respects compatible joins:* $\hat{F}(k)$ *and* $\hat{F}(k')$ *are compatible for all compatible* $k, k' \in Ext(\Theta)$ *and we have* $\hat{F}(k \vee k') = \hat{F}(k) \vee \hat{F}(k')$.

Proposition 4.7. *The* consistency condition *is equivalent to the* gluing condition.

Proof. From $k' \leq k$ we know that k' and k are compatible and $k \vee k' = k$. Suppose that $\hat{F}$ satisfies the gluing condition, we have that $\hat{F}(k) = \hat{F}(k \vee k') = \hat{F}(k) \vee \hat{F}(k')$ but $\hat{F}(k') \leq \hat{F}(k) \vee \hat{F}(k')$ therefore $\hat{F}(k') \leq \hat{F}(k)$.

Suppose that $\hat{F}$ satisfies the consistency condition. Let k and k' be compatible, then $k', k \leq k \vee k'$ and by the consistency condition $\hat{F}(k'), \hat{F}(k) \leq \hat{F}(k \vee k')$. Then $\hat{F}(k') \vee \hat{F}(k) \leq \hat{F}(k \vee k')$, which jointly with $\text{dom}(k \vee k') = \text{dom}(k) \cup \text{dom}(k')$ implies that $\hat{F}(k \vee k') = \hat{F}(k) \vee \hat{F}(k')$ since the two domains must coincide $\text{dom}\left(\hat{F}(k \vee k')\right) = \text{dom}\left(\hat{F}(k) \vee \hat{F}(k')\right)$. □

We denote the set of extended causal function for $(\Theta, \underline{O})$ as $\text{ExtCausFun}\left(\Theta, \underline{O}\right)$. Compatibility with the join of the space of histories is equivalent to being causal with respect to Definition 4.6 for tight and causally complete spaces:

Theorem 4.8. *Let* Θ *be a tight causally complete space, and let* $\underline{O}$ *be a family of non-empty sets of outputs. The extended function* $\hat{F} \in ExtFun\left(\Theta, \underline{O}\right)$ *which are causal are exactly those which satisfy the consistency condition. Indeed for every consistent* $\hat{F}$ *we can find a unique* $Prime\left(\hat{F}\right) \in CausFun\left(\Theta, \underline{O}\right)$ *such that* $Ext\left(Prime\left(\hat{F}\right)\right) = \hat{F}$*:*

$$Prime\left(\hat{F}\right) := h \mapsto \hat{F}(h)_{tip_\Theta(h)} \tag{4.3}$$

Proof. Let $f \in \text{CausFun}\left(\Theta, \underline{O}\right)$ we want to show that $\text{Ext}(f)$ satisfies the consistency condition. Let $k \in \text{Ext}(\Theta)$ then tightness implies that $h_{k',\omega} = h_{k,\omega}$ for all $\omega \in \text{dom}(k') \subseteq \text{dom}(k)$, where $k' \leq k$ are two extended input histories. The output value at each $\omega \in \text{dom}(k')$ is then the same for extended output histories $\text{Ext}(f)(k')$ and $\text{Ext}(f)(k)$: $\text{Ext}(f)(k')_\omega = \text{Ext}(f)(k)_\omega$, proving that $\text{Ext}(f)(k') \leq \text{Ext}(f)(k)$. This holds for every extended function proving their consistency.

Consider now an extended causal function $\hat{F}$ which satisfies the consistency condition. From $\hat{F}$ we construct a $\text{Prime}\left(\hat{F}\right) \in \text{CausFun}\left(\Theta, \underline{O}\right)$ given by:

$$\text{Prime}\left(\hat{F}\right) := h \mapsto \hat{F}(h)_{\text{tip}_\Theta(h)}$$

We see that $\text{Prime}\left(\hat{F}\right)$ is causal by simply being a function of the right type, i.e $\Theta \to \underline{O}$. It remains to show that $\text{Ext}\left(\text{Prime}\left(\hat{F}\right)\right) = \hat{F}$. Note that for any $k \in \text{Ext}(\Theta)$ and $\omega \in \text{dom}(k)$ we have $h_{k,\omega} \leq k$, so by consistency $\hat{F}(h_{k,\omega}) \leq \hat{F}(k)$. In particular $\hat{F}(h_{k,\omega})_\omega = \hat{F}(k)_\omega$ for $\omega \in \text{dom}\left(h_{k,\omega}\right) \cap \text{dom}(k)$. By the definition of the operators Ext and Prime respectively acting on causal functions and extended causal functions, we have that

$$\hat{F}(h_{k,\omega})_\omega = \text{Prime}\left(\hat{F}\right)(h_{k,\omega}) = \text{Ext}\left(\text{Prime}\left(\hat{F}\right)\right)(k)_\omega$$

concluding that $\hat{F}(k)_\omega = \text{Ext}\left(\text{Prime}\left(\hat{F}\right)\right)(k)_\omega$ and therefore $\hat{F}(k) = \text{Ext}\left(\text{Prime}\left(\hat{F}\right)\right)(k)$ for all $k \in \text{Ext}(\Theta)$. Also, the definition of $\text{Ext}(f)$ implies that $\text{Ext}(f)(h)_{\text{tip}_\Theta(h)} = f(h)$, proving the uniqueness claim. □

Given a function f which is causal for a tight causally complete space Θ, we can restrict the extended $\text{Ext}(f)$ to the maximal extended input history and get a causal IO map for Θ (Definition 4.2). Conversely, any causal joint IO function arises from a unique choice of $f \in \text{CausFun}\left(\Theta, \underline{O}\right)$:

Proposition 4.8. *Let Θ be a tight causally complete space and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. For every $f \in \text{CausFun}\left(\Theta, \underline{O}\right)$, the restriction of $\text{Ext}(f)$ to the maximal extended input histories $\prod_{\omega \in E^\Theta} I_\omega$ is a joint function for the operational scenario $(E^\Theta, \underline{I}^\Theta, \underline{O})$ which is causal for Θ. Conversely, any joint IO function F for $(E^\Theta, \underline{I}^\Theta, \underline{O})$ which is causal for Θ arises as the restriction of an extended causal function $\text{Ext}(f)$ to maximal extended input histories, where $f \in \text{CausFun}\left(\Theta, \underline{O}\right)$ can be defined as follows:*

$$f(h) := F(k)_{tip_\Theta(h)} \text{ for any minimal ext.input history } k \text{ s.t } h \leq k \tag{4.4}$$

Proof. Because of Θ satisfying the free-choice condition we have that:

$$\prod_{\omega \in E^\Theta} I^\Theta_\omega \subseteq \text{Ext}(\Theta)$$

and $\text{Ext}(f)$ for $f \in \text{CausFun}\left(\Theta, \underline{O}\right)$ is defined on every extended history. The restriction of $\text{Ext}(f)$ to the maximal extended histories defines a joint IO function for the operational scenario $(E^\Theta, \underline{I}^\Theta, \underline{O})$. To show that it is causal we must check that for every $h \in \Theta$ and $k, k' \in \prod_{\omega \in E^\Theta}$ such that $h \leq k, k'$ then $\text{Ext}(f)_{\text{tip}_\Theta(h)}(k) = \text{Ext}(f)_{\text{tip}_\Theta(h)}(k')$. By tightness we know that $h_{k,\text{tip}_\Theta(h)} = h$, so that

$$\text{Ext}(f)_{\text{tip}_\Theta(h)}(k) = f(h)$$

independently of k.

Conversely, consider a joint IO function F which is causal for Θ. We intend to extend F to a function $\hat{F} \in \text{ExtFun}\left(\Theta, \underline{O}\right)$ as follows:

$$\hat{F}(k) := F(\hat{k})\big|_{\text{dom}(k)} \text{ for any maximal ext. input history } \hat{k} \text{ s.t. } k \leq \hat{k}$$

We must show that the function $\hat{F}$ is well-defined. Given a $k \in \text{Ext}(\Theta)$ and an $\omega \in \text{dom}(k)$, we have $h_{k,\omega} \leq k \leq \hat{k}$ for all choices of $\hat{k}$ in the definition of $\hat{F}(k)$ above. Causality of the joint IO function F for Θ then implies that the value $F_\omega(\hat{k})$ be the same for all such choices of $\hat{k}$, making $\hat{F}(k)_\omega$ well-defined for all $\omega \in \text{dom}(k)$; as a consequence, $\hat{F}$ is well defined as a whole.

If f is defined as in Equation 4.4, we have $\hat{F}(h)_{\text{tip}_\Theta(h)} = f(h)$, so that $\hat{F}$ satisfies the first condition in Theorem 4.8. Furthermore, for any two compatible $k, k' \in \text{Ext}(\Theta)$ and any maximal extended input history $\hat{k}$ such that $k \vee k' \leq \hat{k}$, we have that:

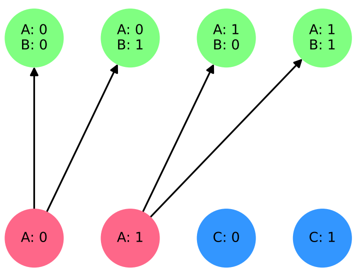
$$\begin{aligned} \hat{F}(k \vee k') &= F(\hat{k})|_{\text{dom}(k \vee k')} \\ \hat{F}(k) &= F(\hat{k})|_{\text{dom}(k)} \\ \hat{F}(k') &= F(\hat{k})|_{\text{dom}(k')} \end{aligned}$$

From the above, it immediately follows that $\hat{F}(k \vee k') = \hat{F}(k) \vee \hat{F}(k')$: this means that $\hat{F}$ also satisfies the second condition in Theorem 4.8, allowing us to conclude that $\hat{F} = \text{Ext}(f)$. Restricting $\hat{F} = \text{Ext}(f)$ to the maximal extended input histories yields back F, because on such histories k we have $\hat{F}(k) = F(k)$ by definition of $\hat{F}$. □

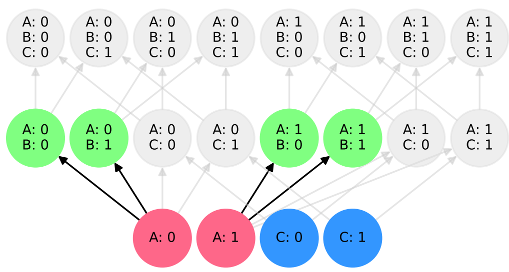
The definitions that we have seen so far are a bit conceptually convoluted. However, this subsection delivers the important task of giving a rigorous equivalence between the functions defined via histories and what is typically understood as being a causal function mapping joint inputs to joint outputs.

4.1.1 Example: causally definite space

As a concrete example, we look at a space induced by causal order total (A, B) ∨ discrete (C) with a choice of binary inputs for all events.



Θ_{33}



$\text{Ext}(\Theta_{33})$

A generic function $F : \{0, 1\}^3 \to \{0, 1\}^3$ with binary outputs takes following form:

$$F(k_A, k_B, k_C) = \begin{pmatrix} F_A(k_A, k_B, k_C) \\ F_B(k_A, k_B, k_C) \\ F_C(k_A, k_B, k_C) \end{pmatrix}$$

where the binary functions $F_\omega : \{0, 1\}^{\{A,B,C\}} \to \{0, 1\}$ assign outputs at each of the events $\omega \in \Omega$. In the case of causal order total (A, B) ∨ discrete (C), we expect joint IO functions F which are causal for Θ_{33} to take the following simplified form, for generic G_A, G_B and G_C:

$$F(k_A, k_B, k_C) = \begin{pmatrix} G_A(k_A) \\ G_B(k_A, k_B) \\ G_C(k_C) \end{pmatrix}$$

This is because A and C are independent events but the output of B can depend on A. Indeed, we show that Definition 4.2 implies the above form for $F(k)$:

- The component $F_A(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{A\!:\!k_A\} \leq k$, for each choice of $k_A \in \{0,1\}$:

$$\begin{aligned} & F_A(\{A\!:\!k_A,\ B\!:\!0,\ C\!:\!0\}) = F_A(\{A\!:\!k_A,\ B\!:\!0,\ C\!:\!1\}) \\ = \quad & F_A(\{A\!:\!k_A,\ B\!:\!1,\ C\!:\!0\}) = F_A(\{A\!:\!k_A,\ B\!:\!1,\ C\!:\!1\}) \end{aligned}$$

  This means that $F_A(k_A, k_B, k_C) = G_A(k_A)$ for a generic function $G_A : \{0,1\} \to \{0,1\}$.

- The component $F_B(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{A\!:\!k_A,\ B\!:\!k_B\} \leq k$, for each choice of $k_A, k_B \in \{0,1\}$:

$$F_B(\{A\!:\!k_A,\ B\!:\!k_B,\ C\!:\!0\}) = F_B(\{A\!:\!k_A,\ B\!:\!k_B,\ C\!:\!1\})$$

  This means that $F_B(k_A, k_B, k_C) = G_B(k_A, k_B)$ for a generic function $G_B : \{0,1\}^2 \to \{0,1\}$.

- The component $F_C(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{C\!:\!k_C\} \leq k$, for each choice of $k_C \in \{0,1\}$:

$$\begin{aligned} & F_C(\{A\!:\!0,\ B\!:\!0,\ C\!:\!k_C\}) = F_C(\{A\!:\!0,\ B\!:\!1,\ C\!:\!k_C\}) \\ = \quad & F_C(\{A\!:\!1,\ B\!:\!0,\ C\!:\!k_C\}) = F_C(\{A\!:\!1,\ B\!:\!1,\ C\!:\!k_C\}) \end{aligned}$$

  This means that $F_C(k_A, k_B, k_C) = G_C(k_C)$ for a generic function $G_C : \{0,1\} \to \{0,1\}$.

Given such a joint IO function F, the causal function $f \in \mathrm{CausFun}(\Theta_{33}, \{0,1\})$ defined by Equation 4.4 takes the following form:

$$\begin{aligned} \{A\!:\!k_A\} & \overset{f}{\mapsto} G_A(k_A) \\ \{A\!:\!k_A,\ B\!:\!k_B\} & \overset{f}{\mapsto} G_B(k_A, k_B) \\ \{C\!:\!k_C\} & \overset{f}{\mapsto} G_C(k_C) \end{aligned}$$

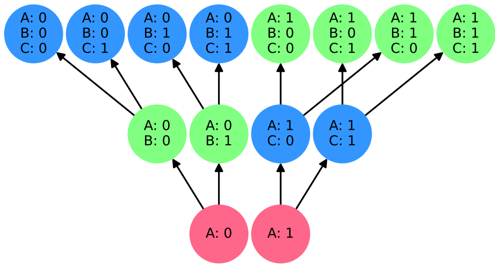
The extended causal function $\mathrm{Ext}(f)$ then takes the following form:

$$\begin{aligned} \{A\!:\!k_A\} & \overset{\mathrm{Ext}(f)}{\mapsto} \{A\!:\!G_A(k_A)\} \\ \{C\!:\!k_C\} & \overset{\mathrm{Ext}(f)}{\mapsto} \{C\!:\!G_C(k_C)\} \\ \{A\!:\!k_A,\ B\!:\!k_B\} & \overset{\mathrm{Ext}(f)}{\mapsto} \{A\!:\!G_A(k_A),\ B\!:\!G_B(k_A, k_B)\} \\ \{A\!:\!k_A,\ C\!:\!k_C\} & \overset{\mathrm{Ext}(f)}{\mapsto} \{A\!:\!G_A(k_A),\ C\!:\!G_C(k_C)\} \\ \{A\!:\!k_A,\ B\!:\!k_B,\ C\!:\!k_C\} & \overset{\mathrm{Ext}(f)}{\mapsto} \{A\!:\!G_A(k_A),\ B\!:\!G_B(k_A, k_B),\ C\!:\!G_C(k_C)\} \end{aligned}$$

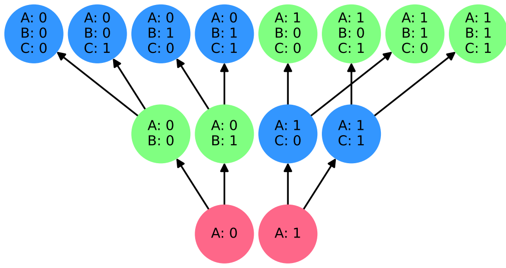
The last line of the definition of $\mathrm{Ext}(f)$ above is its restriction to the maximal extended input histories, which coincides with the original definition of F.

4.1.2 Example: causally indefinite space

We consider another tight and causally complete space, in this case exhibiting dynamical causal order, the space for the causal switch:



Θ_{101}



$\text{Ext}\,(\Theta_{101})$

Recall that this space is tight, but not order-induced: the definite causal order between events B and C depends on the input choice at event A. Lets consider again a generic function $F : \{0,1\}^3 \to \{0,1\}^3$ with binary outputs:

$$F(k_\text{A}, k_\text{B}, k_\text{C}) = \begin{pmatrix} F_\text{A}(k_\text{A}, k_\text{B}, k_\text{C}) \\ F_\text{B}(k_\text{A}, k_\text{B}, k_\text{C}) \\ F_\text{C}(k_\text{A}, k_\text{B}, k_\text{C}) \end{pmatrix}$$

Definition 4.2 gives the following constraints on a joint IO function F which is causal for Θ_{101}:

- The component $F_\text{A}(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{\text{A}\!:\!k_\text{A}\} \leq k$, for each choice of $k_\text{A} \in \{0,1\}$:

  $$\begin{aligned} & F_\text{A}\left(\{\text{A}\!:\!k_\text{A},\ \text{B}\!:\!0,\ \text{C}\!:\!0\}\right) = F_\text{A}\left(\{\text{A}\!:\!k_\text{A},\ \text{B}\!:\!0,\ \text{C}\!:\!1\}\right) \\ = \ & F_\text{A}\left(\{\text{A}\!:\!k_\text{A},\ \text{B}\!:\!1,\ \text{C}\!:\!0\}\right) = F_\text{A}\left(\{\text{A}\!:\!k_\text{A},\ \text{B}\!:\!1,\ \text{C}\!:\!1\}\right) \end{aligned}$$

  This means that $F_\text{A}(k_\text{A}, k_\text{B}, k_\text{C}) = G_\text{A}(k_\text{A})$ for a generic function $G_\text{A} : \{0,1\} \to \{0,1\}$.

- The component $F_\text{B}(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{\text{A}\!:\!0,\ \text{B}\!:\!k_\text{B}\} \leq k$, for each choice of $k_\text{B} \in \{0,1\}$:

  $$F_\text{B}\left(\{\text{A}\!:\!0,\ \text{B}\!:\!k_\text{B},\ \text{C}\!:\!0\}\right) = F_\text{B}\left(\{\text{A}\!:\!0,\ \text{B}\!:\!k_\text{B},\ \text{C}\!:\!1\}\right)$$

  This means that $F_\text{B}(0, k_\text{B}, k_\text{C}) = G_{\text{B},0}(k_\text{B})$ for a generic function $G_{\text{B},0} : \{0,1\} \to \{0,1\}$.

- The component $F_\text{B}(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{\text{A}\!:\!1,\ \text{B}\!:\!k_\text{B},\ \text{C}\!:\!k_\text{C}\} \leq k$, for each choice of $k_\text{B}, k_\text{C} \in \{0,1\}$. This doesn't impose any constraints, as the only such k is $k = \{\text{A}\!:\!1,\ \text{B}\!:\!k_\text{B},\ \text{C}\!:\!k_\text{C}\}$ itself. This means that $F_\text{B}(1, k_\text{B}, k_\text{C}) = G_{\text{B},1}(k_\text{B}, k_\text{C})$ for a generic function $G_{\text{B},1} : \{0,1\}^2 \to \{0,1\}$.

- The component $F_\text{C}(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{\text{A}\!:\!1,\ \text{C}\!:\!k_\text{C}\} \leq k$, for each choice of $k_\text{C} \in \{0,1\}$:

  $$F_\text{C}\left(\{\text{A}\!:\!0,\ \text{B}\!:\!0,\ \text{C}\!:\!k_\text{C}\}\right) = F_\text{C}\left(\{\text{A}\!:\!0,\ \text{B}\!:\!1,\ \text{C}\!:\!k_\text{C}\}\right)$$

  This means that $F_\text{C}(1, k_\text{B}, k_\text{C}) = G_{\text{C},1}(k_\text{C})$ for a generic function $G_{\text{C},1} : \{0,1\} \to \{0,1\}$.

- The component $F_C(k)$ must have the same value for all $k \in \prod_{\omega \in E} I_\omega$ such that $\{A:0,\ B:k_B,\ C:k_C\} \leq k$, for each choice of $k_B, k_C \in \{0,1\}$. This doesn't impose any constraints, as the only such k is $k = \{A:0,\ B:k_B,\ C:k_C\}$ itself. This means that $F_C(0, k_B, k_C) = G_{B,0}(k_B, k_C)$ for a generic function $G_{C,0} : \{0,1\}^2 \to \{0,1\}$.

Putting all constraints above together, we get the following characterisation of a generic F which is causal for Θ_{101}, for generic functions $G_A, G_{B,0}, G_{B,1}, G_{C,0}, G_{C,1}$:

$$F(0, k_B, k_C) = \begin{pmatrix} G_A(0) \\ G_{B,0}(k_B) \\ G_{C,0}(k_B, k_C) \end{pmatrix} \qquad F(1, k_B, k_C) = \begin{pmatrix} G_A(1) \\ G_{B,1}(k_B, k_C) \\ G_{C,1}(k_C) \end{pmatrix}$$

Given one such joint IO function F, the causal function $f \in \mathrm{CausFun}(\Theta_{33}, \{0,1\})$ defined by Equation 4.4 takes the following form:

$$\begin{aligned} \{A:k_A\} &\overset{f}{\mapsto} G_A(k_A) \\ \{A:0,\ B:k_B\} &\overset{f}{\mapsto} G_{B,0}(k_B) \\ \{A:1,\ C:k_C\} &\overset{f}{\mapsto} G_{C,1}(k_C) \\ \{A:0,\ B:k_B,\ C:k_C\} &\overset{f}{\mapsto} G_{C,0}(k_B, k_C) \\ \{A:1,\ B:k_B,\ C:k_C\} &\overset{f}{\mapsto} G_{B,1}(k_B, k_C) \end{aligned}$$

The extended causal function $\mathrm{Ext}(f)$ then takes the following form:

$$\begin{aligned} \{A:k_A\} &\overset{\mathrm{Ext}(f)}{\mapsto} \{A:G_A(k_A)\} \\ \{A:0,\ B:k_B\} &\overset{\mathrm{Ext}(f)}{\mapsto} \big\{A:G_A(0),\ B:G_{B,0}(k_B)\big\} \\ \{A:1,\ C:k_C\} &\overset{\mathrm{Ext}(f)}{\mapsto} \big\{A:G_A(1),\ C:G_{C,1}(k_C)\big\} \\ \{A:0,\ B:k_B,\ C:k_C\} &\overset{\mathrm{Ext}(f)}{\mapsto} \big\{A:G_A(0),\ B:G_{B,0}(k_B),\ C:G_{C,0}(k_B, k_C)\big\} \\ \{A:1,\ B:k_B,\ C:k_C\} &\overset{\mathrm{Ext}(f)}{\mapsto} \big\{A:G_A(1),\ B:G_{B,1}(k_B, k_C),\ C:G_{C,1}(k_B)\big\} \end{aligned}$$

The last two lines of the definition of $\mathrm{Ext}(f)$ above are its restriction to the maximal extended input histories, which coincides with the original definition of F.

4.2 Causal functions for general tight spaces

For a tight and causally complete space, the assignment of causal functions is unproblematic. This is because each history defines a unique tip event (causal completion) and to every event and a maximal extended history we can univocally assign a history having it as tip event which is compatible with the extended history (tightness). So far we have understood global causal functions for tight and causally complete spaces, we are still far from the definition of a presheaf of contexts, but first we need to dive into causally incomplete and non-tight to check how to globally define causal data there. In this section we drop the causal completeness requirement

Spaces are not causally complete precisely when multiple events can act as tips of a single history. The general notion of *causal functions* has to take into account this multiplicity of the tips.

Definition 4.9 (Causal functions for tight spaces)**.** *Let* Θ *be a tight space and let* $\underline{O}$ *be a family of non-empty sets of outputs. The* causal functions $\mathit{CausFun}\left(\Theta, \underline{O}\right)$ *for space* Θ *and outputs* $\underline{O}$ *are the functions mapping each history in* Θ *to the output values for its tip events:*

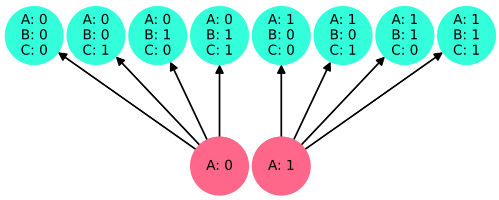
$$\mathit{CausFun}\left(\Theta, \underline{O}\right) := \prod_{h \in \Theta} \prod_{\omega \in \mathit{tips}_\Theta(h)} O_\omega$$

For causally complete spaces we have that $\text{tips}_\Theta(h) = \{\text{tip}_\Theta(h)\}$ and Definition 4.9 becomes equivalent to Definition 4.3 because

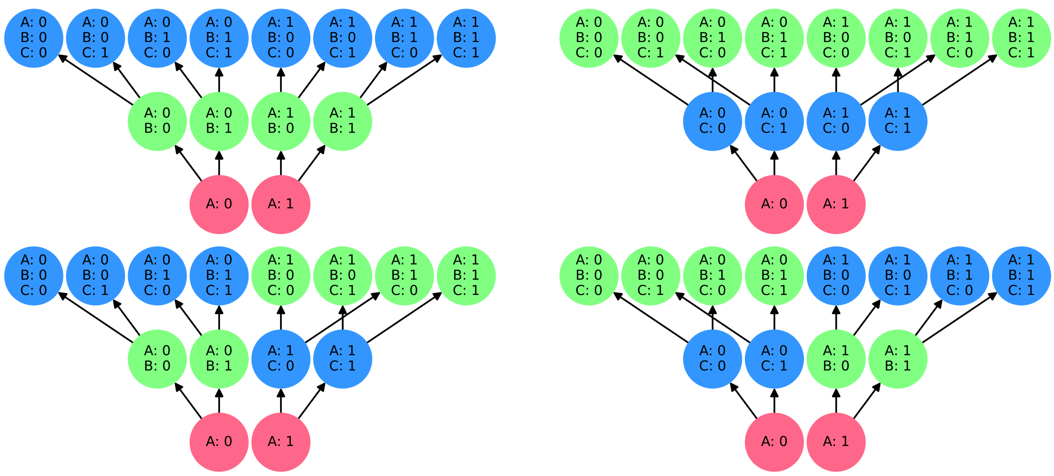
$$\prod_{\omega \text{tips}_\Theta(h)} O_\omega = \prod_{\omega \in \{\text{tip}_\Theta(h)\}} O_\omega \simeq O_{\text{tip}_\Theta(h)}$$

Remember that in the previous chapter we have defined a completion of a causally incomplete space to be a minimal causally complete space which induces at least as many causal constrained as the original space. The functions which are causal for some completion of Θ will be causal for Θ itself: it might be tempting to think of causal incompleteness as specifying a coarse-grained causal description to be made precise by the data assigned to causally complete subspaces (containing more causal restrictions). However, not all casual functions on a causally incomplete space arise in this way. There are some function which do not admit a fine-grained and causally complete explanation.

As a concrete example, consider the tight space $\Theta = \text{Hist}\left(\Omega, \{0, 1\}\right)$ of in input histories induced by the indefinite causal order $\Omega = \text{total}\left(\text{A}, \{\text{B}, \text{C}\}\right)$ with binary inputs. The space satisfies the free-choice condition, and the 8 maximal extended input histories have $\{\text{B}, \text{C}\}$ as their tip events.



Recall that there are four causal completions for this space: two where B and C are unconditionally totally ordered, and two where they are totally ordered conditionally to the input at A.



Definition 4.9 allows us to straightforwardly count the number of causal functions for the causally complete and causally incomplete spaces. The causally incomplete space Θ has the following number of causal functions:

$$\prod_{h\in\Theta} 2^{|\mathrm{tips}_\Theta(h)|} = 2^{\sum_{h\in\Theta}|\mathrm{tips}_\Theta(h)|} = 2^{2\cdot 1+8\cdot 2} = 2^{18} = 262144$$

Each of the four causal completions has the following number of causal functions:

$$\prod_{h\in\Theta} 2^{|\mathrm{tips}_\Theta(h)|} = 2^{\sum_{h\in\Theta} 1} = 2^{|\Theta|} = 2^{14} = 16384$$

However, some causal functions are common to multiple causal completions, so only 50176 of the 262144 causal functions on Θ arise from one of its completions as defined by the following definition:

Definition 4.10. *Let Θ, Θ' be spaces of input histories such that $\Theta' \leq \Theta$. Let $f \in \mathit{CausFun}\left(\Theta, \underline{O}\right)$ be a causal function on Θ and let $f' \in \mathit{CausFun}\left(\Theta', \underline{O'}\right)$ be a causal function on Θ'. We say that f* arises from *f' if the extended causal function $\mathit{Ext}\left(f'\right)$ restricts to the extended causal function $\mathit{Ext}\left(f\right)$:*

$$f \text{ arises from } f' \Leftrightarrow \mathit{Ext}\left(f'\right)|_{\mathit{Ext}(\Theta)} = \mathit{Ext}\left(f\right)$$

where we have used the fact that $\Theta' \leq \Theta$ is defined as $\mathit{Ext}\left(\Theta'\right) \supseteq \mathit{Ext}\left(\Theta\right)$.

Proposition 4.10. *Let $\Theta', \Theta \in \mathit{Spaces}_{FC}\left(\underline{I}\right)$ be such that $\Theta' \leq \Theta$; define $E := E^{\Theta} = E^{\Theta'}$. Let $\underline{O} = (O_\omega)_{\omega\in E}$ and $\underline{O'} = (O'_\omega)_{\omega\in E}$ be families of non-empty output sets such that $O'_\omega \subseteq O_\omega$ for all $\omega \in E$. Then the following is an injection:*

$$\begin{array}{rccc} i_{\Theta',\underline{O'};\Theta,\underline{O}}: & \mathit{CausFun}\left(\Theta', \underline{O'}\right) & \hookrightarrow & \mathit{CausFun}\left(\Theta, \underline{O}\right) \\ & f' & \mapsto & \mathit{Prime}\left(\mathit{Ext}\left(f'\right)|_{\mathit{Ext}(\Theta)}\right) \end{array}$$

We can use the injection above to identify the causal functions for Θ' and $\underline{O'}$ with a subset of the causal functions for Θ and $\underline{O}$. This is safe, because the injections are stable under composition:

$$i_{\Theta',\underline{O'};\Theta,\underline{O}} \circ i_{\Theta'',\underline{O''};\Theta',\underline{O'}} = i_{\Theta'',\underline{O''};\Theta,\underline{O}}$$

Proof. If $\Theta' \leq \Theta$ then $\mathrm{Ext}\left(\Theta\right) \subseteq \mathrm{Ext}\left(\Theta'\right)$. If $F(k \vee_{\Theta'} k') = F(k) \vee F(k')$ then F also satisfies compatibility in Θ and $F(k \vee_\Theta k') = F(k) \vee F(k')$. From this we conclude that if $f', g' \in \mathrm{CausFun}\left(\Theta, \underline{O}\right)$ then $\mathrm{Ext}\left(f'\right)|_{\mathrm{Ext}(\Theta)}$ and $\mathrm{Ext}\left(g'\right)|_{\mathrm{Ext}(\Theta)}$ are consistent extended functions. By Theorem 4.8 we get that

$$\begin{array}{rl} & \mathrm{Prime}\left(\mathrm{Ext}\left(f'\right)|_{\mathrm{Ext}(\Theta)}\right) = \mathrm{Prime}\left(\mathrm{Ext}\left(g'\right)|_{\mathrm{Ext}(\Theta)}\right) \\ \Rightarrow & \mathrm{Ext}\left(f'\right)|_{\mathrm{Ext}(\Theta)} = \mathrm{Ext}\left(g'\right)|_{\mathrm{Ext}(\Theta)} \end{array}$$

Because $\Theta', \Theta \in \mathrm{Spaces}_{\mathrm{FC}}\left(\underline{I}\right)$, then $\max \mathrm{Ext}\left(\Theta\right) = \max \mathrm{Ext}\left(\Theta'\right)$. In particular $\mathrm{Ext}\left(f'\right)(k) = \mathrm{Ext}\left(g'\right)(k)$ for all $k \in \prod_{\omega\in E^\Theta} I_\omega$. Since every $h \in \Theta'$ satisfies $h \leq k$ for some maximal extended input history k, the consistency condition in turn implies that $f'(h) = g'(h)$ for all $h \in \Theta'$, proving that $i_{\Theta',\underline{O}';\Theta,\underline{O}}$ is an injection. These injections are stable under composition:

$$\begin{aligned} & \text{Prime}\left(\text{Ext}\left(\text{Prime}\left(\text{Ext}\left(f''\right)|_{\text{Ext}(\Theta')}\right)\right)\Big|_{\text{Ext}(\Theta)}\right) \\ = \; & \text{Prime}\left(\left(\text{Ext}\left(f''\right)|_{\text{Ext}(\Theta')}\right)\Big|_{\text{Ext}(\Theta)}\right) \\ = \; & \text{Prime}\left(\text{Ext}\left(f''\right)|_{\text{Ext}(\Theta)}\right) \end{aligned}$$

□

For causally incomplete spaces of histories we differentiate between *causally separable* and *causally inseparable* functions accordingly:

Definition 4.11 (Causally separable functions)**.** *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty output sets. A causal function* $f \in \mathit{CausFun}\left(\Theta, \underline{O}\right)$ *is said to be* causally separable *if it arises from* $f' \in \mathit{CausFun}\left(\Theta', \underline{O}\right)$ *for some causally complete* $\Theta' \leq \Theta$*, and* causally inseparable *otherwise.*

The space $\Theta = \text{Hist}\left(\text{total}\left(\text{A}, \{\text{B}, \text{C}\}\right), \{0, 1\}\right)$ is causally incomplete and a simple example of a non-separable function is given by the following "controlled swap function":

$$\begin{aligned} \{\text{A}\!:\!k_\text{A}\} & \overset{\text{cswap}}{\mapsto} \{\text{A}\!:\!k_\text{A}\} \\ \{\text{A}\!:\!0,\ \text{B}\!:\!k_\text{B},\ \text{C}\!:\!k_\text{C}\} & \overset{\text{cswap}}{\mapsto} \{\text{B}\!:\!k_\text{B},\ \text{C}\!:\!k_\text{C}\} \\ \{\text{A}\!:\!1,\ \text{B}\!:\!k_\text{B},\ \text{C}\!:\!k_\text{C}\} & \overset{\text{cswap}}{\mapsto} \{\text{B}\!:\!k_\text{C},\ \text{C}\!:\!k_\text{B}\} \end{aligned}$$

The controlled swap function requires true bipartite signalling, where events B and C are delocalised even conditional to the input at A. Indeed, when the input at A is 1:

1. the output at B depends on the input at C, which must therefore be in B's past;

2. the output at C depends of the input at B, which must therefore be in C's past.

As a consequence, the controlled swap function cannot arise from any causal function f on any causally complete subspace $\Theta' \leq \Theta$ (i.e. we cannot have $\text{Ext}\left(f\right)|_{\text{Ext}(\Theta)} = \text{Ext}\left(\text{cswap}\right)$).

To see why, consider the extended input history $\{\text{A}\!:\!1,\ \text{B}\!:\!0,\ \text{C}\!:\!0\}$, with tips $\{\text{B}, \text{C}\}$. Any causal completion Θ' of our space Θ must necessarily include as an extended input history one of the following partial functions:

1. $\{\text{A}\!:\!1,\ \text{C}\!:\!0\}$, obtained by removing B from the domain of $\{\text{A}\!:\!1,\ \text{B}\!:\!0,\ \text{C}\!:\!0\}$

2. $\{\text{A}\!:\!1,\ \text{B}\!:\!0\}$, obtained by removing C from the domain of $\{\text{A}\!:\!1,\ \text{B}\!:\!0,\ \text{C}\!:\!0\}$

In either case, the causal function $\text{cswap} \in \text{CausFun}\left(\Theta, \{0, 1\}\right)$ cannot arise from any $f' \in \text{CausFun}\left(\Theta', \{0, 1\}\right)$, because $\text{Ext}\left(f'\right)$ cannot satisfy the consistency condition. If $\{\text{A}\!:\!1,\ \text{C}\!:\!0\} \in \text{Ext}\left(\Theta'\right)$, we are forced to make the following inconsistent assignments:

- from $\{A\!:\!1,\ C\!:\!0\} \leq \{A\!:\!1,\ B\!:\!0,\ C\!:\!0\}$, we must have:

$$\mathrm{Ext}\,(f')_{C}\,(\{A\!:\!1,\ C\!:\!0\}) = \mathrm{Ext}\,(f')_{C}\,(\{A\!:\!1,\ B\!:\!0,\ C\!:\!0\}) = 0$$

- from $\{A\!:\!1,\ C\!:\!0\} \leq \{A\!:\!1,\ B\!:\!1,\ C\!:\!0\}$, we must have:

$$\mathrm{Ext}\,(f')_{C}\,(\{A\!:\!1,\ C\!:\!0\}) = \mathrm{Ext}\,(f')_{C}\,(\{A\!:\!1,\ B\!:\!1,\ C\!:\!0\}) = 1$$

If instead $\{A\!:\!1,\ B\!:\!0\} \in \mathrm{Ext}\,(\Theta')$, we are forced to make the following inconsistent assignments:

- from $\{A\!:\!1,\ B\!:\!0\} \leq \{A\!:\!1,\ B\!:\!0,\ C\!:\!0\}$, we must have:

$$\mathrm{Ext}\,(f')_{B}\,(\{A\!:\!1,\ B\!:\!0\}) = \mathrm{Ext}\,(f')_{B}\,(\{A\!:\!1,\ B\!:\!0,\ C\!:\!0\}) = 0$$

- from $\{A\!:\!1,\ B\!:\!0\} \leq \{A\!:\!1,\ B\!:\!0,\ C\!:\!1\}$, we must have:

$$\mathrm{Ext}\,(f')_{B}\,(\{A\!:\!1,\ B\!:\!0\}) = \mathrm{Ext}\,(f')_{B}\,(\{A\!:\!1,\ B\!:\!0,\ C\!:\!1\}) = 1$$

The information above, proving that cswap is causally inseparable, can be summarised as follows. There is an extended input history $k = \{A\!:\!1,\ B\!:\!0,\ C\!:\!0\}$ such that for all events $\omega \in \mathrm{dom}\,(k)$ the function Ext (cswap) could not be extended to $k|_{\mathrm{dom}(k)\setminus\{\omega\}}$:

1. if $\omega = B$, $k|_{\mathrm{dom}(k)\setminus\{\omega\}} = \{A\!:\!1,\ C\!:\!0\}$
2. if $\omega = C$, $k|_{\mathrm{dom}(k)\setminus\{\omega\}} = \{A\!:\!1,\ B\!:\!0\}$

This is because for each ω there is an extended input history $k'_\omega \in \mathrm{Ext}\,(\Theta)$ with $k|_{\mathrm{dom}(k)\setminus\{\omega\}} \leq k'_\omega$ and an event $\xi_\omega \in \mathrm{dom}\,(k)\setminus\{\omega\}$ such that $\mathrm{Ext}\,(\mathrm{cswap})\,(k)_{\xi_\omega} \neq \mathrm{Ext}\,(\mathrm{cswap})\,(k'_\omega)_{\xi_\omega}$:

1. if $\omega = B$, we can choose $k'_\omega = \{A\!:\!1,\ B\!:\!1,\ C\!:\!0\}$ and $\xi_\omega = C$
2. if $\omega = C$, we can choose $k'_\omega = \{A\!:\!1,\ B\!:\!0,\ C\!:\!1\}$ and $\xi_\omega = B$

We refer to such a triple $\left(k, (k'_\omega)_\omega, (\xi_\omega)_\omega\right)$ as an 'inseparability witness', as it proves that the controlled swap cannot arise from any causal function f' on any causally complete subspace $\Theta' \leq \Theta$. We will now generalise and formalise our discussion thus far, and show that causally inseparable functions are exactly those with (at least) one such inseparability witness.

Definition 4.12. *An* inseparability witness *for a causal function* $f \in \mathit{CausFun}\left(\Theta, \underline{O}\right)$ *is a triple* $(k, (k'_\omega)_{\omega\in\mathrm{dom}(k)}, (\xi_\omega)_{\omega\in\mathrm{dom}(k)})$ *where:*

- $k \in \mathit{Ext}\,(\Theta)$ *is an extended input history;*
- *for every* $\omega \in \mathrm{dom}\,(k)$*,* $k'_\omega \in \mathit{Ext}\,(\Theta)$ *is such that* $k|_{\mathrm{dom}(k)\setminus\{\omega\}} \leq k'_\omega$*;*

- *for every* $\omega \in \mathrm{dom}(k)$, $\xi_\omega \in \mathrm{dom}(k) \setminus \{\omega\}$ *is such that* $\mathit{Ext}(f)(k)_{\xi_\omega} \neq \mathit{Ext}(f)(k'_\omega)_{\xi_\omega}$.

Lemma 4.13. *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty output sets. If* $f \in \mathit{CausFun}(\Theta, \underline{O})$ *is a causal function and* $(k, \underline{k'}, \underline{\xi})$ *is an inseparability witness for* f*, then* $|\mathrm{dom}(k)| \geq 2$ *and for all* $\omega \in \mathrm{dom}(k)$ *the partial function* $h_\omega := k|_{\mathrm{dom}(k)\setminus\{\omega\}}$ *is not an extended input history for* Θ.

Proof. We necessarily have $|\mathrm{dom}(k)| \geq 1$, and for any $\omega \in \mathrm{dom}(k)$ we also have $\xi_\omega \in \mathrm{dom}(k) \setminus \{\omega\}$, so necessarily $|\mathrm{dom}(k)| \geq 2$. The partial function h_ω cannot be a an extended input history for Θ, because otherwise the consistency condition on $\mathrm{Ext}(f)$ would imply the following, contradicting the definition of ξ_ω:

$$\begin{aligned} \mathrm{Ext}(f)_{\xi_\omega}(h_\omega) &= \mathrm{Ext}(f)_{\xi_\omega}(k) \\ \mathrm{Ext}(f)_{\xi_\omega}(h_\omega) &= \mathrm{Ext}(f)_{\xi_\omega}(k'_\omega) \end{aligned}$$

□

Lemma 4.14. *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty output sets. Let* $f \in \mathit{CausFun}(\Theta, \underline{O})$ *be a causal function and* $(k, \underline{k'}, \underline{\xi})$ *be an inseparability witness for* f*. If* f *arises from* $f' \in \mathit{CausFun}(\Theta', \underline{O'})$ *for some* $\Theta' \leq \Theta$*, then* $(k, \underline{k'}, \underline{\xi})$ *is an inseparability witness for* f'.

Proof. Be definition, $\Theta' \leq \Theta$ is the same as $\mathrm{Ext}(\Theta') \supseteq \mathrm{Ext}(\Theta)$, so that k and all k'_ω are extended histories for Θ'. Also by definition, f arising from f' is the same as $\mathrm{Ext}(f')|_{\mathrm{Ext}(\Theta)} = \mathrm{Ext}(f)$, so that:

$$\mathrm{Ext}(f')_{\xi_\omega}(k) = \mathrm{Ext}(f)_{\xi_\omega}(k) \neq \mathrm{Ext}(f)_{\xi_\omega}(k'_\omega) = \mathrm{Ext}(f')_{\xi_\omega}(k'_\omega)$$

Hence $(k, \underline{k'}, \underline{\xi})$ is an inseparability witness for f'. □

Theorem 4.15. *A causal function* $f \in \mathit{CausFun}(\Theta, \underline{O})$ *is causally inseparable if and only if it has an inseparability witness.*

Proof. In one direction, let $(k, \underline{k'}, \underline{\xi})$ be an inseparability witness for f. If $\Theta' \leq \Theta$ is causally complete, then there exists an $\omega_k \in \mathrm{dom}(k)$ such that $h_{\omega_k} := k|_{\mathrm{dom}(k)\setminus\{\omega_k\}} \in \mathrm{Ext}(\Theta')$, because Lemma 4.13 forces $|\mathrm{dom}(k)| \geq 2$. Then f cannot arise from some $f' \in \mathrm{CausFun}(\Theta', \underline{O})$: if it did, Lemma 4.14 would imply that $(k, \underline{k'}, \underline{\xi})$ is an inseparability witness for f', and Lemma 4.13 would in turn force $h_{\omega_k} \notin \mathrm{Ext}(\Theta')$.

In the other direction, assume that $f \in \mathrm{CausFun}(\Theta, \underline{O})$ does not have an inseparability witness. For each $k \in \mathrm{Ext}(\Theta)$ and each $\omega \in \mathrm{dom}(k)$, we define $h_{k,\omega} := k|_{\mathrm{dom}(k)\setminus\{\omega\}}$. The absence of an inseparability witness for f means that for all $k \in \mathrm{Ext}(\Theta)$ there is a $\omega_k \in \mathrm{dom}(k)$ such that for all $k' \in \text{Ext}(\Theta)$ with $h_{k,\omega_k} \leq k'$ and all $\xi \in \text{dom}(k) \setminus \{\omega\}$ we have $\text{Ext}(f)_\xi(k) = \text{Ext}(f)_\xi(k')$. This allows us to consistently extend the definition of $\text{Ext}(f)$ to h_{k,ω_k} for all k:

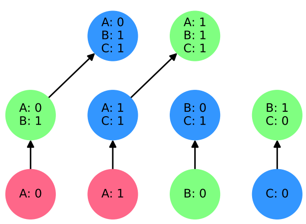
$$\text{Ext}(f')(h_{k,\omega_k}) := \bigwedge \left\{ \text{Ext}(f)(k') \,\middle|\, k' \in \text{Ext}(\Theta), h_{k,\omega_k} \leq k' \right\}$$

The $\vee$-closure of $\Theta \cup \left\{ h_{k,\omega_k} \,\middle|\, k \in \text{Ext}(\Theta) \right\}$ is the space of extended input histories $\text{Ext}(\Theta')$ for a causally complete sub-space $\Theta' \leq \Theta$. Furthermore, the definition of $\text{Ext}(f')(h_{k,\omega_k})$ above can be completed—by gluing over compatible joins—to that of an extended causal function $f' \in \text{CausFun}(\Theta', \underline{O})$ such that f arises from f', proving that f is causally separable. □

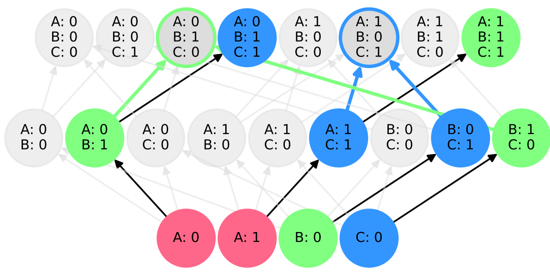
4.3 Causal functions for non-tight spaces

Having investigated the effects of causal incompleteness in the definition of causal functions, we now proceed to relax the tightness assumption: for each extended input history $k \in \text{Ext}(\Theta)$ and $\omega \in \text{dom}(k)$, we are no longer guaranteed that the input history $h \leq k$ with $\omega \in \text{tips}_\Theta(h)$ will be unique. This imposes additional constraints to the definition of causal functions $f \in \text{CausFun}(\Theta)$: if $h, h' \leq k$ are distinct input histories such that $\omega \in \text{tips}_\Theta(h) \cap \text{tips}_\Theta(h')$, then the outputs $f(h)_\omega$ and $f(h')_\omega$ at event ω must coincide.

As a concrete example, consider the following causally complete non-tight space:



Θ_{21}



$\text{Ext}(\Theta_{21})$ with highlights

There are two violations of tightness in this space:

- The extended input history $\{A{:}0,\ B{:}1,\ C{:}0\}$ has two sub-histories with event B as their tip event: $\{A{:}0,\ B{:}1\}$ and $\{B{:}1,\ C{:}0\}$.
- The extended input history $\{A{:}1,\ B{:}0,\ C{:}1\}$ has two sub-histories with event C as their tip event: $\{A{:}1,\ C{:}1\}$ and $\{B{:}0,\ C{:}1\}$.

The two histories are highlighted in $\text{Ext}(\Theta_{21})$ above with a border of the same colour as the corresponding tip event conflicts. For $\text{Ext}(f)_B$ to be well-defined on the (maximal) extended input history $\{A{:}0,\ B{:}1,\ C{:}0\}$, the outputs associated by a causal function f to the two input histories $\{A{:}0,\ B{:}1\}$ and $\{B{:}1,\ C{:}0\}$ must coincide:

$$\begin{aligned} \text{Ext}(f)_B(\{A{:}0,\ B{:}1,\ C{:}0\}) &= f(\{A{:}0,\ B{:}1\}) \\ &= f(\{B{:}1,\ C{:}0\}) \end{aligned}$$

Analogously, for $\mathrm{Ext}(f)_{\mathrm{C}}$ to be well-defined on the (maximal) extended input history $\{\mathrm{A}:1,\ \mathrm{B}:0,\ \mathrm{C}:1\}$, the outputs associated by a causal function f to the two input histories $\{\mathrm{A}:1,\ \mathrm{C}:1\}$ and $\{\mathrm{B}:0,\ \mathrm{C}:1\}$ must coincide:

$$\begin{aligned} \mathrm{Ext}(f)_{\mathrm{B}}(\{\mathrm{A}:1,\ \mathrm{B}:0,\ \mathrm{C}:1\}) &= f(\{\mathrm{A}:1,\ \mathrm{C}:1\}) \\ &= f(\{\mathrm{B}:0,\ \mathrm{C}:1\}) \end{aligned}$$

At first sight, such constrains make it seem like the definition of causal functions is no longer 'free', but this is not actually the case: instead of a constrained mapping of individual input histories to outputs at their tip event(s), we can think of a causal function on a non-tight space as a free mapping of equivalence classes of input histories to outputs on a common tip event. In the case of space Θ_{21} above, we have 10 input histories arranged into 8 pairs of an equivalence class and a common tip event for the histories therein:

1. the singleton $\{\{\mathrm{A}:1,\ \mathrm{B}:1,\ \mathrm{C}:1\}\}$ with tip event B
2. the singleton $\{\{\mathrm{A}:0,\ \mathrm{B}:1,\ \mathrm{C}:1\}\}$ with tip event C
3. the pair $\{\{\mathrm{A}:0,\ \mathrm{B}:1\}, \{\mathrm{B}:1,\ \mathrm{C}:0\}\}$ with common tip event B
4. the pair $\{\{\mathrm{A}:1,\ \mathrm{C}:1\}, \{\mathrm{B}:0,\ \mathrm{C}:1\}\}$ with common tip event C
5. the singleton $\{\{\mathrm{A}:0\}\}$ with tip event A
6. the singleton $\{\{\mathrm{A}:1\}\}$ with tip event A
7. the singleton $\{\{\mathrm{B}:0\}\}$ with tip event B
8. the singleton $\{\{\mathrm{C}:0\}\}$ with tip event C

Causal functions on Θ_{21} are then given by a free choice of output for each equivalence class: for example, the binary case $\mathrm{CausFun}(\Theta_{21}, \{0,1\})$ features $2^8 = 256$ causal functions.

We now formalise the discussion thus far into a definition of causal functions valid for arbitrary spaces of input histories, and generalise the results of Propositions 4.8 and Theorem 4.8. We start by defining the machinery necessary to formulate the constraints associated with lack of tightness, provide a constrained definition on input histories, and prove that it is equivalent to a free definition.

Definition 4.16. *Let Θ be a space of input histories. For each $\omega \in E^{\Theta}$, the* tip histories *for ω are the input histories which have ω as a tip event:*

$$TipHists_{\Theta}(\omega) := \left\{ h \in \Theta \,\middle|\, \omega \in tips_{\Theta}(h) \right\} \tag{4.5}$$

Definition 4.17. *Let Θ be a space of input histories. For any $\omega \in E^{\Theta}$, we say that two histories $h, h' \in \Theta$ are* constrained at ω*, written $h \sim_{\omega} h'$, if they both have ω as a tip event and the consistency condition from Definition 4.7 forces all extended functions $\hat{F} \in ExtFun\left(\Theta, \{0,1\}\right)$ to output the same value for ω on h and h':*

$$\hat{F}(h)_{\omega} = \hat{F}(h')_{\omega}$$

Remark 4.18. *In Definition 4.17 above, we could have replaced $\{0,1\}$ with any family of non-empty output sets $\underline{O}$ such that $|O_{\omega}| \geq 2$. Equivalently, we could have universally quantified over all $\hat{F} \in ExtFun\left(\Theta, \underline{O}\right)$ for all $\underline{O}$, but this would have made it unnecessarily harder to apply the definition.*

Corollary 4.19. *Let Θ be a space of input histories. Θ is tight if and only if $h \sim_{\omega} h'$ always implies $h = h'$.*

Proof. For each $k \in \mathrm{Ext}\left(\Theta\right)$ and each $\omega \in \mathrm{dom}\left(k\right)$, the tightness requirement is exactly that there is a unique $h \in \mathrm{TipHists}_{\Theta}\left(\omega\right)$ with $h \leq k$. □

Definition 4.20. *Let Θ be a space of input histories and let $\underline{O} = (O_{\omega})_{\omega \in E^{\Theta}}$ be a family of non-empty sets of outputs. The* causal functions *$CausFun\left(\Theta, \underline{O}\right)$ for space Θ and outputs $\underline{O}$ are the functions mapping each history in Θ to the output values for its tip events, subject to the the additional requirement that $f(h)_{\omega} = f(h')_{\omega}$ for any input histories h, h' which are constrained at an event ω:*

$$CausFun\left(\Theta, \underline{O}\right) := \left\{ f \in \prod_{h \in \Theta} \prod_{\omega \in tips_{\Theta}(h)} O_{\omega} \;\middle|\; h \sim_{\omega} h' \Rightarrow f(h)_{\omega} = f(h')_{\omega} \right\} \tag{4.6}$$

Observation 4.20. *Let Θ be a space of input histories and let $\underline{O} = (O_{\omega})_{\omega \in E^{\Theta}}$ be a family of non-empty sets of outputs. For any $\omega \in E^{\Theta}$, let $TipEq_{\Theta}\left(\omega\right)$ be the set of equivalence classes for $\sim_{\omega}$:*

$$TipEq_{\Theta}\left(\omega\right) := TipHists_{\Theta}\left(\omega\right) / \sim_{\omega} = \left\{ [h]_{\sim_{\omega}} \;\middle|\; h \in TipHists_{\Theta}\left(\omega\right) \right\} \tag{4.7}$$

There is a bijection between the causal functions in $CausFun\left(\Theta, \underline{O}\right)$ and the functions freely mapping each event $\omega \in E^{\Theta}$ and each equivalence class $[h]_{\sim_{\omega}} \in TipEq_{\Theta}\left(\omega\right)$ to the common output value at ω for all input histories in $[h]_{\sim_{\omega}}$:

$$\begin{array}{rcl} CausFun\left(\Theta, \underline{O}\right) & \longleftrightarrow & \prod_{\omega \in E^{\Theta}} \left(O_{\omega}\right)^{TipEq_{\Theta}(\omega)} \\ f & \mapsto & \left((\omega, [h]_{\sim_{\omega}}) \mapsto f(h)_{\omega}\right) \\ \left(h \mapsto \left(g\left(\omega, [h]_{\sim_{\omega}}\right)\right)_{\omega \in tips_{\Theta}(h)}\right) & \leftarrow\!\shortmid & g \end{array} \tag{4.8}$$

With Definition 4.20 in hand, we are finally in a position to fully generalise Definition 4.5, Theorem 4.8 and Proposition 4.8. The changes necessary to achieve this are small: in causally incomplete spaces, we must account for the possibility that causal functions will produce output values for multiple tip events, for non-tight spaces we must account for the non-uniqueness of the $h \in \mathrm{TipHists}_{\Theta}\left(\omega\right)$ such that $h \leq k$, used by the definition of extended causal functions. Both changes cause no trouble to our original proofs, which go through essentially unchanged.

Definition 4.21. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. For each causal function $f \in CausFun(\Theta, \underline{O})$, define the corresponding* extended causal function *$Ext(f) \in ExtFun(\Theta, \underline{O})$ as follows:*

$$Ext(f)(k) := \left(f\left(h_{k,\omega}\right)_\omega\right)_{\omega \in \mathrm{dom}(k)} \text{ for all } k \in Ext(\Theta) \tag{4.9}$$

where $h_{k,\omega}$ is any input history in $h \in TipHists_\Theta(\omega)$ such that $h \leq k$. We refer to $Ext(f)(k)$ as the extended output history *corresponding to extended input history k. We write $ExtCausFun(\Theta, \underline{O})$ for the subset of $ExtFun(\Theta, \underline{O})$ consisting of the extended causal functions.*

Observation 4.21. *The function $Ext(f)$ is well-defined because the definition of the causal function f implies that $f\left(h_{k,\omega}\right)_\omega$ is the same for any choice of $h_{k,\omega}$.*

Theorem 4.22. *Let Θ be a space of input histories, and let $\underline{O}$ be a family of non-empty sets of outputs. The extended function $\hat{F} \in ExtFun(\Theta, \underline{O})$ which are causal are exactly those which satisfy the consistency condition. Indeed for every consistent $\hat{F}$ we can find a unique $Prime(\hat{F}) \in CausFun(\Theta, \underline{O})$ such that $Ext(Prime(\hat{F})) = \hat{F}$:*

$$Prime\left(\hat{F}\right) := h \mapsto (\hat{F}(h)_\omega)_{\omega \in tips_\Theta(h)} \tag{4.10}$$

Proof. Let $f \in \mathrm{CausFun}(\Theta, \underline{O})$ we want to show that $\mathrm{Ext}(f)$ satisfies the consistency condition. Let $k, k' \in \mathrm{Ext}(\Theta)$ such that $k \leq k'$. By definition of $\mathrm{Ext}(f)$ we have that

$$\mathrm{Ext}(f)(k) := (f(h_{k,\omega})_\omega)_{\omega \in \mathrm{dom}(k)}$$

where $h_{k,\omega}$ is any input history $h \in \mathrm{TipHists}_\Theta(\omega)$ such that $h \leq k$. Similarly:

$$\mathrm{Ext}(f)(k') := (f(h'_{k',\omega})_\omega)_{\omega \in \mathrm{dom}(k')}$$

Let $\omega \in \mathrm{dom}(\mathrm{Ext}(f)(k)) \cap \mathrm{dom}(\mathrm{Ext}(f)(k'))$. Since $h_{k,\omega}, h_{k',\omega} \leq k'$, it is not any more true that $h_{k,\omega} = h_{k',\omega}$ however we know that $h_{k,\omega} \sim_\omega h_{k',\omega}$ and therefore

$$\mathrm{Ext}(f)(k)_\omega = \mathrm{Ext}(f)(k')_\omega$$

proving that $\mathrm{Ext}(f)(k') \leq \mathrm{Ext}(f)(k)$. This holds for every extended function proving their consistency.

Consider now an extended causal function $\hat{F}$ which satisfies the consistency condition. From $\hat{F}$ we construct a $\mathrm{Prime}(\hat{F}) \in \mathrm{CausFun}(\Theta, \underline{O})$ given by:

$$\mathrm{Prime}\left(\hat{F}\right) := h \mapsto (\hat{F}(h)_\omega)_{\omega \in \mathrm{tips}_\Theta(h)}$$

We see that Prime $\left(\hat{F}\right)$ is causal by simply being a function mapping each event $\omega \in E^{\Theta}$ and each equivalence class $[h]_{\sim_\omega} \in \mathrm{TipEq}_{\Theta}(\omega)$ to the common output value at ω by taking

$$\mathrm{Prime}\left(\hat{F}\right) \mapsto \left((\omega, [h]_{\sim_\omega}) \mapsto \mathrm{Prime}\left(\hat{F}\right)(h)_\omega\right)$$

which is well defined because $\hat{F}$ is an extended causal function.

It remains to show that $\mathrm{Ext}\left(\mathrm{Prime}\left(\hat{F}\right)\right) = \hat{F}$. Note that for any $k \in \mathrm{Ext}(\Theta)$ and $\omega \in \mathrm{dom}(k)$ we have $h_{k,\omega} \leq k$ for some $h_{k,\omega} \in \mathrm{TipHists}_{\Theta}(\omega)$ such that $h \leq k$, so by consistency $\hat{F}(h_{k,\omega}) \leq \hat{F}(k)$. In particular $\hat{F}(h_{k,\omega})_\omega = \hat{F}(k)_\omega$ for $\omega \in \mathrm{dom}\left(h_{k,\omega}\right) \cap \mathrm{dom}(k)$. By the definition of the operators Ext and Prime respectively acting on causal functions and extended causal functions, we have that

$$\hat{F}(h_{k,\omega})_\omega = \mathrm{Prime}\left(\hat{F}\right)(h_{k,\omega}) = \mathrm{Ext}\left(\mathrm{Prime}\left(\hat{F}\right)\right)(k)_\omega$$

concluding that $\hat{F}(k)_\omega = \mathrm{Ext}\left(\mathrm{Prime}\left(\hat{F}\right)\right)(k)_\omega$ and therefore $\hat{F}(k) = \mathrm{Ext}\left(\mathrm{Prime}\left(\hat{F}\right)\right)(k)$ for all $k \in \mathrm{Ext}(\Theta)$. □

We can finally express extended causal functions as 'gluing' of the output values of causal functions, via compatible joins.

Proposition 4.22. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. For every $f \in CausFun\left(\Theta, \underline{O}\right)$, we have:*

$$Ext(f) = k \mapsto \bigvee_{h \in k\downarrow \cap \Theta} f(h) \tag{4.11}$$

Proof. By Theorem 4.22, the extended functions which are causal are the one satisfying the consistency condition. By Proposition 4.7 such causal function satisfy the gluing condition. Let $k, k' \in \mathrm{ExtFun}(\Theta)$ be consistent histories, since $\mathrm{Ext}(f)$ is causal we have that $\mathrm{Ext}(f)(k \vee k') = \mathrm{Ext}(f)(k) \vee \mathrm{Ext}(f)(k')$. Writing an extended history as the joint of compatible histories yields the proposition. □

Proposition 4.22. *Let Θ be a space of input histories satisfying the free-choice condition and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. For every $f \in CausFun\left(\Theta, \underline{O}\right)$, the restriction of $Ext(f)$ to the maximal extended input histories is a joint IO function for the operational scenario $(E^\Theta, \underline{I^\Theta}, \underline{O})$ which is causal for Θ. Conversely, any joint IO function F which is causal for Θ arises as the restriction of $Ext(f)$ to maximal extended input histories, where $f \in CausFun\left(\Theta, \underline{O}\right)$ can be defined as follows for all $\omega \in tips_\Theta(h)$:*

$$f(h)_\omega := F(k)_\omega \text{ for any maximal ext. input history } k \text{ s.t. } h \leq k \tag{4.12}$$

Proof. The proof is essentially the same as for Proposition 4.8: we only need to check that nothing goes wrong when dropping causal completeness and tightness. The proof for the first part of the statement only makes use of the free-choice condition, which we have now explicitly required, so it goes through unaltered. The second part of the statement was modified to account for the more general definition of causal functions on causally incomplete spaces, potentially involving outputs at multiple tip events. Aside from this modification, and the explicit assumption of the free-choice condition, the proof for the second part of the statement also goes through unaltered. □

4.4 The presheaf of (extended) causal functions

Deterministic causal data are joint input and output functions satisfying causal constraints. Global causal IO functions satisfy all causal constraints associated with a given space. As such, they can be suitable for a classical description, where the outcome of the measurement is assigned non-contextually but we want to relax this assumption and use sheaf theory to describe empirical models which can arise from assigning distribution of local causal data defined on certain contexts but transcending global compatibility.

From the perspective of sheaf theory, data is associated with the open set of some topological space with an appropriate definition of *restriction* which is used to define the notion of compatibility. The reader should recall that in the standard sheaf theoretic approach (that we described in Chapter 1) the possible global assignments are identified with particular values of some hidden variable. The essence of non-locality or contextuality is to be found in the incompatibility between the global hidden mechanisms and the assignments realised in each individual context. We will treat global sections similarly, with the crucial difference that we are now considering data respecting additional causality requirements.

For the standard case of contextuality and non-locality, which as we will show later on is fully recovered by our framework, the assignment of data is a sheaf, i.e locally assigned deterministic functions can be always 'glued' together in a unique way. In our case, this will still hold for the tight spaces but fails due to the lack of tightness revealing a phenomenon known as deterministic casually-induced contextuality. Post-composing the assignment of deterministic causal data with the probability monad will in general define a presheaf permitting the emergence of general contextual phenomena.

For the machinery of sheaf theory to become available to us, we must first endow our spaces of input histories with a suitable topology. Because a space of input histories Θ is a partial order, a natural choice of topology is given by taking its lower sets $\Lambda(\Theta)$ to be the open sets.

This is the dual of the Alexandrov topology, where the upper sets are taken to be the open sets, and all techniques applicable to Alexandrov topologies naturally dualise to lowerset topologies. In particular, we make the following standard observations:

- The points of Θ, i.e. the input histories $h \in \Theta$, can be identified with downsets $h\downarrow \in \Lambda(\Theta)$.
- The downsets of input histories are exactly the lowersets $U \in \Lambda(\Theta)$ which are $\cup$-prime, i.e. those which cannot be written as $U = V \cup W$ for some lowersets $V, W \neq U$.
- The order on Θ can be reconstructed from the inclusion order on $\Lambda(\Theta)$, by observing that $h \leq h'$ if and only if $h\downarrow \subseteq h'\downarrow$.

The extended input histories $k \in \mathrm{Ext}(\Theta)$ can themselves be identified with certain lowersets, namely with the intersection $k\downarrow \cap\, \Theta$ of their downset in $\mathrm{Ext}(\Theta)$ with the space Θ. This identification is both injective and order-preserving, generalising the previous identification of $h \in \Theta$ with $h\downarrow$:

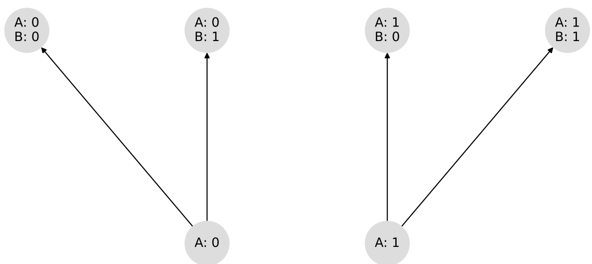
$$\begin{array}{rcl}(\mathrm{Ext}(\Theta), \leq) & \hookrightarrow & (\Lambda(\Theta), \subseteq) \\ k & \mapsto & k\downarrow \cap\, \Theta\end{array}$$

The locale of lowersets $\Lambda(\Theta)$ provides an equivalent way to talk about input histories, extended input histories and their order. We adopt lowersets as our default topology for spaces of (extended) input histories as to any other poset inheriting the order from partial functions, and proceed to show that causality for extended functions is the same as continuity.

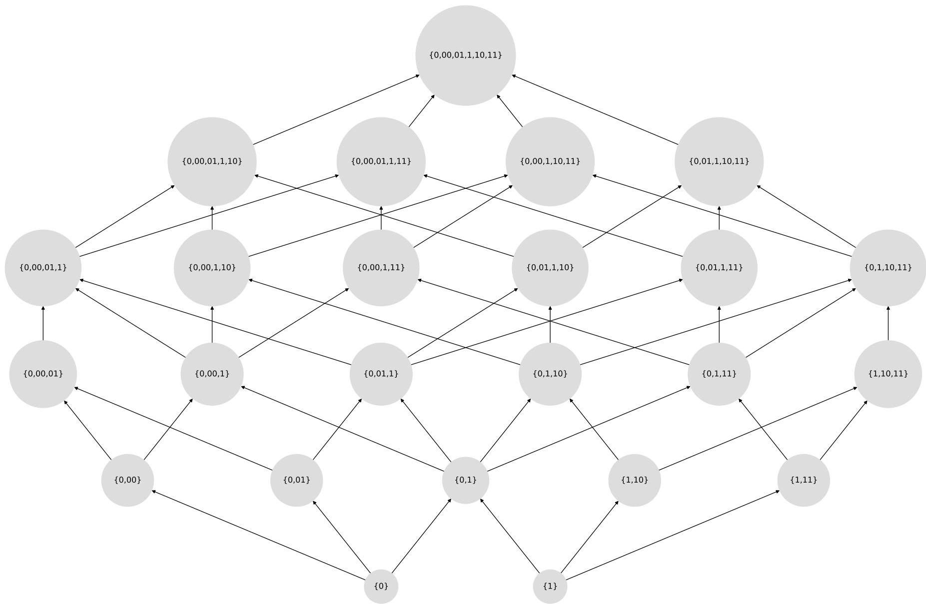
Definition 4.23. *When talking about spaces of (extended) input histories as topological spaces, we take them endowed with the lowerset topology. Unless otherwise specified, when talking about subsets $S \subseteq PFun\left(\underline{Y}\right)$ we take them endowed with the partial order of $PFun\left(\underline{Y}\right)$ and the lowerset topology.*

In the following example, we showcase a poset of open sets arising from an order induced space of input histories.

Example 4.24. *Consider the space $Hist\,(total\,(A, B)\,, \{0, 1\})$. Recall that the space of histories is therefore given by*



the lattice of lowersets is the following (not including the minimum, the empty set of histories):



In the Hasse diagram above each vertex identifies a context, a downward closed set of histories. The labels identify, in a more succinct way, the histories in the lowerset. For example, the symbol $\{0, 1\}$ *denotes the histories* $\{A\!:\!0\}$ *and* $\{A\!:\!1\}$ *, while labels containing* $\{ab\}$ *for* $a, b \in \{0, 1\}$ *denote the two events histories* $\{A\!:\!a,\ B\!:\!b\}$ *. The label* $\{0, 00\}$ *represent the context where both inputs for A and B are chosen to be* 0 *and contains the histories* $\{A\!:\!0\}$ *and* $\{A\!:\!0,\ B\!:\!0\}$ *. The open set* $\{0, 1\}$*, on the other hand, describes the lowerset where the B is ignored but where A still has the choice of performing either* 0 *or* 1*. A more exotic context is given by* $\{0, 01, 1, 10\}$*. In this case, we are interested in the cases in which the choices of A and B are anti-correlated in their maximal histories. A set like* $\{1, 00\}$ *is not a valid context as the histories* $\{A\!:\!1\}$ *and* $\{A\!:\!0,\ B\!:\!0\}$ *are not consistent and do not form a lowerset of the set of histories.*

We start by proving an elementary property about lowerset topologies:

Proposition 4.24. *Let* X, Y *be partial orders endowed with the lowerset topology. A function* $f : X \to Y$ *is continuous if and only if it is order preserving.*

Proof. Let $f : X \to Y$ be order preserving and take $\lambda \in \Lambda(Y)$. We need to show that $f^{-1}(\lambda)$ is a lowerset. Let $x \in f^{-1}(\lambda)$ and $x' \leq x$ then by order preservation $f(x') \leq f(x)$, λ is a lowerset and therefore $f(x') \in \lambda$.

For the other direction assume that $f : X \to Y$ is continuous. Let $x, x' \in X$ such that $x \leq x'$. Consider $f(x')\!\downarrow \in \Lambda(Y)$. Since f is continuous we know that $f^{-1}f(x')\!\downarrow$ is a lowerset of X, this allows us to conclude that $f(x) \leq f(x')$. □

Proposition 4.24 can be used to show that continuity with respect to the lowerset topology provides an alternative, mathematically elegant characterisation for extended causal functions.

Observation 4.24. *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty output sets. The extended functions on* Θ *are a subset of the functions from* $Ext(\Theta)$ *to* $PFun(\underline{O})$*:*

$$ExtFun(\Theta, \underline{O}) = \prod_{k \in Ext(\Theta)} \prod_{\omega \in \mathrm{dom}(k)} O_\omega \quad \subseteq \quad Ext(\Theta) \to PFun(\underline{O})$$

In fact, they are exactly the functions $\hat{F} : Ext(\Theta) \to PFun(\underline{O})$ *which commute with the domain map, in the following sense:*

$$\mathrm{dom}\left(\hat{F}(k)\right) = \mathrm{dom}(k)$$

Theorem 4.25 (Causality as continuity)**.** *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty output sets. An extended function* $\hat{F} \in ExtFun(\Theta, \underline{O})$ *is causal if and only if it is continuous as a function* $\hat{F} : Ext(\Theta) \to PFun(\underline{O})$ *where both* $Ext(\Theta)$ *and* $PFun(\underline{O})$ *are equipped with the lowerset topology.*

Proof. Causality of $\hat{F}$ is equivalent to the consistency condition by Theorem 4.8. The consistency condition implies that if $k' \leq k$ then $\hat{F}(k') \leq \hat{F}(k)$ as partial functions. $\hat{F}$ is order preserving and by Proposition 4.24 continuous. □

We now proceed to define the presheaf of causal data. The first question that we need to investigate is whether an arbitrary open set of the topological space $\Xi \in \Lambda(\Theta)$ is itself a space of input histories for the scenario with events E^Ξ. This stability would guarantee that the assignments of causal functions described above naturally extend to all open sets of the topology. The following propositions show that the open sets are valid spaces of histories and that they inherit tightness and tip events from the parent spaces.

Proposition 4.25. *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty sets of outputs. Any lowerset* $\lambda \in \Lambda(\Theta)$ *is a space of input histories, with* $Ext(\lambda) \in \Lambda(Ext(\Theta))$.

Proof. Taking lowerset does not introduce new histories, from this we conclude that every history in λ is prime. We have that $\mathrm{Ext}(\lambda) \subseteq \mathrm{Ext}(\Theta)$ by definition of Ext, and $\mathrm{Ext}(\lambda)$ is a lowerset because λ is a lowerset. □

Recall that for a non-tight space there may be two different histories $h, h' \in \Theta$ and $k \in \max \mathrm{ExtHist}(\Theta)$ such that $h, h' \leq k$ and $h \sim_\omega h'$. For example consider the non-tight space Θ_{21}, we see that the histories $\{A:0, B:1\} \sim_B \{B:1, C:0\}$ and $\{A:0, B:1\}, \{B:1, C:0\} \leq \{A:0, B:1, C:0\}$.

Proposition 4.25. *Let* Θ *be a space of input histories and let* $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ *be a family of non-empty sets of outputs. Let* $\lambda \in \Lambda(\Theta)$ *be a lowerset:*

- *For all $h \in \lambda$, we have $tips_\lambda(h) = tips_\Theta(h)$.*
- *For all $h, h' \in \lambda$, we have that $h \sim_\omega h'$ in λ implies $h \sim_\omega h'$ in Θ. Hence, if Θ is tight then so is λ.*

Proof. Because λ is a lowerset, all sub-histories in Θ of a history $h \in \lambda$ are also sub-histories of h in λ. This means that the tip events of h in Θ are the same as the tip events of h in λ, and in particular that λ is causally complete whenever Θ is. In turn, this means that any $h \in \lambda$ such that $h \leq k$ and $\omega \in \mathrm{tips}_\lambda(h)$ is also a $h \in \Theta$ such that $h \leq k$ and $\omega \in \mathrm{tips}_\Theta(h)$, for every $k \in \mathrm{Ext}(\lambda) \subseteq \mathrm{Ext}(\Theta)$. This immediately implies that λ is tight whenever Θ is, and it also implies that $h, h' \in \lambda$ are constrained at ω in λ only if they are in Θ. □

Taken together, the preservation of tips and constraints for input histories immediately implies that causal functions restrict to causal functions, motivating a very straightforward definition of the presheaf of causal functions.

Corollary 4.26. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. Let $f \in CausFun(\Theta, \underline{O})$ be a causal function for Θ, and $\lambda \in \Lambda(\Theta)$ be a lowerset. Then $f|_\lambda \in CausFun(\lambda, \underline{O})$.*

Proof. From Proposition 4.25 we know that $\mathrm{tips}_\lambda(h) = \mathrm{tips}_\Theta(h)$ for every $\lambda \in \Lambda(\Theta)$. Therefore $f|_\lambda$ is of the type:

$$f|_\lambda \in \prod_{h \in \lambda} \prod_{\omega \in \mathrm{tips}_\lambda(h)} O_\omega$$

Proposition 4.25 also states that $h \sim_\omega h'$ in λ implies $h \sim_\omega h'$ in Ω, which in turn implies that $f(h)_\omega = f(h')_\omega$, because $f \in \mathrm{CausFun}(\Theta, \underline{O})$. Hence we have $f|_\lambda \in \mathrm{CausFun}(\lambda, \underline{O})$. □

Definition 4.27. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. The* presheaf of causal functions *$CausFun(\Lambda(\Theta), \underline{O})$ for space Θ and outputs $\underline{O}$ is the presheaf on the topological space Θ defined as follows:*

- *to lowersets $\lambda \in \Lambda(\Theta)$, the open sets of Θ, it associates the causal functions for λ:*

$$CausFun(\Lambda(\Theta), \underline{O})(\lambda) := CausFun(\lambda, \underline{O}) \tag{4.13}$$

- *to inclusions $\lambda' \subseteq \lambda$ of lowersets, it associates ordinary function restriction:*

$$CausFun(\Lambda(\Theta), \underline{O})(\lambda, \lambda') := f \mapsto f|_{\lambda'} \tag{4.14}$$

Theorem 4.28. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. CausFun $\left(\Lambda(\Theta), \underline{O}\right)$ is a well-defined separated presheaf. Compatible families are families of functions which are compatible in the sense of Definition 3.3. Their gluing is given by their compatible join in the sense of Definition 3.4 whenever the compatible join is causal, and no gluing exists otherwise.*

Proof. CausFun $\left(\Lambda(\Theta), \underline{O}\right)(\lambda)$ is well-defined by Proposition 4.25, the restrictions are well-defined by Corollary 4.26. Because restrictions are ordinary function restrictions, a family of functions is compatible if any pair of functions in the family agree on the intersections of their domains, agreeing with Definition 3.3. A gluing for a family of compatible functions, if it exists, must be a function defined on the union of their domains, which agrees with each function on its domain: the only option is the compatible join according to Definition 3.4. If the compatible join is causal, then it is the gluing of the family; otherwise, no gluing can exist. □

If Θ is a tight space, then the assignment of causal functions forms a sheaf, as stated by the following theorem:

Theorem 4.29. *Let Θ be a space of input histories. If Θ is tight, then CausFun $\left(\Lambda(\Theta), \underline{O}\right)$ is a sheaf.*

Proof. This is a straightforward consequence of the fact that if Θ is tight so is every $\lambda \in \Lambda(\Theta)$. The assignment of causal functions in unconstrained and given a compatible family of causal functions $\{f_i\}_i$ for a family of lowersets $\{\lambda_i\}_i$ where $\lambda_i \in \Lambda(\Theta)$, there is a unique causal f for $\bigcup_i \lambda_i$ obtained by gluing the compatible functions. □

The extended causal functions on a space of input histories can also be arranged into a presheaf, which is naturally isomorphic to the presheaf of causal functions via the Ext bijection. The reason for explicitly defining such a presheaf is that the output data of extended causal functions are already 'glued together', providing outputs for all events in the domain of any extended input history. This will prove more convenient when connecting distributions over (extended) causal functions to the conditional probability distributions used by other works on causality.

Definition 4.30. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. The* presheaf of extended causal functions *ExtCausFun $\left(\Lambda(\Theta), \underline{O}\right)$ for space Θ and outputs $\underline{O}$ is the presheaf on the topological space Θ defined as follows:*

- *to lowersets $\lambda \in \Lambda(\Theta)$, it associates the extended causal functions for λ:*

$$\mathit{ExtCausFun}\left(\Lambda(\Theta), \underline{O}\right)(\lambda) := \mathit{ExtCausFun}\left(\lambda, \underline{O}\right) \tag{4.15}$$

- *to inclusions $\lambda' \subseteq \lambda$ of lowersets, it associates the following restrictions:*

$$\mathit{ExtCausFun}\left(\Lambda(\Theta), \underline{O}\right)(\lambda, \lambda') := \mathit{Ext}(f) \mapsto \mathit{Ext}(f)|_{\mathit{Ext}(\lambda)} \tag{4.16}$$

Proposition 4.30. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. $ExtCausFun\left(\Lambda(\Theta), \underline{O}\right)$ is a well-defined presheaf. The family of bijections $Ext = \left(Ext : CausFun\left(\lambda, \underline{O}\right) \to ExtCausFun\left(\lambda, \underline{O}\right)\right)_{\lambda \in \Lambda(\Theta)}$ defines a natural isomorphism of presheaves $Ext : CausFun\left(\Lambda(\Theta), \underline{O}\right) \cong ExtCausFun\left(\Lambda(\Theta), \underline{O}\right)$:*

$$Ext(f)|_{Ext(\lambda)} = Ext\left(f|_\lambda\right) \tag{4.17}$$

As a consequence, $ExtCausFun\left(\Lambda(\Theta), \underline{O}\right)$ is always a separated presheaf and it is a sheaf when Θ is a tight space.

Proof. Proving Equality 4.17 is equivalent to showing that the following diagram commutes

$$\begin{array}{ccc} \text{CausFun}\left(\lambda, \underline{O}\right) & \xrightarrow{\text{Ext}} & \text{ExtCausFun}\left(\lambda, \underline{O}\right) \\ \big\downarrow{\scriptstyle \_|_{\lambda'}} & & \big\downarrow{\scriptstyle \_|_{\text{Ext}(\lambda')}} \\ \text{CausFun}\left(\lambda', \underline{O}\right) & \xrightarrow{\text{Ext}} & \text{ExtCausFun}\left(\lambda', \underline{O}\right) \end{array}$$

since Ext is a bijection this shows that the presheaves are naturally isomorphic. The claim that the assignment of extended causal functions forms a separated presheaf/sheaf follows directly from the natural isomorphism. To prove the equation we first notice that the restriction $\text{Ext}(f)|_{\text{Ext}(\lambda)}$ is well-defined because $\text{Ext}(\lambda) \subseteq \text{Ext}(\Theta)$. Let us chase the diagram by evaluating the function $f \in \text{CausFun}\left(\lambda, \underline{O}\right)$ on a generic input history $k \in \text{Ext}(\lambda)$. On the left side of the equation we have:

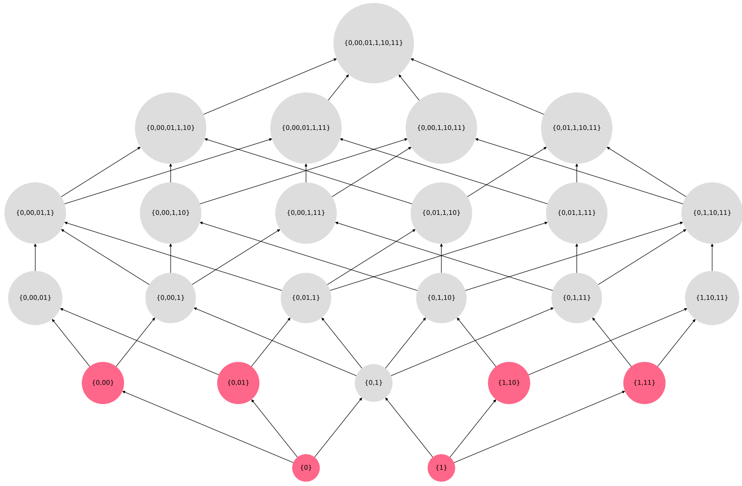
$$\text{Ext}(f)|_{\text{Ext}(\lambda)}(k) = \text{Ext}(f)(k)$$

On the right hand side:

$$\text{Ext}\left(f|_\lambda\right)(k) = \bigvee_{h \in k\downarrow \cap \lambda} f|_\lambda(h) = \bigvee_{h \in k\downarrow} f(h) = \text{Ext}(f)(k)$$

The first and the last equalities are a consequence of Proposition 4.22, while the middle follows from λ being a lowerset. □

Example 4.31. *Consider the running example of the space given by $Hist(total(A, B), \{0, 1\})$. Assigning causal data for all the contexts is made easy by the fact that $CausFun\left(\Lambda(\Theta), \underline{O}\right)$ is a sheaf. It is therefore enough to assign data to the atoms of the lattice in Example 4.24:*



The causal data assigned to any context can be univocally recovered from this assignment to the atomic components. Consider for example the context $\{0, 00, 01\}$ *describing the following lowerset of histories:*

$$\lambda'' = \{\{A\!:\!0\}, \{A\!:\!0,\, B\!:\!0\}, \{A\!:\!0,\, B\!:\!1\}\}$$

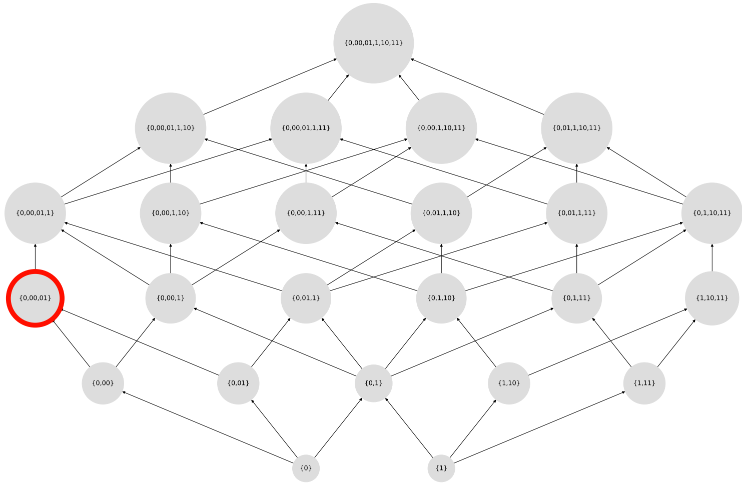
This can be recovered from gluing compatible data defined over $\lambda = \{\{A\!:\!0\}, \{A\!:\!0,\, B\!:\!0\}\}$ *and* $\lambda' = \{\{A\!:\!0\}, \{A\!:\!0,\, B\!:\!1\}\}$. *For example:*

$$f \in CausFun(\lambda, O) \qquad f = \begin{cases} f(\{A\!:\!0\}) & \mapsto 0 \\ f(\{A\!:\!0,\, B\!:\!0\}) & \mapsto 1 \end{cases}$$

$$f' \in CausFun(\lambda', O) \qquad f = \begin{cases} f(\{A\!:\!0\}) & \mapsto 0 \\ f(\{A\!:\!0,\, B\!:\!1\}) & \mapsto 1 \end{cases}$$

Clearly the restriction of $f|_A = f'|_A$ *to the lowerset* $A = \{\{A\!:\!0\}\}$ *and gluing* f *and* f' *as defined by Theorem 4.28 gives a causal function for the context* λ*:*

$$f \vee f' = \begin{cases} f(\{A\!:\!0\}) & \mapsto 0 \\ f(\{A\!:\!0,\, B\!:\!1\}) & \mapsto 1 \\ f(\{A\!:\!0,\, B\!:\!0\}) & \mapsto 1 \end{cases}$$



4.5 Causal distributions

In this subsection we show how to recover non-deterministic behaviour by post-composing the sheaf of causal functions with the distribution monad and discuss the relationship between covers and empirical models. First, recall the notion of the distribution monad introduced in Chapter 1:

Definition 4.32. *The* distribution monad $\mathcal{D}$ *is the following mapping on sets and functions:*

- *If X is a set, $\mathcal{D}(X)$ is the set of probability distributions over X with finite support:*

$$\mathcal{D}(X) := \left\{ d : X \to \mathbb{R}^+ \;\middle|\; \sum_{x \in X} d(x) = 1, \mathit{supp}(d) \text{ is finite} \right\} \tag{4.18}$$

  where the support of a distribution is the set of points over which it is non-zero:

$$\mathit{supp}(d) := \{ x \in X \,|\, d(x) \neq 0 \} \tag{4.19}$$

- *If $f : X \to Y$ is a function between sets, $\mathcal{D}(f)$ is the function $\mathcal{D}(X) \to \mathcal{D}(Y)$ defined as the linear extension of f to probability distributions with finite support:*

$$\mathcal{D}(f) := d \mapsto \sum_{x \in X} d(x) \delta_{f(x)} \tag{4.20}$$

  where $\delta_y \in \mathcal{D}(Y)$ is the delta distribution at y:

$$\delta_y := y' \mapsto \begin{cases} 1 \text{ if } y' = y \\ 0 \text{ otherwise} \end{cases} \tag{4.21}$$

Proposition 4.32. *Let X be a topological space and let $\mathcal{T}(X) \subseteq \mathcal{P}(X)$ be it collection of open sets. Let P, P' be presheaves on some topological space X. If $\phi : P \cong P'$ are naturally isomorphic, then $\mathcal{D}\phi : \mathcal{D}P \cong \mathcal{D}P'$ are also naturally isomorphic, where we defined:*

$$\mathcal{D}\phi := (\mathcal{D}(\phi_U))_{U \in \mathcal{T}(X)}$$

Proof. The distribution monad is a functor and therefore preserves the bijection $\phi_U : P(U) \to P'(U)$:

$$\begin{aligned} \mathcal{D}\left(\phi_U^{-1} \circ \phi_U\right) &= \mathcal{D}\left(id_{P(U)}\right) = id_{\mathcal{D}P(U)} \\ \mathcal{D}\left(\phi_U \circ \phi_U^{-1}\right) &= \mathcal{D}\left(id_{P'(U)}\right) = id_{\mathcal{D}P'(U)} \end{aligned}$$

Furthermore, the bijections commute with restrictions:

$$\begin{aligned} \mathcal{D}(\phi_V) \circ \mathcal{D}P(U,V) &= \mathcal{D}(\phi_V \circ P(U,V)) \\ &= \mathcal{D}(P(U,V) \circ \phi_U) = \mathcal{D}P(U,V) \circ \mathcal{D}(\phi_U) \end{aligned}$$

We conclude that $\mathcal{D}\phi$ is a natural isomorphism between $\mathcal{D}P$ and $\mathcal{D}P'$. □

Proposition 4.32 allows us to define the causal distributions by using either the sheaf of causal functions or the sheaf of extended causal functions. To simplify our upcoming definition of empirical models, we choose to extend causal functions as our base for causal distributions.

Definition 4.33. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty input sets. The* presheaf of causal distributions *for Θ is defined as follows:*

$$CausDist\left(\Lambda(\Theta), \underline{O}\right) := \mathcal{D}\, ExtCausFun\left(\Lambda(\Theta), \underline{O}\right) \tag{4.22}$$

We also define the following notation for the individual sets of distributions:

$$CausDist\left(\lambda, \underline{O}\right) := \mathcal{D}\left(ExtCausFun\left(\lambda, \underline{O}\right)\right) \tag{4.23}$$

Proposition 4.33. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty output sets. The restrictions of the presheaf $CausDist\left(\Lambda(\Theta), \underline{O}\right)$ act by marginalisation on probability distributions $d \in CausDist\left(\lambda, \underline{O}\right)$:*

$$d|_{\lambda'} = Ext(f') \mapsto \sum_{f \text{ s.t. } f|_{\lambda'} = f'} d(Ext(f)) \tag{4.24}$$

In words, the probability assigned by the marginal $d|_{\lambda'} \in CausDist\left(\lambda', \underline{O}\right)$ to a generic extended causal function $Ext(f') \in ExtCausFun\left(\lambda', \underline{O}\right)$ is the sum of the probabilities assigned by d to all extended causal functions $Ext(f) \in ExtCausFun\left(\lambda, \underline{O}\right)$ which restrict to $Ext(f')$.

Proof. From the Definition of $\mathcal{D}$ we get:

$$\mathcal{D}\left(\mathrm{Ext}(f) \mapsto \mathrm{Ext}(f)|_{\mathrm{Ext}(\lambda')}\right) : \mathcal{D}(\mathrm{Ext}(f)) \to \mathcal{D}\left(\mathrm{Ext}(f)|_{\mathrm{Ext}(\lambda)}\right)$$

$$d \mapsto \sum_{\text{Ext}(f)} d(\text{Ext}(f))\delta_{\text{Ext}(f)|_{\text{Ext}(\lambda')}}$$

We can rearrange the sum in order to make the dependence on arbitrary $\text{Ext}(f') \in \text{ExtCausFun}(\lambda, \underline{O})$ explicit:

$$\sum_{\text{Ext}(f)} d(\text{Ext}(f))\, \delta_{\text{Ext}(f)|_{\text{Ext}(\lambda')}} = \sum_{\text{Ext}(f')} \left(\sum_{\text{Ext}(f) \text{ s.t. } \text{Ext}(f)|_{\text{Ext}(\lambda')} = \text{Ext}(f')} d(\text{Ext}(f)) \right) \delta_{\text{Ext}(f')}$$

According to Proposition 4.30 an equivalent condition for $\text{Ext}(f)|_{\text{Ext}(\lambda')} = \text{Ext}(f')$ is that $f|_{\lambda'} = f'$:

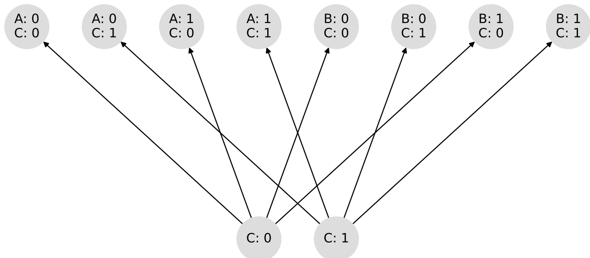
$$\sum_{\text{Ext}(f')} \left(\sum_{\text{Ext}(f) \text{ s.t. } \text{Ext}(f)|_{\text{Ext}(\lambda')} = \text{Ext}(f')} d(\text{Ext}(f)) \right) \delta_{\text{Ext}(f')} = \sum_{\text{Ext}(f')} \left(\sum_{f \text{ s.t. } f|_{\lambda'} = f'} d(\text{Ext}(f)) \right) \delta_{\text{Ext}(f')}$$

□

Definition 4.34. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. A* standard empirical model *e is a compatible family $e = (e_{k\downarrow})_{k \in \max Ext(\Theta)}$ for the presheaf of causal distributions:*

$$e_{k\downarrow} \in CausDist(k\downarrow, \underline{O})$$

Example 4.35. *Consider the following empirical model for the causal fork space:*



ABC	*000*	*001*	*010*	*011*	*100*	*101*	*110*	*111*
000	1/4	1/4	0	0	0	0	1/4	1/4
001	0	0	1/4	1/4	1/4	1/4	0	0
010	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
011	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
100	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
101	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
110	1/4	0	0	1/4	0	1/4	1/4	0
111	1/4	0	0	1/4	0	1/4	1/4	0

In this tabular form, each row $i_A i_B i_C$ corresponds to a maximal extended input history $\{A:i_A, B:i_B, C:i_C\} \in StdCov(\Theta)$, while each column $o_A o_B o_C$ corresponds to an associated extended output history $\{A:o_A, B:o_B, C:o_C\}$. For now, however, each line $i_A i_B i_C$ of the same empirical model has to be defined explicitly as a distribution on extended causal functions:

$$\mathcal{D}(ExtCausFun(\{A:i_A, B:i_B, C:i_C\}\downarrow, \{0,1\}))$$

Equivalently, we look at distributions on causal functions, which are freely characterised:

$$\mathcal{D}\left(CausFun\left(\{A\!:\!i_A,\ B\!:\!i_B,\ C\!:\!i_C\}\downarrow, \{0,1\}\right)\right)$$

Because the space Θ *is both tight and causally complete, the causal functions on* $\{A\!:\!i_A,\ B\!:\!i_B,\ C\!:\!i_C\}\downarrow$ *take the following form:*

$$\prod_{h \leq \{A i_A,\ B i_B,\ C i_C\}} O_{tip_\Theta(h)}$$

where we used the fact that $tip_{\{A i_A,\ B i_B,\ C i_C\}\downarrow}(h) = tip_\Theta(h)$ *to simplify the expression. The input histories* $h \leq \{A\!:\!i_A,\ B\!:\!i_B,\ C\!:\!i_C\}$ *are exactly:*

- $h = \{C\!:\!i_C\}$ *with tip event C*
- $h = \{C\!:\!i_C,\ A\!:\!i_A\}$ *with tip event A*
- $h = \{C\!:\!i_C,\ B\!:\!i_B\}$ *with tip event B*

Hence the causal functions in $CausFun\left(\{A\!:\!i_A,\ B\!:\!i_B,\ C\!:\!i_C\}\downarrow, \{0,1\}\right)$ *are:*

$$f_{o_A o_B o_C | i_A i_B i_C} := \begin{cases} \{C\!:\!i_C\} & \mapsto o_C \\ \{C\!:\!i_C,\ A\!:\!i_A\} & \mapsto o_A \\ \{C\!:\!i_C,\ B\!:\!i_B\} & \mapsto o_B \end{cases}$$

Using these functions, we can reconstruct the desired distribution for each row of the empirical model above. For example, the second row is indexed by the maximal extended input history $\{A\!:\!0,\ B\!:\!0,\ C\!:\!1\}$ *and it corresponds to the following distribution on extended causal functions:*

$$\frac{1}{4}\delta_{Ext\left(f_{010|001}\right)} + \frac{1}{4}\delta_{Ext\left(f_{011|001}\right)} + \frac{1}{4}\delta_{Ext\left(f_{100|001}\right)} + \frac{1}{4}\delta_{Ext\left(f_{101|001}\right)}$$

Doing this for all rows yields the following standard empirical model:

$$e_{\{A i_A,\ B i_B,\ C i_C\}\downarrow} := \begin{cases} \frac{1}{4} \sum_{o_A \oplus o_B = i_C} \sum_{o_C} \delta_{Ext\left(f_{o_A o_B o_C | i_A i_B i_C}\right)} \\ \frac{1}{8} \sum_{o_A} \sum_{o_B} \sum_{o_C} \delta_{Ext\left(f_{o_A o_B o_C | i_A i_B i_C}\right)} \end{cases}$$

What is standard about the definition of the empirical models given above? As discussed earlier, the literature on causality and causal inequalities is accustomed to tables such as the one provided in Example 4.35. To the different joint inputs we simply associate a probability distribution over the joint outputs. The key is to observe that $\left\{ k\downarrow \,\middle|\, k \in \max \mathrm{ExtHist}\left(\Theta, \underline{O}\right) \right\}$ is an example of an 'open cover' for the topological space Θ, and to explore what happens if we define empirical models on other such open covers. What we derive is a 'hierarchy of contextuality', corresponding to different operational assumptions. Three covers are of particular interest.

- The ‘standard cover’ $\left\{ k\downarrow \middle| k \in \max \mathrm{ExtHist}\left(\Theta, \underline{O}\right) \right\}$, accommodating generic causal distributions on joint outputs conditional to the maximal extended input histories. It models settings where it is, at the very least, possible to define conditional distributions when all events are taken together. Empirical models on the standard cover are the standard empirical models defined above.

- The ‘classical cover’ $\{\Theta\}$ is the ‘coarsest’ cover, lying at the top of the hierarchy. It models settings admitting a deterministic causal hidden variable explanation. Empirical models on the classical cover can be restricted to every other open cover: the empirical models arising this way are known as ‘non-contextual’.

- The ‘solipsistic cover’ $\left\{ h\downarrow \middle| h \in \max \mathrm{Hist}\left(\Theta, \underline{O}\right) \right\}$ is the ‘finest’ cover, lying at the bottom of the hierarchy. It models settings more restrictive than those modelled by the standard cover, where it might only be possible to define distributions over the events in the past of some event.

Starting from any cover C, we can obtain a coarser cover C' by fusing certain contexts—open sets of Θ—together into their union. This corresponds to the operational requirement that distributions be simultaneously definable on multiple histories.

We say that an empirical model is ‘contextual’ if it doesn’t arise by restriction from an empirical model on the classical cover. We now proceed to formalise all of the above.

Definition 4.36. *Let X be a topological space and $\mathcal{T}(X) \subseteq \mathcal{P}(X)$ be its topology. An* open cover*, or simply a* cover*, for X is an antichain in the partial order $\mathcal{T}(X)$, i.e. a collection $C \subseteq \mathcal{T}(X)$ of open sets which are incomparable:*

$$\forall U, V \in C.\ V \leq U \Rightarrow V = U$$

such that

$$\bigvee_{U \in C} U = X$$

If C and C' are covers on X, we say that C' is finer *than C, written $C' \leq C$, if the following holds:*

$$C' \leq C \Leftrightarrow \forall V \in C'.\, \exists U \in C.\ \text{s.t. } V \subseteq U \tag{4.25}$$

Equivalently, we say that C is coarser *than C'. Note that $\leq$ is a partial order on covers for X, known as the* refinement order.

A comparison with the standard sheaf theoretic approach of [6] is in order. In the standard sheaf theoretic literature—differently from our approach—there is no canonical cover associated with the topologies. The impossibility of defining the standard cover arises because the set of measurements X is always and exclusively endowed with the discrete topology $\mathcal{P}(X)$.

Consider the space $\Theta = \text{Hist}(\text{indiscrete}(A), I_A)$ where $I_A = \{0, 1\}$ and the space $\Theta = \text{Hist}(\text{indiscrete}(A, B), (I_A, I_B))$ where $I_A = I_B = \{0\}$. The spaces $\Lambda(\Theta)$ and $\Lambda(\Theta')$ are isomorphic as locales; they are both equivalent to the powerset $\mathcal{P}(\{0, 1\})$. This notwithstanding, the standard cover for the two spaces is different: the unique maximal extended history for Θ' is given by $\{\mathrm{A}:0,\ \mathrm{B}:0\}$ while Θ has two maximal extended histories: $\{\mathrm{A}:0\}$ and $\{\mathrm{A}:1\}$.

We see, therefore, that the points of the topological space (the prime histories) have additional structure informed by causality, which allows us to single out a standard cover canonically. In the original framework, spacelike separability is reflected by choosing a suitable cover for the discrete topology. The standard framework cannot cope with causal assumptions beyond nonlocality: everything has to be embeddable in the powerset locale for some set of measurements. Measurement contexts are always exclusively characterised by the choice of a cover and not by changing the underlying topology.

The set of measurements, which in the sheaf theoretic literature is expressed as the disjoint union $\coprod_{i \in \mathcal{I}} I_i$ is, translated in our language is the set of histories $\text{Hist}\left(\text{discrete}(\mathcal{I}), \underline{I}\right)$. The maximal contexts, normally described as $\prod_{i \in \mathcal{I}} I_i$ can be now identified as the family of histories $k \in \max \text{Ext}\left(\text{Hist}\left(\text{discrete}(\mathcal{I}), \underline{I}\right)\right)$. Empirical models are defined to be compatible distributions of functions, assigning joint inputs to joint outputs:

$$\prod_{i \in \mathcal{I}} I_i \to \mathcal{D}\left(\prod_{i \in \mathcal{I}} O_i\right)$$

This is isomorphic (as we will formally shown in the next chapter) to assigning compatible data $e_k \in \text{CausDist}\left(k\downarrow, \underline{O}\right)$ for all $k\downarrow$ where k is a maximal extended history for the scenario.

Definition 4.37. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. Let e be an empirical model. We say that e is* non-contextual *if it arises as restriction $e = \hat{e}|_{\text{dom}(e)}$ of a classical empirical model $\hat{e} \in EmpMod\left(ClsCov(\Theta), \underline{O}\right)$; otherwise, we say that e is* contextual. *If e is a standard empirical model, we adopt* local *as a synonym of non-contextual, and* non-local *as a synonym of contextual.*

Definition 4.38. *Let Θ be a space of input histories. The* standard cover *on Θ is the following open cover:*

$$StdCov(\Theta) := \left\{ k\downarrow \,\middle|\, k \in \max ExtHist\left(\Theta, \underline{O}\right) \right\} \tag{4.26}$$

The solipsistic cover *on Θ is the following open cover:*

$$SolCov(\Theta) := \left\{ h\downarrow \,\middle|\, h \in \max Hist\left(\Theta, \underline{O}\right) \right\} \tag{4.27}$$

The classical cover *on Θ is the following open cover:*

$$ClsCov(\Theta) := \{\Theta\} \tag{4.28}$$

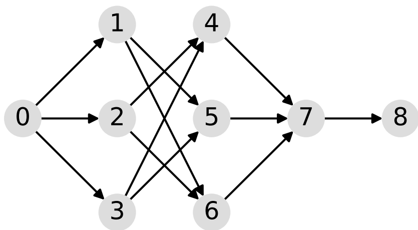
The hierarchy of covers *for Θ is the set $Covers(\Theta)$ of open covers ordered by refinement $\preceq$.*

Proposition 4.38. *Let Θ be a space of input histories. The partial order Covers (Θ) is a lattice, with the solipsistic cover SolCov (Θ) as its unique minimum and the classical cover ClsCov (Θ) as its unique maximum. In particular:*

$$\mathit{SolCov}\,(\Theta) \leq \mathit{StdCov}\,(\Theta) \leq \mathit{ClsCov}\,(\Theta)$$

Proof. Let C be an arbitrary cover. We have that $C \preceq \mathrm{ClsCov}\,(\Theta)$: every $\lambda \in \Lambda(\Theta)$, $\lambda \subseteq \Theta$. To show that the solipsistic cover lies at the bottom of the hierarchy, notice that for every $h \in \Theta$ there exists $\lambda_h \in C$ such that $h \in \lambda_h$ and $h \downarrow \subseteq \lambda_h$. This holds for every maximal history entailing $\mathrm{SolCov}\,(\Theta) \preceq C$. □

A our first and simplest example, we look at the open covers on the discrete space with 1 event and ternary inputs Hist (A, $\{0, 1, 2\}$): we chose this particular example because it is simple enough that all covers can be enumerated explicitly, but at the same time supports an interesting contextual empirical model (on cover #7 below). There are 9 open covers for this space, arranged in the following hierarchy.



Because Hist (A, $\{0, 1, 2\}$) = ExtHist (A, $\{0, 1, 2\}$), the standard cover and solipsistic cover coincide for this example.

- Cover #0 (standard/solipsistic cover) contains the following lowersets:

$$\big\{\big\{\,\{\mathrm{A}\!:\!0\}\,\big\}, \big\{\,\{\mathrm{A}\!:\!1\}\,\big\}, \big\{\,\{\mathrm{A}\!:\!2\}\,\big\}\big\}$$

- Cover #1 contains the following lowersets:

$$\big\{\big\{\,\{\mathrm{A}\!:\!0\}\,\big\}, \big\{\,\{\mathrm{A}\!:\!1\}\,, \{\mathrm{A}\!:\!2\}\,\big\}\big\}$$

- Cover #2 contains the following lowersets:

$$\big\{\big\{\,\{\mathrm{A}\!:\!1\}\,\big\}, \big\{\,\{\mathrm{A}\!:\!0\}\,, \{\mathrm{A}\!:\!2\}\,\big\}\big\}$$

- Cover #3 contains the following lowersets:

$$\big\{\big\{\,\{\mathrm{A}\!:\!2\}\,\big\}, \big\{\,\{\mathrm{A}\!:\!0\}\,, \{\mathrm{A}\!:\!1\}\,\big\}\big\}$$

- Cover #4 contains the following lowersets:

$$\{\{\{A:0\},\{A:1\}\},\{\{A:1\},\{A:2\}\}\}$$

- Cover #5 contains the following lowersets:

$$\{\{\{A:0\},\{A:2\}\},\{\{A:1\},\{A:2\}\}\}$$

- Cover #6 contains the following lowersets:

$$\{\{\{A:0\},\{A:1\}\},\{\{A:0\},\{A:2\}\}\}$$

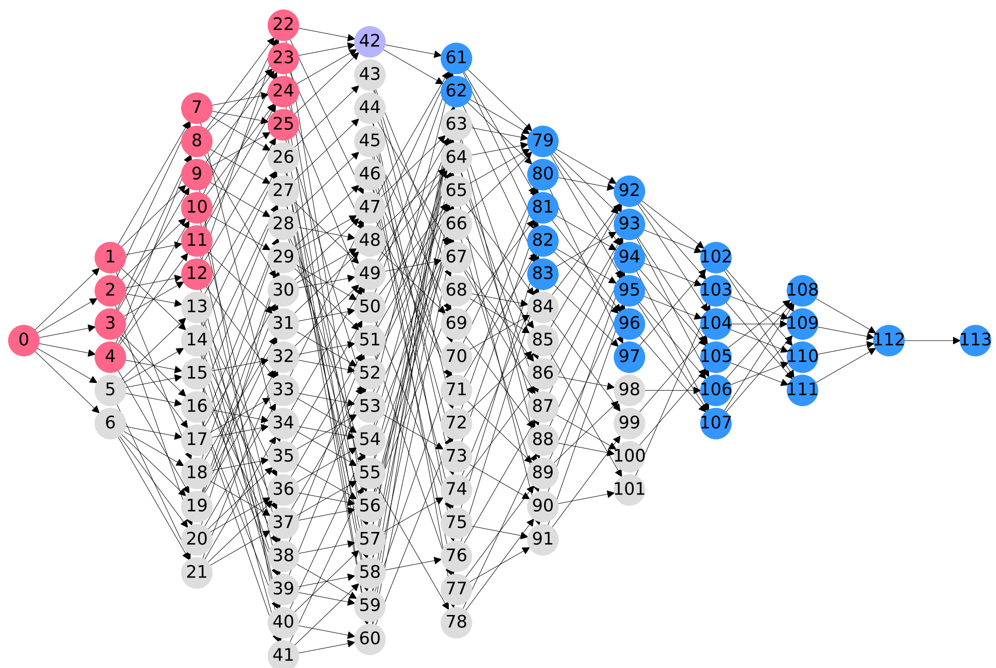
- Cover #7 contains the following lowersets:

$$\{\{\{A:0\},\{A:1\}\},\{\{A:0\},\{A:2\}\},\{\{A:1\},\{A:2\}\}\}$$

- Cover #8 (global cover) contains the following lowersets:

$$\{\{\{A:0\},\{A:1\},\{A:2\}\}\}$$

As our second example, we look at the covers for the no-signalling space $\text{Hist}(\text{discrete}(A,B),\{0,1\})$ on 2 events with binary inputs. This space has 114 open covers, arranged into the following hierarchy.



The standard cover #42 is coloured violet in the hierarchy and it takes the following form:

$$\{\{\{A:0\},\{B:0\}\},\{\{A:0\},\{B:1\}\},\{\{A:1\},\{B:0\}\}\}$$

The refinements of the standard cover are coloured red in the hierarchy above. They include the solipsistic cover #0, which takes the following form:

$$\{\{\{A:0\}\},\{\{A:1\}\},\{\{B:0\}\},\{\{B:1\}\}\}$$

The closest refinements of the standard cover are obtained by removing one of its 4 open sets. For example, cover #22 takes the following form:

$$\{\{\{A:0\},\{B:0\}\},\{\{A:0\},\{B:1\}\},\{\{A:1\},\{B:0\}\}\}$$

The coarsenings of the standard cover are coloured blue in the hierarchy above. They include the classical cover #113, which takes the following form:

$$\{\{\{A:0\},\{A:1\},\{B:0\},\{B:1\}\}\}$$

The closest coarsenings of the standard cover are obtained by adding both input histories for either event A (cover #61) or event B (cover #62). For example, cover #61 takes the following form:

$$\{\{\{A:0\},\{A:1\}\},\{\{A:0\},\{B:0\}\},\{\{A:0\},\{B:1\}\},\{\{A:1\},\{B:0\}\},\{\{A:1\},\{B:1\}\}\}$$

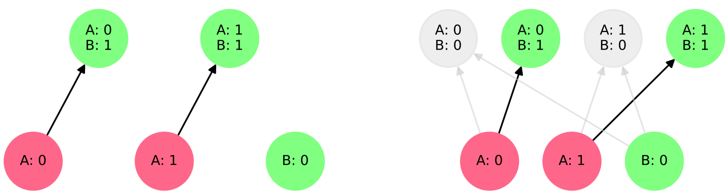
Finally, there are covers which don't lie either below or above the standard cover. The minimal covers unrelated to the standard cover are #5 and #6: these covers add both input histories for either event A (cover #6) or event B (cover #5) to the solipsistic cover, much as covers #61 and #62 did for the standard cover. For example, cover #5 takes the following form:

$$\{\{\{A:0\}\},\{\{A:1\}\},\{\{B:0\},\{B:1\}\}\}$$

The maximal covers unrelated to the standard cover are #98, #99, #100 and #101. The take the following form, for all $i_A, i_B \in \{0,1\}$:

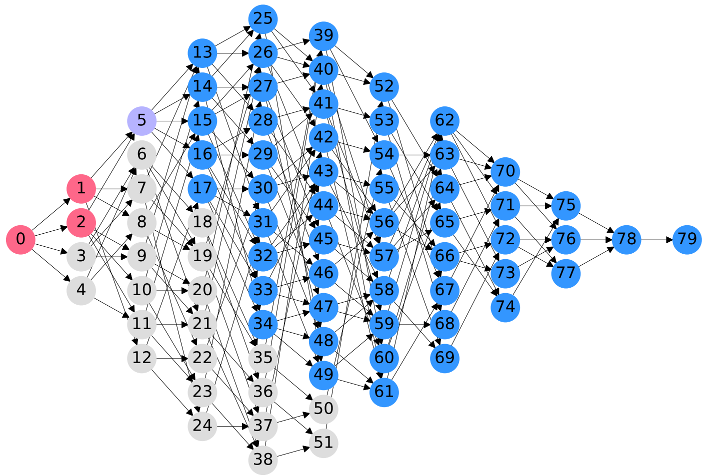
$$\{\{\{A:0\},\{A:1\},\{B:i_B\}\},\{\{A:i_A\},\{B:0\},\{B:1\}\}\}$$

As our third example, we look at the following space, one of the four spaces lying in the middle layer of the hierarchy of causally complete spaces on 2 events with binary inputs:



Θ $\qquad$ $\mathrm{Ext}(\Theta)$

This space has 80 open covers, arranged into the following hierarchy.



The standard cover #5 is coloured violet in the hierarchy and takes the following form:

$$\{\{\{\mathrm{A}\!:\!0\},\{\mathrm{B}\!:\!0\}\},\{\{\mathrm{A}\!:\!0\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!0\}\},\{\{\mathrm{A}\!:\!1\},\{\mathrm{B}\!:\!0\}\},\{\{\mathrm{A}\!:\!1\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!1\}\}\}$$

The refinements of the standard cover are coloured red in the hierarchy above. They include the solipsistic cover #0, which takes the following form:

$$\{\{\{\mathrm{B}\!:\!0\}\},\{\{\mathrm{A}\!:\!0\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!0\}\},\{\{\mathrm{A}\!:\!1\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!1\}\}\}$$

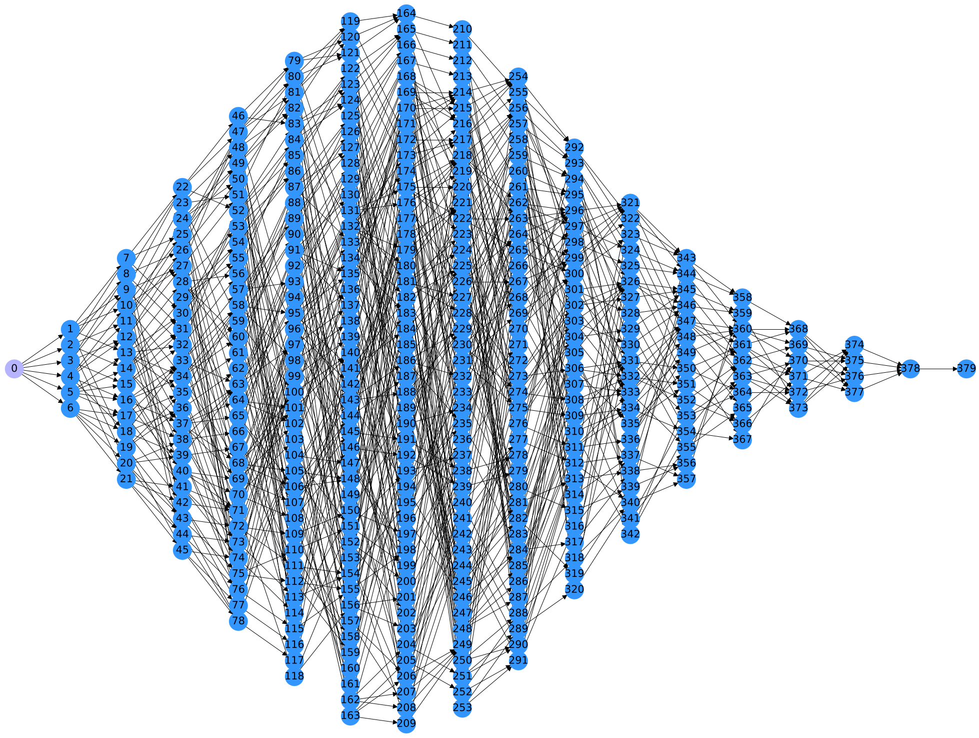
The two covers #1 and #2 lying between the solipsistic and standard cover take the following form, for $i_A \in \{0, 1\}$

$$\{\{\{\mathrm{A}\!:\!i_A\},\{\mathrm{B}\!:\!0\}\},\{\{\mathrm{A}\!:\!0\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!0\}\},\{\{\mathrm{A}\!:\!1\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!1\}\}\}$$

The coarsenings of the standard cover are coloured blue in the hierarchy above. They include the classical cover #79, which takes the following form:

$$\{\{\{\mathrm{A}\!:\!0\},\{\mathrm{A}\!:\!1\},\{\mathrm{B}\!:\!0\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!0\},\{\mathrm{B}\!:\!1,\ \mathrm{A}\!:\!1\}\}\}$$

As our fourth and final example, we look at the totally ordered space $\mathrm{Hist}(\mathrm{total}(A, B), \{0, 1\})$ on 2 events with binary inputs. This space has 380 open covers, arranged into the following hierarchy.



Because $\mathrm{Hist}(\mathrm{total}(A, B), \{0, 1\}) = \mathrm{ExtHist}(\mathrm{total}(A, B), \{0, 1\})$, the standard and solipsistic covers coincide in this example:

$$\{\{\{\mathrm{A}:0\}, \{\mathrm{B}:0,\ \mathrm{A}:0\}\}, \{\{\mathrm{A}:0\}, \{\mathrm{B}:1,\ \mathrm{A}:0\}\}, \{\{\mathrm{A}:1\}, \{\mathrm{B}:0,\ \mathrm{A}:1\}\}, \{\{\mathrm{A}:1\}, \{\mathrm{B}:1,\ \mathrm{A}:1\}\}\}$$

The classical cover takes the following form:

$$\{\{\{\mathrm{A}:0\}, \{\mathrm{A}:1\}, \{\mathrm{B}:0,\ \mathrm{A}:0\}, \{\mathrm{B}:1,\ \mathrm{A}:0\}, \{\mathrm{B}:0,\ \mathrm{A}:1\}, \{\mathrm{B}:1,\ \mathrm{A}:1\}\}\}$$

The immediate refinements of the standard cover take one of two possible forms. Covers #1, #2, #4 and #5 take the following form, for $i_A, i_B \in \{0, 1\}$:

$$\{\{\{\mathrm{A}:0\}, \{\mathrm{B}:0,\ \mathrm{A}:0\}\}, \{\{\mathrm{A}:0\}, \{\mathrm{B}:1,\ \mathrm{A}:0\}\}, \{\{\mathrm{A}:1\}, \{\mathrm{B}:0,\ \mathrm{A}:1\}\}, \{\{\mathrm{A}:0\}, \{\mathrm{A}:1\}, \{\mathrm{A}:i_A,\ \mathrm{B}:i_B\}\}\}$$

Covers #3 and #6 are bit-flips of each other, taking the following form:

$$\{\{\{\mathrm{A}:0\}, \{\mathrm{B}:0,\ \mathrm{A}:0\}\}, \{\{\mathrm{A}:0\}, \{\mathrm{B}:1,\ \mathrm{A}:0\}\}, \{\{\mathrm{A}:1\}, \{\mathrm{B}:0,\ \mathrm{A}:1\}, \{\mathrm{B}:1,\ \mathrm{A}:1\}\}\}$$

$$\{\{\{\mathrm{A}:1\}, \{\mathrm{B}:0,\ \mathrm{A}:1\}\}, \{\{\mathrm{A}:1\}, \{\mathrm{B}:1,\ \mathrm{A}:1\}\}, \{\{\mathrm{A}:0\}, \{\mathrm{B}:0,\ \mathrm{A}:0\}, \{\mathrm{B}:1,\ \mathrm{A}:0\}\}\}$$

Having seen a few examples of covers, we now move to the definition of empirical models for an arbitrary cover C. These are a straightforward generalisation of those for the standard cover: they are simply families of distributions on causal functions for each lowerset $\lambda \in C$ in the cover.

Definition 4.39. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. If C is a cover of Θ, an* empirical model *e on C is a compatible family $e = (e_\lambda)_{\lambda \in C}$ for the presheaf of causal distributions $CausDist\left(\Lambda(\Theta), \underline{O}\right)$. We write $EmpMod\left(C, \underline{O}\right)$ for the empirical models on a cover C of Θ, with outputs valued in $\underline{O}$.*

Definition 4.40. *We define the following terminology for empirical models defined on the canonical covers:*

- *A* standard empirical model *is an empirical model on the standard cover*
- *A* solipsistic empirical model *is an empirical model on the solipsistic cover*
- *A* classical empirical model *is an empirical model on the classical cover.*

It is important to observe that technically speaking, from the way we have defined empirical models, they retain the information about the underlying space of histories. The cover can be recovered as $C = \mathrm{dom}(e)$ and similarly Θ can be recovered as the union $\bigcup C$ of the open sets in the cover.

For a familiar example of an empirical model on a cover different from the standard one, we look at cover #7 for the space $\Theta := \mathrm{Hist}(\mathrm{A}, \{0, 1, 2\})$ previously discussed:

$$C := \left\{ \left\{ \{\mathrm{A}\!:\!0\}, \{\mathrm{A}\!:\!1\} \right\}, \left\{ \{\mathrm{A}\!:\!0\}, \{\mathrm{A}\!:\!2\} \right\}, \left\{ \{\mathrm{A}\!:\!1\}, \{\mathrm{A}\!:\!2\} \right\} \right\}$$

Each lowerset in this cover takes the form $\lambda_{i,i'} := \left\{ \{\mathrm{A}\!:\!i\}, \{\mathrm{A}\!:\!i'\} \right\}$ for distinct $i, i' \in \{0, 1, 2\}$ and it has the following binary-valued causal functions, for all $o, o' \in \{0, 1\}$:

$$f_{oo'|ii'} := \begin{cases} \{\mathrm{A}\!:\!i\} & \mapsto o \\ \{\mathrm{A}\!:\!i'\} & \mapsto o' \end{cases}$$

We define the following empirical model $e^{tri} \in \mathrm{EmpMod}(C, \{0, 1\})$, sometimes known as the 'contextual triangle' and originally due to [133, 85]:

$$\begin{aligned} e^{tri}_{\lambda_{01}} &:= \tfrac{1}{2}\delta_{\mathrm{Ext}(f_{01|01})} + \tfrac{1}{2}\delta_{\mathrm{Ext}(f_{10|01})} \\ e^{tri}_{\lambda_{02}} &:= \tfrac{1}{2}\delta_{\mathrm{Ext}(f_{01|02})} + \tfrac{1}{2}\delta_{\mathrm{Ext}(f_{10|02})} \\ e^{tri}_{\lambda_{12}} &:= \tfrac{1}{2}\delta_{\mathrm{Ext}(f_{01|12})} + \tfrac{1}{2}\delta_{\mathrm{Ext}(f_{10|12})} \end{aligned}$$

We can represent this empirical model in tabular form, with rows indexed by $i\,i'$ and columns indexed by $o\,o'$:

	00	01	10	11
01	0	1/2	1/2	0
02	0	1/2	1/2	0
12	0	1/2	1/2	0

As the name suggests, this empirical model is an example of a 'contextual' empirical model: these are the models which cannot be explained 'classically'. Classical empirical models for a space Θ are, by definition, probability distributions on extended causal functions defined on the entire space:

$$\text{CausDist}\left(\text{ClsCov}\left(\Theta\right), \underline{O}\right) = \mathcal{D}\left(\text{ExtCausFun}\left(\Theta, \underline{O}\right)\right)$$

As a consequence, any empirical model $e \in \text{EmpMod}\left(C, \underline{O}\right)$ which arises as restriction $\hat{e}|_{\text{dom}(e)}$ of some classical empirical model $\hat{e} \in \text{EmpMod}\left(\text{ClsCov}\left(\Theta\right), \underline{O}\right)$ admits a *deterministic causal hidden variable model (HVM)*: the observed probabilities are fully explained by some probabilistic mixture of causal functions defined globally on Θ. Empirical model admitting such a deterministic causal HVM are known as 'non-contextual' (or 'local', in the special case of the standard cover).

Definition 4.41. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. Let e be an empirical model. We say that e is* non-contextual *if it arises as restriction $e = \hat{e}|_{\text{dom}(e)}$ of a classical empirical model $\hat{e} \in \text{EmpMod}\left(\text{ClsCov}\left(\Theta\right), \underline{O}\right)$; otherwise, we say that e is* contextual. *If e is a standard empirical model, we adopt* local *as a synonym of non-contextual, and* non-local *as a synonym of contextual.*

Observation 4.41. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. Let e be an empirical model on a cover C, let C' be a finer cover and let $e' := e|_{C'}$ be the restriction of e to C'. If $e = \hat{e}|_C$ is non-contextual, then $e' = \hat{e}|_{C'}$ is non-contextual. Hence, if $e|_{C'}$ is contextual, then e is contextual.*

The contextual triangle empirical model e^{tri} previously defined on cover #7 of space $\Theta :=$ Hist (A, {0, 1, 2}) is a known example of a contextual empirical model. The causal functions in CausFun (Θ, {0, 1}) form the following set:

$$\prod_{h \in \Theta} O_{\text{tip}_\Theta(h)}$$

Specifically, there are 8 causal functions, taking the following form for $(o_0, o_1, o_2) \in \{0,1\}^3$:

$$g_{o_0 o_1 o_2} := \begin{cases} \{\text{A}:0\} & \mapsto o_0 \\ \{\text{A}:1\} & \mapsto o_1 \\ \{\text{A}:2\} & \mapsto o_2 \end{cases}$$

Classical empirical models for Θ then take the following form, for probability distributions $d \in \mathcal{D}\left(\{0,1\}^3\right)$:

$$e^{(d)} := \sum_{o_0} \sum_{o_1} \sum_{o_2} d(o_0, o_1, o_2) \delta_{\text{Ext}\left(g_{o_0 o_1 o_2}\right)}$$

The restrictions of the 8 causal functions for Θ to the lowersets in cover #7 take the following form:

$$g_{o_0 o_1 o_2}\big|_{\lambda_{i,i'}} = f_{o_i o_{i'}|ii'}$$

Hence, the restrictions of the classical empirical models to cover #7 take the following form:

$$e^{(d)}\Big|_{\lambda_{i,i'}} = \sum_{o_0}\sum_{o_1}\sum_{o_2} d(o_0, o_1, o_2)\delta_{\mathrm{Ext}\left(f_{o_i o_{i'}|ii'}\right)}$$

We can represent the generic restriction of a classical model in tabular form:

	00	01	10	11
01	$d(000) + d(001)$	$d(010) + d(011)$	$d(100) + d(101)$	$d(110) + d(111)$
02	$d(000) + d(010)$	$d(001) + d(011)$	$d(100) + d(110)$	$d(101) + d(111)$
12	$d(000) + d(100)$	$d(001) + d(101)$	$d(010) + d(110)$	$d(011) + d(111)$

For a non-contextual empirical model for cover #7 of Hist $(\mathrm{A}, \{0, 1, 2\})$, the table above shows that the difference between the sum of the elements in the first and fourth columns and the sum of the elements in the second and third column—that is, the sum of the output correlation coefficients over the 3 input contexts—is bounded below by -1:

$$\begin{array}{rl} & (d(000) + d(001)) + (d(110) + d(111)) + (d(000) + d(010)) \\ + & (d(101) + d(111)) + (d(000) + d(100)) + (d(011) + d(111)) \\ - & (d(010) + d(011)) - (d(100) + d(101)) - (d(001) + d(011)) \\ - & (d(100) + d(110)) - (d(001) + d(101)) - (d(010) + d(110)) \\ = & 3d(000) - d(001) - d(010) - d(011) \\ - & d(100) - d(101) - d(110) + 3d(111) \\ = & 4(d(000) + d(111)) - 1 \geq -1 \end{array}$$

For the contextual triangle empirical model e^{tri}, the same number comes to -3 instead, proving that the empirical model is contextual.

In the conclusions of Chapter 3, we briefly mentioned the possibility of providing a characterisation of the switch spaces. We conclude with a theorem explicitly proving that standard empirical models defined on switch spaces cannot exhibit any non-locality.

Theorem 4.42. *Let $\Theta \in$ CSwitchSpaces $(\underline{I})$ be a non-empty causal switch space and let $e \in$ EmpMod $(\mathit{StdCov}(\Theta), \underline{O})$ be a standard empirical model on Θ. Then e is local, i.e. it arises as a restriction of a classical empirical model $\hat{e} \in$ EmpMod $(\mathit{ClsCov}(\Theta), \underline{O})$ to the standard cover StdCov (Θ).*

The proof of this theorem can be found in [65]. It requires results about the relationship between the compositional property of spaces of input histories and the assignment of causal data that were omitted from this thesis for reasons of conciseness.

4.6 Conclusions

In this chapter, we constructed the presheaf of causal distributions CausDist $(\Theta, \underline{O})$ for arbitrary spaces of input histories. We characterise the 'deterministic' causal data definable on a general space as the class of continuous functions mapping extended histories to partial assignments of outputs on the set of events. We have seen that this assignment forms a sheaf when the underlying space is tight.

We provided a topological description of the contexts arising from the spaces of input histories and explained the significance of open covers in the definition of empirical models. We show that the hierarchy of covers is reflected in a hierarchy of contextuality; the data assigned to the open sets covered by the coarsest cover—the classical cover—describes models arising from classical causal mechanisms correlating the joint inputs and outputs.

With this chapter, we conclude the topological description of causality. By extending the sheaf theoretic framework, we captured the intuition that causality is to be understood by looking at the structure of the classical contexts characterising an operational description of a protocol. In the standard approaches to causal modelling, the underlying assumption is that explanations are always to be found on the 'global cover'. This may not be the case for theories allowing a contextual fragmentation of the observable quantities. Spaces of input histories allow presenting causal assumptions as topological spaces imposing compatibilities between these globally incoherent pieces of observable reality.

We will see more examples of empirical models in Chapter 6, but before doing this, we present an external and geometrical description of the casual compatibility imposed by choosing a topology and a cover to explain empirical data. For this part of the investigation, we coined the term 'geometry of causality', which can be seen as a generalisation of the convex geometrical techniques employed in studying non-signalling correlations and contextuality.

Chapter 5

The geometry of causality

Empirical models define distributions of casual data assigned to open covers of a topological space. The order between contexts constrains the 'shape' of causal data; our study has described these topological spaces of contexts for various operational assumptions. This 'internal' perspective describes causal assumptions as cohesive principles unifying a 'fragmented' contextual description.

In the preliminary Chapter 1, we reviewed the geometrical study of non-locality using polytopes and inequalities. In this chapter, we reproduce a generalised framework in which the constraints between contexts are expressed geometrically as linear equation bounding polytopes of compatible empirical models. Historically, introducing the sheaf-theoretic perspective followed the geometrical understanding of the correlations. In our case, this process will be reversed; only after describing the general framework can we recast it in geometrical terms.

This approach follows footsteps that are anterior to the study of the convex-geometrical structure of non-locality. As already mentioned, the idea of representing conditional distributions as constrained geometrical objects dates back to the work of Boole [30] on 'conditions of possible experience'. Little would he have believed that the violations of such conditions (for example, embodied in Bell's inequalities) would be essential for understanding the most fundamental account of natural phenomena.

We associate a polytope of compatible conditional distributions to every space of input histories endowed with an open cover, a 'causaltope'. Different standard causaltopes (obtained considering the standard cover) for the same empirical scenario will be embedded in the same real space, allowing us to decompose the empirical scenario by defining a notion of 'causal separability', which will be paramount to our understanding of contextual causality.

5.1 Polytopes

The primary objects underlying our geometrical perspective are provided by convex subsets of points embedded into a high-dimensional real space. These objects are usually referred to as *polytopes*.

Definition 5.1 ($\mathcal{V}$-polytope)**.** *A* polytope *$P \subseteq \mathbb{R}^d$ is the convex hull a finite set of points*

Definition 5.2 ($\mathcal{H}$-polyhedra [145])**.** *An $\mathcal{H}$-polyhedron denotes an intersection of closed half-spaces: a set $P \subseteq \mathbb{R}^d$ presented in the form:*

$$P = \{\underline{x} \in \mathbb{R}^d : A\underline{x} \leq \underline{z}\} \text{ for some } A \in \mathbb{R}^{m \times d}, \underline{z} \in \mathbb{R}^m$$

The fact that the two definitions given above coincide is often known as the fundamental theorem for convex polytopes or the Weyl-Minkowski theorem:

Theorem 5.3 (Weyl-Minkowski theorem [145])**.** *A subset $P \subseteq \mathbb{R}^d$ is the convex hull of a finite point set (a $\mathcal{V}$-polytope) if and only if it is a bounded intersection of half-spaces (an $\mathcal{H}$-polytope).*

The half-space description can provide a minimal canonical description of the polytope where every facet is associated with an inequality. Such a description is practical when explicitly using inequalities such as Bell inequalities or Causal inequalities to discriminate models that are not in the polytope, but it is not easily applicable for the high dimensional cases described by our work. For convenience, which will become evident as our work progresses, in the rest of this chapter we take a slightly different perspective and define polytopes more loosely, as follows:

Definition 5.4. *A* polytope *is a bounded subset $K \subset \mathbb{R}^J$ defined by the joint solutions $\underline{x} \in \mathbb{R}^J$ to a system $A\underline{x} = \underline{b}$ of linear equations and a system $C\underline{x} \leq \underline{d}$ of linear inequalities.*

This definition gives the descriptive freedom to allow to freely 'slice' polytopes, but it comes at the cost of losing the canonicity of the half-space description:

- Some of the equations or inequalities could be redundant.
- Equations are not necessary: $\underline{a}^T \underline{x} = b$ can be replaced by $\underline{a}^T \underline{x} \leq b$ and $\underline{a}^T \underline{x} \geq b$.
- Inequalities can pair up into equations, as above.

The vector space $\mathbb{R}^J$ should be understood as the finite-dimensional vector space formed by functions $J \to \mathbb{R}$ under pointwise addition and scalar multiplication:

Definition 5.5. *For any finite set J, we define $\mathbb{R}^J$ as the finite-dimensional real vector space formed by functions $J \to \mathbb{R}$ under pointwise addition and scalar multiplication. We adopt the Kronecker delta functions as the standard basis for this space:*

$$\underline{\delta}_i := j \mapsto \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

If $\underline{x} \in \mathbb{R}^J$, we write x_j for the j-th component of $\underline{x}$ in the standard basis, for every $j \in J$:

$$x_j := \underline{x}(j) \in \mathbb{R}$$

We take $\mathbb{R}^J$ to be equipped with the inner product for the standard basis:

$$\underline{x}^T \underline{y} := \sum_{j \in J} x_j y_j$$

We also take $\mathbb{R}^J$ to be equipped with the product order:

$$\underline{x} \leq \underline{y} \overset{def}{\Leftrightarrow} \forall j \in J.\, x_j \leq y_j$$

For every $n \in \mathbb{N}$, we write $\mathbb{R}^n$ to denote $J := \{1, ..., n\}$, where $n = 0$ means $J = \emptyset$.

The importance of the inner product structure is that it allows describing equations using matrices so that:

$$A\underline{x} = \underline{b} \Leftrightarrow \forall r \in R.\, \sum_{j \in J} A_{rj} x_j = b_r$$

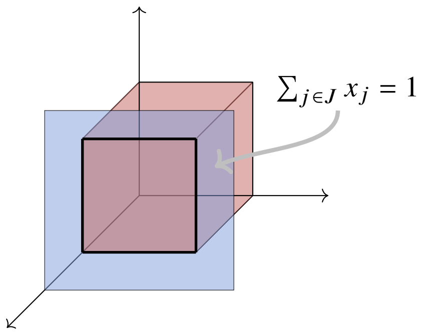
The choice of product order between the vectors allows to extend the matrix representation to linear inequalities:

$$A\underline{x} \leq \underline{b} \Leftrightarrow \forall r \in R.\, \sum_{j \in J} A_{rj} x_j \leq b_r$$

We provide a couple of examples of elementary polytopes that will be important in our discussion, with their associated description in terms of equalities and inequalities:

Example 5.6 (Standard hypercube)**.** *The* standard hypercube $[0, 1]^J \subset \mathbb{R}^J$ *is defined by the inequalities $x_j \leq 1$ and $x_j \geq 0$ for all $j \in J$. In matrix form, the linear constraints take the form:*

$$\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ -1 & & \\ & \ddots & \\ & & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

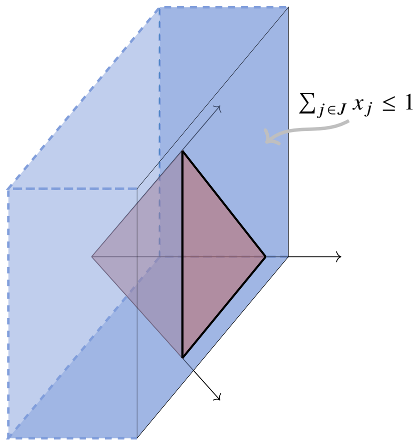


More generally, for every $\underline{u} \in \mathbb{R}^J$ with $u_j > 0$ for all $j \in J$, we can define the standard hypercuboid $\prod_j [0, u_j] \subset \mathbb{R}^J$, *where the upper-bounding inequalities $x_j \leq 1$ for the unit hypercube are replaced by $x_j \leq u_j$:*

$$\begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ -1 & & \\ & \ddots & \\ & & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq \begin{pmatrix} u_1 \\ \vdots \\ u_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Example 5.7 (Standard simplex). *The standard simplex $\Delta^J \subset \mathbb{R}^J$ is defined by the equations $x_j \leq 0$ and a single upper-bound $\sum_{j\in J} x_j \leq 1$. The following matrix represents the linear equations:*

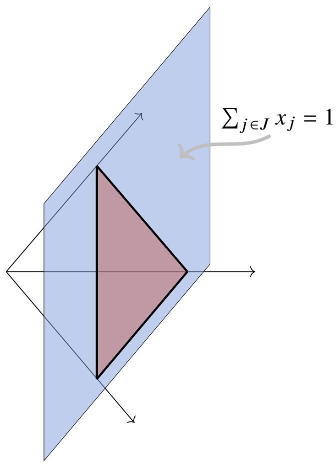
$$\begin{pmatrix} 1 & \dots & 1 \\ -1 & & \\ & \ddots & \\ & & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$



The examples described above are all defined by inequalities alone; they also have the same dimensionality as the ambient space. The *polytope of probability distributions* however, gives us an example which is naturally defined by both equalities and inequalities:

Example 5.8 (Probability distributions). *Let J be a finite set and consider the space $\mathcal{D}(J) \subset \mathbb{R}^J$ of probability distributions over the set J. The probabilities must be a non-negative real numbers $x_j \geq 0$ such that sum to unity $\sum_{j\in J} x_j = 1$. Since we have a mixture of equalities and inequalities, we represent the linear constraints by the following matrices:*

$$\begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = 1 \quad \begin{pmatrix} -1 & & \\ & \ddots & \\ & & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \leq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$



For an indexed family of polytopes, we can form their product polytope:

Example 5.9 (Product polytope). *From a family of polytopes $K^{(y)} \subset R^{J^{(y)}}$ indexed by a non-empty finite set Y, each defined by its own system of linear equations $A^{(y)}\underline{x}^{(y)} = \underline{b}^{(y)}$ and linear inequalities $B^{(y)}\underline{d}^{(y)} = \underline{b}^{(y)}$, we can construct the associated* product polytope $\prod_{y\in Y} K^{(y)}$ *embedded in $\prod_{y\in Y} \mathbb{R}^{J^{(f)}} = \mathbb{R}^{\sqcup_{y\in Y} J^{(y)}}$ where the disjoint union $\sqcup_{y\in Y} J^{(y)}$ is formally defined as follows:*

$$\bigsqcup_{y\in Y} J^{(y)} := \left\{ (y, j) \,\middle|\, y \in Y, j \in J^{(y)} \right\} \tag{5.1}$$

In terms of matrices, the product polytope is defined by combining equations and inequalities for each factor in a block-diagonal way:

$$\begin{pmatrix} A^{(1)} & 0 & \dots & 0 \\ 0 & A^{(2)} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & A^{(m)} \end{pmatrix} \begin{pmatrix} \underline{x}^{(1)} \\ \vdots \\ \underline{x}^{(m)} \end{pmatrix} = \begin{pmatrix} \underline{b}^{(1)} \\ \vdots \\ \underline{b}^{(m)} \end{pmatrix} \qquad \begin{pmatrix} C^{(1)} & 0 & \dots & 0 \\ 0 & C^{(2)} & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & C^{(m)} \end{pmatrix} \begin{pmatrix} \underline{x}^{(1)} \\ \vdots \\ \underline{x}^{(m)} \end{pmatrix} \leq \begin{pmatrix} \underline{d}^{(1)} \\ \vdots \\ \underline{d}^{(m)} \end{pmatrix}$$

Which is equivalent to the juxtaposition of all the equations and inequalities as follows

$$\forall y \in Y.\ A^{(y)}\underline{x}^{(y)} = \underline{b}^{(y)} \qquad \forall y \in Y.\ C^{(y)}\underline{x}^{(y)} \leq \underline{d}^{(y)} \tag{5.2}$$

Where we index the coordinates of vectors $\underline{x} \in \mathbb{R}^{\sqcup_{y\in Y} J^{(y)}}$ as $x_j^{(y)}$, for $y \in Y$ and $j \in J^{(y)}$, and we have defined:

$$\underline{x}^{(y)} := \left(x_j^{(y)}\right)_{j\in J^{(y)}} \in \mathbb{R}^{J^{(y)}}$$

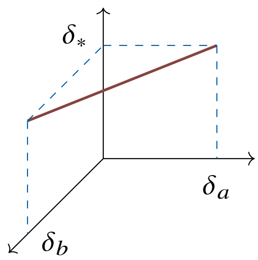
Conditional probability distributions give an important example of a product of polytope. Consider a finite non-empty set Y, and a family of finite sets indexed by Y, which will be denoted by $\underline{J} = \left(J^{(y)}\right)_{y\in Y}$. The polytope of conditional distributions is given by the convex set $\prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right)$. As a product polytope, it can be explicitly described using the following equalities and inequalities:

$$\begin{pmatrix} \underline{1}^T & 0 & \dots & 0 \\ 0 & \underline{1}^T & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \underline{1}^T \end{pmatrix} \begin{pmatrix} \underline{x}^{(1)} \\ \vdots \\ \underline{x}^{(m)} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \qquad \begin{pmatrix} -I & 0 & \dots & 0 \\ 0 & -I & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & -I \end{pmatrix} \begin{pmatrix} \underline{x}^{(1)} \\ \vdots \\ \underline{x}^{(m)} \end{pmatrix} \leq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Example 5.10. *Let $Y = \{0,1\}$ and consider the conditional sets of outcomes $J^{(0)} = \{*\}$ and $J^{(1)} = \{a,b\}$. The ambient vector space $\mathbb{R}^{\sqcup_{y\in Y} J^{(y)}} \simeq \mathbb{R}^{|J^{(0)}|+|J^{(1)}|} = \mathbb{R}^3$. The linear equalities and inequalities defining the polytope are:*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_* \\ x_a \\ x_b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_* \\ x_a \\ x_b \end{pmatrix} \leq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

The equations bound the segment which embedded in $\mathbb{R}^{\sqcup_{y\in Y} J^{(y)}}$ connects $(\delta_*, \delta_a, 0)$ to $(\delta_*, 0, \delta_b)$.



Observation 5.10. *There is a unique minimal affine subspace $\mathbb{A}(K) \subset \mathbb{R}^J$ which contains a given polytope $K \subset \mathbb{R}^J$, and the following procedure can explicitly identify it:*

1. *Replace every pair of inequalities in the form $\underline{c}^T\underline{x} \leq d$ and $-\underline{c}^T\underline{x} \leq -d$ with the corresponding equation $\underline{c}^T\underline{x} = d$.*

2. Turn the system of equations into reduced row echelon form (RREF), removing zero rows.

The affine subspace $\mathbb{A}(K)$ is the space of solutions for the resulting system of equations, and there is a bijection between affine subspaces and systems of equations in RREF without zero rows. The polytope K is a regular closed subset of $\mathbb{A}(K)$: the topological dimension of K is the dimension of $\mathbb{A}(K)$, which is equal to $|J|$ minus the number of non-zero rows in the system of equations in RREF.

The standard hypercube $k = [0,1]^J$ and the standard simplex $K = \Lambda^J$ all have dimensions $|J|$, therefore $\mathbb{A}(K) = \mathbb{R}^J$. The polytope $\mathcal{D}(J)$ of probability distributions has dimension $|J| - 1$, with the following minimal affine subspace:

$$\mathbb{A}(\mathcal{D}(J)) = \left\{ \underline{x} \in \mathbb{R}^J \,\middle|\, \sum_{j \in J} x_j = 1 \right\}$$

The polytope $\prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)$ of conditional probability distributions has dimension $\sum_{y \in Y} (|J^{(y)}| - 1) = \left(\sum_{y \in Y} |J^{(y)}|\right) - |Y|$, with the following minimal affine subspace:

$$\mathbb{A}\left(\prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)\right) = \left\{ \underline{x} \in \mathbb{R}^{\sqcup_{y \in Y} J^{(y)}} \,\middle|\, \forall y \in Y.\ \sum_{j \in J^{(y)}} x_j^{(y)} = 1 \right\}$$

Note that the system of equations presented for this last example was already in RREF, without zero rows. Polytopes of conditional distributions are the fundamental building blocks allowing us to define a general theory of polytopes of conditional distributions constrained by equalities, which is the aim of the next section.

5.2 Constrained conditional probability distributions

Definition 5.11. *Let $K \subset \mathbb{R}^J$ be a polytope. We say that a polytope $K' \subset \mathbb{R}^J$ is obtained by* slicing *from K if it takes the form $K' = K \cap W$ for some affine subspace $W \subset \mathbb{R}^J$. If we wish to specify the subspace, we say that K' is obtained by* slicing K with W. *We adopt the following notation for it:*

$$Slice_W(K) := K \cap W \tag{5.3}$$

Proposition 5.11. *Let $K \subset \mathbb{R}^J$ be a polytope, defined by a system of equations $A\underline{x} = \underline{b}$ and inequalities $C\underline{x} \leq \underline{d}$. Let $W \subset \mathbb{R}^J$ be an affine subspace, defined by a system of equations $A'\underline{x} = \underline{b'}$. Then $Slice_W(K) \subset \mathbb{R}^J$ is a polytope, defined by the equations and inequalities for K together with the equations for W:*

$$\left(\frac{A}{A'}\right)\underline{x} = \left(\frac{b}{b'}\right) \qquad\qquad C\underline{x} \leq \underline{d}$$

Proof. We have that $\underline{x}$ lies in the affine subspace W if and only if $A'\underline{x} = \underline{b'}$. Therefore $\underline{x} \in K \cap W$ if and only if $A\underline{x} = \underline{b}$, $A'\underline{x} = \underline{b'}$ and $C\underline{x} \leq \underline{d}$. □

Proposition 5.11. *Let $K \subset \mathbb{R}^J$ be a polytope and let $V, W \subset \mathbb{R}^J$ be affine subspaces. Slicing $Slice_W(K)$ with V is the same as slicing K with $W \cap V$:*

$$Slice_V(Slice_W(K)) = Slice_{V \cap W}(K)$$

We say that 'slicing is closed under iteration'.

Proof. This is simply associativity of intersection:

$$\text{Slice}_V(\text{Slice}_W(K)) = V \cap (W \cap K) = (V \cap W) \cap K = \text{Slice}_{V \cap W}(K)$$

□

We can think about the polytope of conditional probability distribution on some Y to be obtained by slicing the standard hypercube with a family of normalisation equations stating that the distributions conditional to each $y \in Y$ must sum to 1:

Definition 5.12. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. The corresponding* normalisation equations *are defined as follows:*

$$\forall y \in Y. \sum_{j \in J^{(y)}} x_j^{(y)} = 1$$

We write $NormEqs\left(\underline{J}\right)$ for the affine subspace of $\mathbb{R}^{\sqcup_{y \in Y} J^{(y)}}$ defined by the equations.

Even though to construct our 'causaltopes' we will directly start from polytopes of conditional distributions; it is nevertheless instructive to think about polytope of conditional distributions on the parametrised event set $\underline{J} = (J^{(y)})$ as being obtained by slicing the product of the $|J^{(y)}|$ hypercubes by the normalisation equations:

Proposition 5.12. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. The polytope of conditional probability distributions $\prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)$ is obtained by slicing the standard hypercube $[0,1]^{\sqcup_{y \in Y} J^{(y)}} = \prod_{y \in Y} [0,1]^{J^{(y)}}$ with the normalisation equations:*

$$\prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right) = [0,1]^{\sqcup_{y \in Y} J^{(y)}} \cap NormEqs\left(\underline{J}\right)$$

In particular, we have $\mathbb{A}\left(\prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)\right) = NormEqs\left(\underline{J}\right)$.

Proof. Taking the normalisation equations together with the defining inequalities for $[0,1]^{\sqcup_{y \in Y} J^{(y)}} \cap \text{NormEqs}\left(\underline{J}\right)$ yields the following system of equations and inequalities:

$$\forall y \in Y. \sum_{j \in J^{(y)}} x_j^{(y)} = 1 \qquad\qquad \forall y \in Y. \forall j \in J^{(y)}. 0 \le x_j^{(y)} \le 1$$

The normalisation equation $\sum_{j\in J^{(y)}} x_j^{(y)} = 1$ together with the inequalities $0 \le x_j^{(y)}$ for all $j \in J^{(y)}$ implies the inequalities $x_j^{(y)} \le 1$ for all $j \in J^{(y)}$, making them redundant. We are thus left with the defining system of equations and inequalities for the polytope of conditional probability distributions $\prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right)$, as claimed. □

Proposition 5.12 describes the polytope of conditional probability distributions as a slice of the standard hypercube by an affine subspace. The 'causaltopes' we are about to present are obtained by further slicing the polytope of conditional probability distributions by appropriate affine subspaces described by the causality equations. There is another polytope of interest that can be associated with conditional probability distributions, the one obtained by slicing the standard hypercube be the 'quasi-normalisation equations':

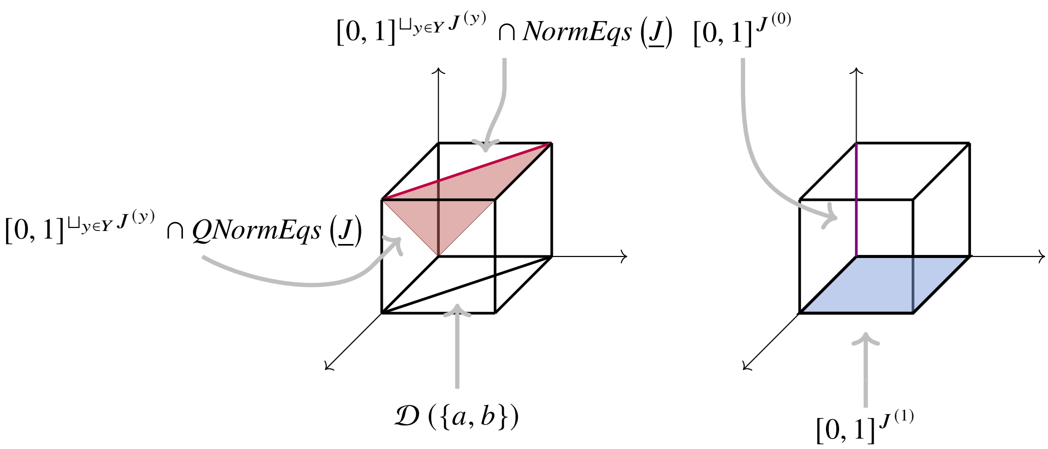
Definition 5.13. *Let Y be a finite non-empty set, with a total order $Y = \{y_1, ..., y_n\}$ fixed on it. Let $\underline{J} = \left(J^{(y)}\right)_{y\in Y}$ be a family of finite non-empty sets. The corresponding* quasi-normalisation equations *are defined as follows:*

$$\forall i \in \{1, ..., n-1\}. \sum_{j\in J^{(y_i)}} x_j^{(y_i)} = \sum_{j\in J^{(y_{i+1})}} x_j^{(y_{i+1})}$$

We write $QNormEqs\left(\underline{J}\right)$ for the affine subspace of $\mathbb{R}^{\sqcup_{y\in Y} J^{(y)}}$ defined by the equations. The polytope of quasi-normalised conditional distributions *is defined by slicing the standard hypercube $[0,1]^{\sqcup_{y\in Y} J^{(y)}}$ with the quasi-normalisation equations:*

$$[0,1]^{\sqcup_{y\in Y} J^{(y)}} \cap QNormEqs\left(\underline{J}\right)$$

Example 5.14. *Consider Example 5.10, the polytope of conditional distributions is described by the red segment of the left picture, while the shaded triangular area gives the quasi normalised conditional distributions. The figure on the right represent the three relevant hypercubes: $[0,1]^{J^{(1)}}$ which shaded in blue and associated to the free assignment of values between $[0,1]$ to a two event set, the one dimensional $[0,1]^{J^{(0)}}$ for a single element set and highlighted in purple, and the full product $[0,1]^{\sqcup_{y\in\{0,1\}} J^{(y)}}$.*



Proposition 5.14. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. The affine subspace QNormEqs $\left(\underline{J}\right)$ is independent of the specific choice of total order for Y, and hence so is the polytope of quasi-normalised conditional distributions.*

Proof. By reflexive-transitive closure, the quasi-normalisation equations are equivalent to the following set of equations, which is independent of the choice of total order on Y

$$\forall y, y' \in Y. \sum_{j \in J^{(y)}} x_j^{(y)} = \sum_{j \in J^{(y')}} x_j^{(y')}$$

□

Polytopes of quasi-normalised conditional distributions possess important features. In particular, any point in such polytope can be uniquely rescaled to represent a point in the associated conditional distribution polytope.

Proposition 5.14. *For each quasi-normalised conditional distribution $\underline{u}$, there exists a unique mass $\left(\underline{u}\right) \in [0, 1]$ and a distribution $\underline{e} \in \prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)$ such that:*

$$\underline{u} = mass\left(\underline{u}\right) \underline{e}$$

We refer to mass $\left(\underline{u}\right)$ as the mass *of the quasi-normalised distribution $\underline{u}$. If mass $\left(\underline{u}\right) > 0$, the distribution $\underline{e}$ is furthermore unique.*

Proof. Let $y_o \in Y$ be any element and define the mass $m := \sum_{j \in J^{(y_0)}} c_j^{(y_0)}$. The quasi-normalisation equations imply that

$$\forall y \in Y. \quad \sum_{j \in J^{(y_i)}} c_j^{(y_i)} = \sum_{j \in J^{(y_0)}} c_j^{(y_0)} = m$$

If $m = 0$ then positivity implies that $\underline{u} = m\underline{e}$ for all conditional distributions $\underline{e} \in \prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)$. If $m > 0$, then the following $\underline{e}$ is a conditional probability distribution:

$$\underline{e} := \frac{1}{m}\underline{u}$$

We have $m\underline{e} = \underline{u}$ by definition, and $m\underline{e}' = \underline{u}m\underline{e}$ implies $\underline{e} = \underline{e}'$ because $m \neq 0$. Setting mass $\left(\underline{u}\right) := m$ completes the proof. □

The 'causaltopes' defined in this work are obtained by slicing polytopes of conditional probability distributions with linear subspaces. Because slicing is the same as applying constraints, we refer to these as 'constrained' conditional probability distributions. We proceed by proving some useful results about constrained probability distributions.

Definition 5.15. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y\in Y}$ be a family of finite non-empty sets. A polytope of* constrained *conditional probability distributions is one in the following form, for some linear subspace $W \subseteq \mathbb{R}^{\sqcup_{y\in Y} J^{(y)}}$:*

$$CCPD\left(W, \underline{J}\right) := Slice_W\left(\prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right)\right) \subseteq \prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right) \tag{5.4}$$

The corresponding polytope of constrained quasi-normalised conditional probability distributions takes the following form:

$$\begin{aligned} CCPD_{QNorm}\left(W, \underline{J}\right) &:= Slice_W\left([0,1]^{\sqcup_{y\in Y} J^{(y)}} \cap QNormEqs\left(\underline{J}\right)\right) \\ &= Slice_{W\cap QNormEqs(\underline{J})}\left([0,1]^{\sqcup_{y\in Y} J^{(y)}}\right) \end{aligned} \tag{5.5}$$

Polytopes of constrained (quasi-normalised) conditional probability distributions have the same inclusion hierarchy as the 'minimal' linear subspaces that define them:

Proposition 5.15. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y\in Y}$ be a family of finite non-empty sets. Let $CCPD\left(V, \underline{J}\right)$ and $CCPD\left(U, \underline{J}\right)$ be polytopes of constrained conditional probability distributions. Write $\langle CCPD\left(V, \underline{J}\right)\rangle \subseteq V$ and $\langle CCPD\left(U, \underline{J}\right)\rangle \subseteq U$ for the linear subspaces spanned by linear combinations of vectors in $CCPD\left(V, \underline{J}\right)$ and $CCPD\left(U, \underline{J}\right)$ respectively. The following statements hold:*

1. *if $V \subset U$ then $CCPD\left(V, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$*
2. *$CCPD\left(\langle CCPD\left(V, \underline{J}\right)\rangle, \underline{J}\right) = CCPD\left(V, \underline{J}\right)$*
3. *if $CCPD_{QNorm}\left(V, \underline{J}\right) \subseteq CCPD_{QNorm}\left(U, \underline{J}\right)$ then $CCPD\left(V, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$*
4. *if $CCPD\left(V, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$ then $\langle CCPD\left(V, \underline{J}\right)\rangle \subseteq \langle CCPD\left(V, \underline{J}\right)\rangle$*
5. *if $CCPD\left(V, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$ then $CCPD_{QNorm}\left(V, \underline{J}\right) \subseteq CCPD_{QNorm}\left(U, \underline{J}\right)$*

Proof. We prove point by point:

1. If $V \subseteq U$ then:
$$V \cap \prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right) \subseteq U \cap \prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right)$$
That is, $\mathrm{CCPD}\left(V, \underline{J}\right) \subseteq \mathrm{CCPD}\left(U, \underline{J}\right)$.

2. The previous point, together with $\langle \mathrm{CCPD}\left(V, \underline{J}\right)\rangle \subseteq V$, proves that:
$$\mathrm{CCPD}\left(\langle \mathrm{CCPD}\left(V, \underline{J}\right)\rangle, \underline{J}\right) \subseteq \mathrm{CCPD}\left(V, \underline{J}\right)$$
The equality then follows from the observation that $\mathrm{CCPD}\left(V, \underline{J}\right) \subseteq \langle \mathrm{CCPD}\left(V, \underline{J}\right)\rangle$.

3. If $\mathrm{CCPD}_{\mathrm{QNorm}}\left(V, \underline{J}\right) \subseteq \mathrm{CCPD}_{\mathrm{QNorm}}\left(U, \underline{J}\right)$, then

$$\mathrm{CCPD}_{\mathrm{QNorm}}\left(V, \underline{J}\right) \cap \mathrm{NormEqs}\left(\underline{J}\right) \subseteq \mathrm{CCPD}_{\mathrm{QNorm}}\left(U, \underline{J}\right) \cap \mathrm{NormEqs}\left(\underline{J}\right)$$

   That is, again, $\mathrm{CCPD}\left(V, \underline{J}\right) \subseteq \mathrm{CCPD}\left(U, \underline{J}\right)$.

4. If $\mathrm{CCPD}\left(V, \underline{J}\right) \subseteq \mathrm{CCPD}\left(U, \underline{J}\right)$ then:

$$\left\langle \mathrm{CCPD}\left(V, \underline{J}\right)\right\rangle \subseteq \left\langle \mathrm{CCPD}\left(U, \underline{J}\right)\right\rangle$$

5. Hence, if $\mathrm{CCPD}\left(V, \underline{J}\right) \subseteq \mathrm{CCPD}\left(U, \underline{J}\right)$ then:

$$\mathrm{CCPD}_{\mathrm{QNorm}}\left(\left\langle \mathrm{CCPD}\left(V, \underline{J}\right)\right\rangle, \underline{J}\right) \subseteq \mathrm{CCPD}_{\mathrm{QNorm}}\left(\left\langle \mathrm{CCPD}\left(U, \underline{J}\right)\right\rangle, \underline{J}\right)$$

   That is, $\mathrm{CCPD}_{\mathrm{QNorm}}\left(V, \underline{J}\right) \subseteq \mathrm{CCPD}_{\mathrm{QNorm}}\left(U, \underline{J}\right)$.

□

Polytopes of constrained conditional probability distributions are closed under meet (i.e. under intersection).

Proposition 5.15. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. Let $CCPD\left(V, \underline{J}\right)$ and $CCPD\left(U, \underline{J}\right)$ be polytopes of constrained conditional probability distributions. Then:*

$$CCPD\left(V, \underline{J}\right) \cap CCPD\left(U, \underline{J}\right) = CCPD\left(V \cap U, \underline{J}\right)$$

Proof.

$$\begin{aligned} & \mathrm{CCPD}\left(V, \underline{J}\right) \cap \mathrm{CCPD}\left(U, \underline{J}\right) \\ = & \left(V \cap \prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)\right) \cap \left(U \cap \prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)\right) \\ = & \left(V \cap U \cap \prod_{y \in Y} \mathcal{D}\left(J^{(y)}\right)\right) = \mathrm{CCPD}\left(V \cap U, \underline{J}\right) \end{aligned}$$

□

As expected, constrained conditional probability distributions can be recovered from constrained quasi-normalised conditional probability distributions by applying the normalisation equations.

Proposition 5.15. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. For every linear subspace $W \subseteq \mathbb{R}^{\sqcup_{y \in Y} J^{(y)}}$, we always have:*

$$CCPD\left(W, \underline{J}\right) = Slice_{NormEqs\left(\underline{J}\right)}\left(CCPD_{QNorm}\left(W, \underline{J}\right)\right)$$

For every $y \in Y$, we define a normalisation equation for the distribution conditional to y:

$$\mathit{NormEqs}\left(\underline{J}\right)^{(y)} := \left\{ \underline{x} \in \mathbb{R}^{\sqcup_{y\in Y} J^{(y)}} \,\middle|\, \sum_{j\in J^{(y)}} x_j^{(y)} = 1 \right\}$$

Then for any individual choice of $y \in Y$ we also have:

$$\mathit{CCPD}\left(W,\underline{J}\right) = \mathit{Slice}_{\mathit{NormEqs}\left(\underline{J}\right)^{(y)}}\left(\mathit{CCPD}_{\mathit{QNorm}}\left(W,\underline{J}\right)\right)$$

Proof. Both claims follow from the closure of slicing under iteration. For the first claim, we use the following observation:

$$\mathrm{QNormEqs}\left(\underline{J}\right) \cap \mathrm{NormEqs}\left(\underline{J}\right) = \mathrm{NormEqs}\left(\underline{J}\right)$$

Now observe the $\sum_{j\in J^{(y)}} x_j^{(y)} = 1$ together with $\sum_{j\in J^{(y)}} x_j^{(y)} - \sum_{j\in J^{(y')}} x_j^{(y')} = 0$ implies $\sum_{j\in J^{(y')}} x_j^{(y')} = 1$, for every $y' \in Y$. Hence we get the following, which in turn proves the second claim:

$$\mathrm{QNormEqs}\left(\underline{J}\right) \cap \mathrm{NormEqs}\left(\underline{J}\right)^{(y)} = \mathrm{NormEqs}\left(\underline{J}\right)$$

□

Given two nested polytopes $\mathrm{CCPD}\left(V,\underline{J}\right) \subset \mathrm{CCPD}\left(U,\underline{J}\right)$ of constrained conditional probability distributions, a key task in our work will be to find the largest 'fraction' of a distribution $\underline{u} \in \mathrm{CCPD}\left(U,\underline{J}\right)$ which is 'supported' by $\mathrm{CCPD}\left(V,\underline{J}\right)$, i.e. to find a decomposition $\underline{u} = \underline{v} + \underline{w}$ where $\underline{v} \in \mathrm{CCPD}_{\mathrm{QNorm}}\left(V,\underline{J}\right)$, $\underline{w} \in \mathrm{CCPD}_{\mathrm{QNorm}}\left(U,\underline{J}\right)$ and the mass of $\underline{v}$ is as large as possible. The following result—a consequence of our polytopes being defined by linear constraints—significantly simplifies this task by removing the need to explicitly enforce $\underline{w} \in \mathrm{CCPD}_{\mathrm{QNorm}}\left(U,\underline{J}\right)$ in our linear programs.

Proposition 5.15. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y\in Y}$ be a family of finite non-empty sets. Let $\mathit{CCPD}\left(V,\underline{J}\right) \subset \mathit{CCPD}\left(U,\underline{J}\right)$ be two nested polytopes of constrained conditional probability distributions and let $\underline{u} \in \mathit{CCPD}\left(U,\underline{J}\right)$. If $\underline{v} \in \mathit{CCPD}\left(V,\underline{J}\right)$ is such that $\underline{v} \leq \underline{u}$, then necessarily:*

$$\underline{u} - \underline{v} \in \mathit{CCPD}\left(U,\underline{J}\right)$$

Proof. Because $\underline{u}, \underline{v} \in \prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right)$ and $\underline{v} \leq \underline{u}$, we have that the difference $\underline{u} - \underline{v} \in \prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right)$ is itself a conditional probability distribution. Because $\underline{u} \in \mathrm{CCPD}\left(U,\underline{J}\right)$ and $\underline{v} \in \mathrm{CCPD}\left(V,\underline{J}\right) \subseteq \mathrm{CCPD}\left(U,\underline{J}\right)$, then $\underline{u}, \underline{v} \in U$ and hence the difference $\underline{u} - \underline{v}$ satisfies the constraints imposed by U. We conclude that:

$$\underline{u} - \underline{v} \in \prod_{y\in Y} \mathcal{D}\left(J^{(y)}\right) \cap U = \mathrm{CCPD}\left(U,\underline{J}\right)$$

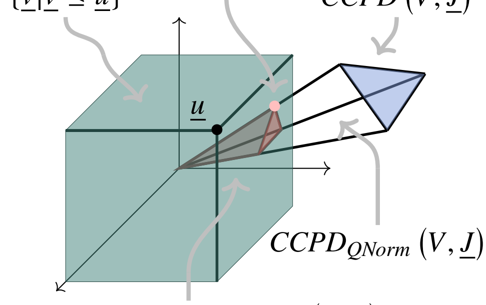
□

For a vector laying in a polytope of constrained conditional probability distributions, it therefore makes sense to find components supported by smaller CCPDs. In particular, we can define the notion of component, maximal component and the associated fractions, which are well defined by Proposition 5.14.

Definition 5.16. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. Let $\underline{u} \in CCPD\left(U, \underline{J}\right)$ be a constrained conditional probability distribution. For any sub-polytope $CCPD\left(V, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$, we give the following definitions:*

- *A* component *of $\underline{u}$ in $CCPD\left(V, \underline{J}\right)$ is any $\underline{v} \in CCPD_{QNorm}\left(V, \underline{J}\right)$ such that $\underline{v} \leq \underline{u}$.*
- *A* maximal *component of $\underline{u}$ in $CCPD\left(V, \underline{J}\right)$ is one of maximal mass.*
- *The* supported fraction *of $\underline{u}$ in $CCPD\left(V, \underline{J}\right)$ is the mass of a maximal component of $\underline{u}$ in $CCPD\left(V, \underline{J}\right)$.*

Colloquially, we say that $\underline{u}$ is X% supported by $CCPD\left(V, \underline{J}\right)$ to mean that the supported fraction of $\underline{u}$ in $CCPD\left(V, \underline{J}\right)$ is $\frac{X}{100}$. The following is a schematic representation of the notion of components and maximal components. Note that in particular we have not depicted the $CCPD\left(U, \underline{J}\right)$ containing the smaller $CCPD\left(V, \underline{J}\right)$. The green box represents the cone of vectors which are smaller than $\underline{u}$. The pink shaded area is the set of all components.



The description of polytopes of constrained conditional probability distributions in terms of slicing of conditional probability distributions allows to describe the maximal component supported by a sub-CCPD by means of linear programming:

Observation 5.16. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. Let $\underline{u} \in CCPD\left(U, \underline{J}\right)$ be a constrained conditional probability distribution and let $CCPD\left(V, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$ be a sub-polytope. Let V be defined explicitly by a system of linear equations:*

$$V = \left\{ \underline{x} \in \mathbb{R}^{\sqcup_{y \in Y} J^{(y)}} \,\middle|\, A\underline{x} = \underline{0} \right\}$$

The maximal components $\underline{v}$ *of* $\underline{u}$ *in CCPD* $\left(V, \underline{J}\right)$ *are the solutions to the following linear program (LP):*

$$\begin{array}{rl} \textit{maximise} & \textit{mass}\left(\underline{v}\right) \\ \textit{subject to:} & \underline{v} \in \textit{QNormEqs}\left(\underline{J}\right) \\ & \underline{v} \in V \\ & \underline{v} \geq 0 \\ & \underline{v} \leq \underline{u} \end{array} \tag{5.6}$$

The following form makes the mass and quasi-normalisation equations explicit for any choice of total order $\{y_1, ..., y_n\}$ *on* Y*:*

$$\begin{array}{rl} \textit{maximise} & \sum_{j \in J^{(y_1)}} v_j^{(y_1)} \\ \textit{subject to:} & \forall i \in \{1, ..., n-1\}. \ \sum_{j \in J^{(y_i)}} v_j^{(y_i)} = \sum_{j \in J^{(y_{i+1})}} v_j^{(y_{i+1})} \\ & A\underline{v} = \underline{0} \\ & \underline{v} \geq 0 \\ & \underline{v} \leq \underline{u} \end{array} \tag{5.7}$$

More generally, we will be interested in finding the largest fraction of a (constrained) conditional probability distribution supported 'jointly' by multiple sub-polytopes. This is the same as being supported by the convex hull of the sub-polytopes, but with the caveat that the convex hull of constrained conditional probability distributions need not be a polytope of constrained conditional probability distributions. In particular, we have no way to apply Definition 5.16 or Observation 5.16 to such a convex hull.

Definition 5.17. *Let* Y *be a finite non-empty set and let* $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ *be a family of finite non-empty sets. Let* $\underline{u} \in$ *CCPD* $\left(U, \underline{J}\right)$ *be a constrained conditional probability distribution. For any family* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ *of sub-polytopes CCPD* $\left(V^{(z)}, \underline{J}\right) \subseteq$ *CCPD* $\left(U, \underline{J}\right)$*, we give the following definitions:*

- *A* decomposition *of* $\underline{u}$ *in* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ *is any family* $\left(\underline{v}^{(z)}\right)_{z \in Z}$ *of distributions* $\underline{v}^{(z)} \in \textit{CCPD}_{QNorm}\left(V^{(z)}, \underline{J}\right)$*, the* components*, such that* $\sum_{z \in Z} \underline{v}^{(z)} \leq \underline{u}$*.*

- *The* mass *of a decomposition* $\left(\underline{v}^{(z)}\right)_{z \in Z}$ *is the sum of the masses of the individual components:*

$$\textit{mass}\left(\left(\underline{v}^{(z)}\right)_{z \in Z}\right) := \sum_{z \in Z} \textit{mass}\left(\underline{v}^{(z)}\right)$$

- *A* maximal decomposition *of* $\underline{u}$ *in* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ *is one of maximal mass.*

- *The* supported fraction *of* $\underline{u}$ *in* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ *is the mass of a maximal decomposition of* $\underline{u}$ *in* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$*.*

Colloquially, we say that $\underline{u}$ *is* X% *supported by* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ *to mean that the supported fraction of* $\underline{u}$ *in* $\left(\textit{CCPD}\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ *is* $\frac{X}{100}$*.*

By iterating the result of Proposition 5.15 we can obtain the following corollary:

Corollary 5.18. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. Let $CCPD\left(U, \underline{J}\right)$ be a polytope of constrained conditional probability distributions and let $\left(CCPD\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ be a family of sub-polytopes $CCPD\left(V^{(z)}, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$. Let $\underline{u} \in CCPD\left(U, \underline{J}\right)$ be a constrained conditional probability distribution. If $\left(\underline{v}^{(z)}\right)_{z \in Z}$ is a decomposition of $\underline{u}$ in $\left(CCPD\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$, then necessarily:*

$$\underline{u} - \sum_{z \in Z} \underline{v}^{(z)} \in CCPD\left(U, \underline{J}\right)$$

Proof. This follows by iterating Proposition 5.15 for each component in the decomposition. □

The above corollary allows to define general decompositions of $\underline{u}$ with respect to sub-CCPDs in the form of a linear program:

Observation 5.18. *Let Y be a finite non-empty set and let $\underline{J} = \left(J^{(y)}\right)_{y \in Y}$ be a family of finite non-empty sets. Let $\underline{u} \in CCPD\left(U, \underline{J}\right)$ be a constrained conditional probability distribution and let $\left(CCPD\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ be a family of sub-polytopes $CCPD\left(V^{(z)}, \underline{J}\right) \subseteq CCPD\left(U, \underline{J}\right)$. Let each $V^{(z)}$ be defined explicitly by a system of linear equations:*

$$V^{(z)} = \left\{ \underline{x} \in \mathbb{R}^{\sqcup_{y \in Y} J^{(y)}} \,\middle|\, A^{(z)} \underline{x} = \underline{0} \right\}$$

The maximal components $\left(\underline{v}^{(z)}\right)_{z \in Z}$ of $\underline{u}$ in $\left(CCPD\left(V^{(z)}, \underline{J}\right)\right)_{z \in Z}$ are the solutions to the following linear program (LP):

$$\begin{array}{rl} \textit{maximise} & mass\left(\left(\underline{v}^{(z)}\right)_{z \in Z}\right) \\ \textit{subject to:} & \forall z \in Z.\ \underline{v}^{(z)} \in QNormEqs\left(\underline{J}\right) \\ & \forall z \in Z.\ \underline{v}^{(z)} \in V^{(z)} \\ & \underline{v}^{(z)} \geq 0 \\ & \sum_{z \in Z} \underline{v}^{(z)} \leq \underline{u} \end{array} \tag{5.8}$$

Making the mass and linear systems explicit, we get:

$$\begin{array}{rl} \textit{maximise} & \sum_{z \in Z} mass\left(\underline{v}^{(z)}\right) \\ \textit{subject to:} & \forall z \in Z.\ \underline{v}^{(z)} \in QNormEqs\left(\underline{J}\right) \\ & \forall z \in Z.\ A^{(z)} \underline{v}^{(z)} = 0 \\ & \underline{v}^{(z)} \geq 0 \\ & \sum_{z \in Z} \underline{v}^{(z)} \leq \underline{u} \end{array} \tag{5.9}$$

5.3 Causaltopes

This section shows how to think about sets of empirical models for a given cover as polytopes of constrained conditional probability distributions. Recall from Chapter 4 that an empirical model e on a cover C is a compatible family $(e_\lambda)_{\lambda \in C}$ for the presheaf of causal distributions $\text{CausDist}\left(\Theta, \underline{O}\right)$:

$$e_\lambda \in \text{CausDist}\left(\lambda, \underline{O}\right) = \mathcal{D}\left(\text{ExtCausFun}\left(\lambda, \underline{O}\right)\right)$$

The empirical model assigns a probability distribution on the extended causal functions to each context. As such, empirical models on a given cover inherit the convex structure of the individual set of distributions by taking context-wise convex combinations:

$$(x \cdot e + (1-x) \cdot e')_\lambda := x \cdot e_\lambda + (1-x) \cdot e'_\lambda$$

where $e, e' \in \text{EmpMod}\left(C, \underline{O}\right), x \in [0,1]$ and $\lambda \in C$ ranges over the contexts specified by the cover C.

The cover can always be extracted from the empirical model as $C = \text{dom}\,(e)$, so it is not strictly necessary to explicitly state 'on C' when talking about an empirical model e. Similarly, the space itself can be recovered from the cover as the union $\bigcup C$ of the open sets. As such, these objects remember the structure of the space, which makes them different from conditional probability distributions. We were formally imprecise when referring to these empirical models as points in some polytope. For the case of empirical models on the 'standard cover' we show that there exists a bijection between $\text{ExtCausFun}\left(k\downarrow, \underline{O}\right)$ and the space of conditional probability distributions $\mathcal{D}\left(\prod_{\omega \in \text{dom}(k\downarrow)} O_\omega\right)$. Later we will generalise this observation to general covers.

Theorem 5.19. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. For any $k \in Ext\,(\Theta)$, the following function is a bijection:*

$$\begin{array}{ccc} \mathit{ExtCausFun}\left(k\downarrow, \underline{O}\right) & \longrightarrow & \prod\limits_{\omega \in \text{dom}(k)} O_\omega \\ \mathit{Ext}\,(f) & \mapsto & \mathit{Ext}\,(f)\,(k) \end{array} \tag{5.10}$$

As a consequence, the following function is a convex-linear bijection:

$$\begin{array}{ccc} \mathit{CausDist}\left(k\downarrow, \underline{O}\right) & \longrightarrow & \mathcal{D}\left(\prod\limits_{\omega \in \text{dom}(k)} O_\omega\right) \\ d & \mapsto & \lceil d \rceil := \sum\limits_{\mathit{Ext}(f)} d(\mathit{Ext}\,(f))\delta_{\mathit{Ext}(f)(k)} \end{array} \tag{5.11}$$

We refer to $\lceil d \rceil$ as the top-element distribution *for $d \in \mathit{CausDist}\left(k\downarrow, \underline{O}\right)$. We furthermore adopt the following notation for its inverse:*

$$\begin{array}{ccc} \mathit{CausDist}\left(k\downarrow, \underline{O}\right) & \longleftarrow & \mathcal{D}\left(\prod\limits_{\omega \in \text{dom}(k)} O_\omega\right) \\ \lfloor p \rfloor_k & \leftarrow\!\shortmid & p \end{array} \tag{5.12}$$

Proof. We appeal to the consistency condition to infer that for every $h \in k\downarrow$ we have that $\text{Ext}\,(f)\,(h) = \text{Ext}\,(f)\,(k)|_{\text{dom}(h)}$ for all $h \in k\downarrow$, so the function $\text{Ext}\,(f) \mapsto \text{Ext}\,(f)\,(k)$ is injective. We can also observe surjectivity by observing that for a given $o \in \prod_{\omega \in \mathrm{dom}(k)} O_\omega$, setting $f(h) := o|_{\mathrm{dom}(h)}$ defines a causal function (Definition 4.20) since we have that $f(h)_\omega = o_\omega = f(h')_\omega$ whenever $\omega \in \mathrm{dom}(h) \cap \mathrm{dom}(h')$. We have that $\mathrm{Ext}(f)(k) = o$ and therefore $\mathrm{Ext}(f) \mapsto \mathrm{Ext}(f)(k)$ is surjective.

Since $\mathrm{Ext}(f) \mapsto \mathrm{Ext}(f)(k)$ is a bijection, then application of the distribution monad induces a convex-linear bijection:

$$\mathcal{D}\left(\mathrm{Ext}(f) \mapsto \mathrm{Ext}(f)(k)\right) = d \mapsto \sum_{\mathrm{Ext}(f)} d(\mathrm{Ext}(f))\delta_{\mathrm{Ext}(f)(k)} = \lceil d \rceil$$

□

If we want to extend Theorem 5.19 to arbitrary contexts, we need to deal with the fact that a lowerset for a space of input histories might contain incompatible histories. Recall from the previous chapter that $\sim_\omega$ is an equivalence relation, and let us denote the set of all such equivalence classes by $\mathrm{TipEq}_\omega(\lambda)$:

$$\mathrm{TipEq}_\lambda(\omega) := \left\{ [h]_{\sim_\omega} \,\middle|\, h \in \mathrm{TipHists}_\lambda(\omega) \right\}$$

We have already shown that λ is tight precisely when all equivalence classes contain a single history. When $\lambda = k\downarrow$, then there is a unique equivalence class associated to every $\omega \in \mathrm{dom}(k)$. This is because any two histories are consistent. We generalise Theorem 5.19 by requiring the output to be independently defined for every equivalence class:

Theorem 5.20. *Let Θ be a space of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. For any $\lambda \in \Lambda(\Theta)$, the following function is a bijection:*

$$\begin{array}{rcl} \mathit{ExtCausFun}\left(\lambda, \underline{O}\right) & \longrightarrow & \prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{\mathit{TipEq}_\lambda(\omega)} \\ \mathit{Ext}(f) & \mapsto & \lceil \mathit{Ext}(f) \rceil \end{array} \tag{5.13}$$

where took $\mathrm{dom}(\lambda) := \bigcup_{h \in \lambda} \mathrm{dom}(h)$ *and we defined:*

$$\lceil \mathit{Ext}(f) \rceil := \omega \mapsto (\mathit{Ext}(f)(h)_\omega)_{[h]_{\sim_\omega} \in \mathit{TipEq}_\lambda(\omega)} \tag{5.14}$$

As a consequence, the following function is a convex-linear bijection:

$$\begin{array}{rcl} \mathit{CausDist}\left(\lambda, \underline{O}\right) & \longrightarrow & \mathcal{D}\left(\prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{\mathit{TipEq}_\lambda(\omega)}\right) \\ d & \mapsto & \lceil d \rceil := \sum_{\mathit{Ext}(f)} d(\mathit{Ext}(f))\delta_{\lceil \mathit{Ext}(f) \rceil} \end{array} \tag{5.15}$$

We furthermore adopt the following notation for its inverse:

$$\begin{array}{rcl} \mathit{CausDist}\left(\lambda, \underline{O}\right) & \longleftarrow & \mathcal{D}\left(\prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{\mathit{TipEq}_\lambda(\omega)}\right) \\ \lfloor p \rfloor_\lambda & \mapsfrom & p \end{array} \tag{5.16}$$

Proof. By definition, $\mathrm{Ext}(f)(h')_\omega$ takes a constant value for all $h' \in [h]_{\sim_\omega}$, making $\lceil \mathrm{Ext}(f) \rceil$ well-defined. The correspondence is bijective (because extended causal functions are in bijection with causal functions) and analogously to Theorem 5.19, the bijection lifts to a bijection between the corresponding spaces of distributions: the latter bijection is defined by taking convex-linear combinations, and hence it is a convex-linear function. □

For example, consider the space given by a single event $\Omega = \mathrm{indiscrete}(A)$, such that $I_A = \{0, 1\}$. The space of histories is then $\Theta = \{\{\mathrm{A}:0\}, \{\mathrm{A}:1\}\}$. We select the classical cover $C = \{\Theta\}$, in which the two histories are incompatible. Then, Theorem 5.20 shows that the ExtCausFun $(\Theta, \underline{O})$ are in bijective correspondence with $\{0, 1\}^2$: a choice of joint outcome for the two incomparable histories, as expected.

If we apply the bijections to every element of a cover we describe a conditional probability space providing the ambient space for our correlations. This procedure applies the causality constraints inside every context but does not impose them across the contextual data. This describes the space of 'pseudo-empirical models'.

Definition 5.21. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $C \in Covers(\Theta)$ be any cover. The polytope of* pseudo-empirical models *on C is defined to be the following polytope of conditional probability distributions:*

$$PEmpMods(C, \underline{O}) := \prod_{\lambda \in C} \mathcal{D}\left(\prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{TipEq_\lambda(\omega)} \right) \tag{5.17}$$

We adopt the following shorthand for the embedding vector space, which is spanned by all linear combinations of the pseudo-empirical models:

$$\left\langle PEmpMods(C, \underline{O}) \right\rangle := \mathbb{R}^{\bigsqcup_{\lambda \in C} \prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{TipEq_\lambda(\omega)}} \tag{5.18}$$

The passage from pseudo-empirical models to empirical models is done by providing the description of the relevant constraints, i.e the 'causality equations', which guarantee that the marginals of the probability distributions associated with various lowersets in the cover agree on their common sub-contexts. The intuition is simple: we need to impose as 'slicing equalities' the requirement that marginalising two elements of a cover to a common sub-context always yields the same result.

To do so, we first need to define the marginalisation as a map between the probability distributions associated with each context. Pseudo-empirical models are conditional probability distributions, and hence we adopt the notation from the previous subsection to describe them:

$$\underline{u} = \left(\underline{u}^{(\lambda)}\right)_{\lambda \in C}$$

For a given λ, the components $u_o^{(\lambda)}$ are indexed by functions/families $o \in \prod_{\omega \in \text{dom}(\lambda)} (O_\omega)^{\text{TipEq}_\lambda(\omega)}$, the components of which are in turn indexed as follows:

$$o = \left(o_{\omega,[h]_{\sim_\omega}}\right)_{\omega \in \text{dom}(\lambda),\ [h]_{\sim_\omega} \in \text{TipEq}_\lambda(\omega)}$$

Observation 5.21. *Let Θ be a space of input histories and let $\lambda, \mu \in \Lambda(\Theta)$ be two lowersets. If $\mu \subseteq \lambda$, then the following is a well-defined injection:*

$$\begin{array}{rccc} \mathit{TipEq}_{\mu \subseteq \lambda}(\omega): & \mathit{TipEq}_\mu(\omega) & \to & \mathit{TipEq}_\lambda(\omega) \\ & [h]_{\sim_\omega} & \mapsto & [h]_{\sim_\omega} \end{array}$$

Definition 5.22. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $C \in \mathit{Covers}(\Theta)$ be any cover. For every $\lambda \in C$ and every $\mu \in \Lambda(\Theta)$ such that $\mu \subseteq \lambda$, the* output history restriction *from λ to μ is defined as follows:*

$$\begin{array}{rccc} \rho_{\lambda,\mu}: & \prod_{\omega \in \text{dom}(\lambda)} (O_\omega)^{\mathit{TipEq}_\lambda(\omega)} & \longrightarrow & \prod_{\omega \in \text{dom}(\mu)} (O_\omega)^{\mathit{TipEq}_\mu(\omega)} \\ & o & \mapsto & \left((\omega, [h]_{\sim_\omega}) \mapsto o_{\omega,[h]_{\sim_\omega}}\right) \end{array} \tag{5.19}$$

Formally, $o_{\omega,[h]_{\sim_\omega}}$ stands for $o_{\omega,\mathit{TipEq}_{\mu\subseteq\lambda}(\omega)([h]_{\sim_\omega})}$. The restriction extends convex-linearly to a output history distribution restriction *between the corresponding spaces of probability distributions:*

$$\begin{array}{rccc} \mathcal{D}(\rho_{\lambda,\mu}): & \mathcal{D}\left(\prod_{\omega \in \text{dom}(\lambda)} (O_\omega)^{\mathit{TipEq}_\lambda(\omega)}\right) & \longrightarrow & \mathcal{D}\left(\prod_{\omega \in \text{dom}(\mu)} (O_\omega)^{\mathit{TipEq}_\mu(\omega)}\right) \\ & d & \mapsto & \left(o \mapsto \sum_{o' \text{ s.t. } \rho_{\lambda,\mu}(o')=o} d(o')\right) \end{array} \tag{5.20}$$

The causality equations are given by the affine subspaces in which the restriction to a common lowerset $\mu \subseteq \lambda, \lambda'$ agree. Providing all causality equations to describe the causaltopes is highly redundant and we will show later on how in certain cases this description can be simplified.

Definition 5.23. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $C \in \mathit{Covers}(\Theta)$ be any cover. The* causality equations *are indexed by all $\mu \in \Lambda(\Theta)$ and all $\lambda, \lambda' \in C$ such that $\mu \subseteq \lambda$ and $\mu \subseteq \lambda'$. For one such triple μ, λ, λ', we equate the output history distribution restrictions from λ to μ and from λ' to μ:*

$$\mathit{CausEqs}(C, \underline{O})_{\mu,\lambda,\lambda'} := \left\{ \underline{u} \in \left\langle \mathit{PEmpMods}(C, \underline{O}) \right\rangle \,\middle|\, \underline{u}^{(\lambda)}\Big|_\mu = \underline{u}^{(\lambda')}\Big|_\mu \right\} \tag{5.21}$$

where we have adopted the following shorthand for the restriction:

$$\underline{u}^{(\lambda)}\Big|_\mu := \mathcal{D}(\rho_{\lambda,\mu})\left(\underline{u}^{(\lambda)}\right) \tag{5.22}$$

We write $\mathit{CausEqs}(C, \underline{O})$ for the linear subspace of $\left\langle \mathit{PEmpMods}(C, \underline{O}) \right\rangle$ spanned jointly by all causality equations:

$$\mathit{CausEqs}(C, \underline{O}) := \bigcap_{\mu \in \Lambda(\Theta)} \bigcap_{\lambda \in C \cap \mu\uparrow} \bigcap_{\lambda' \in C \cap \mu\uparrow} \mathit{CausEqs}(C, \underline{O})_{\mu,\lambda,\lambda'} \tag{5.23}$$

where $\mu\uparrow$ is the upset of μ in the partial order $\Lambda(\Theta)$ formed by lowersets under inclusion.

Proposition 5.23. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $C \in Covers(\Theta)$ be any cover. For every $\mu \in \Lambda(\Theta)$, let $\lambda_{\mu,1}, ..., \lambda_{\mu,n_\mu}$ be a total order on the $\lambda \in C$ such that $\mu \subseteq \lambda$. Then we have:*

$$CausEqs\left(C, \underline{O}\right) = \bigcap_{\mu \in \Lambda(\Theta), n_\mu \geq 1} \bigcap_{i=1}^{n_\mu - 1} CausEqs\left(C, \underline{O}\right)_{\mu, \lambda_{\mu,i}, \lambda_{\mu,i+1}} \tag{5.24}$$

Proof. Fix $\mu \in \Lambda(\Theta)$. If $n_\mu = 0$, then there are no equations associated with μ, so we can restrict our attention to the μ s.t. $n_\mu \geq 1$. Consider the following subspace:

$$\bigcap_{\lambda \in C \cap \mu\uparrow} \bigcap_{\lambda' \in C \cap \mu\uparrow} \text{CausEqs}\left(C, \underline{O}\right)_{\mu, \lambda, \lambda'}$$

The linear constraints are exactly those enforcing $\underline{u}^{(\lambda)}\big|_\mu = \underline{u}^{(\lambda')}\big|_\mu$ for all $\lambda, \lambda' \in C \cap \mu\uparrow$. If we impose a total order on $C \cap \mu\uparrow$, the exact same constraints can be enforced by a chain of $n_\mu - 1$ equations, as follows:

$$\bigcap_{i=1}^{n_\mu - 1} \text{CausEqs}\left(C, \underline{O}\right)_{\mu, \lambda_{\mu,i}, \lambda_{\mu,i+1}}$$

This concludes our proof. □

Proposition 5.23. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $StdCov(\Theta) \in Covers(\Theta)$ be the standard cover. For every $h \in Ext(\Theta)$, let $k_{h,1}, ..., k_{h,n_h}$ be a total order on the $k \in Ext(\Theta)$ such that $k \leq h$, i.e. such that $h\downarrow \subseteq k\downarrow$. Then we have:*

$$CausEqs\left(C, \underline{O}\right) = \bigcap_{h \in Ext(\Theta), n_h \geq 1} \bigcap_{i=1}^{n_h - 1} CausEqs\left(C, \underline{O}\right)_{h\downarrow, k_{h,i}\downarrow, k_{h,i+1}\downarrow} \tag{5.25}$$

Proof. We build upon the result of Proposition 5.23. Consider $\mu \in \Lambda(\Theta)$ with $n_\mu \geq 1$ and note that the associated $\lambda_{\mu,i} \in C$ take the form $k_{\mu,i}\downarrow$ for some $k_{\mu,i} \in \text{Ext}(\Theta)$. Consider any $i \in \{1, ..., n_\mu - 1\}$, so that $\mu \subseteq k_{\mu,i}\downarrow \cap k_{\mu,i+1}\downarrow$: because the intersection of a downset is a downset, we have the following, for some $h_{\mu,i} \in \text{Ext}(\Theta)$:

$$\mu \subseteq h_{\mu,i}\downarrow \subseteq k_{\mu,i}\downarrow \cap k_{\mu,i+1}\downarrow$$

Since we included the relevant equations for all such $h_{\mu,i}$, we can infer the equations for μ by composing restrictions:

$$\underline{u}^{(\lambda)}\Big|_\mu = \left(\underline{u}^{(\lambda)}\Big|_{h_{\mu,i}\downarrow}\right)\Bigg|_\mu = \left(\underline{u}^{(\lambda')}\Big|_{h_{\mu,i}\downarrow}\right)\Bigg|_\mu = \underline{u}^{(\lambda')}\Big|_\mu$$

This concludes our proof. □

The definition of the marginalisation allowed us to describe the affine spaces defining the causal equations. With these ingredients, we can finally describe *causaltopes*, a portmanteau of causal polytopes.

Definition 5.24. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $C \in Covers(\Theta)$ be any cover. The associated* causaltope *is defined to be the following space of constrained conditional probability distributions:*

$$\begin{aligned} Caus\left(C, \underline{O}\right) &:= CCPD\left(CausEqs\left(C, \underline{O}\right), \left(\prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{TipEq_\lambda(\omega)}\right)_{\lambda \in C}\right) \\ &= CausEqs\left(C, \underline{O}\right) \cap \prod_{\lambda \in C} \mathcal{D}\left(\prod_{\omega \in \mathrm{dom}(\lambda)} (O_\omega)^{TipEq_\lambda(\omega)}\right) \end{aligned} \tag{5.26}$$

Observation 5.24. *When $C = StdCov(\Theta)$ is the standard cover, we refer to the associated causaltope as a* standard causaltope*, taking the following simplified form:*

$$Caus_{std}\left(\Theta, \underline{O}\right) := CCPD\left(CausEqs_{std}\left(C, \underline{O}\right), \left(\prod_{\omega \in \mathrm{dom}(k)} O_\omega\right)_{k \in \max Ext(\Theta)}\right)$$

We write $CausEqs_{std}\left(\Theta, \underline{O}\right)$ for the causal equations on the standard cover.

Theorem 5.25. *Let Θ be a space of input histories, let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs and let $C \in Covers(\Theta)$ be any cover. Then the following is a convex-linear bijection:*

$$\begin{aligned} EmpMod\left(C, \underline{O}\right) &\leftrightarrow Caus\left(C, \underline{O}\right) \\ \underline{e} &\mapsto \left(\lceil e_\lambda \rceil\right)_{\lambda \in C} \end{aligned} \tag{5.27}$$

Proof. By applying Theorem 5.20 to the individual components $e_\lambda \mapsto \lceil e_\lambda \rceil$ we conclude that the map is a convex-linear injection, where convex-linearity follows from the fact that all components of the empirical model are weighted equally in convex combinations. To prove that the map is surjective, we consider any arbitrary $\underline{u} \in \mathrm{Caus}\left(C, \underline{O}\right)$ and define:

$$e_\lambda := \left\lfloor \underline{u}^\lambda \right\rfloor_\lambda$$

The causality equations guarantee that restrictions from cover lowersets to arbitrary lowersets coincide:

$$\lceil e_\lambda \rceil|_\mu = \lceil e_{\lambda'} \rceil|_\mu$$

We now show that the restrictions of probability distributions above are exactly the same as the restrictions of empirical model components. To do so, it suffices to expand the top-element distribution and its restriction into their definitions:

$$\begin{aligned} \lceil e_\lambda \rceil|_\mu &= \mathcal{D}\left(\rho_{\lambda,\mu}\right)\left(\sum_{\mathrm{Ext}(f)} e_\lambda\left(\mathrm{Ext}\left(f\right)\right) \delta_{\lceil \mathrm{Ext}(f) \rceil}\right) \\ &= \sum_{\mathrm{Ext}(f)} e_\lambda\left(\mathrm{Ext}\left(f\right)\right) \mathcal{D}\left(\rho_{\lambda,\mu}\right)\left(\delta_{\lceil \mathrm{Ext}(f) \rceil}\right) \\ &= \sum_{\mathrm{Ext}(f)} e_\lambda\left(\mathrm{Ext}\left(f\right)\right) \delta_{\rho_{\lambda,\mu}(\lceil \mathrm{Ext}(f) \rceil)} \\ &= \sum_{\mathrm{Ext}(f)} e_\lambda\left(\mathrm{Ext}\left(f\right)\right) \delta_{\lceil \mathrm{Ext}(f) \rceil|_\mu} \\ &= \mathrm{Ext}\left(f'\right) \mapsto \sum_{f \text{ s.t. } f|_\mu = f'} e_\lambda\left(\mathrm{Ext}\left(f\right)\right) \end{aligned}$$

The last line is the definition of restriction for empirical model components from λ to μ, completing our proof. □

Theorem 5.25 provides an equivalent geometric characterisation for empirical models, as constrained conditional probability distributions on joint outputs. From this moment onwards, we will freely confuse the topological and geometric picture, referring to the points of causaltopes as 'empirical models'.

5.4 Standard causaltopes

In Chapter 4, we explained that different choices of covers correspond to different assumptions of classicality; an empirical model for the classical cover corresponds to a distribution over classical (global) causal mechanisms. The standard cover is the more unconstrained assignment of distributions compatible with a given space of input histories. The covers which are finer than the standard cover, such as the solipsistic cover, describe empirical models where only partial information about the correlations between timelike histories is recorded.

Therefore, the standard cover is what we usually consider when talking about an operational protocol: a description of all the correlations between joint inputs and outputs. When discussing the compatibility of an empirical model to space of input histories, we usually work with standard empirical models. In this case, the geometric characterisation provided above coincides with the literature on indefinite causality. Empirical models are distributions $u_o^{(i)} \in [0, 1]$ on joint outputs $o \in \prod_{\omega \in E^\Omega} O_\omega$ conditional to joint inputs $i \in \prod_{\omega \in E^\Theta} I_\omega^\Theta$.

Standard causaltopes, therefore, represent the arena where everything that can be known about the behaviour of a protocol is recorded and where the assumption of non-disturbance between measurements is only expressed by causal compatibility with no other assumption about the particular structure of the contexts. Standard causaltopes are where we can give substance to the notion of causal discovery in a canonical way. In this section, we define the vocabulary underlying this type of analysis, i.e. we will explain what is meant by *casual components*, *causal decomposition* and *casual fractions*. We first show that the hierarchy of spaces of input histories is reflected in the containment structure of the causaltopes.

Proposition 5.25. *Let $\Theta \leq \Theta'$ be a spaces of input histories such that $E^{\Theta'} = E^\Theta$ and $I^{\Theta'} = I^\Theta$ and $\max Ext(\Theta) = \max Ext(\Theta')$ (e.g. because they both satisfy the free choice condition). Let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. The standard causaltope for Θ is always contained in the standard causaltope for Θ':*

$$CausEqs_{std}(\Theta, \underline{O}) \subseteq CausEqs_{std}(\Theta', \underline{O})$$

$$Caus_{std}(\Theta, \underline{O}) \subseteq Caus_{std}(\Theta', \underline{O})$$

Proof. Because Θ and Θ' have the same events and inputs and because $\max \text{Ext}(\Theta) = \max \text{Ext}(\Theta')$, the two spaces have the same pseudo-empirical models:

$$\prod_{k \in \max \text{Ext}(\Theta)} \mathcal{D}\left(\prod_{\omega \in \text{dom}(k)} O_\omega \right)$$

where we used the fact that $\text{TipEq}_{k\downarrow}(\omega)$ is always a singleton. Hence, comparing the associated linear sub-spaces of causality equations makes sense. By Proposition 5.23, the causality equations for the standard cover are generated by extended input histories: since $\Theta' \leq \Theta$ is defined to mean $\text{Ext}(\Theta') \supseteq \text{Ext}(\Theta)$, the causality equations for Θ' are a superset of those for Θ. This concludes our proof. □

The standard causaltope for the space of histories obtained by the indiscrete and the discrete preorder represent the maximal and the minimal element of the hierarchy of nested polytopes:

Observation 5.25. *For any non-empty set E of events and any family of non-empty input sets $\underline{I} = (I_e)_{e \in E}$, the standard causaltope for the indiscrete space $Hist(indiscrete(E), \underline{I})$ is the polytope of pseudo-empirical models:*

$$Caus_{std}(\Theta_{ind}, \underline{O}) = PEmpMods(StdCov(\Theta_{ind}), \underline{O})$$

where we have defined the shorthand $\Theta_{ind} := Hist(indiscrete(E), \underline{I})$.

Observation 5.25. *Let Θ be a spaces of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. The standard causaltope for the discrete space $Hist(discrete(E^\Theta), \underline{I^\Theta})$ is always contained in the causaltope for Θ:*

$$Caus_{std}\left(Hist\left(discrete\left(E^\Theta\right), \underline{I^\Theta}\right), \underline{O}\right) \subseteq Caus_{std}(\Theta, \underline{O})$$

For any non-empty set E of events and any family of non-empty input sets $\underline{I} = (I_e)_{e \in E}$, we refer to $Caus_{std}(Hist(discrete(E), \underline{I}), \underline{O})$ as the no-signalling causaltope.

The following definitions provide 'causally flavoured' variants of Definition 5.16 and Definition 5.17. In particular, we adopt a special name for the fraction supported by the no-signalling causaltope.

Definition 5.26. *Let Θ be a spaces of input histories and let $\underline{O} = (O_\omega)_{\omega \in E^\Theta}$ be a family of non-empty sets of outputs. Let $\underline{u} \in Caus_{std}(\Theta, \underline{O})$ be a standard empirical model. For any $\Theta' \leq \Theta$ such that $E^{\Theta'} = E^\Theta$ and $I^{\Theta'} = I^\Theta$ and $\max Ext(\Theta) = \max Ext(\Theta')$, we give the following definitions:*

- *A* component *of $\underline{u}$ in Θ' is a component of $\underline{u}$ in the sub-polytope of constrained conditional probability distributions $Caus_{std}(\Theta', \underline{O})$ according to Definition 5.16.*
- *A* maximal *component of $\underline{u}$ in Θ' is one of maximal mass.*

- *The* causal fraction *of $\underline{u}$ in Θ' is the mass of a maximal component of $\underline{u}$ in Θ'.*

Colloquially, we say that $\underline{u}$ is X% supported by Θ' to mean that the supported fraction of $\underline{u}$ in Θ' is $\frac{X}{100}$. The no-signalling fraction *of $\underline{u}$ is the causal fraction of $\underline{u}$ in the discrete space Hist $\left(discrete\left(E^{\Theta}\right), \underline{I^{\Theta}}\right)$.*

Definition 5.27. *Let Θ be a spaces of input histories and let $\underline{O} = (O_{\omega})_{\omega \in E^{\Theta}}$ be a family of non-empty sets of outputs. Let $\underline{u} \in Caus_{std}\left(\Theta, \underline{O}\right)$ be a standard empirical model. For any family $\left(\Theta^{(z)}\right)_{z \in Z}$ of sub-spaces $\Theta^{(z)} \leq \Theta$ such that $E^{\Theta^{(z)}} = E^{\Theta}$ and $I^{\Theta^{(z)}} = I^{\Theta}$ and $\max Ext\left(\Theta^{(z)}\right) = \max Ext\left(\Theta\right)$, we give the following definitions:*

- *A* (causal) decomposition *of $\underline{u}$ over the sub-spaces $\left(\Theta^{(z)}\right)_{z \in Z}$ is a decomposition of $\underline{u}$ in $\left(Caus_{std}\left(\Theta^{(z)}, \underline{O}\right)\right)_{z \in Z}$ according to Definition 5.17.*
- *A* maximal (causal) decomposition *of $\underline{u}$ over the sub-spaces $\left(\Theta^{(z)}\right)_{z \in Z}$ is one of maximal mass.*
- *The* causal fraction *of $\underline{u}$ over the sub-spaces $\left(\Theta^{(z)}\right)_{z \in Z}$ is the mass of a maximal decomposition of $\underline{u}$ in $\left(\Theta^{(z)}\right)_{z \in Z}$.*

Colloquially, we say that $\underline{u}$ is X% supported by $\left(\Theta^{(z)}\right)_{z \in Z}$ to mean that the causal fraction of $\underline{u}$ over the sub-spaces $\left(\Theta^{(z)}\right)_{z \in Z}$ is $\frac{X}{100}$.

Consider a standard empirical model $\underline{u}$ for a causally incomplete space Θ. A key question in the study of indefinite causality is whether the $\underline{u}$ admits an explanation in terms of 'dynamical' definite causal structure, i.e. whether it is 100% jointly supported by some causally complete sub-spaces of Θ. This leads to the definition of the qualitative notion of 'causal (in)separability' and the associated quantitative notion of 'causally (in)separable fraction'.

Definition 5.28. *Let Θ be a spaces of input histories and let $\underline{O} = (O_{\omega})_{\omega \in E^{\Theta}}$ be a family of non-empty sets of outputs. Let $\underline{u} \in Caus_{std}\left(\Theta, \underline{O}\right)$ be a standard empirical model. We give the following definitions:*

- *The* causally separable fraction *of $\underline{u}$ is the causal fraction of $\underline{u}$ over the causal completions of Θ, i.e. over the maximal causally complete subspaces of Θ. If Θ is causally complete, the causally separable fraction is always 1.*
- *The* causally inseparable fraction *of $\underline{u}$ is 1 minus its causally separable fraction.*
- *The empirical model $\underline{u}$ is* causally separable *if it has causally separable fraction 1, and it is* causally inseparable *otherwise.*

Colloquially, we say that $\underline{u}$ has X% causally (in)separable fraction if its causally (in)separable fraction is $\frac{X}{100}$.

We are now ready to define *contextual causality* for empirical models over the standard cover. Probing causal inseparability for a causally incomplete space Θ already signals that the model cannot be decomposed in sub-normalised empirical models on the causal completions of Θ. However, causal inseparability alone does not guarantee the impossibility of providing an objective assignment of causal orders between the events. Here we need to be particularly careful. Causal inseparability for a space can also be a consequence of event delocalisation. To circumvent these problems, we restrict our definition to the empirical models which are entirely supported by the causal completion for the indiscrete space.

Recall that it follows from Theorem 3.25 that we only need to check the switch spaces to conclude if a model is causally separable for the indiscrete space. Moreover, Theorem 4.42 explains that switch spaces cannot exhibit any non-locality. An empirical model which is causally separable for $\text{Caus}_{std}\left(\Theta_{ind}, \underline{O}\right)$ can therefore be described entirely using separable causal functions, with no inseparability involved.

Definition 5.29 (Contextual causality)**.** *Let $e \in Caus_{std}\left(\Theta_{ind}, \underline{O}\right)$ be a standard empirical model which is causally separable for the indiscrete space $\Theta_{ind} = Hist\left(indiscrete\left(E\right), \underline{I}\right)$. We say that e exhibits* contextual causality *if there exists a refinement $\Theta' \leq \Theta_{ind}$ such that $e \in Caus_{std}\left(\Theta', \underline{O}\right)$ but e is causally inseparable with respect to Θ'.*

5.5 Conclusions

In this final chapter, we developed a geometric description complementary to the topological one: causal correlations become the points of causaltopes: convex polytopes obtained by slicing the set of conditional probability distributions with certain causality equations. Our methods can be seen as a generalisation of the geometric techniques used in the theory independent study of no-signalling correlations, and they present a finer-grained picture of causal separability than the one painted by the literature on causal inequalities [31, 101, 1].

Specifically, we can quantify the device-independent explainability of conditional probability distributions relative to arbitrary putative causal structures, incorporating constraints such as space-like separation of parties or dynamical no-signalling. The more general, relative nature of our definition of causal separability allows us to define new witnesses for indefinite causal order by exploiting the experimental legitimacy of imposing some causal constraints even in the presence of indefinite causality.

The advantage of a geometrical perspective is not limited to causal inference: combined with geometric tools from the Abramsky-Brandenburger framework, it allows us to quantitatively investigate the correlation between indefinite causality and non-locality/contextuality. This theoretical framework gives rise to novel methods to certify the non-classicality of causation, of particular interest in scenarios where quantum theory is endowed with the possibility of superposing the causal order of quantum channels. Unlike previous literature on the topic, however, the phenomenology involved in our certification of indefinite causality is entirely theory independent.

In the next chapter, after the theoretical tour de force, we will finally put our framework into use and calculate a selection of causal decompositions aimed at showcasing the information about the causal structure which can be deduced from the empirical models in a theory independent way.

Chapter 6

Examples of causal decompositions

A table of probabilities which assigns distributions of joint outputs to joint inputs for an operational scenario $(E, \underline{I}, \underline{O})$ can be thought of as providing the theory-independent empirical content associated with some protocol. Such behaviours are in bijective correspondence with the points of the causaltope $\mathrm{Caus}_{std}\left(\mathrm{Hist}\left(\mathrm{discrete}\left(E\right), \underline{I}\right), \underline{O}\right)$ (Observation 5.25).

In such cases, the order between the events could be indefinite, or the event themselves be delocalised; Nevertheless, we can always use linear programming to calculate the fractions supported by polytopes entailing finer causal assumptions. In particular, we can test if some empirical behaviour is causally separable with respect to an ambient space Θ. If Θ is causally definite, this would be asking whether specific causal constraints are satisfied; otherwise, we can encode in Θ a more coarse grained causal assumption which is nevertheless causally incomplete to narrow down the putative causal explanations. We can construct a notion of causal separability that is strictly finer than the one used in previous literature on indefinite causal order [101, 31, 1], where Θ is fixed to the indiscrete space.

In this chapter, we discuss the empirical models for a selection of examples of interest. All empirical models are for the standard cover so that any non-classicality arises from non-locality rather than other forms of contextuality.

All models have binary inputs and outputs $I_\omega = O_\omega = \{0, 1\}$ at each event, unless otherwise specified. For convenience, we will describe our scenarios in terms of agents performing operations at the events, always following the same convention: Alice acts at event A, Bob acts at event B, Charlie acts at event C, Diane acts at event D, Eve acts at event E and Felix acts at event F.

The examples in this chapter are all instances of the linear programs described in Chapter 5; the optimisation has been performed using the standard methods offered by the SciPy[140] module 'scipy.linalg.linprog'. We are currently in the process of developing fully-fledged software for the causal analysis of empirical models. Specifically, we aim to automatise tasks such as generating causal equations for a given space of input histories and calculating various quantities of interest, including the causal fraction and other more fine-grained causal decompositions.

6.1 A Classical Switch Empirical Model

In this example, Alice classically controls the order of Bob and Charlie, as follows:

- Alice flips one of two biased coins, depending on her input: when her input is 0, her output is 75% 0 and 25% 1; when her input is 1, her output is 25% 0 and 75% 1 instead.
- Bob and Charlie are in a quantum switch, controlled in the Z basis and with $|0\rangle$ as a fixed input: on output $a \in \{0, 1\}$, Alice feeds state $|a\rangle$ into the control system of the switch, determining the relative causal order of Bob and Charlie.
- Bob and Charlie both apply the same quantum instrument: they measure the incoming qubit they receive in the Z basis, obtaining their output, and then encode their input into the Z basis of the outgoing qubit.
- Both the control qubit and the outgoing qubit of the switch are discarded: even without Alice controlling the switch in the Z basis, discarding the control qubit would be enough to make the control classical.

The description above results in the following empirical model on 3 events:

ABC	000	001	010	011	100	101	110	111
000	3/4	0	0	0	1/4	0	0	0
001	3/4	0	0	0	0	0	1/4	0
010	0	3/4	0	0	1/4	0	0	0
011	0	3/4	0	0	0	0	1/4	0
100	1/4	0	0	0	3/4	0	0	0
101	1/4	0	0	0	0	0	3/4	0
110	0	1/4	0	0	3/4	0	0	0
111	0	1/4	0	0	0	0	3/4	0

To better understand the table above, we focus on the second row, corresponding to input 001:

1. Alice's input is 0, so her output is 75% 0 and 25% 1. This means that the probabilities of outputs 0__ in row 001 of the empirical model must sum to 75%, and the probabilities of output 1__ must sum to 25%.
2. Conditional to Alice's output being 0, the output is 000 with 100% probability:
   (a) Bob goes first and receives the input state $|0\rangle$ for the switch: he measures the state in the Z basis, obtaining output 0 with 100% probability. Because his input is 0, he then prepares the state $|0\rangle$, which he forwards into the switch.
   (b) Charlie goes second and receives the state $|0\rangle$ prepared by Bob: he measures the state in the Z basis, obtaining output 0 with 100% probability. Because his input is 1, he then prepares the state $|1\rangle$, which he forwards into the switch.

(c) Charlie's state $|1\rangle$ comes out of the switch, and is discarded.

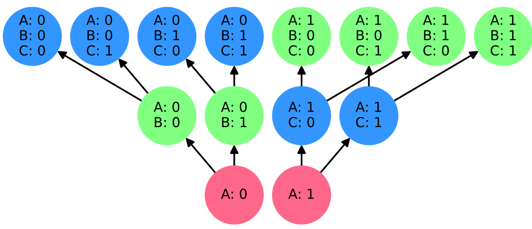
3. Conditional to Alice's output being 1, the output is 110 with 100% probability:

    (a) Charlie goes first and receives the input state $|0\rangle$ for the switch: he measures the state in the Z basis, obtaining output 0 with 100% probability. Because his input is 1, he then prepares the state $|1\rangle$, which he forwards into the switch.

    (b) Bob goes second and receives the state $|1\rangle$ prepared by Charlie: he measures the state in the Z basis, obtaining output 1 with 100% probability. Because his input is 0, he then prepares the state $|0\rangle$, which he forwards into the switch.

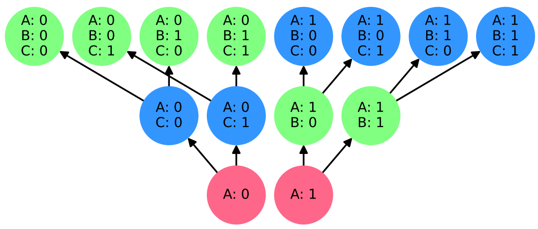
    (c) Bob's state $|0\rangle$ comes out of the switch, and is discarded.

This empirical model is causally separable. A maximum fraction of 75% is supported by the switch space where Alice choosing 0 makes Bob precede Charlie and a maximum fraction of 25% is supported by the switch space where Alice choosing 0 makes Charlie precede Bob, with a fraction of 0% supported by both spaces (i.e. no overlap). Below we show the two spaces, the corresponding causal fraction, and the (renormalised) component of the empirical model supported by each space:



Causal fraction: 75%

ABC	000	001	010	011	100	101	110	111
000	1	0	0	0	0	0	0	0
001	1	0	0	0	0	0	0	0
010	0	1	0	0	0	0	0	0
011	0	1	0	0	0	0	0	0
100	0	0	0	0	1	0	0	0
101	0	0	0	0	0	0	1	0
110	0	0	0	0	1	0	0	0
111	0	0	0	0	0	0	1	0



Causal fraction: 25%

ABC	000	001	010	011	100	101	110	111
000	0	0	0	0	1	0	0	0
001	0	0	0	0	0	0	1	0
010	0	0	0	0	1	0	0	0
011	0	0	0	0	0	0	1	0
100	1	0	0	0	0	0	0	0
101	1	0	0	0	0	0	0	0
110	0	1	0	0	0	0	0	0
111	0	1	0	0	0	0	0	0

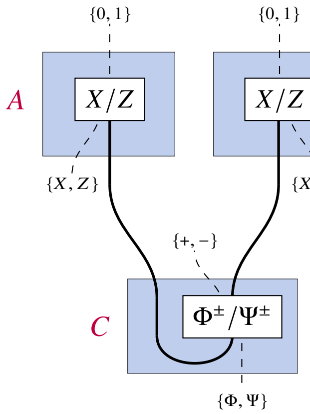
6.2 A Causal Fork Empirical Model

In this example, Charlie produces one of the four Bell basis states and forwards one qubit each to Alice and Bob, who measure it in either the Z or X basis:

1. On input $c \in \{0, 1\}$ Charlie prepares the 2-qubit state $|0c\rangle$. He then performs a XX parity measurement, resulting in one of $|\Phi^{\pm}\rangle$ states (if his input was 0) or one of $|\Psi^{\pm}\rangle$ states (if his input was 1), all with 50% probability. He forwards this state to Alice and Bob, one qubit each.

2. Alice and Bob perform a Z basis measurement on input 0 and an X basis measurement on input 1, and use the measurement outcome as their output.

The following figure summarises the experiment:



The description above results in the following empirical model on 3 events:

ABC	000	001	010	011	100	101	110	111
000	1/4	1/4	0	0	0	0	1/4	1/4
001	0	0	1/4	1/4	1/4	1/4	0	0
010	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
011	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
100	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
101	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
110	1/4	0	0	1/4	0	1/4	1/4	0
111	1/4	0	0	1/4	0	1/4	1/4	0

To better understand the process, we restrict our attention to the rows where Charlie has input 0, corresponding to Alice and Bob receiving the Bell basis states $|\Phi^{\pm}\rangle$:

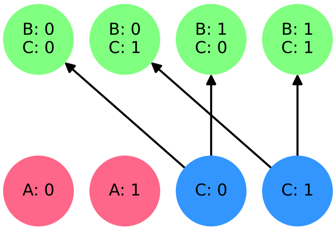
ABC	000	001	010	011	100	101	110	111
000	1/4	1/4	0	0	0	0	1/4	1/4
010	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
100	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
110	1/4	0	0	1/4	0	1/4	1/4	0

When Charlie's output is 0 (left below) Alice and Bob receive the Bell basis state $|\Phi^{+}\rangle$: they get perfectly correlated outputs when they both measure in Z or both measure in X, and uncorrelated uniformly distributed outputs otherwise. When Charlie's output is 1 (right below) Alice and Bob receive the Bell basis state $|\Phi^{-}\rangle$: they get perfectly correlated outputs when they both measure in Z, perfectly anti-correlated outputs when they both measure in X, and uncorrelated uniformly distributed outputs otherwise.

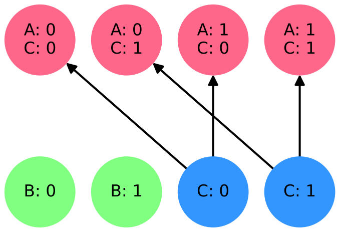
ABC	000	010	100	110
000	1/4	0	0	1/4
010	1/8	1/8	1/8	1/8
100	1/8	1/8	1/8	1/8
110	1/4	0	0	1/4

ABC	001	011	101	111
000	1/4	0	0	1/4
010	1/8	1/8	1/8	1/8
100	1/8	1/8	1/8	1/8
110	0	1/4	1/4	0

Rather interestingly, the empirical model for this experiment is 100% supported by two incompatible spaces of input histories, both in equivalence class 33 (see Figure 3.4 (p.110)): the space $\Theta_{A \vee (C \to B)}$ induced by causal order $A \vee (C \to B)$ (left below) an the space $\Theta_{B \vee (C \to A)}$ induced by causal order $B \vee (C \to A)$. In other words, the empirical data is compatible both with absence of signalling from C to A (left below) and with absence of signalling from C to B (right below).

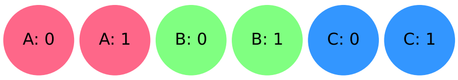


Causal fraction: 100%



Causal fraction: 100%

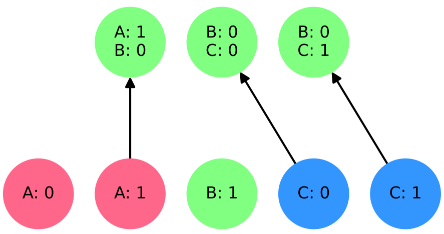
What makes this empirical model even more interesting is that its no-signalling fraction is 0%: no part of it can be explained without signalling from C to at least one of A or B.



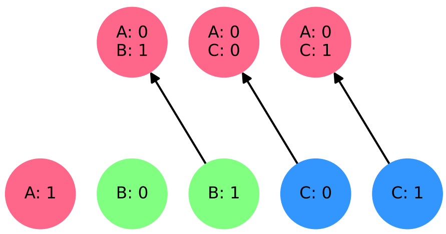
Causal fraction: 0%

Since the discrete space $\Theta_{A \vee B \vee C}$ is the meet of the two order-induced spaces $\Theta_{A \vee (C \to B)}$ and $\Theta_{B \vee (C \to A)}$, we now have an example of an empirical model which is fully supported by two spaces of input histories but not supported at all by their meet. In particular, this shows that the intersection of two causaltopes is not necessarily the causaltope for the meet of the underlying spaces.

The empirical model does happen to be 100% supported by two unrelated non-tight subspaces of $\Theta_{A \vee (C \to B)}$ and $\Theta_{B \vee (C \to A)}$ respectively, both falling into equivalence class 2:



Causal fraction: 100%

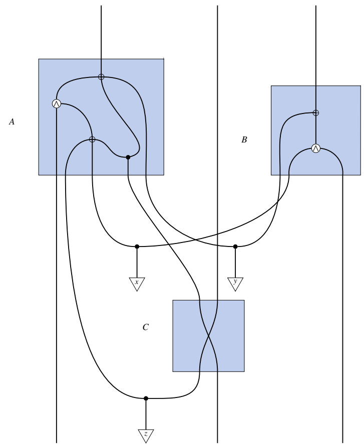


Causal fraction: 100%

Unlike $\Theta_{A \vee (C \to B)}$ and $\Theta_{B \vee (C \to A)}$, these two spaces have exactly the same standard causaltope, of dimension 27. Because it is only 1 dimension larger than the no-signalling causaltope, this is the minimal supporting causaltope for our empirical model.

The empirical model is local for both space $\Theta_{\mathrm{A}\vee(\mathrm{C}\to\mathrm{B})}$ and space $\Theta_{\mathrm{B}\vee(\mathrm{C}\to\mathrm{A})}$: for example, below is a decomposition as a uniform mixture of 8 causal functions for $\Theta_{\mathrm{B}\vee(\mathrm{C}\to\mathrm{A})}$. The black dots are classical copies, the $\oplus$ dots are classical XORs and the $\wedge$ dots classical ANDs.

$$\frac{1}{8} \sum_{(x,y,z)\in\{0,1\}^3}$$

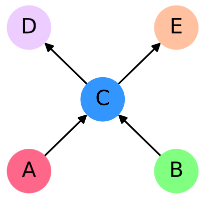


However, we know from Figure 3.4 (p.49) that spaces in equivalence class 2 have exactly the same causal functions as the discrete space, in equivalence class 0. Since the empirical model has a no-signalling fraction of 0%, it immediately follows that it has a local fraction of 0% in its minimal supporting causaltope, i.e. that it is maximally non-local there. To recap, this example bears many gifts:

- It shows that there are empirical models 100% supported by multiple spaces but 0% supported by their meet; in particular, it shows that the intersection of causaltopes is not necessarily the causaltope for the meet of the underlying spaces.
- Further to the previous point, it shows that there are causaltopes whose intersection is not the causaltope for any space.
- It shows that there can be unrelated spaces with equal causaltopes, differing from the no-signalling causaltope.
- It provides an empirical model whose minimally supporting space is non-tight, providing additional evidence for the importance of non-tight spaces in the study of causality.
- It shows that the notions of non-locality and contextuality depend on a specific choice of causal constraints, by providing an empirical model which is local in for a space and maximally non-local for a sub-space.

6.3 A Causal Cross Empirical Model

In this example, Charlie receives qubits from Alice and Bob and forwards them to Diane and Eve, choosing whether to forward the qubits as $A \to D, B \to E$ or as $A \to E, B \to D$. Below is the 'cross' causal order that naturally supports this example:



More specifically, the parties act as follows:

1. Alice and Bob encode their input into the Z basis of one qubit each, which they then forward to Charlie. Their output is trivial, constantly set to 0.

2. Charlie receives the two qubits from Alice and Bob, and decides how to forward them based on his input:

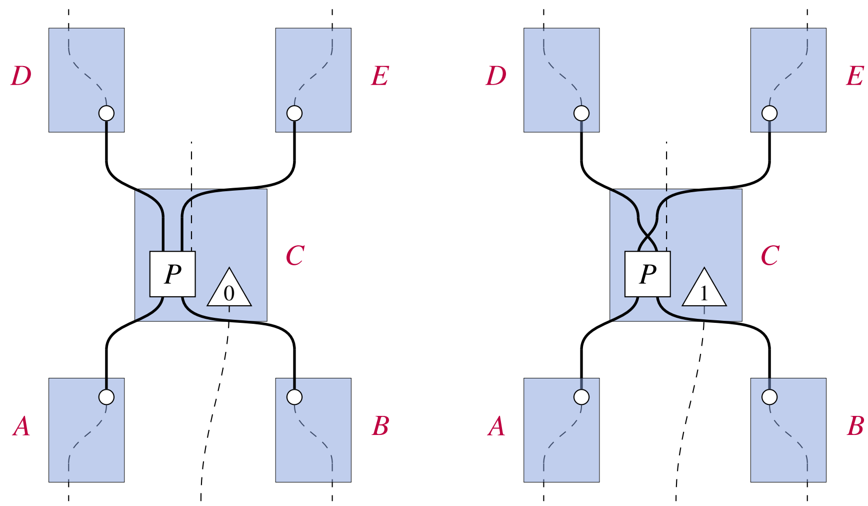
   - On input 0, Charlie forwards Alice's qubit to Diane and Bob's qubit to Eve.
   - On input 1, Charlie forwards Alice's qubit to Eve and Bob's qubit to Diane.

3. Diane and Eve have trivial input, with only 0 as an option. They measure the qubit they receive in the Z basis and use the outcome as their output.

We will consider two version of this protocol: one where Charlie measures the parity of the qubits he receives, and one where he doesn't perform any measurement and trivially outputs 0. The version where Charlie measures the parity corresponds to the following empirical model e; note that the outputs of Alice and Bob, as well as the inputs of Diane and Eve, are fixed to 0.

ABCDE	00000	00001	00010	00011	00100	00101	00110	00111
00000	1	0	0	0	0	0	0	0
00100	1	0	0	0	0	0	0	0
01000	0	0	0	0	0	1	0	0
01100	0	0	0	0	0	0	1	0
10000	0	0	0	0	0	0	1	0
10100	0	0	0	0	0	1	0	0
11000	0	0	0	1	0	0	0	0
11100	0	0	0	1	0	0	0	0

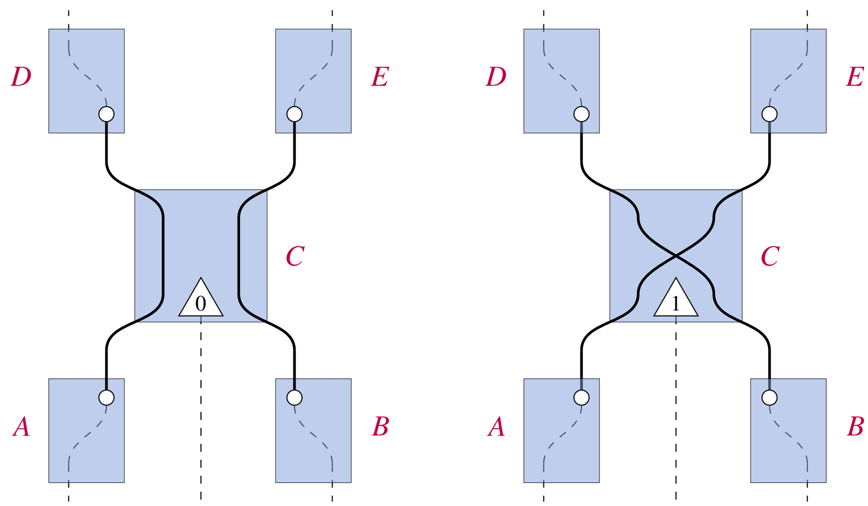
The figures below exemplify this full scenario:



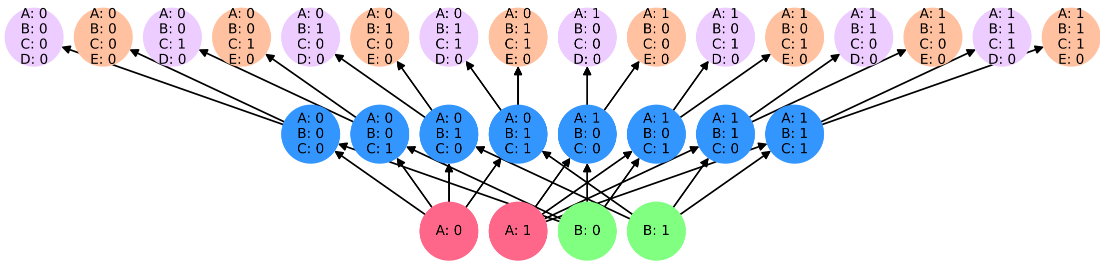
The version where Charlie doesn't measures the parity corresponds to the following simplified empirical model e'; note that the outputs of Alice, Bob and Charlie, as well as the inputs of Diane and Eve, are fixed to 0.

ABCDE	00000	00001	00010	00011
00000	1	0	0	0
00100	1	0	0	0
01000	0	1	0	0
01100	0	0	1	0
10000	0	0	1	0
10100	0	1	0	0
11000	0	0	0	1
11100	0	0	0	1

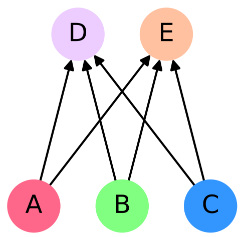
The figures below exemplify this latter, simplified scenario:



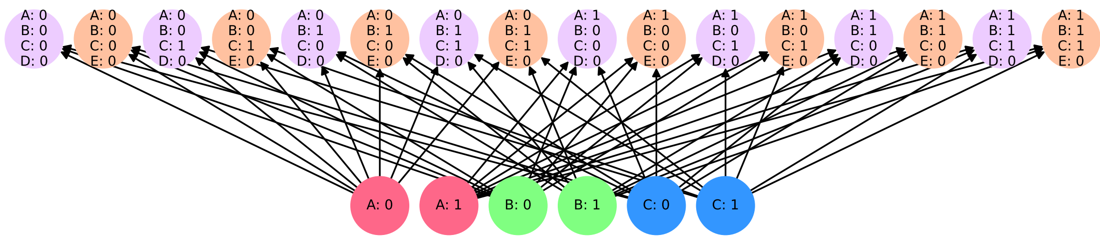
By construction, both empirical models are 100% supported by the space of input histories induced the cross causal order. In fact, they are both deterministic, and hence they correspond to causal functions for the space.



In the second version of the experiment, Charlie doesn't learn anything about Alice and Bob's inputs: his trivial output can be explained without signalling from either one of Alice or Bob. Indeed, the simplified empirical model e' is 100% supported by the space of input histories induced by the following '$K_{3,2}$' causal order.

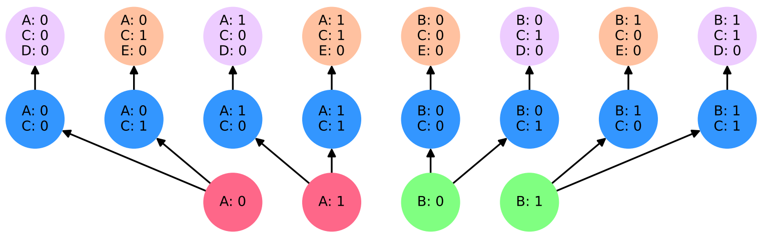


The space of input histories is explicitly depicted below.



In fact, the causaltopes for the spaces induced by the cross causal order and the $K_{3,2}$ causal order coincide when the output $O_C = \{0\}$ for Charlie is trivial.

We can also construct an entirely different space where Charlie's output is independent of Alice and Bob's input, by exploiting the additional constraints on causal functions afforded by lack of tightness. Indeed, the empirical model e' is also 100% supported by the non-tight space below: no-signalling from Alice to Charlie is enforced by the $\{\mathrm{B}:b,\ \mathrm{C}:c\}$ histories on the right, while no-signalling from Bob to Charlie is enforced by the $\{\mathrm{A}:a,\ \mathrm{C}:c\}$ histories on the left.



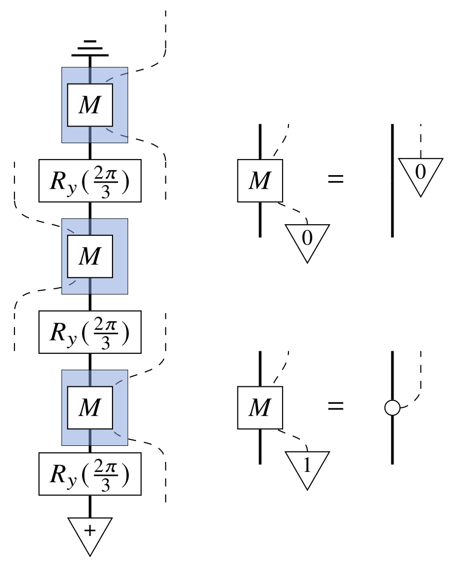
This space of input histories is a subspace of the space induced by the cross causal order, but is unrelated to the one for the $K_{3,2}$ causal order. However, the 16-dimensional causaltope for this non-tight space sits inside the causaltope for the two order-induced spaces: even though the spaces are unrelated, every standard empirical model for the non-tight space is also a standard empirical model for the space induced by the $K_{3,2}$ order.

6.4 The Leggett-Garg Empirical Model

To disprove macro-realistic explanations of quantum mechanical phenomena, the authors of [84] propose the following experiment on a 2-level quantum system, i.e. a qubit, which evolves in time by rotating about the Y axis at a constant angular rate. Writing $\Delta t > 0$ for the minimum time over which qubit evolution performs a $\frac{2\pi}{3}$ Y rotation, the experiment proceeds as follows:

1. The qubit is prepared in the $|+\rangle$ state at time t_0 and left alone to evolve.
2. At time $t_1 := t_0 + \Delta t$, known to us as event A, the qubit is either left alone (input 0 at A, with output fixed to 0) or a non-demolition measurement in the Z basis is performed on it (input 1 at A, with meas. outcome as output). The qubit is again left alone to evolve.
3. At time $t_2 := t_1 + \Delta t$, known to us as event B, the qubit is either left alone (input 0 at B, with output fixed to 0) or a non-demolition measurement in the Z basis is performed on it (input 1 at B, with meas. outcome as output). The qubit is again left alone to evolve.
4. At time $t_3 := t_2 + \Delta t$, known to us as event C, the qubit is either discarded (input 0 at C, with output fixed to 0) or a demolition measurement in the Z basis is performed on it (input 1 at C, with meas. outcome as output).

The figure below exemplifies the scenario we have just described:



The description above results in the following empirical model on 3 events:

ABC	000	001	010	011	100	101	110	111
000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
001	0.933	0.067	0.000	0.000	0.000	0.000	0.000	0.000
010	0.067	0.000	0.933	0.000	0.000	0.000	0.000	0.000
011	0.017	0.050	0.700	0.233	0.000	0.000	0.000	0.000
100	0.500	0.000	0.000	0.000	0.500	0.000	0.000	0.000
101	0.125	0.375	0.000	0.000	0.375	0.125	0.000	0.000
110	0.125	0.000	0.375	0.000	0.375	0.000	0.125	0.000
111	0.031	0.094	0.281	0.094	0.094	0.281	0.094	0.031

The Leggett-Garg inequalities provides bounds, valid in macro-realistic interpretations, for the sum of the expected ± 1-valued parity of outputs when the ± 1-valued parity of inputs is $+1$:

$$-1 \leq \mathbb{E}(-1^{o_A \oplus o_B \oplus o_C} | 011) + \mathbb{E}(-1^{o_A \oplus o_B \oplus o_C} | 101) + \mathbb{E}(-1^{o_A \oplus o_B \oplus o_C} | 110) \leq 3$$

The authors then observe that, in the experiment they propose, the sum of such expected parities is $-\frac{3}{2}$, violating the lower bound and thus excluding a macro-realistic explanation. Indeed, we can restrict ourselves to the relevant rows of the empirical model:

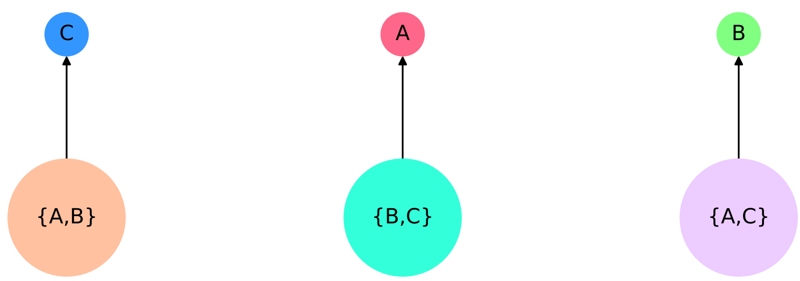
ABC	000	001	010	011	100	101	110
011	0.017	0.050	0.700	0.233	0.000	0.000	0.000
101	0.125	0.375	0.000	0.000	0.375	0.125	0.000
110	0.125	0.000	0.375	0.000	0.375	0.000	0.125

The sum of the expected parity of outputs is then computed as follows:

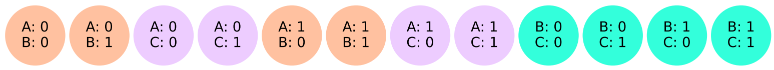
$$\begin{aligned} & \mathbb{E}(-1^{o_A \oplus o_B \oplus o_C} | 011) + \mathbb{E}(-1^{o_A \oplus o_B \oplus o_C} | 101) + \mathbb{E}(-1^{o_A \oplus o_B \oplus o_C} | 110) \\ = & (0.017 - 0.050 - 0.700 + 0.233) \\ & +(0.125 - 0.375 - 0.375 + 0.125) \\ & +(0.125 - 0.375 - 0.375 + 0.125) \\ = & -\tfrac{1}{2} - \tfrac{1}{2} - \tfrac{1}{2} = -\tfrac{3}{2} \end{aligned}$$

By construction, this empirical model is 100% supported by the space of inputs histories for the total order total (A, B, C). As a consequence, it is necessarily non-contextual/local for this space: for an explicit decomposition as a convex combination of 12 causal functions, see Subsection 4.5.4 of [65].

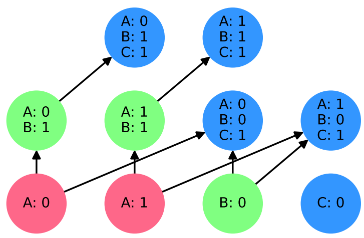
The constraints specified by Equation (1) of [84] are in fact causal constraint, stating that a macro-realist model has to be supported by the following 3 indefinite causal orders.



By construction, the Leggett-Garg empirical model is 100% supported by the leftmost causal order (of which total (A, B, C) is a sub-order). However, it is only 56.7% supported by the middle causal order and 62.5% supported by the right causal order, showing that it violates the causal constraints imposed by the macro-realist assumption. The meet of the spaces of input histories induced by the three indefinite causal orders above is the non-causally-complete, non-tight space depicted below:



The meet of the causaltopes for the spaces induced by the three indefinite causal orders is the same as the causaltope for the meet space: they both have dimension 26, and hence coincide with the no-signalling causaltope. Imposing the three constraints together is thus the same as imposing no-signalling, and the Leggett-Garg empirical model has a 30.25% no-signalling fraction. It also has varying causal fractions over other total orders: 62.50% over total (A, C, B), 56.70% over total (B, A, C) and total (B, C, A), 45.87% over total (C, B, A) and 37.95% over total (C, A, B). The (unique) minimal supporting space is depicted below, with a 35-dim causaltope.



This space captures the causal constraints—additional with respect to the total order total (A, B, C)—associated with the absence of measurement on input 0:

- The presence of history $\{C{:}0\}$ states that there is no signalling from A nor B to C when no measurement is performed at C.
- The presence of history $\{B{:}0\}$ states that there is no signalling from A to B when no measurement is performed at B.

The causal functions involved in the deterministic causal HVM for the empirical model over total (A, B, C) are also causal for the minimal supporting space, hence the empirical model is non-contextual/local there as well.

The initial aim of Leggett and Garg was to prove a result on the lines of non-locality for multiple space-like events and contextuality for single events. When considering sequential scenarios, however, a moment of thought can convince us that non-locality cannot be at stake here; there always exists a mechanism which feeds forward the information about the measurements performed in the past, simulating any correlation which is causally compatible with the causal order. With respect to the standard cover, there always exists a distribution of fine-tuned hidden mechanisms which are causal (no-signalling to the past) but classical. Sequential contextuality in the sheaf-theoretic sense cannot be achieved for a totally ordered sequence of events.

Our analysis showed that the Leggett-Garg results say something different, something which has little to say with a time-oriented generalisation of contextuality. The Leggett-Garg inequalities single out measurements which are always mutually non-disturbing (so, in principle, compatible with a no-signalling multipartite protocol); quantum correlations violate these inequalities for the simple reason that measurement is a disturbing mechanism. Leggett-Garg unequivocally shows that no non-disturbing (macroscopically realist) measurement models can account for the correlations obtained in quantum experiments but has little to say about contextuality.

Using similar ideas to uphold notions of contextuality in time needs a redefinition of the notion of contextuality. In this work, we have been conservative insofar as admitting as a global hidden variable mechanisms anything compatible with some overarching causal structure; we, therefore, started by seeing contextuality as the impossibility for any carefully crafted strategy to simulate the correlations. There is no a priori limit on the ontic structure of the theory, if not the impossibility of signalling to the past. Our approach is a generalisation of Bell's idea of non-locality. However, it is only a possible path, one which, to our belief, tries to preserve the adversarial ethos of the original arguments. A possibility to save contextuality in time would be to justify a limitation of the allowed hidden explanations of some empirical behaviour. Many approaches to contextuality believe this to be a justified requirement by arguing that the operational structure of quantum theory needs to be preserved at the ontic level. The aforementioned possibility has been argued by [89] to prove specific quantum advantages for sequential protocols. It is an interesting question if approaches of this type can be recast in the sheaf-theoretic form. Prima facie implementing these restrictions would entail a constrained assignment of causal functions to context, making the mathematics at play significantly harder. Any such restriction would also entail an additional degree of dependency on quantum theory which would need a careful conceptual reassessment of the notion of contextuality, perhaps on the line of the research developed by Spekkens [95, 127, 126]. A detailed analysis of the feasibility of such a research project is left for future work.

6.5 An OCB Empirical Model

In this example, we look consider two agents, Alice and Bob, acting within the context of the process matrix described by Oreshkov, Costa and Brukner in [101]:

$$W^{A_1A_2B_1B_2} = \frac{1}{4}[1^{A_1A_2B_1B_2} + \frac{1}{\sqrt{2}}(\sigma_z^{A_2}\sigma_z^{B_1} + \sigma_z^{A_1}\sigma_x^{B_1}\sigma_z^{B_2})]$$

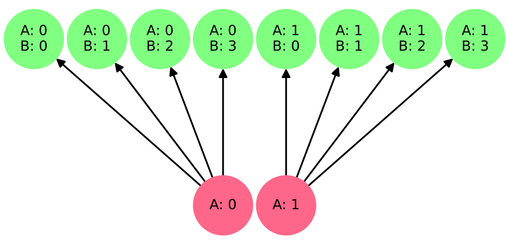
The two agents perform the following local instruments:

- Alice measures the incoming qubit in the Z basis, producing an output $x \in \{0,1\}$. She then encodes her input $a \in \{0,1\}$ into the Z basis of the outgoing qubit.
- Bob has input $(b, b') \in \{0,1\}^2$:
  - If $b' = 0$, Bob measures the incoming qubit in the X basis, obtaining a measurement outcome $z \in \{+,-\}$: if $z = +$, Bob prepares the outgoing qubit in $|b\rangle$; if $z = -$, Bob prepares the outgoing qubit in $|1-b\rangle$ instead. Regardless of the value of z, the output $y \in \{0,1\}$ of Bob is set constantly to $y = 0$.
  - If $b' = 1$, Bob measures the incoming qubit in the Z basis and uses the measurement outcome as his output $y \in \{0,1\}$. He prepares the outgoing qubit in $|0\rangle$, regardless of the value of b.

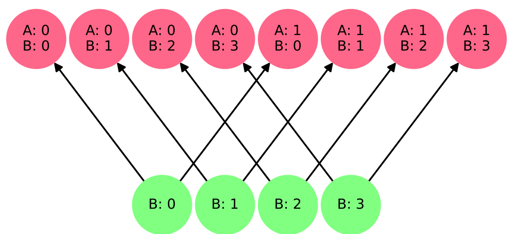
The description above results in the following empirical model on 2 events:

ABB	00	01	10	11
000	$1/4+\sqrt{2}/8$	$1/4+\sqrt{2}/8$	$1/4-\sqrt{2}/8$	$1/4-\sqrt{2}/8$
001	$1/4-\sqrt{2}/8$	$1/4-\sqrt{2}/8$	$1/4+\sqrt{2}/8$	$1/4+\sqrt{2}/8$
010	$1/4+\sqrt{2}/16$	$1/4-\sqrt{2}/16$	$1/4+\sqrt{2}/16$	$1/4-\sqrt{2}/16$
011	$1/4+\sqrt{2}/16$	$1/4-\sqrt{2}/16$	$1/4+\sqrt{2}/16$	$1/4-\sqrt{2}/16$
100	$1/4+\sqrt{2}/8$	$1/4+\sqrt{2}/8$	$1/4-\sqrt{2}/8$	$1/4-\sqrt{2}/8$
101	$1/4-\sqrt{2}/8$	$1/4-\sqrt{2}/8$	$1/4+\sqrt{2}/8$	$1/4+\sqrt{2}/8$
110	$1/4-\sqrt{2}/16$	$1/4+\sqrt{2}/16$	$1/4-\sqrt{2}/16$	$1/4+\sqrt{2}/16$
111	$1/4-\sqrt{2}/16$	$1/4+\sqrt{2}/16$	$1/4-\sqrt{2}/16$	$1/4+\sqrt{2}/16$

This is our first example of causally inseparable empirical model: the maximum causal fraction achieved over the causaltopes for the spaces induced by total (A, B) and total (B, A) (the maximal causally complete spaces on 2 events $\{A, B\}$) is around 93.9%. The particular decomposition achieving this fraction in the convex hull has components with the following causal fractions over the two individual causaltopes:



Hist (total (A, B), $\{0,1\}$)
Causal fraction: 29.3%



Hist (total (B, A), $\{0,1\}$)
Causal fraction: 64.6%

There is no ambiguity in the allocation above: the component for each causaltope has causal fraction 0% in the other causaltope. The two unnormalised components are shown below:

ABB	00	01	10	11
000	0.086	0.061	0.073	0.073
001	0.073	0.073	0.061	0.086
010	0.146	0.000	0.146	0.000
011	0.146	0.000	0.146	0.000
100	0.146	0.000	0.073	0.073
101	0.073	0.073	0.000	0.146
110	0.000	0.146	0.000	0.146
111	0.000	0.146	0.000	0.146

ABB	00	01	10	11
000	0.280	0.366	0.000	0.000
001	0.000	0.000	0.366	0.280
010	0.131	0.162	0.192	0.162
011	0.131	0.162	0.192	0.162
100	0.280	0.366	0.000	0.000
101	0.000	0.000	0.366	0.280
110	0.162	0.131	0.162	0.192
111	0.162	0.192	0.162	0.131

The maximum causal fraction supported by total (A, B) is 29.3% and the maximum causal fraction supported by total (B, A) is 64.6%. The maximum causal fraction supported by the no-signalling polytope is also 29.3%: shown below, this component is necessarily different from the 29.3% component shown on the left above, because it is supported by both total (A, B) and total (B, A).

ABB	00	01	10	11
000	0.000	0.146	0.073	0.073
001	0.073	0.073	0.146	0.000
010	0.146	0.000	0.000	0.146
011	0.146	0.000	0.146	0.000
100	0.000	0.146	0.073	0.073
101	0.073	0.073	0.146	0.000
110	0.146	0.000	0.000	0.146
111	0.146	0.000	0.146	0.000

6.6 The BFW Empirical Model

We now look at the empirical model introduced by Baumeler, Feix and Wolf in [19, 20], described by the authors as the 50%-50% mixture of a circular 'identity' classical process and a circular 'bitflip' classical process, for three agents Alice, Bob and Charlie.

ABC	000	001	010	011	100	101	110	111
000	$1/2$	0	0	0	0	0	0	$1/2$
001	0	0	0	$1/2$	$1/2$	0	0	0
010	0	$1/2$	0	0	0	0	$1/2$	0
011	0	0	$1/2$	0	0	$1/2$	0	0
100	0	0	$1/2$	0	0	$1/2$	0	0
101	0	$1/2$	0	0	0	0	$1/2$	0
110	0	0	0	$1/2$	$1/2$	0	0	0
111	$1/2$	0	0	0	0	0	0	$1/2$

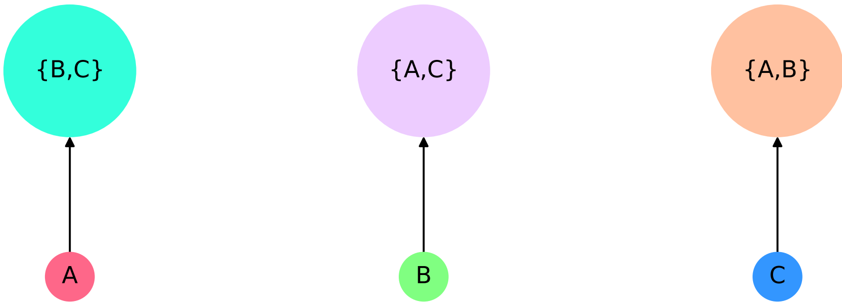
Specifically, the empirical model above is the 50%-50% mixture of the following causally inseparable functions (cf. Subsubsection 4.5.5 of [65]) for the causally incomplete space induced by the causal order indiscrete (A, B, C):

ABC	000	001	010	011	100	101	110	111
000	1	0	0	0	0	0	0	0
001	0	0	0	0	1	0	0	0
010	0	1	0	0	0	0	0	0
011	0	0	0	0	0	1	0	0
100	0	0	1	0	0	0	0	0
101	0	0	0	0	0	0	1	0
110	0	0	0	1	0	0	0	0
111	0	0	0	0	0	0	0	1

ABC	000	001	010	011	100	101	110	111
000	0	0	0	0	0	0	0	1
001	0	0	0	1	0	0	0	0
010	0	0	0	0	0	0	1	0
011	0	0	1	0	0	0	0	0
100	0	0	0	0	0	1	0	0
101	0	1	0	0	0	0	0	0
110	0	0	0	0	1	0	0	0
111	1	0	0	0	0	0	0	0

This is an example of a maximally causally inseparable empirical model: it has 0% support over all causally complete spaces on 3 events.

Interestingly, however, the BFW empirical model is 100% supported by each of the following 3 indefinite causal orders: this shows that either one of the 3 parties can be taken to act first, as long as the other two parties remain in indefinite causal order. In contrast, the two individual causally inseparable functions have 0% support over each of the 3 indefinite causal orders.



The meet of the associated spaces of input histories is the discrete space on 3 events: the associated no-signalling causaltope is 26-dimensional and supports 0% of the BFW empirical model. The intersection of the associated causaltopes, on the other hand, has dimension 38, and it supports 100% of the BFW empirical model.

We now look more in detail at the 3 indefinite causal order explanations: without loss of generality, we take Alice to act first. The absence of any support by the 2 total orders where Alice acts first means that no part of the indefinite causal order between Bob and Charlie is explainable by a fixed causal structure. Furthermore, the absence of any support by the 2 non-trivial switch orders where Alice acts first means that no part of the indefinite causal order between Bob and Charlie is controlled by Alice's input. Therefore, the only remaining explanation is that the indefinite causal order between Bob and Charlie is somehow correlated to Alice's outputs. To verify that this is indeed the case, we consider the scenarios corresponding to a fixed input choice by Alice (input 0 left below, input 1 right below), where the output of Alice has been discarded.

BC	00	01	10	11
00	1/2	0	0	1/2
01	1/2	0	0	1/2
10	0	1/2	1/2	0
11	0	1/2	1/2	0

BC	00	01	10	11
00	0	1/2	1/2	0
01	0	1/2	1/2	0
10	1/2	0	0	1/2
11	1/2	0	0	1/2

Unsurprisingly, the two restricted empirical models above are both causally separable. Perhaps surprisingly, they are both 100% supported by the no-signalling causaltope for Bob and Charlie. To understand whether Alice's output determines a fixed causal order between Bob and Charlie, we look at the empirical models obtained by conditioning on each of Alice's outputs (output 0 left below, output 1 right below), which have 50%-50% distribution independently of her input.

ABC	000	001	010	011
000	1	0	0	0
001	0	0	0	1
010	0	1	0	0
011	0	0	1	0
100	0	0	1	0
101	0	1	0	0
110	0	0	0	1
111	1	0	0	0

ABC	100	101	110	111
000	0	0	0	1
001	1	0	0	0
010	0	0	1	0
011	0	1	0	0
100	0	1	0	0
101	0	0	1	0
110	1	0	0	0
111	0	0	0	1

The empirical models above are deterministic and correspond to causal functions on the space of input histories determined by the order total $(A, \{\mathrm{B}, \mathrm{C}\})$. However, the two functions are causally inseparable, and hence so are the two empirical models. When Alice's input and output are both 0, i.e. in the first 4 rows of the empirical model left above, Bob and Charlie's outputs are related to their inputs in a way which requires bi-directional signalling:

$$\begin{aligned} o_B &= i_C \\ o_C &= \neg(i_B \wedge i_C) \end{aligned}$$

This evidence supports the intuition that causal inseparability for the BFW model stems from a cyclic order structure, as its very definition seems to suggest.

6.7 A Quantum Switch with Entangled Control

In this example we consider a single quantum switch between Alice and Bob, where the switch control is one qubit of a maximally entangled 2-qubit pair, with Charlie measuring the control qubit of the switch at the end and Diane measuring the other qubit in the entangled pair. Specifically, for angles $\gamma_0, \gamma_1 \in [0, \pi)$ and $\alpha \in [0, \sqrt{3}/6)$:

1. A 2-qubit Bell state $|\Phi^+\rangle$ is created: one qubit is sent to the control of a quantum switch, the other is sent to Charlie.

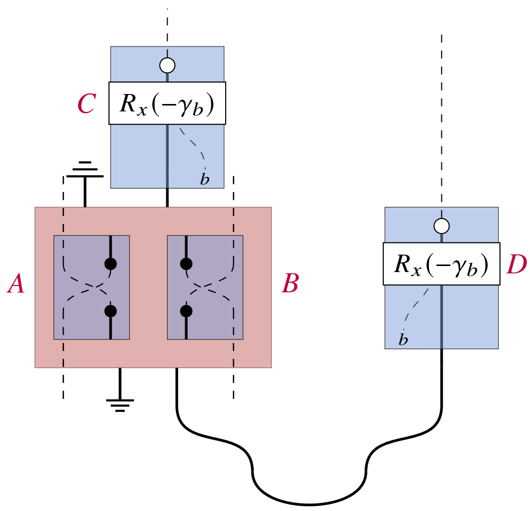
2. Alice and Bob are in a quantum switch, with one of the two $|\Phi^+\rangle$ qubits as its control and the maximally mixed state as the state for its input qubit:

   - Alice performs an X measurement on the incoming qubit and uses the measurement outcome as her output. Alice encodes her input into the X basis of the outgoing qubit, and then applies a Y rotation by an angle α to the qubit before forwarding it.
   - Bob performs an X measurement on the incoming qubit and uses the measurement outcome as his output. Bob encodes his input into the X basis of the outgoing qubit.

3. The output qubit of the switch is discarded. Charlie receives the control qubit, which he measures in a basis chosen as follows: on input 0, he first applies a X rotation by $-\gamma_0$ and then measures in the Z basis; on input 1, he first applies a X rotation by $-\gamma_1$ and then measures in the Z basis.

4. Diane receives the second qubit of the entangled state $|\Phi^+\rangle$, which she measures in a basis chosen in the same manner as Charlie: on input 0, she first applies a X rotation by $-\gamma_0$ and then measures in the Z basis; on input 1, she first applies a X rotation by $-\gamma_1$ and then measures in the Z basis.

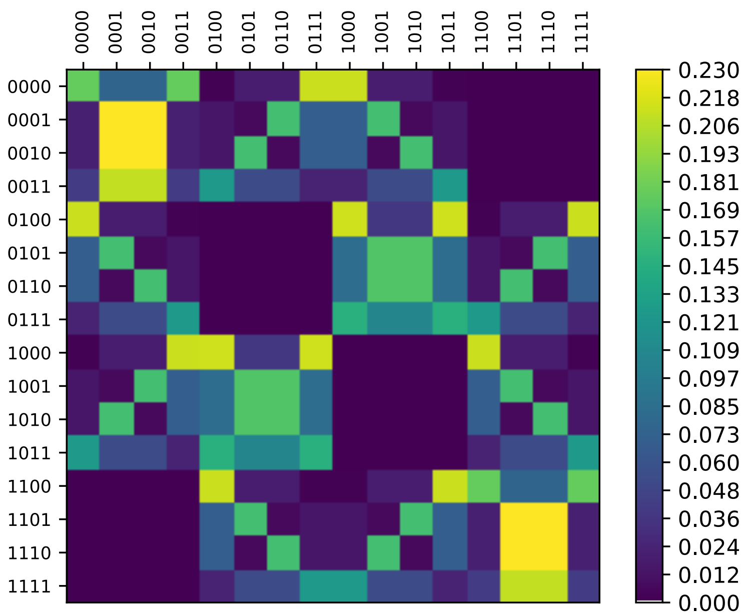
The figure below exemplifies the scenario we have just described, in the case where $\alpha = 0$:



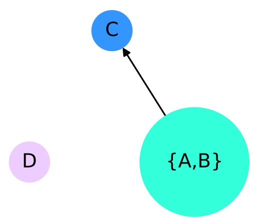
The description above results in the following empirical model on 4 events, for parameters $\alpha = 0$, $\gamma_0 = \frac{47\pi}{256}$ and $\gamma_1 = \frac{162\pi}{256}$.

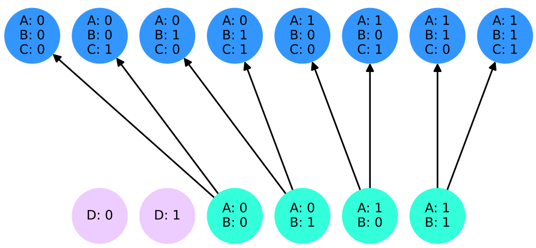
ABCD	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	0.176	0.074	0.074	0.176	0.002	0.019	0.019	0.211	0.211	0.019	0.019	0.002	0.000	0.000	0.000	0.000
0001	0.020	0.230	0.230	0.020	0.014	0.006	0.161	0.068	0.068	0.161	0.006	0.014	0.000	0.000	0.000	0.000
0010	0.020	0.230	0.230	0.020	0.014	0.161	0.006	0.068	0.068	0.006	0.161	0.014	0.000	0.000	0.000	0.000
0011	0.041	0.209	0.209	0.041	0.123	0.052	0.052	0.022	0.022	0.052	0.052	0.123	0.000	0.000	0.000	0.000
0100	0.211	0.019	0.019	0.002	0.000	0.000	0.000	0.000	0.213	0.037	0.037	0.213	0.002	0.019	0.019	0.211
0101	0.068	0.161	0.006	0.014	0.000	0.000	0.000	0.000	0.083	0.167	0.167	0.083	0.014	0.006	0.161	0.068
0110	0.068	0.006	0.161	0.014	0.000	0.000	0.000	0.000	0.083	0.167	0.167	0.083	0.014	0.161	0.006	0.068
0111	0.022	0.052	0.052	0.123	0.000	0.000	0.000	0.000	0.146	0.104	0.104	0.146	0.123	0.052	0.052	0.022
1000	0.002	0.019	0.019	0.211	0.213	0.037	0.037	0.213	0.000	0.000	0.000	0.000	0.211	0.019	0.019	0.002
1001	0.014	0.006	0.161	0.068	0.083	0.167	0.167	0.083	0.000	0.000	0.000	0.000	0.068	0.161	0.006	0.014
1010	0.014	0.161	0.006	0.068	0.083	0.167	0.167	0.083	0.000	0.000	0.000	0.000	0.068	0.006	0.161	0.014
1011	0.123	0.052	0.052	0.022	0.146	0.104	0.104	0.146	0.000	0.000	0.000	0.000	0.022	0.052	0.052	0.123
1100	0.000	0.000	0.000	0.000	0.211	0.019	0.019	0.002	0.002	0.019	0.019	0.211	0.176	0.074	0.074	0.176
1101	0.000	0.000	0.000	0.000	0.068	0.161	0.006	0.014	0.014	0.006	0.161	0.068	0.020	0.230	0.230	0.020
1110	0.000	0.000	0.000	0.000	0.068	0.006	0.161	0.014	0.014	0.161	0.006	0.068	0.020	0.230	0.230	0.020
1111	0.000	0.000	0.000	0.000	0.022	0.052	0.052	0.123	0.123	0.052	0.052	0.022	0.041	0.209	0.209	0.041

Since the table is large and the entries are irregular, below is a heat-map representation of the same empirical model, making its block structure more readily apparent:

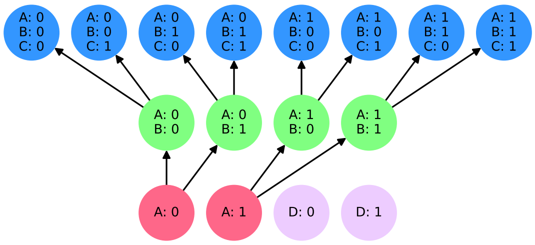


By construction, the empirical model is 100% supported by the causally incomplete space of input histories shown right below, induced by the indefinite causal order shown left below.

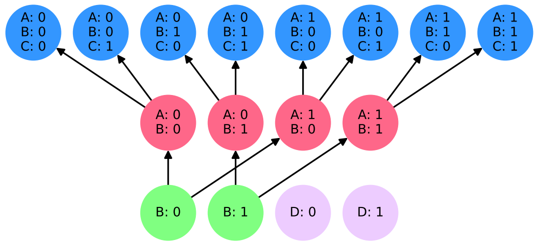




The empirical model is causally inseparable for the space right above, with a causal fraction of around 81.169% over its two causal completions. The two completions are induced by the causal orders total $(A, B, C) \vee$ discrete (D) and total $(B, A, C) \vee$ discrete (D), and their causaltopes are 128-dimensional.

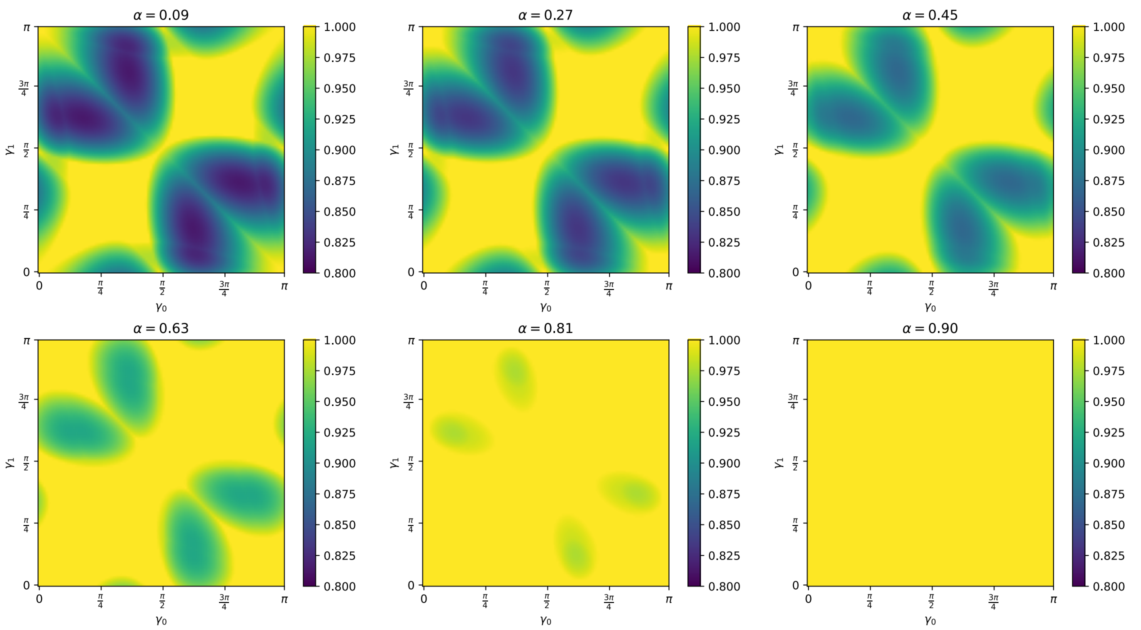


Causal fraction: 39.164%



Causal fraction: 42.005%

The causal fractions can be slightly shifted one way or the other: around 0.601% out of the 39.164% on the left can also be explained by the space on the right, and around 1.420% out of the 42.005% on the right can also be explained by the space on the left. Below we plot the overall causally separable fraction as a function of the γ_0 and γ_1 measurement angles used by both Charlie and Diane, for a selection of increasing values of the angle α used by Alice:



The maximum achievable causal inseparability—that is, one minus the minimum achievable causally separable fraction—decreases as α increases: the manifold of empirical models spanned by $(\gamma_0, \gamma_1) \in [0, \pi)^2$ steadily retreats into the convex hull of the two causaltopes, and the model becomes causally separable for all values of the measurement angles around $\alpha \approx 0.9$ or beyond.

The model becomes causally separable if the no-signalling constraint to Diane is dropped: the model is 100% supported by the spaces of input histories induced by causal orders total (A, B, C, D) and total (B, A, C, D), two of the four causal completions of total $(\{\mathrm{A}, \mathrm{B}\}, C, D)$. From this observation we can conclude that the empirical model exhibits contextual causality (Definition 5.29).

6.8 Another Quantum Switch with Entangled Control

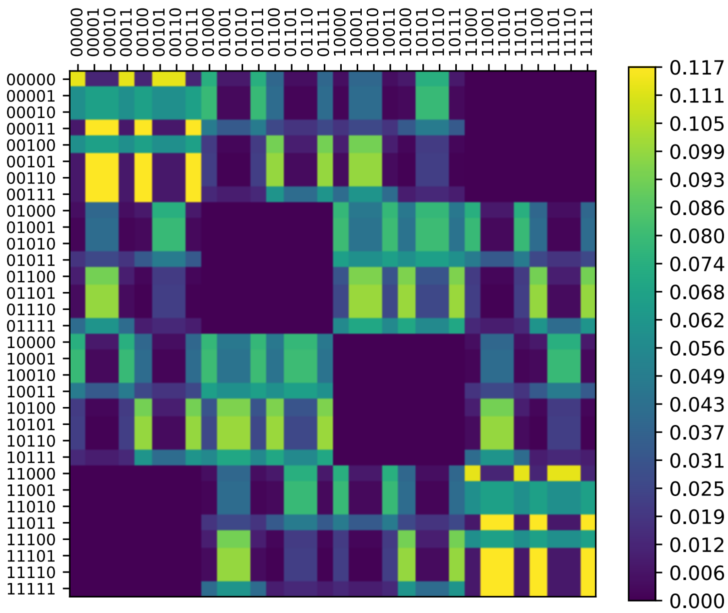
In this example we analyse a variation of the previous example where the control qubit for the quantum switch is now part of a 3-partite state, this time a GHZ state in the X basis:

$$\frac{1}{2} \sum_{b_0 \oplus b_1 \oplus b_2 = 0} |b_0 b_1 b_2\rangle$$

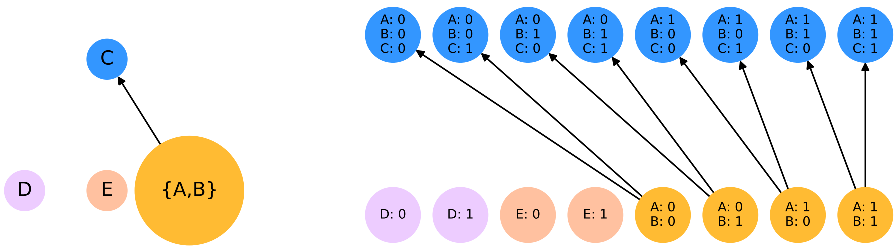
The setup is identical to that of the previous example, except that now there are two qubits entangled with the control qubit, one measured by Diane—as in the previous example—and one measured by Eve. Note that the choice of creating the GHZ state in the X basis is consistent with the entangled state used in the previous example:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}} \sum_{b_0 \oplus b_1 = 0} |b_0 b_1\rangle$$

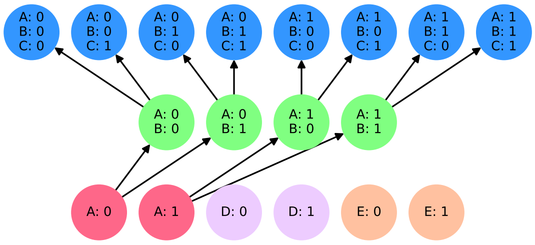
The description above results in the following empirical model on 4 events, for parameters $\alpha = 0$, $\gamma_0 = \frac{3\pi}{5}$ and $\gamma_1 = \frac{7\pi}{25}$. Since the table is very large and the entries are irregular, we resort to a heat-map representation of the empirical model, making its block structure more readily apparent



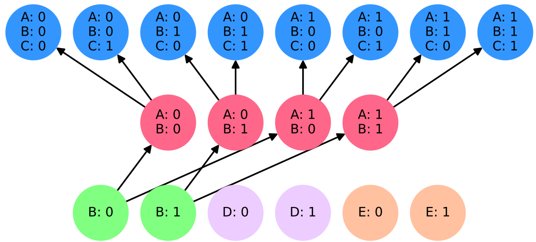
By construction, the empirical model is 100% supported by the causally incomplete space of input histories shown right below, induced by the indefinite causal order shown left below.



The empirical model above is causally inseparable for the space right above, with a causally separable fraction of around 70.582% over its two causal completions. The two completions are induced by the causal orders total $(A, B, C) \vee$ discrete (D, E) and total $(B, A, C) \vee$ discrete (D, E), and their causaltopes are 386-dimensional.

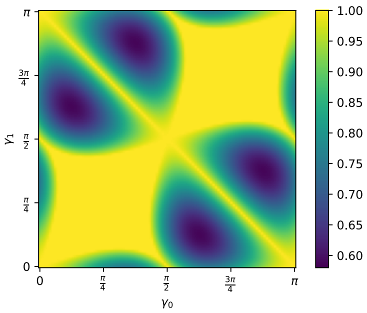


Causal fraction: 35.431%

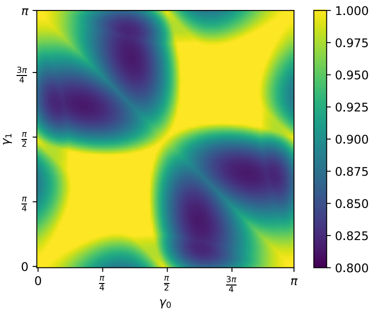


Causal fraction: 35.151%

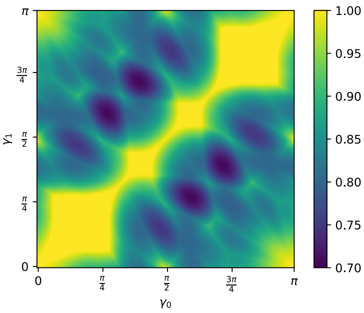
On the right below, we plot the causally separable fraction as a function of the γ_0 and γ_1 for the 3-partite GHZ case at $\alpha = 0$. On the centre below, we reproduce the causally separable fraction of the 2-partite Bell case at $\alpha = 0$, for comparison.



local fraction, Bell state



single switch, Bell state



single switch, GHZ state

Note how the GHZ variant has significantly lower minimum causally separable fraction, as well as significantly smaller flat plateaus of causal separability. The two figures on the centre and right above closely resemble Figures 1(a) and 1(b) of [5], showing the local fractions for measurements of the Bell and GHZ state: the local fraction for the Bell state is reproduced on the left above, for ease of comparison. This suggests a strong relationship between causal inseparability in our examples and contextuality of the underlying entangled states, a phenomenon which will be explored more in detail in Example 6.10 later on.

The model becomes causally separable if the no-signalling constraint to Diane and Eve is dropped: the model is 100% supported by the spaces of input histories induced by causal orders total (A, B, C, D, E) and total (B, A, C, D, E), two of the four causal completions of total $(\{\mathrm{A}, \mathrm{B}\}, C, D, E)$. From this observation we can conclude that the empirical model exhibits contextual causality (Definition 5.29).

6.9 Two Entangled Quantum Switches

In this example we consider two quantum switches—one between Alice and Bob, the other between Charlie and Diane—with entangled controls, where Eve measures the control of the Alice/Bob switch and Felix measures the control of the Charlie/Diane switch. Specifically, for angles $\gamma_0, \gamma_1 \in [0, \pi)$:

1. A 2-qubit Bell state $|\Phi^+\rangle$ is created: one qubit is sent to the control of the Alice/Bob switch, the other is sent to the control of the Charlie/Diane switch.

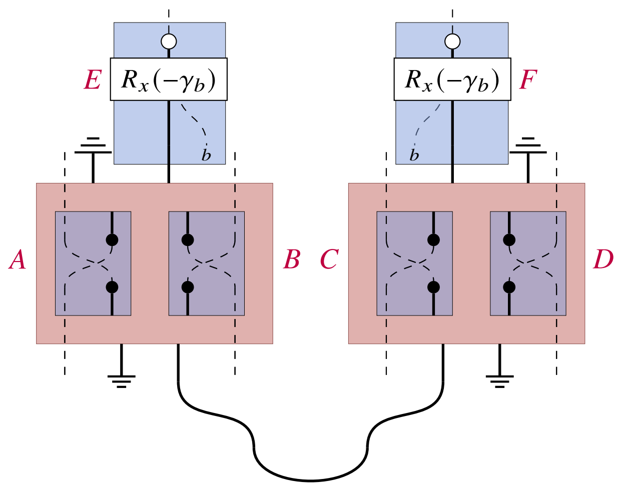
2. Alice and Bob are in the left quantum switch, with the first of the two $|\Phi^+\rangle$ qubits as its control and the maximally mixed state as the state for its input qubit.

3. Charlie and Diane are in the right quantum switch, with second of the two $|\Phi^+\rangle$ qubits as its control and the maximally mixed state as the state for its input qubit.

4. Alice, Bob, Charlie and Diane all do the same thing: they perform an X measurement on the incoming qubit, using the measurement outcome as their individual output, and then encode their individual input into the X basis of the outgoing qubit.

5. The output qubit of each switch is discarded. Eve and Felix receive the control qubit of the left and right switch respectively, and do the same thing: on input 0, they apply a X rotation by $-\gamma_0$ and then measure in the Z basis; on input 1, they first apply a X rotation by $-\gamma_1$ and then measure in the Z basis.

The figure below exemplifies the scenario we have just described:



We consider the empirical model on 6 events obtained for parameters $\gamma_0 = \frac{\pi}{5}$ and $\gamma_1 = \frac{3\pi}{5}$. Since the table is very large, we resort to a heat-map representation of the empirical model, making its block structure more readily apparent; this is shown in Figure 6.1 on p.211. By construction, the empirical model is 100% supported by the causally incomplete space of input histories shown right below, induced by the indefinite causal order shown left below.

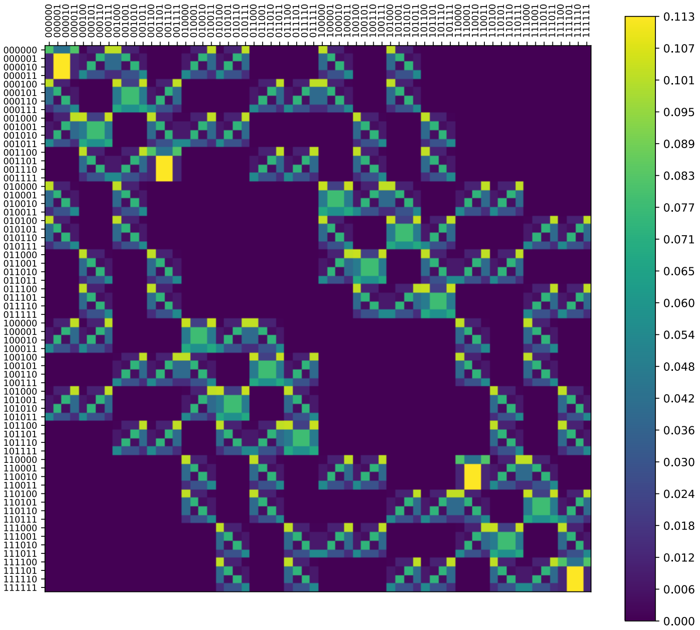
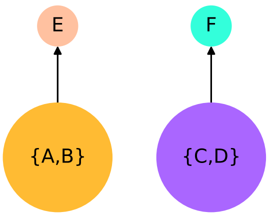
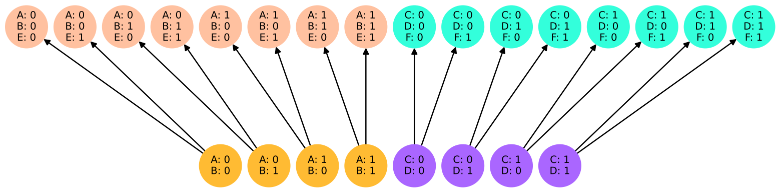
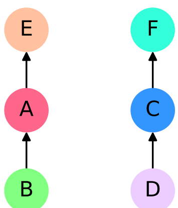


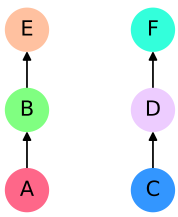
Figure 6.1: Empirical model for two entangled switches, with parameters $\gamma_0 = \frac{\pi}{5}$ and $\gamma_1 = \frac{3\pi}{5}$.





The empirical model is causally inseparable for the space right above, with a causally separable fraction of around 86.936% over two of its four causal completions and 0% over the other two. The two causal completions supporting a non-zero causal fraction of the empirical model are the spaces of input histories induced by the two definite causal orders on 6 events shown below, and their causaltopes are 1848-dimensional.

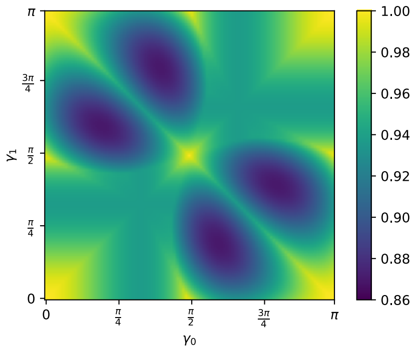




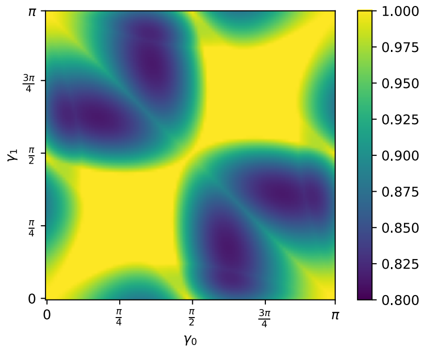
Causal fraction: 43.468%

Causal fraction: 43.468%

The causal fractions shown above are unambiguous: the component for each one of the two spaces is 0% supported by the other. On the left below, we plot the causally separable fraction as a function of the γ_0 and γ_1 measurement angles used by both Eve and Felix. On the right below, we reproduce the causally separable fraction of the Bell state single switch example at $\alpha = 0$, for comparison.



double switch, Bell state



single switch, Bell state

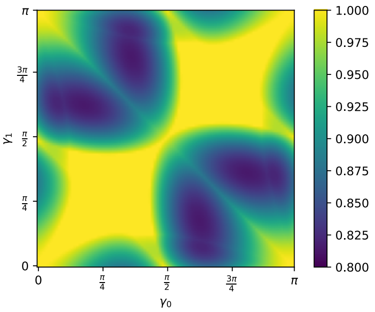
Note how the double switch has slightly higher minimum causally separable fraction over the (γ_0, γ_1) parameter space—86.936% instead of 81.169%—when compared to the single switch at $\alpha = 0$. However, also note how the double switch is causally inseparable at all angle values other than multiples of $\pi/2$, while the single switch shows flat plateaus of causal separability.

The causal separability of the empirical model increases if the no-signalling constraint between the two triples of agents is dropped: the remaining 13.063% is fully explained by the total orders total (C, D, F, A, B, E) and total (D, C, F, B, A, E). Note that an exhaustive search over all switch spaces would have been infeasible in this case: for 6 events, they are defined by 16511297126400 distinct sub-spaces of dimension around 2000, within an ambient space of dimension $\left(2^6\right)^2 = 4096$. We therefore conclude that the empirical model is causally separable with respect to the indiscrete preorder and exhibits contextual causality.

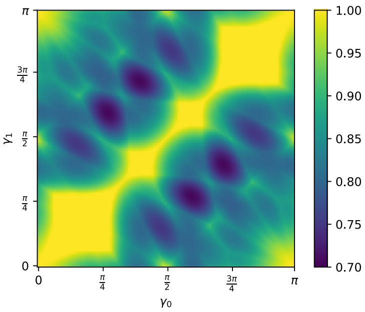
This example highlights the importance of defining causal separability relative to an ambient causal order (or, more generally, an ambient space of input histories): when no-signalling constraints between the two triples of agents are enforced—as could very much be in a real-world experimental scenario—then the model is causally inseparable, witnessing indefinite causal order. If no constraints are enforced, on the other hand, then the model becomes causally separable, and cannot be used as a witness of indefinite causal order. Hence, a relative notion of causal separability affords concrete additional opportunities in the exploration of indefinite causality.

6.10 Two Contextually Controlled Classical Switches.

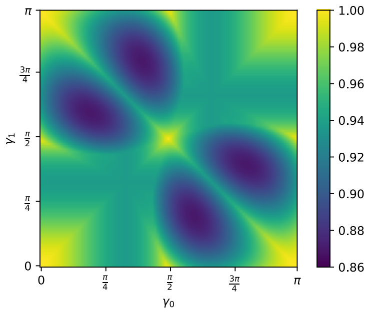
In the previous three examples, we have provided evidence of a connection between causal inseparability and contextuality: the figures below, depicting the causally separable fractions for the three examples as a function of the two measurement angles, closely resemble Figures 1(a) and 1(b) from [5], depicting the contextual fractions for analogous measurements of the Bell and GHZ state.



single switch, Bell state



single switch, GHZ state

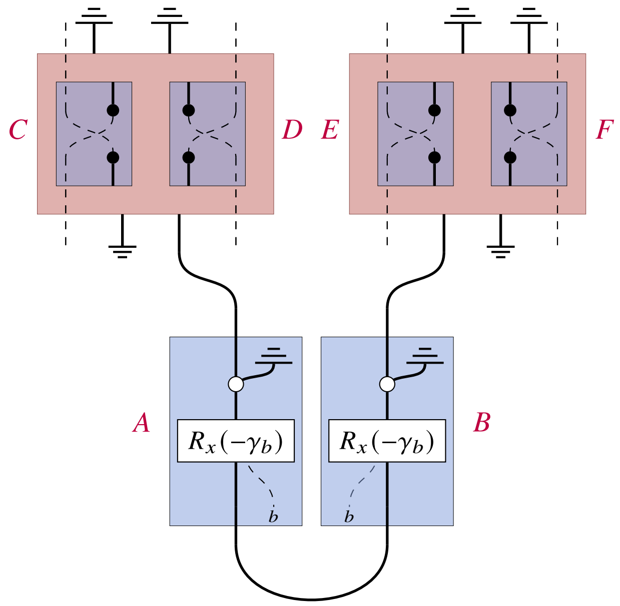


double switch, Bell state

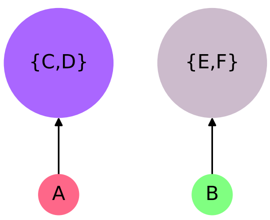
In this example, we make the connection between contextuality and causal inseparability explicit, by correlating the causal order of the parties in two bipartite quantum switches with the outcomes of a Bell experiment. This is, to our knowledge, the first demonstration of the phenomenon of *contextual causality*, where causal structure is correlated to contextual information, in such a way that non-locality/contextuality implies causal inseparability. Indeed, the causally separable fraction for the empirical model in this example is exactly the local fraction for the Bell experiment; in future work, we will also explore ways to characterise the connection in more general settings—such as those from the previous three examples—where the correlation between contextuality and causal structure might be imperfect. The experiment proceeds as follows, with Alice and Bob's output values fixed to 0:

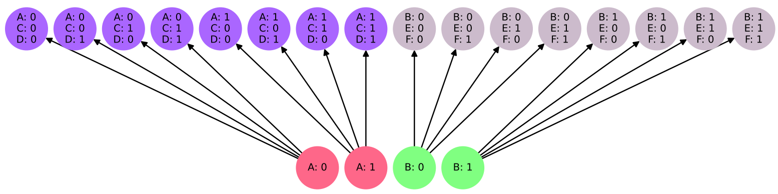
1. Two qubits in a Bell state $|\Phi^+\rangle$ are shared by Alice and Bob at the start of the protocol.
2. On input 0, Alice applies a $-\gamma_0$ X rotation to her qubit, then decoheres it in the Z basis; on input 1, Alice applies a $-\gamma_1$ X rotation to her qubit, then decoheres it in the Z basis. Alice forwards the qubit to the control of the switch between Charlie and Diane.
3. On input 0, Bob applies a $-\gamma_0$ X rotation to his qubit, then decoheres it in the Z basis; on input 1, Bob applies a $-\gamma_1$ X rotation to his qubit, then decoheres it in the Z basis. Bob forwards the qubit to the control of a switch between Eve and Felix.
4. Inside the switches, Charlie, Diane, Eve and Felix all do the same thing: they perform a Z measurement on the incoming qubit, using the measurement outcome as their individual output, and then encode their individual input into the Z basis of the outgoing qubit.
5. Both the outgoing qubit and the control qubit of the two switches are discarded.

The figure below exemplifies the scenario we have just described:



The description above gives rise to an empirical model on 6 events: for the moment, we focus on the case $\gamma_0 = 0, \gamma_1 = \frac{2\pi}{3}$. Since the table is very large, we resort to a heat-map representation of the empirical model, making its block structure more readily apparent; this is shown in Figure 6.2 on p.215. By construction, the empirical model is 100% supported by the causally incomplete space of input histories shown right below, induced by the indefinite causal order shown left below.





The empirical model is causally inseparable for the space right above: it has a causally separable fraction of 75%, coinciding with the local fraction for the Bell empirical model used to control the switches. The space has $2^2 \cdot 2^2 = 16$ causal completions: these are all possible combinations of the $2^2 = 4$ switch spaces on events $\{A, C, D\}$ having A as first event and the $2^2 = 4$ switch spaces on events $\{B, E, F\}$ having B as first event. Of the 16 causal completions, only the 6 spaces shown below support a non-zero fraction of the empirical model: each supports exactly 12.5%, with no fraction in common between spaces, for a total of $6 \cdot 12.5\% = 75\%$.

In the first space, Charlie succeeds Diane when Alice's input is 0 and precedes her when Alice's input is 1, while Eve succeeds Felix when Bob's input is 0 and precedes him when Bob's input is 1.

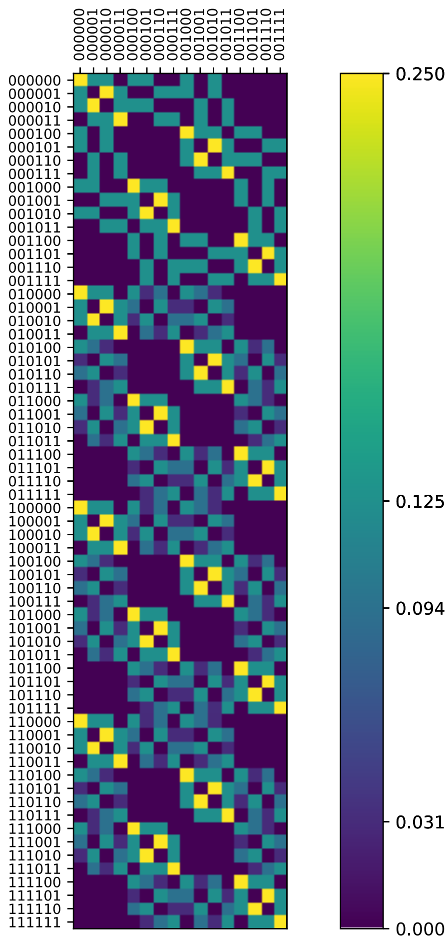
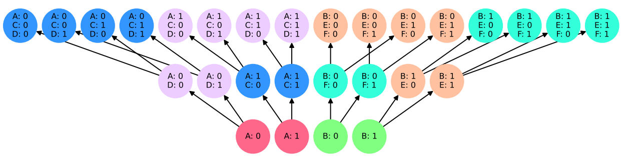
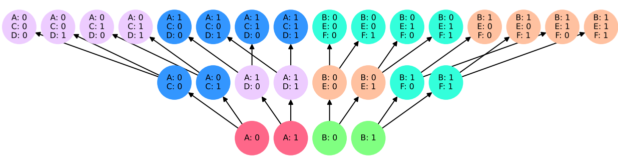


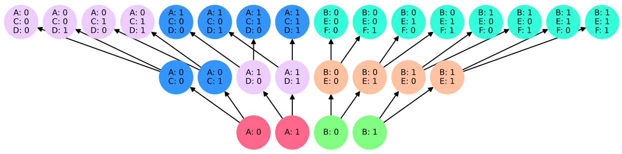
Figure 6.2: Empirical model for two contextually controlled switches.



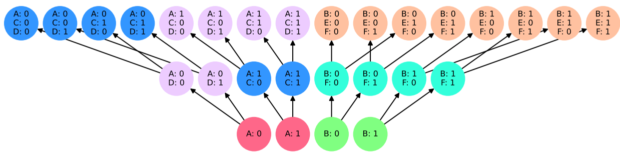
In the second space, Charlie precedes Diane when Alice's input is 0 and succeeds her when Alice's input is 1, while Eve precedes Felix when Bob's input is 0 and succeeds him when Bob's input is 1.



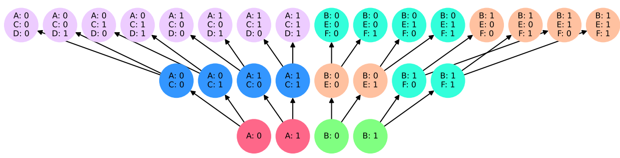
In the third space, Eve always precedes Felix, while Charlie precedes Diane when Alice's input is 0 and succeeds her when Alice's input is 1:



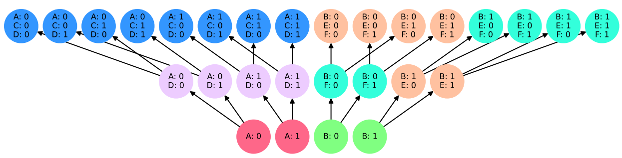
In the fourth space, Eve always succeeds Felix, while Charlie succeeds Diane when Alice's input is 0 and precedes her when Alice's input is 1:



In the fifth space, Charlie always precedes Diane, while Eve precedes Felix when Bob's input is 0 and succeeds him when Bob's input is 1:

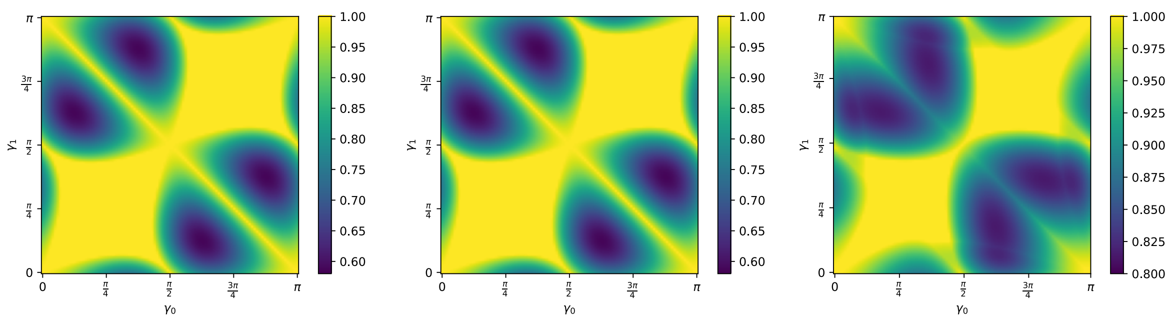


In the sixth and final space, Charlie always succeeds Diane, while Eve succeeds Felix when Bob's input is 0 and precedes him when Bob's input is 1:



Because its construction relies on contextuality of the inputs to the switches, we expect the empirical model to be causally separable once the no-signalling constraint between Alice and Bob is removed.

Below (left) we plot the causally separable fraction as a function of the angles $\gamma_0, \gamma_1 \in [0, \pi]$, which we compare to the local fraction for the Bell state (centre) and to the analogous landscape for the single switch (right). The causally separable fraction for this example matches the local fraction for the underlying Bell state exactly, providing unequivocal evidence for contextual causality in this setting.



contextually controlled switches — local fraction, Bell state — single switch, Bell state

Bibliography

[1] Alastair A. Abbott, Christina Giarmatzi, Fabio Costa, and Cyril Branciard. Multipartite causal correlations: Polytopes and inequalities. *Physical Review A*, 94(3), 2016.

[2] Alastair A. Abbott, Julian Wechs, Fabio Costa, and Cyril Branciard. Genuinely multipartite noncausality. *Quantum*, 1:39, 2017.

[3] Alastair A. Abbott, Julian Wechs, Dominic Horsman, Mehdi Mhalla, and Cyril Branciard. Communication through coherent control of quantum channels. *Quantum*, 4:333, 2020.

[4] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, cohomology and paradox. *Leibniz International Proceedings in Informatics (LIPIcs)*, 41:211-228, 2015.

[5] Samson Abramsky, Rui Soares Barbosa, and Shane Mansfield. Contextual fraction as a measure of contextuality. *Physical Review letters*, 119(5):050504, 2017.

[6] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036, 2011.

[7] Samson Abramsky and Bob Coecke. Categorical quantum mechanics. *Handbook of quantum logic and quantum structures*, 2:261–325, 2009.

[8] Samson Abramsky and Ross Duncan. A categorical quantum logic. *Mathematical Structures in Computer Science*, 16(3):469–489, 2006.

[9] Jonathan Allcock, Nicolas Brunner, Marcin Pawlowski, and Valerio Scarani. Recovering part of the boundary between quantum and nonquantum correlations from information causality. *Physical Review A*, 80(4):040103, 2009.

[10] John-Mark A. Allen, Jonathan Barrett, Dominic C. Horsman, Ciarán M. Lee, and Robert W. Spekkens. Quantum Common Causes and Quantum Causal Models. *Physical Review X*, 7(3), 2017.

[11] Mateus Araújo, Adrien Feix, Fabio Costa, and Časlav Brukner. Quantum circuits cannot control unknown operations. *New Journal of Physics*, 16(9):093026, 2014.

[12] Mateus Araújo, Adrien Feix, Miguel Navascués, and Časlav Brukner. A purification postulate for quantum mechanics with indefinite causal order. *Quantum*, 1:10, 2017.

[13] Mateus Araújo, Cyril Branciard, Fabio Costa, Adrien Feix, Christina Giarmatzi, and Caslav Brukner. Witnessing causal nonseparability. *New Journal of Physics*, 17(10):102001, 2015.

[14] Guido Bacciagaluppi. Leggett-Garg Inequalities, Pilot Waves and Contextuality. *arXiv preprint arXiv:1409.4104*, 2014.

[15] Jonathan Barrett and Adrian Kent. Non-Contextuality, Finite Precision Measurement and the Kochen–Specker Theorem. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 35(2):151–176, 2004.

[16] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu, and David Roberts. Nonlocal correlations as an information-theoretic resource. *Physical Review A*, 71(2), 2005.

[17] Jonathan Barrett, Robin Lorenz, and Ognyan Oreshkov. Quantum causal models. *arXiv preprint arXiv:1906.10726*, 2019.

[18] Jonathan Barrett, Robin Lorenz, and Ognyan Oreshkov. Cyclic quantum causal models. *Nature communications*, 12(1):1–15, 2021.

[19] Ämin Baumeler, Adrien Feix, and Stefan Wolf. Maximal incompatibility of locally classical behavior and global causal order in multiparty scenarios. *Physical Review A*, 90:042106, 2014.

[20] Ämin Baumeler and Stefan Wolf. Perfect signaling among three parties violating predefined causal order. In *2014 IEEE International Symposium on Information Theory*, pages 526–530. IEEE, 2014.

[21] Ämin Baumeler and Stefan Wolf. The space of logically consistent classical processes without causal order. *New Journal of Physics*, 18(1):013036, 2016.

[22] Jessica Bavaresco, Mateus Araújo, Časlav Brukner, and Marco Túlio Quintino. Semi-device-independent certification of indefinite causal order. *Quantum*, 3:176, 2019.

[23] Alessandro Bisio and Paolo Perinotti. Theoretical framework for higher-order quantum theory. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2225):20180706, 2019.

[24] Richard Blute and Marc Comeau. Von neumann categories. *Applied Categorical Structures*, 23(5):725–740, 2015.

[25] Niels Bohr. Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Physical Review*, 48:696–702, 1935.

[26] Niels Bohr. Causality and Complementarity. *Philosophy of Science*, 4(3):289–298, 1937.

[27] Niels Bohr. On the notions of causality and complementarity. *Dialectica*, 2(3-4):312–319, 1948.

[28] Niels Bohr. VI - Discussion with Einstein on Epistemological Problems in Atomic Physics. In Jørgen Kalckar, editor, *Foundations of Quantum Physics II (1933–1958)*, volume 7 of *Niels Bohr Collected Works*, pages 339–381. Elsevier, 1996.

[29] Luca Bombelli, Joohan Lee, David Meyer, and Rafael D. Sorkin. Space-time as a causal set. *Physical Review letters*, 59(5):521, 1987.

[30] George Boole. XII. On the theory of probabilities. *Philosophical Transactions of the Royal Society of London*, 152:225–252, 1862.

[31] Cyril Branciard, Mateus Araújo, Adrien Feix, Fabio Costa, and Časlav Brukner. The simplest causal inequalities and their violation. *New Journal of Physics*, 18(1):013008, 2015.

[32] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio. Bilocal versus nonbilocal correlations in entanglement-swapping experiments. *Physical Review A*, 85:032119, 2012.

[33] Costantino Budroni, Adán Cabello, Otfried Gühne, Matthias Kleinmann, and Jan Åke Larsson. Kochen-Specker Contextuality. *arXiv preprint arXiv:2102.13036*, 2021.

[34] Adán Cabello. Converting Contextuality into Nonlocality. *Physical Review Letters*, 127(7), 2021.

[35] Eric G. Cavalcanti. Classical Causal Models for Bell and Kochen-Specker Inequality Violations Require Fine-Tuning. *Physical Review X*, 8:021018, 2018.

[36] Eric G. Cavalcanti and Raymond Lal. On Modifications of Reichenbach's Principle of Common Cause in Light of Bell's Theorem. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424018, 2014.

[37] Rafael Chaves, Christian Majenz, and David Gross. Information–theoretic implications of quantum causal structures. *Nature Communications*, 6(1):5766, 2015.

[38] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Probabilistic theories with purification. *Physical Review A*, 81(6):062348, 2010.

[39] Giulio Chiribella, Giacomo Mauro D'Ariano, Paolo Perinotti, and Benoit Valiron. Quantum computations without definite causal structure. *Physical Review A*, 88(2), 2013.

[40] Giulio Chiribella and Hlér Kristjánsson. Quantum Shannon theory with superpositions of trajectories. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475(2225):20180903, 2019.

[41] Marios Christodoulou and Carlo Rovelli. On the possibility of laboratory evidence for quantum superposition of geometries. *Physics Letters B*, 792:64–68, 2019.

[42] Bob Coecke, Chris Heunen, and Aleks Kissinger. Categories of quantum and classical channels. *Quantum Information Processing*, 15(12):5179–5209, 2016.

[43] Bob Coecke and Raymond Lal. Time asymmetry of probabilities versus relativistic causal structure: An arrow of time. *Physical Review Letters*, 108(20):200403, 2012.

[44] Bob Coecke and Raymond Lal. Causal categories: Relativistically interacting processes. *Foundations of Physics*, 43(4):458–501, 2013.

[45] Daniel Collins, Nicolas Gisin, Sandu Popescu, David Roberts, and Valerio Scarani. Bell-Type Inequalities to Detect Truen-Body Nonseparability. *Physical Review Letters*, 88(17), 2002.

[46] Fabio Costa and Sally Shrapnel. Quantum causal modelling. *New Journal of Physics*, 18(6):063032, 2016.

[47] Giacomo Mauro D'Ariano, Giulio Chiribella, and Paolo Perinotti. *Quantum Theory from First Principles: An Informational Approach*. Cambridge University Press, 2017.

[48] Julio I. de Vicente. Simple conditions constraining the set of quantum correlations. *Physical Review A*, 92(3):032103, 2015.

[49] Qingxiuxiong Dong, Shojun Nakayama, Akihito Soeda, and Mio Murao. Controlled quantum operations and combs, and their applications to universal controllization of divisible unitary operations. *arXiv preprint arXiv:1911.01645*, 2019.

[50] Hippolyte Dourdent, Alastair A. Abbott, Nicolas Brunner, Ivan Šupić, and Cyril Branciard. Semi-device-independent Certification of Causal Nonseparability with Trusted Quantum Inputs. *Quantum*, 3:176, 2021.

[51] Ehtibar N. Dzhafarov, Janne V. Kujala, and Victor H. Cervantes. Contextuality-by-Default: A Brief Overview of Ideas, Concepts, and Terminology. In Harald Atmanspacher, Thomas Filk, and Emmanuel Pothos, editors, *Quantum Interaction*, pages 12–23. Springer International Publishing, 2016.

[52] Andreas Döring and Chris Isham. What is a thing?: Topos theory in the foundations of physics. *Lecture Notes in Physics*, page 753–937, 2010.

[53] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.

[54] Abert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, 1935.

[55] Arthur Fine. Hidden Variables, Joint Probability, and the Bell Inequalities. *Physical Review Letters*, 48:291–295, 1982.

[56] Thomas C. Fraser and Elie Wolfe. Causal compatibility inequalities admitting quantum violations in the triangle structure. *Physical Review A*, 98(2), 2018.

[57] Tobias Fritz. Beyond Bell's theorem: correlation scenarios. *New Journal of Physics*, 14(10):103001, 2012.

[58] Tobias Fritz. Beyond bell's theorem ii: Scenarios with arbitrary causal structure. *Communications in Mathematical Physics*, 341(2):391–434, 2016.

[59] Nicolas Gisin, Jean-Daniel Bancal, Yu Cai, Patrick Remy, Armin Tavakoli, Emmanuel Zambrini Cruzeiro, Sandu Popescu, and Nicolas Brunner. Constraints on nonlocality in networks from no-signaling and independence. *Nature Communications*, 11(1):2378, 2020.

[60] Stefano Gogioso. A process-theoretic church of the larger hilbert space. *arXiv preprint arXiv:1905.13117*, 2019.

[61] Stefano Gogioso and Nicola Pinzani. The sheaf-theoretic structure of definite causality. *Electronic Proceedings in Theoretical Computer Science*, 343:301–324, sep 2021.

[62] Stefano Gogioso and Nicola Pinzani. Classification of causally complete spaces on 3 events with binary inputs. *To appear*, 2023.

[63] Stefano Gogioso and Nicola Pinzani. The Combinatorics of Causality. *To appear*, 2023.

[64] Stefano Gogioso and Nicola Pinzani. The Geometry of Causality. *To appear*, 2023.

[65] Stefano Gogioso and Nicola Pinzani. The Topology of Causality. *To appear*, 2023.

[66] Stefano Gogioso and Carlo Maria Scandolo. Categorical probabilistic theories. *EPTCS*, 266:367–385, 2018.

[67] Stefano Gogioso, Maria E. Stasinou, and Bob Coecke. Functorial Evolution of Quantum Fields. *Frontiers in Physics*, 9:534265, 2021.

[68] Koon Tong Goh, J ędrzej Kaniewski, Elie Wolfe, Tamás Vértesi, Xingyao Wu, Yu Cai, Yeong-Cherng Liang, and Valerio Scarani. Geometry of the set of quantum correlations. *Physical Review A*, 97:022104, 2018.

[69] Kaumudibikash Goswami, Christina Giarmatzi, Michael Kewming, Fabio Costa, Cyril Branciard, Jacquiline Romero, and Andrew G. White. Indefinite Causal Order in a Quantum Switch. *Physical Review Letters*, 121(9):090503, 2018.

[70] Yelena Guryanova, Ralph Silva, Anthony J. Short, Paul Skrzypczyk, Nicolas Brunner, and Sandu Popescu. Exploring the limits of no backwards in time signalling. *Quantum*, 3:211, 2019.

[71] Lucien Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Phys. Rev. Lett.*, 71:1665–1668, 1993.

[72] Lucien Hardy. Quantum Gravity Computers: On the Theory of Computation with Indefinite Causal Structure. *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle*, page 379–401, 2009.

[73] James Hefford and Aleks Kissinger. On the pre- and promonoidal structure of spacetime. *arXiv preprint arXiv:2206.09678*, 2022.

[74] Joe Henson, Raymond Lal, and Matthew F Pusey. Theory-independent limits on correlations from generalized Bayesian networks. *New Journal of Physics*, 16(11):113043, 2014.

[75] Timothée Hoffreumon and Ognyan Oreshkov. The Multi-round Process Matrix. *Quantum*, 5:384, 2021.

[76] Tom Holden-Dye and Sandu Popescu. Indefinite global time. *arXiv preprint arXiv:2101.08739*, 2021.

[77] Paweł Horodecki and Ravishankar Ramanathan. The relativistic causality versus no-signaling paradigm for multi-party correlations. *Nature Communications*, 10(1):1701, 2019.

[78] Chris J. Isham. *Topos Methods in the Foundations of Physics*, page 187–206. Cambridge University Press, 2011.

[79] Adrian Kent. Noncontextual Hidden Variables and Physical Measurements. *Physical Review Letters*, 83(19):3755–3757, 1999.

[80] Aleks Kissinger, Matty Hoban, and Bob Coecke. Equivalence of relativistic causal structure and process terminality. *arXiv preprint arXiv:1708.04118*, 2017.

[81] Aleks Kissinger and Sander Uijlen. A categorical semantics for causal structure. In *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–12. IEEE, 2017.

[82] Raymond Klibansky and Niels Bohr. VIII - Quantum Physics and Philosophy–Causality and Complementarity. In Jørgen Kalckar, editor, *Foundations of Quantum Physics II (1933–1958)*, volume 7 of *Niels Bohr Collected Works*, pages 385–394. Elsevier, 1996.

[83] Ravi Kunjwal and Robert W. Spekkens. From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism. *Physical Review Letters*, 115:110403, 2015.

[84] Anthony Leggett and Anupam Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters*, 54:857–860, 1985.

[85] Yeong-Cherng Liang, Robert W. Spekkens, and Howard M. Wiseman. Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. *Physics Reports*, 506(1):1–39, 2011.

[86] Robin Lorenz and Jonathan Barrett. Causal and compositional structure of unitary transformations. *Quantum*, 5:511, 2021.

[87] David B. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics*, 18(7):1399–1404, 1977.

[88] Shane Mansfield and Rui Soares Barbosa. Extendability in the Sheaf-theoretic Approach: Construction of Bell Models from Kochen-Specker Models. arXiv preprint arXiv:1402.4827, 2014.

[89] Shane Mansfield and Elham Kashefi. Quantum advantage from sequential-transformation contextuality. *Physical Review letters*, 121(23):230401, 2018.

[90] Fotini Markopoulou. The Internal Description of a Causal Set: What the Universe Looks Like from the Inside. *Communications in Mathematical Physics*, 211(3):559–583, 2000.

[91] Chiara Marletto and Vladko Vedral. Gravitationally induced entanglement between two massive particles is sufficient evidence of quantum effects in gravity. *Physical Review Letters*, 119(24), 2017.

[92] Owen J E Maroney and Christopher G. Timpson. Quantum- vs. Macro- Realism: What does the Leggett-Garg Inequality actually test? *arXiv preprint arXiv:1412.6139*, 2014.

[93] Keye Martin and Prakash Panangaden. Domain theory and general relativity. In *New structures for physics*, pages 687–703. Springer, 2010.

[94] Keye Martin and Prakash Panangaden. Spacetime geometry from causal structure and a measurement. In Samson Abramsky and Michael Mislove, editors, *Mathematical Foundations of Information Flow: Clifford Lectures*, volume 71, pages 213–232. American Mathematical Society, 2012.

[95] Michael D. Mazurek, Matthew F. Pusey, Ravi Kunjwal, Kevin J. Resch, and Robert W. Spekkens. An experimental test of noncontextuality without unphysical idealizations. *Nature Communications*, 7(1):ncomms11780, 2016.

[96] N. David Mermin. A Kochen-Specker Theorem for Imprecisely Specified Measurement, 1999.

[97] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the Set of Quantum Correlations. *Physical Review Letters*, 98(1), 2007.

[98] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

[99] Daniel K. L. Oi. Interference of Quantum Channels. *Physical Review Letters*, 91:067902, 2003.

[100] Ognyan Oreshkov. Time-delocalized quantum subsystems and operations: on the existence of processes with indefinite causal structure in quantum mechanics. *Quantum*, 3:206, 2019.

[101] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. Quantum correlations with no causal order. *Nature Communications*, 3:1092, 2012.

[102] Ognyan Oreshkov and Christina Giarmatzi. Causal and causally separable processes. *New Journal of Physics*, 18(9):093020, 2016.

[103] Nikola Paunković and Marko Vojinović. Causal orders, quantum circuits and spacetime: distinguishing between definite and superposed causal orders. *Quantum*, 4:275, 2020.

[104] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[105] Judea Pearl and Erik Cavalcanti. Classical causal models cannot faithfully explain Bell nonlocality or Kochen-Specker contextuality in arbitrary scenarios. *Quantum*, 5:518, 2021.

[106] Jacques Pienaar and Časlav Brukner. A graph-separation theorem for quantum causal models. *New Journal of Physics*, 17(7):073020, 2015.

[107] Nicola Pinzani and Stefano Gogioso. Giving Operational Meaning to the Superposition of Causal Orders. *Electronic Proceedings in Theoretical Computer Science*, 340:256–278, 2021.

[108] Nicola Pinzani, Stefano Gogioso, and Bob Coecke. Categorical semantics for time travel. In *2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 1–20. IEEE, 2019.

[109] Itamar Pitowsky. *From George Boole To John Bell — The Origins of Bell's Inequality*, pages 37–49. Springer Netherlands, Dordrecht, 1989.

[110] Itamar Pitowsky. George Boole's 'Conditions of Possible Experience' and the Quantum Puzzle. *The British Journal for the Philosophy of Science*, 45(1):95–125, 1994.

[111] Itamar Pitowsky. *Quantum Probability — Quantum Logic*. Lecture Notes in Physics. Springer Berlin Heidelberg, 2014.

[112] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[113] Sandu Popescu and Daniel Rohrlich. Causality & Nonlocality as Axioms for Quantum Mechanics. In Geoffrey Hunter, Stanley Jeffers, and Jean-Pierre Vigier, editors, *Causality and Locality in Modern Physics*, page 383, 1998.

[114] John Power. Premonoidal categories as categories with algebraic structure. *Theoretical Computer Science*, 278(1):303–321, 2002. Mathematical Foundations of Programming Semantics 1996.

[115] Alejandro Pozas-Kerstjens, Nicolas Gisin, and Armin Tavakoli. Full Network Nonlocality. *Physical Review Letters*, 128(1), 2022.

[116] Lorenzo M. Procopio, Amir Moqanaki, Mateus Araújo, Fabio Costa, Irati Alonso Calafell, Emma G. Dowd, Deny R. Hamel, Lee A. Rozema, Časlav Brukner, and Philip Walther. Experimental superposition of orders of quantum gates. *Nature Communications*, 6(1):7913, 2015.

[117] Tom Purves and Anthony J. Short. Nonclassically causal correlations without backwards-in-time signaling. *Physical Review A*, 99:022101, 2019.

[118] Tom Purves and Anthony J. Short. Quantum Theory Cannot Violate a Causal Inequality. *Physical Review Letters*, 127(11), 2021.

[119] Hans Reichenbach. *The direction of time*. University of California Berkeley Press, 1956.

[120] Marc-Olivier Renou, Elisa Bäumer, Sadra Boreiri, Nicolas Brunner, Nicolas Gisin, and Salman Beigi. Genuine Quantum Nonlocality in the Triangle Network. *Physical Review Letters*, 123:140401, 2019.

[121] Alfred Arthur Robb. *A Theory of Time and Space*. Cambridge University Press, 1914.

[122] Giulia Rubino, Lee A. Rozema, Daniel Ebler, Hlér Kristjánsson, Sina Salek, Philippe Allard Guérin, Alastair A. Abbott, Cyril Branciard, Časlav Brukner, Giulio Chiribella, and Philip Walther. Experimental quantum communication enhancement by superposing trajectories. *Physical Review Research*, 3(1):013093, 2021.

[123] Giulia Rubino, Lee A. Rozema, Adrien Feix, Mateus Araújo, Jonas M. Zeuner, Lorenzo M. Procopio, Časlav Brukner, and Philip Walther. Experimental verification of an indefinite causal order. *Science Advances*, 3(3), 2017.

[124] Giulia Rubino, Lee A. Rozema, Francesco Massa, Mateus Araújo, Magdalena Zych, Časlav Brukner, and Philip Walther. Experimental Entanglement of Temporal Orders. In *Quantum Information and Measurement (QIM) V: Quantum Technologies*, page S3B.3. Optica Publishing Group, 2019.

[125] Sina Salek, Daniel Ebler, and Giulio Chiribella. Quantum communication in a superposition of causal orders. *arXiv preprint arXiv:1809.06655*, 2018.

[126] David Schmid, John H. Selby, Matthew F. Pusey, and Robert W. Spekkens. A structure theorem for generalized-noncontextual ontological models. *arXiv preprint arXiv:2005.07161*, 2020.

[127] David Schmid, John H. Selby, Elie Wolfe, Ravi Kunjwal, and Robert W. Spekkens. Characterization of Noncontextuality in the Framework of Generalized Probabilistic Theories. *PRX Quantum*, 2:010331, 2021.

[128] David Schmid, Robert W. Spekkens, and Elie Wolfe. All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences. *Physical Review A*, 97:062103, 2018.

[129] Peter Selinger. Dagger compact closed categories and completely positive maps: (extended abstract). *Electronic Notes in Theoretical Computer Science*, 170:139–163, 2007. Proceedings of the 3rd International Workshop on Quantum Programming Languages (QPL 2005).

[130] Sally Shrapnel and Fabio Costa. Causation does not explain contextuality. *Quantum*, 2:63, 2018.

[131] Christoph Simon, Časlav Brukner, and Anton Zeilinger. Hidden-Variable Theorems for Real Experiments. *Physical Review Letters*, 86(20):4427–4430, 2001.

[132] Ernst Specker Simon Kochen. The Problem of Hidden Variables in Quantum Mechanics. *Indiana Univ. Math. J.*, 17:59–87, 1968.

[133] Ernst Specker. Die Logik nicht gleichzeitig entscheidbarer Aussagen. volume 14, pages 239–246. 1960.

[134] Robert W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A*, 71(5), 2005.

[135] Sam Staton and Sander Uijlen. Effect algebras, presheaves, non-locality and contextuality. *Information and Computation*, 261:336–354, 2018.

[136] George Svetlichny. Distinguishing three-body from two-body nonseparability by a Bell-type inequality. *Physical Review D*, 35:3066–3069, 1987.

[137] Armin Tavakoli. Quantum Correlations in Connected Multipartite Bell Experiments. *arXiv preprint arXiv:1509.08491*, 2015.

[138] Augustin Vanrietvelde and Giulio Chiribella. Universal control of quantum processes using sector-preserving channels. *Quantum Information and Computation*, 21(15,16):1320–1352, 2021.

[139] Augustin Vanrietvelde, Hlér Kristjánsson, and Jonathan Barrett. Routed quantum circuits. *Quantum*, 5:503, 2021.

[140] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[141] Julian Wechs, Hippolyte Dourdent, Alastair A. Abbott, and Cyril Branciard. Quantum Circuits with Classical Versus Quantum Control of Causal Order. *PRX Quantum*, 2(3), 2021.

[142] Norbert Wiener. A theory of time and space. *Journal of Philosophy, Psychology and Scientific Methods*, 13(22):611–613, 1916.

[143] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. The Inflation Technique for Causal Inference with Latent Variables. *arXiv preprint arXiv:1609.00672*, 2016.

[144] Christopher J. Wood and Robert W. Spekkens. The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning. *New Journal of Physics*, 17:033002, 2015.

[145] Günter M. Ziegler. *Lectures on polytopes*. Springer-Verlag, New York, 1995.