

Analysing causal structures in generalised probabilistic theories

Mirjam Weilenmann^{1,2} and Roger Colbeck²

¹Institute for Quantum Optics and Quantum Information (IQOQI) Vienna, Austrian Academy of Sciences, Boltzmannngasse 3, 1090 Vienna, AT

²Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK

Causal structures give us a way to understand the origin of observed correlations. These were developed for classical scenarios, but quantum mechanical experiments necessitate their generalisation. Here we study causal structures in a broad range of theories, which include both quantum and classical theory as special cases. We propose a method for analysing differences between such theories based on the so-called measurement entropy. We apply this method to several causal structures, deriving new relations that separate classical, quantum and more general theories within these causal structures. The constraints we derive for the most general theories are in a sense minimal requirements of any causal explanation in these scenarios. In addition, we make several technical contributions that give insight for the entropic analysis of quantum causal structures. In particular, we prove that for any causal structure and for any generalised probabilistic theory, the set of achievable entropy vectors form a convex cone.

Given a set of observed variables, some of which may be correlated, a causal structure gives a more detailed picture of how the correlations come about. Depending on the situation, this causal structure may posit the existence of hidden common causes and the nature of these depends on the physical theory. For instance, the experimental violation of a Bell inequality [1] can be explained either by adapting the causal structure within the realm of classical physics (at the expense of resorting to fine-tuning [2]) or by allowing hidden systems to be non-classical.

Causal structures also provide a suitable basis

Mirjam Weilenmann: mirjam.weilenmann@oeaw.ac.at

Roger Colbeck: roger.colbeck@york.ac.uk

for analysing the features of different theories by allowing us to phrase communication and cryptographic protocols in terms of the dependencies among the involved systems. They help us predict the success of players engaged in a protocol when restricted according to different theories, for example, in random access coding and the related principle of Information Causality [3, 4].

The differences between the observable correlations that can be achieved with classical and quantum resources within a given causal structure have been extensively analysed, starting with the derivation of several classical constraints and their quantum violations [5, 6], and progressing to a systematic analysis [7–11]. Less work has been dedicated to understanding the limitations of quantum systems [9, 12, 13] and of the behaviour of theories beyond. For the latter, there have been analyses of the implications of the no-signalling principle on causal structures [14, 15]. More generally, understanding the differences of generalised probabilistic theories (GPTs) with respect to different tasks may inform the search for principles that single out quantum mechanics.

In this work, we introduce a technique for deriving constraints on the observable correlations that are achievable in different causal structures according to different GPTs, with the aim of moving towards a systematic analysis of the differences between such theories. Our approach is based on *measurement entropy* [16, 17] and inspired by entropic approaches to analysing causal structures involving classical and quantum resources [8, 9, 18–22]. That such a generalisation is possible, was not at all clear, since work regarding the definition of entropy in GPTs showed that there is no entropy measure that retains the relevant properties of the von Neumann entropy [16, 17]. In particular, the additivity of entropy under the composition of different systems

is not retained by the proposed measures, which in previous entropic approaches for analysing causal structures was crucial for encoding causal constraints.

One of the key points that allows us to overcome these issues is to explicitly include the conditional entropy in the analysis. Nevertheless, since our final results are stated in terms of the Shannon entropy they can be directly compared to those obtained with previous entropic techniques.

We apply our method to various causal structures, generating a series of entropic constraints that exclude certain causal explanations of observed correlations when restricted by arbitrary GPTs. This allows us to compare different causal structures with respect to the correlations they allow in different theories (in particular, we compare classical, quantum and arbitrary GPTs). In some cases we find the same sets of entropic description regardless of the theory (here known quantum inequalities also apply to GPTs), while in others we can show an entropic separation. For instance, we apply our technique to Information Causality [3], a candidate principle for singling out quantum theory, showing that our method improves upon that of [15], yielding the stronger inequalities of [23]. Although the maximally non-local GPT, *box-world* does not satisfy the notion of Information Causality, we identify minimal notions of causation that are satisfied.

In addition to providing a method for analysing causal structures with GPT resources, we make technical contributions by showing that any set of achievable entropy vectors for the observed variables in a causal structure involving quantum or other generalised probabilistic resources is a convex cone. Previously this had only been shown for the entropy vectors of classical resources [8, 20, 22, 24]. This insight allows for easy comparison of the entropic sets within different theories, and in some cases enables us to prove that a given characterisation is complete by showing that all extremal points are achievable. We also give some insights into the entropic analysis of quantum causal structures.

1 Preliminaries

For every system A in a GPT, there is an associated state space \mathcal{S}_A , a compact convex subset

of a real vector space V and an associated space of effects, \mathcal{F}_A . An *effect* $e \in \mathcal{F}_A$ is a linear map $\mathcal{S}_A \rightarrow [0, 1]$ (thus, e is a vector in the dual space to V). There is a special effect, $u_A \in \mathcal{F}_A$, called the unit effect, with the property that $u_A(s) = 1$ for all $s \in \mathcal{S}_A$. A *measurement* M is a collection of effects whose sum is the unit effect, i.e., we can write $M = \{e^x \in \mathcal{F}_A : \sum_x e^x = u_A\}$. We use \mathcal{E}_A to represent the set of allowed measurements on A . The interpretation of $e^x(s)$ is the probability of outcome x when M is performed on a system in state s .

Consider two measurements on A : $M = \{e^x\}_{x \in \mathcal{R}_M}$ and $N = \{f^y\}_{y \in \mathcal{R}_N}$. If there exists a map $F : \mathcal{R}_M \rightarrow \mathcal{R}_N$ such that

$$\sum_{x \in \mathcal{R}_M : F(x)=y} e^x = f^y \quad \forall y \in \mathcal{R}_N \quad (1)$$

we say that M is a *refinement* of N (equivalently, N is a *coarse-graining* of M).¹ A refinement is *trivial* if for all $x \in \mathcal{R}_M$, $e^x = c(x) f^{F(x)}$ for some $c(x) \in \mathbb{R}_{>0}$. The subset of *fine-grained measurements*, \mathcal{E}_A^* , are those for which there are no non-trivial refinements. Throughout this article we restrict to GPTs where there is at least one finite-outcome fine-grained measurement (in classical and quantum theory this is a restriction to finite-dimensional systems).

Transformations of systems are represented by linear maps between state spaces, $T : \mathcal{S}_A \rightarrow \mathcal{S}_B$ and the set of such transformations is denoted $\mathcal{T}_{A \rightarrow B}$. The set $\mathcal{T}_{A \rightarrow A}$, contains the identity transformation, I_A , and is closed under composition. Furthermore, a transformation followed by a measurement is a valid measurement.

Two systems A and B can be thought of as parts of a single joint system AB . We do not specify precisely what the joint state space is, but a minimal requirement is that if $s_A \in \mathcal{S}_A$ and $s_B \in \mathcal{S}_B$ then $s_A \otimes s_B \in \mathcal{S}_{AB}$. States that can be written as $s_A \otimes s_B$ are called *product* and convex combinations thereof are *separable*. Analogously, measurements $M = \{e^x\}_{x \in \mathcal{R}_M} \in \mathcal{E}_A$ and $N = \{f^y\}_{y \in \mathcal{R}_N} \in \mathcal{E}_B$ can be composed into $L = M \otimes N \in \mathcal{E}_{AB}$ with outcome set $\mathcal{R}_L = \mathcal{R}_M \times \mathcal{R}_N$ and effects $g^{x,y} = e^x \otimes f^y$. Product effects act on product states according to $(e^x \otimes f^y)(s_A \otimes s_B) = e^x(s_A) f^y(s_B)$.

¹If F is bijective, (1) is a *relabelling*. The set of measurements \mathcal{E}_A is assumed to be closed under relabelling and coarse-graining.

This implies that we have non-signalling theories: Suppose $\{e_a^x\}_x \in \mathcal{E}_A$ and $\{e_b^y\}_y \in \mathcal{E}_B$ are measurements for $a = 1, \dots, n_a$ and $b = 1, \dots, n_b$, then, for example,

$$p(y|a, b) = \sum_x (e_a^x \otimes f_b^y)(s_{AB}) = (u_A \otimes f_b^y)(s_{AB}),$$

which is independent of a .

We also assume that there are well-defined reduced states: $\forall s_{AB} \in \mathcal{S}_{AB} \exists s_A \in \mathcal{S}_A$ s.t. $\forall e \in \mathcal{F}_A, e(s_A) = (e \otimes u_B)(s_{AB})$. The post-measurement state on A after a measurement on B with outcome x is

$$s_{A|x} = \frac{(I_A \otimes e^x)(s_{AB})}{e^x(s_B)}. \quad (2)$$

If the system A is classical, then \mathcal{S}_A is a simplex and (up to relabelling) there is only one fine-grained measurement that is not a trivial refinement of another fine-grained measurement. We call this a *standard classical measurement*. Note that classical systems can be represented in any GPT and composing them maintains separability.

Box world [25] is the GPT in which the joint state space of several systems is in one-to-one correspondence with the set of no-signalling distributions amongst those systems, i.e., its state space in this sense the largest possible within the framework.

2 Measurement entropy and its properties

The approach to analysing causal structures that we use in this work is based on measurement entropy. In this section we introduce this and outline some of its properties.

Measurement entropy was first introduced in [17, 23]; we follow the exposition of [23] here. The *measurement entropy*, H_+ , is the minimal Shannon entropy of the outcome distribution after a fine-grained measurement, i.e., for $s_A \in \mathcal{S}_A$,

$$H_+(A) = \inf_{\{e^x\} \in \mathcal{E}_A^*} - \sum_x e^x(s_A) \log_2 e^x(s_A). \quad (3)$$

Several ways to define the conditional measurement entropy have been proposed [17, 23], of which we use the following [23]. For any state

$s_{AB} \in \mathcal{S}_{AB}$ with reduced state $s_B \in \mathcal{S}_B$, the *conditional measurement entropy* is

$$H_+(A|B) = \inf_{\{f^y\} \in \mathcal{E}_B} \sum_y f^y(s_B) H_+(A|_y), \quad (4)$$

where $H_+(A|_y)$ is the entropy of the state on A after a measurement on B with outcome y , $s_{A|y}$. For classical systems these entropies coincide with the Shannon entropy, H .

The measurement entropy satisfies a list of properties that are useful to this work. Some of these have previously been derived in [17, 23], others are new to this work. In the remainder of this section, $\mathcal{S}_A, \mathcal{S}_{AB}$ etc. refer to state spaces within an arbitrary GPT. For the proofs of the first two properties we refer to [23].

Property 1 (Positivity [23]). $H_+(A) \geq 0$ for all $s_A \in \mathcal{S}_A$ and $H_+(A|B) \geq 0$ for all $s_{AB} \in \mathcal{S}_{AB}$.

Property 2 (Reduction to Shannon entropy [17, 23]). Let A and B be classical systems and $s_{AB} \in \mathcal{S}_{AB}$, then $H_+(A) = H(A)$ and $H_+(A|B) = H(A|B)$.

Property 3 (Data processing). Let $s_{AB} \in \mathcal{S}_{AB}$, $T \in \mathcal{T}_{B \rightarrow C}$ and $s_{AC} = (I_A \otimes T)(s_{AB})$. Then $H_+(A|B) \leq H_+(A|C)$.

Proof. If $\{f^j\}$ form a measurement on C , then $\{g^j\}$ form a measurement on B , where $g^j : s_B \mapsto f^j(T(s_B))$. It follows that

$$\begin{aligned} H_+(A|C) &= \inf_{\{f^j\} \in \mathcal{E}_C} \sum_j f^j(s_C) H_+(A|_j) \\ &= \inf_{\{f^j\} \in \mathcal{E}_C} \sum_j g^j(s_B) H_+(A|_j) \\ &\geq \inf_{\{g^j\} \in \mathcal{E}_B} \sum_j g^j(s_B) H_+(A|_j) \\ &= H_+(A|B) \quad \square \end{aligned}$$

Property 4 (Independence). If two systems A and B are independent, i.e., $s_{AB} = s_A \otimes s_B$, then $H_+(A|B) = H_+(A)$.

Proof. If A and B are independent, after any measurement $\{e^j\} \in \mathcal{E}_B$ on B the post-measurement state on A is s_A , independent of the outcome of the measurement. Therefore $H_+(A|B) = H_+(A)$. \square

Property 5 (Classical subsystem inequalities). For a joint state $s_{ABC} \in \mathcal{S}_{ABC}$ with classical subsystems A and B , $H_+(AB|C) \geq H_+(A|C)$.

Proof. For any measurement $\{f^j\} \in \mathcal{E}_C$ we have

$$\begin{aligned} \sum_j f^j(s_C) H(AB)_{s_{AB}^j} &\geq \sum_j f^j(s_C) H(A)_{s_A^j} \\ &\geq H_+(A|C). \end{aligned}$$

Applying this to a sequence of measurements that converge to $H_+(AB|C)$ establishes the claimed result. \square

Property 6 (Subadditivity [17, 23]). *In GPTs for which $M \in \mathcal{E}_A^*$ and $N \in \mathcal{E}_B^*$ imply $M \otimes N \in \mathcal{E}_{AB}^*$ (which holds for locally tomographic theories, including box-world) $H_+(A) + H_+(B) \geq H_+(AB)$.*

We refer to [23] or Appendix A for a proof of Property 6. Note further that Property 6 is the only one that does not hold in arbitrary GPTs.

Property 7 (Lemma C2 of [23]). *For $s_{ABC} \in \mathcal{S}_{ABC}$ where A and B are classical systems, $H_+(A|BC) \geq H_+(AB|C) - H_+(B)$.*

We refer to [23] or Appendix A for a proof of Property 7.

Property 8. *For $s_{ABC} \in \mathcal{S}_{ABC}$, where B is a classical subsystem, $H_+(A|BC) \leq H_+(AB|C) - H_+(B|C)$. If C is also classical then this holds with equality.*

We prove this property in Appendix A. Note that Properties 7 and 8 are both relaxations of the chain rule, $H(A|BC) = H(AB|C) - H(B|C)$, that holds for Shannon and von Neumann entropy.

3 Entropy vector method for causal structures in GPTs

A *causal structure* is a set of nodes arranged in a directed acyclic graph, some of which are labelled observed. Each observed node has a corresponding random variable, while the other, unobserved nodes correspond to resources from a GPT. For a causal structure C we use C^C , C^Q , C^B or C^G depending on whether the resources are classical, quantum, box-world systems or from some unspecified GPT respectively. For each unobserved node we associate a subsystem with each of its outgoing edges. An example for this is displayed in Figure 1.

A direct arrow from a node A in a causal structure to a node Z means that A is a *parent* of Z ; a directed path from A to Z means that A is an *ancestor* of Z . For an unobserved node A , all subsystems associated with its outgoing edges are considered parents/ancestors of each its children/descendants. Given a causal structure, a *coexisting set* of systems [9, 22] is one for which a joint state can be defined. In general, no coexisting set includes all nodes, since there is no joint state of a system and the output obtained from a measurement on it (unless the system is classical).

Our method to generate new inequalities for causal structures with GPT resources begins by considering an *entropy vector* whose components are the entropies and conditional entropies of all coexisting sets. Conditional entropies composed entirely of classical subsystems are excluded because they are linear combinations of other entropies (e.g., $H(X|Y) = H(XY) - H(Y)$).

We then impose a system of linear (in)equalities that are necessary for a vector to be realisable as an entropy vector in a causal structure. These inequalities are constructed using the properties of the measurement entropy explained earlier and strong subadditivity in the cases where the measurement entropy reduces to the Shannon entropy. In the case of locally-tomographic GPTs, such as box-world, there is one additional property (Property 6) that does not hold in all GPTs. Further constraints come from the causal structure: two sets of nodes are independent if they do not share any ancestors in the causal structure. In general, there may be further independencies among the observed variables (see Theorem 22(i) of [14] and Appendix D). This system of inequalities constrains a polyhedral cone, which can be projected to a *marginal cone* that contains no components involving unobserved systems. The projection is performed with a Fourier-Motzkin elimination algorithm [26]. An example that illustrates this procedure in detail is provided at the beginning of Section 5.1.

When dealing with causal structures for which computing all entropy inequalities for the marginal scenarios of interest is computationally impractical or even not possible with the computational resources at hand, due to the scaling of Fourier-Motzkin elimination [27], we can still

derive valid entropy inequalities by marginalising subsets of all valid inequalities.

Furthermore, given a particular observed distribution that we suspect to be incompatible with a causal structure (either for classical theory, quantum theory or boxworld), it is not necessary to go through the marginalisation procedure discussed here. Instead we can look for a certificate of incompatibility using a linear program. This program can be set up by computing the entropy vector for the distribution in question and then adding this as a list of equalities (one for each of its components) to the list of valid entropy inequalities for the causal structure. If the resulting system of linear (in)equalities is infeasible, then the distribution in question is certified as incompatible with the causal structure within the theory under consideration.

For some causal structures, the entropic constraints derived using Property 6 are also valid for GPTs that are not locally tomographic, which is the content of the following proposition.

Proposition 1. *Let C be a causal structure in which there are no nodes with two or more unobserved parents. For any GPT G , any correlations achievable with a finite number of finite-outcome measurements in C^G are achievable in C^B .*

This proposition follows from the insight presented in the proof of Lemma 2 in [28] that any joint measurement on a classical and a GPT system can be written as a measurement on the classical system followed by an outcome-dependent measurement on the other (see also Lemma 4 in Appendix A).

Proof. Since each node has at most one unobserved parent, by Lemma 4 at each node we can assume a standard classical measurement on the classical subsystems followed by a measurement on the GPT subsystem depending on the result. Consider then $s_{A_1 A_2 \dots} \in \mathcal{S}_{A_1 A_2 \dots}^G$ and let $\{e_a^x\}_x \in \mathcal{E}_{A_1}$ for $a \in \{1, 2, \dots, m_a\}$, $\{f_b^y\}_y \in \mathcal{E}_{A_2}$ for $b \in \{1, 2, \dots, m_b\}$ etc. Since the outcome distribution $p(x, y, \dots | a, b, \dots)$ is no-signalling and since (by definition) all no-signalling distributions can be realised by states in box-world, there exists $s'_{A_1 A_2 \dots} \in \mathcal{S}_{A_1 A_2 \dots}^B$ and box-world measurements on A_1, A_2, \dots that give rise to the same correlations. \square

Note that the same argument does not hold if there are multiple unobserved parents at a single

node. This is because some joint measurements cannot be expressed as a measurement on one system followed by a measurement on the other conditioned on the first (cf. Lemma 4), for example a measurement in the Bell basis in quantum mechanics.

The method presented in this section recovers previous entropic approaches for describing classical [8, 20] and quantum [9] causal structures as special cases. In the classical case, the measurement entropy and its conditional version coincide with the Shannon entropy and all variables in a causal structure (observed and unobserved ones) coexist. In this case, our method is equivalent to that of [8, 20].² For the quantum case, the recovery of the method proposed in [9] from ours is less obvious and is explained in Section 4.

When considering different causal structures, convexity of the sets of achievable entropy vectors of the observed variables is useful for their comparison: for instance, it allows us to prove that the achievable entropies in one case are contained in those of another by considering only the extreme points. The following theorem (proven in Appendix B) establishes convexity in general and is therefore an important structural insight.

Theorem 1. *For any causal structure C^G the closure of the set of achievable entropy vectors of the observed variables is a convex cone.*

3.1 Causal structures and post-selection

For a causal structure involving a parentless observed node X that takes values $1, 2, \dots, n$, we can also analyse an adapted causal structure where each descendant of X is split into n -variables and X is dropped, e.g., a descendant Y is split into $Y_{|X=1}, \dots, Y_{|X=n}$. The resulting causal structure is said to be *post-selected* on X (see Figure 5 for an example, and [9, 22] for further details of this procedure). For some causal structures, C , post-selection is necessary for deriving entropy inequalities that distinguish between C^C , C^Q and C^B [9, 18, 29]. When post-selecting on parentless nodes, convexity of the set of achievable entropy vectors of the observed variables also holds by the following corollary of Theorem 1, which is also proven in Appendix B.

²In this case Property 5 imposes that the entropy is strongly subadditive.

Corollary 1. *For any causal structure C^G in which one or more nodes have been split into alternatives by post-selecting on parentless observed nodes, the closure of the set of achievable entropy vectors for the coexisting observed variables is a convex cone.*

4 Quantum causal structures

For quantum causal structures there is an entropy vector method in which the components are the unconditional von Neumann entropies, H , of all coexisting sets [9]. Conditional entropies are not explicitly included, but, because $H(A|B) = H(AB) - H(B)$, relations involving conditional entropy can still be encoded. It is natural to ask whether the technique introduced earlier in this paper could yield different results to the existing entropic approach to quantum causal structures. In this section, we consider this question.

Following the approach outlined earlier in the paper we consider including all entropy inequalities from [9] (see Appendix D for a full description of the method employed in [9]). These are automatically part of our approach, since the unconditional measurement entropy—for which we include all inequalities valid in the theory at hand—coincides with the von Neumann entropy in the quantum case. In addition, we also take into account inequalities for conditional measurement entropies, which are always positive and differ from the conditional von Neumann entropy. Thus, our approach could lead to more restrictive inequalities than the previous quantum one. With the following proposition we show that the previous method for quantum causal structures [9] can be refined in a way that makes the additional variables corresponding to conditional measurement entropies superfluous.

Proposition 2. *Consider a causal structure C^Q and suppose that, in addition to any causal constraints, we use positivity of unconditional entropies, strong subadditivity and additionally impose positivity of conditional entropies for all combinations of variables that occur in a coexisting set. If we then eliminate all variables corresponding to unobserved systems, the resulting entropic inequalities for the observed variables are all valid.*

As a result, all valid inequalities for conditional measurement entropy can be imposed for the conditional von Neumann entropy instead, and can hence be encoded as linear constraints on the (unconditional) von Neumann entropy. In other words, although the conditional von Neumann entropy can be negative for some quantum states, by constraining it to be positive and eliminating unobserved variables, we obtain valid entropy inequalities for the observed variables in C^Q . This was not used in previous entropic analyses of quantum causal structures [9].

Previous quantum methods [9, 22] instead analysed quantum causal structures by considering the von Neumann entropy of coexisting sets and imposing positivity of the entropy, strong-subadditivity as well as the weak monotonicity constraints that for any state ρ_{XYZ} , $H(X|Y) + H(X|Z) \geq 0$ [30]. Weak monotonicity constraints are not needed in the statement of Proposition 2 because they are implied by the positivity of conditional entropies. See Appendix C for the full proof of Proposition 2 and for a complete account of previous quantum methods.

This also gives an important insight into the entropy vector method: the difference between the inequalities that result from using the entropy vector method in the classical [8, 20] and quantum [9] cases is entirely due to the fact that in the quantum case not all variables coexist and does not arise from the different properties of the Shannon and von Neumann entropy (see also Appendix C for further discussion).

For most causal structures of interest we can prove that our refined entropy vector method does not allow us to find any tighter entropy inequalities than that of [9] (see Lemma 8 in Appendix D). Nonetheless, the possibility of using positivity of conditional entropy instead of weak monotonicity simplifies the quantum method even in these cases.

5 Applications: Entropic characterisation of causal structures with GPT resources

In this section we illustrate the techniques introduced above through a series of examples that are chosen to show the different types of result that can occur. We summarize our findings here,

before presenting these cases in detail.

First, in Section 5.1, we consider the case without post-selection and give three types of example:

- A causal structure in which the actual entropic cones for classical, quantum and GPTs are provably the same is given in Section 5.1.1.
- A causal structure in which our methods lead to the same outer approximations to the entropy cones in the classical, quantum and GPTs cases is given in Section 5.1.2.
- Two causal structures for which our methods yield the same outer approximations of the entropy cones in the classical and quantum cases, but a different outer approximation is obtained for GPTs is given in Sections 5.1.3 and 5.1.4. In these cases we are not aware of any GPT correlations that violate the classical/quantum inequalities.

Note that we already know of a case (the triangle causal structure) where we have different outer approximations of the entropy cones in the classical and quantum cases [10]. We are currently unaware of a case without post-selection in which the outer approximations of the entropy cones between two theories are different and in which we know of correlations in one theory that violate those in another. In other words, it remains possible that the gaps we find in the outer approximations are a symptom of the method used, rather than features of the actual entropy cones.

Then, in Section 5.2 we move to the use of post-selection and show that:

- For the bilocality causal structure there are different entropy cones for classical, quantum and locally tomographic GPTs (Section 5.2.1). We also identify distributions that certify that the true entropy cones are indeed different for all three cases.
- We can use our methods to obtain entropic inequalities for the Information Causality scenario in arbitrary GPTs (Section 5.2.2).

5.1 Analysis of causal structures without post-selection

In this section we illustrate that our method for GPTs can recover the classically valid entropy

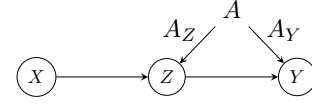


Figure 1: The instrumental causal structure. The nodes labelled X , Y and Z correspond to observations, modelled as random variables. The node A labels a resource-system with subsystems A_Y and A_Z associated to its outgoing edges.

inequalities for some causal structures (the first two examples), and that for other causal structures we recover different inequalities (the last two examples).

5.1.1 Instrumental Scenario

In the classical literature, one well-studied causal structure is the *instrumental causal structure* [31] of Figure 1. The quantum version of this causal structure has also been studied [13, 32].

In this case, the coexisting sets are all subsets of $\{A_Y, A_Z, X\}$, $\{A_Y, X, Z\}$ and $\{X, Y, Z\}$. The second set implies that the entropy vector includes components corresponding to $H_+(A_Y)$, $H_+(X)$, $H_+(Z)$, $H_+(A_Y X)$, $H_+(A_Y Z)$, $H_+(XZ)$, $H_+(A_Y XZ)$, $H_+(A_Y|X)$, $H_+(A_Y|Z)$, $H_+(A_Y|XZ)$, $H_+(A_Y X|Z)$, $H_+(A_Y Z|X)$, $H_+(X|A_Y)$, $H_+(X|A_Y Z)$, $H_+(XZ|A_Y)$, $H_+(Z|A_Y)$, $H_+(Z|A_Y X)$, and similarly for the other two, leading to a vector with 35 components.

For these we impose all entropy inequalities of Properties 1 to 8 as well as the independencies of the subsystems of A and X , such as $H_+(A_Y A_Z|X) = H_+(A_Y A_Z)$ or $H_+(X|A_Y A_Z) = H_+(X)$.

Eliminating all other variables in order to obtain inequalities that only involve the components $(H(X), H(Y), H(Z), H(XY), H(XZ), H(YZ), H(XYZ))$, we obtain the *Shannon inequalities* (positivity of entropy and conditional entropy, and strong subadditivity) for three variables and

$$I(X : YZ) \leq H(Z), \quad (5)$$

which form a polyhedral cone Γ . Valid entropy vectors for distributions compatible with the instrumental scenario for a system A of some locally tomographic GPT are necessarily within Γ . For this causal structure, it is known that being in Γ is necessary and sufficient for being in the closure of the set of valid entropy vectors when

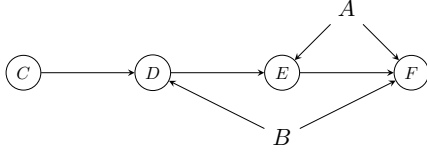


Figure 2: Causal Structure where our outer approximations for classical, quantum and box-world resources coincide.

A is classical or quantum [10, 14, 33]. Since classical systems are a special case of systems in a GPT, it follows that membership of Γ is also sufficient for locally tomographic GPTs. We have hence found all valid entropy inequalities in this scenario for such theories.

According to Proposition 1, (5) holds in any (not necessarily locally tomographic) GPT³. Thus, in the instrumental causal structure, Γ completely characterises the set of achievable entropy vectors independently of whether system A is classical, quantum or any GPT system.

5.1.2 Causal structure of Figure 2

Applied to the causal structure of Figure 2, our method leads to the following entropy inequalities for the observed variables when A and B are taken to be box-world systems:

$$I(C : E|D) = 0 \quad (6)$$

$$I(C : DEF) \leq H(D) \quad (7)$$

$$I(C : EF) \leq H(E) \quad (8)$$

and the Shannon inequalities. In [14] the same inequalities were derived for classical A and B . It follows from Proposition 2 (see below), that these constraints also hold in the quantum case. Thus, for all three theories we obtain the same outer approximation of the respective entropy cones⁴. Violation of any of these inequalities excludes this causal structure as a possible explanation of the observed correlations, irrespective of the nature of the pre-shared resources.

³It is already clear that it holds in any GPT satisfying the premise of Property 6. Note that Proposition 1 implies that the set of correlations in C^B is at least as large as those in C^G , so any valid restriction that holds for correlations in C^B holds also for correlations in C^G .

⁴This may change if we were to take non-Shannon inequalities into account prior to Fourier-Motzkin elimination.

In this example, these outer approximations are not tight: there are further valid entropy inequalities for the classical systems C , D , E and F —so-called non-Shannon inequalities—that lead to tighter approximations, e.g. the following inequality (from [24]):

$$I(D : E|F) + I(D : E|C) + I(F : C) - I(D : E) + I(D : F|E) + I(E : F|D) + I(D : E|F) \geq 0. \quad (9)$$

5.1.3 Causal structure of Figure 3(a)

With resources A and B that are allowed in box-world we obtain the Shannon inequalities and

$$I(C : EF) \leq H(E), \quad (10)$$

$$I(C : DEF) \leq H(D). \quad (11)$$

Classical and quantum resources A and B lead to slightly tighter inequalities, namely the Shannon inequalities and

$$I(C : EF) \leq H(E) \quad (12)$$

$$I(C : DEF) + I(D : F|E) \leq H(D). \quad (13)$$

The question of whether or not there exist box-world distributions that violate one of these inequalities remains open⁵.

5.1.4 Causal structure of Figure 3(b)

With resources A and B that are allowed according to the theory of box-world we obtain the Shannon inequalities and

$$I(C : D) = 0, \quad (14)$$

$$H(F|CE) \leq H(CF|DE), \quad (15)$$

$$I(D : CEF) \leq H(E). \quad (16)$$

Classical and quantum resources lead to slightly tighter inequalities, namely the Shannon inequalities, Equation (14), Inequality (15) and

$$I(D : CEF) \leq H(E|C). \quad (17)$$

Note that the classical case was treated in [14].

⁵In general, the methods used here lead to outer approximations to the entropy cone of a causal structure. In some cases we can show these to be tight, see e.g. [29, 34], but not always. It could thus be the case that while there is a gap between the outer approximations of these cones, there is still no gap between the true cones.

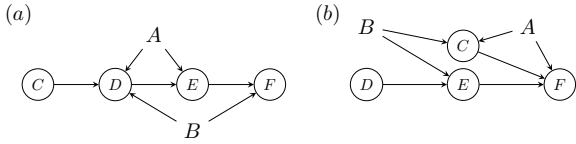


Figure 3: Causal structures where there are different entropic bounds depending on the nature of the unobserved A and B .

In general, such comparisons are interesting for analysing the nature of causation in different theories. In a sense the box-world inequalities can be thought of as minimal requirements for a theory with a reasonable notion of causation. Developing a systematic understanding of this may hint at ways to find a physical principle that singles out quantum correlations in general scenarios.

5.2 Analysis of causal structures with post-selection

In this section we apply our method to post-selected causal structure and show how this allows us to distinguish the correlations obtained in different GPTs. We give two examples for this.⁶

5.2.1 Bilocality

The bilocal causal structure, first analysed in [36, 37], characterises the situation we encounter in scenarios where we rely on entanglement swapping [38], see Figure 4(a). Technologically, this scenario is encountered, for instance, in quantum repeaters [39] or event-ready detection schemes [40].

The entropy vector method provides us with a convenient means to compare the observable correlations when the sources L_1 and L_2 come from different theories. Applying the entropy vector method to the bilocal causal structure with sources from a locally tomographic GPT, apart from the Shannon inequalities we find only

$$H(X_0 Z_0) = H(X_0) + H(Z_0) \quad (18)$$

up to symmetry (exchanging X_0 and X_1 as well as Z_0 and Z_1).

⁶These constraints can be useful even in the classical case, where there is already the inflation technique for deriving incompatibility constraints [11, 35], if the cardinality of the variables is high, for instance.

Applying the entropy vector method to the bilocal causal structure in the quantum case we find the Shannon inequalities as well as

$$H(X_0 Z_0) = H(X_0) + H(Z_0) \quad (19)$$

$$I(X_0 Y_0 : Z_0) \leq H(Y_0 | X_1) \quad (20)$$

$$I(X_1 : Z_1 | Y_0) \leq H(Y_0 | X_0) + H(Y_0 | Z_0) - H(Y_0), \quad (21)$$

up to symmetry. [There are 4 instances of (19) (obtained by exchanging X_0 and X_1 as well as Z_0 and Z_1), 16 of (20) (obtained by exchanging X_0 and X_1 , Y_0 and Y_1 , or Z_0 and Z_1 and by exchanging the roles of X and Z) and 8 of (21) (obtained by exchanging X_0 and X_1 , Y_0 and Y_1 , or Z_0 and Z_1).]

The equality constraints are found in both the quantum and GPT case, while the quantum description is tighter, and in this case the gap is provably significant. An example of a box-world distribution that violates (21) is obtained by taking the systems L_1 and L_2 to be PR-boxes, $A \in \{0, 1\}$ is the uniform input and $X \in \{0, 1\}$ the output on the left, and analogously for C and Z on the right. $B \in \{0, 1\}$ is with probability $\frac{1}{2}$ input into the first and with probability $\frac{1}{2}$ input into the second box and the respective output serves as an input for the other, where Y is equal to the outputs of the two boxes. This distribution also violates the classical inequalities below and is, to the best of our knowledge, the first manifestation of such a violation with a GPT correlation that is proven to be achievable in the bilocal causal structure. [Note that when the violation of the classical inequalities of [20] was discussed, a tripartite box was considered directly (and without proof that it can be generated from a GPT in the bilocal causal structure)⁷.]

For classical sources L_1 and L_2 , we obtain a convex cone constrained by Shannon inequalities and 53 other independent classes of linear inequalities. We list one representative of each of these 53 classes in Appendix E. Note that the (in)equalities obtained in the quantum case (cf. (19)–(21)) are present. A quantum distribution that leads to an entropy vector outside this cone and which is thus not achievable with classical resources in this causal structure is, for instance, the following: Suppose L_1 and L_2 share a singlet state each

⁷Note also that we have a slight difference from [20]: they did not consider B .

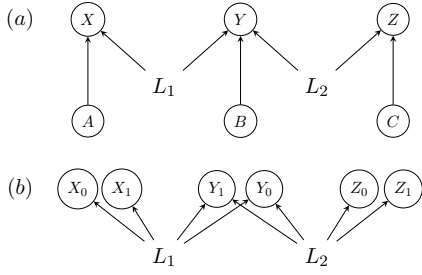


Figure 4: Bilocal causal structure. (a) The nodes labelled A, B, C, X, Y and Z correspond to observations, modelled as random variables. The nodes A and B label a resource-system which in this case we take to be quantum. (b) Post-selected version of the bilocal causal structure, where X_0 stands for $X_{|A=0}$ and similarly for X_1 as well as the corresponding Y and Z .

and at the nodes the output is made according to projective measurements Π_θ in the basis $\{\cos\theta|0\rangle + \sin\theta|1\rangle, \sin\theta|0\rangle - \cos\theta|1\rangle\}$. The angles are chosen as follows: for X_0 , $\theta = x$, for X_1 , $\theta = 3x$, for Z_0 , $\theta = 0$, for Z_1 , $\theta = 2x$ and for Y_0 we consider $\theta = 0$ on the state shared by L_1 and $\theta = (2y_0 + 1)x$ on the other, for Y_1 we consider $\theta = 2x$ on the state shared by L_1 and $\theta = (2y_0 + 1)x$ on the other system, where the output is made up of the outputs of both measurements in both cases and where $x = 0.1$ and y_0 denotes the outcome of the first measurement. This distribution violates several of our inequalities, for instance one inequality from class 31 from Appendix E). Note also that for the entropic inequalities derived in [20] for the bilocal causal structure with classical sources, no quantum violations are known, so, as far as we know, the characterisation reported here is the first entropic one that is provably able to resolve this gap.

5.2.2 Information Causality

Information causality [3] is a candidate principle for singling out quantum theory. Roughly speaking the principle is that that sending n bit of classical information from one party to another cannot give the recipient access to more than n bits of previously unknown information regardless of any pre-shared resources the parties may have. The associated causal structure is shown in Figure 5, and information causality is obeyed

if

$$I(X_1 : Y_{|R=1}) + I(X_2 : Y_{|R=2}) \leq H(Z). \quad (22)$$

This relation is known to hold in both classical and quantum theory, while it is violated in box-world.⁸

Using the technique of the present paper, we find that the following relations hold for underlying box-world systems

$$I(X_1 X_2 : Y_{|R=1} Z) \leq H(Z) \quad (23)$$

$$I(X_1 X_2 : Y_{|R=2} Z) \leq H(Z). \quad (24)$$

These are valid in any GPT (see Proposition 1) for the entropy vectors \mathbf{H} containing the entropies of all 23 subsets of the coexisting sets $\{X_1, X_2, Y_{|R=1}, Z\}$ and $\{X_1, X_2, Y_{|R=2}, Z\}$. Thus, although information causality does not hold in general, some minimal notion of causation remains (beyond no-signalling).

We remark that the information causality scenario in boxworld was also considered in [15], but in a slightly different way. There the relations $I(X_1 : Y_{|R=1}) \leq H(Z)$ and $I(X_2 : Y_{|R=2}) \leq H(Z)$ were postulated. We are able to recover these with our approach in the following way. If, instead of considering all joint entropies of coexisting variables, only the restricted entropy vectors with components $\mathbf{H}_R = (H(X_1), H(X_2), H(Y_{|R=1}), H(Y_{|R=2}), H(Z), H(X_1 Y_{|R=1}), H(X_2 Y_{|R=2}))$ are considered, the inequalities $I(X_1 : Y_{|R=1}) \leq H(Z)$ and $I(X_2 : Y_{|R=2}) \leq H(Z)$ emerge. Since all extremal vertices of the entropy cone of vectors \mathbf{H}_R are achievable (as was shown in [15]) and, according to Corollary 1, the (closure of the) set of achievable entropy vectors \mathbf{H}_R is convex, this is the true entropy cone for this restricted marginal scenario. However, we don't see a clear motivation for excluding the additional observed entropies.

The two inequalities (23) and (24) were already derived in [23] for box-world, but here they emerge systematically from our method (and our results imply that they hold for any GPT). They are the only inequalities (other than Shannon inequalities) that follow from our method as it was introduced above. However, further inequalities hold for this scenario, for instance the non-

⁸In fact, tighter entropy inequalities have since been shown to hold in the quantum (and classical) case [9].

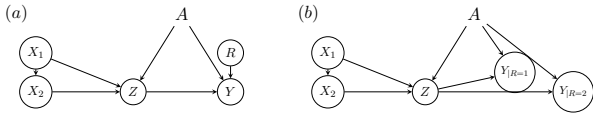


Figure 5: Causal structure of the Information Causality scenario. (a) Alice holds two pieces of information X_1 and X_2 and is allowed to send a message Z to Bob. Bob then has to make a guess of either X_1 or X_2 , depending on the request of a referee, $R = 1$ or $R = 2$. A is a pre-shared resource the parties may use. (b) We divide Y into two variables, $Y_{|R=1}$ and $Y_{|R=2}$, depending on the question R . While for classical A Bob can always compute the value of both $Y_{|R=1}$ and $Y_{|R=2}$, more generally these have to be understood as alternatives, of which only one is generated.

Shannon inequality [24]

$$I(X_2 : Y_{|R=1}|Z) + I(X_2 : Y_{|R=1}|X_1) + I(Z : X_1) - I(X_2 : Y_{|R=1}) + I(X_2 : Z|Y_{|R=1}) + I(Y_{|R=1} : Z|X_2) + I(X_2 : Y_{|R=1}|Z) \geq 0.$$

Because a complete set of non-Shannon inequalities is not known, we do not have a complete characterisation of the entropy cone of vectors \mathbf{H} for this scenario (which may require infinitely many linear inequalities).

6 Limitations and directions to overcome them

As is the case for previous entropic methods [8, 9, 20], there are causal structures for which this method does not imply any entropic constraints for the observed variables (except Shannon inequalities), an example being the triangle causal structure [7, 9–11, 14, 19, 41].⁹ Furthermore, all known strategies that certify incompatibility of entropy vectors relying on GPT resources with classical and quantum scenarios rely on post-selection (cf. Figure 5). If post-selection is necessary for this, for some causal structures (such as the triangle causal structure) current entropic techniques cannot certify this distinction. This is not a severe limitation, since most experimentally interesting causal structures involve measurement settings, which we can post-select on.

Considering entropy vectors rather than the corresponding joint probability distributions

⁹However, the triangle causal structure with inputs can be treated with our method.

gives a computational advantage and provides constraints that are valid independently of the dimension of the involved resources. However, this advantage comes with restricted precision (see for instance [29]). In particular, there are distributions between observed variables that are realisable with box-world resources but not with classical or quantum systems but which the method cannot certify as such. How to overcome this remains an open question, although the ideas in [42] may form a useful starting point (notwithstanding the limitations of the ideas of [42] identified in [43]).

While we found that our method strictly improves on previous entropic methods [15], another promising research avenue is to generalise the inflation technique to GPTs. This research has been started in [11], where it was pointed out that certain inflations are valid for any GPT and hence some general constraints can be derived with it. The question of how these constraints relate to the ones found with our method is left for future work. One key difference is that the method presented in [11] does not distinguish what GPT the latent systems are described by (e.g. whether they are quantum and box-world systems). Considering the latent variables explicitly allows us to make this distinction and to certify that different sets of correlations are produced within different GPTs (e.g. quantum theory and box-world).

Acknowledgements

We thank Peter J. Brown for comments on an earlier version of this manuscript. This work was supported by an EPSRC First grant (grant number EP/P016588/1), the EPSRC’s Quantum Communications Hub (grant number M013472/1) and by the Austrian Science fund (FWF) stand-alone project P 30947. The majority of this work was carried out while MW was based at the University of York.

References

- [1] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1(3):195–200, 1964. ISSN 01923188. DOI: [10.1002/prop.19800281202](https://doi.org/10.1002/prop.19800281202).

- [2] C. J. Wood and R. 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(3):033002, 2015. ISSN 1367-2630. DOI: [10.1088/1367-2630/17/3/033002](https://doi.org/10.1088/1367-2630/17/3/033002).
- [3] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, 2009. ISSN 0028-0836. DOI: [10.1038/nature08400](https://doi.org/10.1038/nature08400).
- [4] S. W. Al-Safi and A. J. Short. Information causality from an entropic and a probabilistic perspective. *Physical Review A*, 84(4):042323, 2011. ISSN 1050-2947. DOI: [10.1103/PhysRevA.84.042323](https://doi.org/10.1103/PhysRevA.84.042323).
- [5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969. ISSN 0031-9007. DOI: [10.1103/PhysRevLett.23.880](https://doi.org/10.1103/PhysRevLett.23.880).
- [6] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going beyond Bell’s theorem. In M. Kafatos, editor, *Bell’s Theorem, Quantum Mechanics and Conceptions of the Universe*, pages 69–72. Kluwer Academic, Dordrecht, The Netherlands, 1989. DOI: [10.1007/978-94-017-0849-4](https://doi.org/10.1007/978-94-017-0849-4).
- [7] T. Fritz. Beyond Bell’s theorem: correlation scenarios. *New Journal of Physics*, 14(10):103001, 2012. ISSN 1367-2630. DOI: [10.1088/1367-2630/14/10/103001](https://doi.org/10.1088/1367-2630/14/10/103001).
- [8] T. Fritz and R. Chaves. Entropic inequalities and marginal problems. *IEEE Transactions on Information Theory*, 59(2):803–817, 2013. ISSN 0018-9448. DOI: [10.1109/TIT.2012.2222863](https://doi.org/10.1109/TIT.2012.2222863).
- [9] R. Chaves, C. Majenz, and D. Gross. Information-theoretic implications of quantum causal structures. *Nature communications*, 6:5766, 2015. ISSN 2041-1723. DOI: [10.1038/ncomms6766](https://doi.org/10.1038/ncomms6766).
- [10] M. Weilenmann and R. Colbeck. Non-Shannon inequalities in the entropy vector approach to causal structures. *Quantum*, 2, 2018. DOI: [10.22331/q-2018-03-14-57](https://doi.org/10.22331/q-2018-03-14-57).
- [11] E. Wolfe, R. W. Spekkens, and T. Fritz. The Inflation Technique for Causal Inference with Latent Variables. *Journal of Causal Inference*, 2016. DOI: [10.1515/jci-2017-0020](https://doi.org/10.1515/jci-2017-0020).
- [12] B. S. Tsirelson. Quantum generalizations of Bell’s inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980. ISSN 1573-0530. DOI: [10.1007/BF00417500](https://doi.org/10.1007/BF00417500).
- [13] T. Van Himbeek, J. Bohr Brask, S. Pironio, R. Ramanathan, A. B. Sainz, and E. Wolfe. Quantum violations in the Instrumental scenario and their relations to the Bell scenario. *Quantum*, 3:186, 2019. ISSN 2521-327X. DOI: [10.22331/q-2019-09-16-186](https://doi.org/10.22331/q-2019-09-16-186).
- [14] J. Henson, R. Lal, and M. F. Pusey. Theory-independent limits on correlations from generalized Bayesian networks. *New Journal of Physics*, 16(11):113043, 2014. ISSN 1367-2630. DOI: [10.1088/1367-2630/16/11/113043](https://doi.org/10.1088/1367-2630/16/11/113043).
- [15] R. Chaves and C. Budroni. Entropic nonsignaling correlations. *Physical Review Letters*, 116(24):240501, 2016. ISSN 0031-9007. DOI: [10.1103/PhysRevLett.116.240501](https://doi.org/10.1103/PhysRevLett.116.240501).
- [16] A. J. Short and J. Barrett. Strong nonlocality: a trade-off between states and measurements. *New Journal of Physics*, 12(3):033034, 2010. ISSN 1367-2630. DOI: [10.1088/1367-2630/12/3/033034](https://doi.org/10.1088/1367-2630/12/3/033034).
- [17] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke. Entropy and information causality in general probabilistic theories. *New Journal of Physics*, 12(3):033024, 2010. ISSN 1367-2630. DOI: [10.1088/1367-2630/12/3/033024](https://doi.org/10.1088/1367-2630/12/3/033024).
- [18] S. L. Braunstein and C. M. Caves. Information-theoretic Bell inequalities. *Physical Review Letters*, 61(6):662–665, 1988. ISSN 0031-9007. DOI: [10.1103/PhysRevLett.61.662](https://doi.org/10.1103/PhysRevLett.61.662).
- [19] B. Steudel and N. Ay. Information-theoretic inference of common ancestors. *Entropy*, 17(4):2304–2327, 2015. ISSN 1099-4300. DOI: [10.3390/e17042304](https://doi.org/10.3390/e17042304).
- [20] R. Chaves and T. Fritz. Entropic approach to local realism and noncontextuality. *Physical Review A*, 85(3):032113, 2012. ISSN 1050-2947. DOI: [10.1103/PhysRevA.85.032113](https://doi.org/10.1103/PhysRevA.85.032113).
- [21] J. Pienaar. Which causal structures might support a quantum–classical gap? *New*

- Journal of Physics*, 19(4):043021, 2017. DOI: [10.1088/1367-2630/aa673e](https://doi.org/10.1088/1367-2630/aa673e).
- [22] M. Weilenmann and R. Colbeck. Analysing causal structures with entropy. *Proceedings of the Royal Society A*, 473(2207), 2017. DOI: [10.1098/rspa.2017.0483](https://doi.org/10.1098/rspa.2017.0483).
- [23] A. J. Short and S. Wehner. Entropy in general physical theories. *New Journal of Physics*, 12(3):033023, 2010. ISSN 1367-2630. DOI: [10.1088/1367-2630/12/3/033023](https://doi.org/10.1088/1367-2630/12/3/033023).
- [24] Z. Zhang and R. W. Yeung. A non-Shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory*, 43(6):1982–1986, 1997. ISSN 00189448. DOI: [10.1109/18.641561](https://doi.org/10.1109/18.641561).
- [25] J. Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3):032304, 2007. ISSN 1050-2947. DOI: [10.1103/PhysRevA.75.032304](https://doi.org/10.1103/PhysRevA.75.032304).
- [26] H. P. Williams. Fourier’s method of linear programming and its dual. *The American Mathematical Monthly*, 93(9):681–695, 1986. DOI: [10.2307/2322281](https://doi.org/10.2307/2322281).
- [27] D. Monniaux. Quantifier elimination by lazy model enumeration. In *International Conference on Computer Aided Verification*, pages 585–599. Springer, 2010. DOI: [10.1007/978-3-642-14295-6_51](https://doi.org/10.1007/978-3-642-14295-6_51).
- [28] J. Cadney and N. Linden. Measurement entropy in generalized nonsignalling theory cannot detect bipartite nonlocality. *Physical Review A*, 86(5):052103, 2012. ISSN 1050-2947. DOI: [10.1103/PhysRevA.86.052103](https://doi.org/10.1103/PhysRevA.86.052103).
- [29] M. Weilenmann and R. Colbeck. Inability of the entropy vector method to certify nonclassicality in linelike causal structures. *Physical Review A*, 94:042112, 2016. DOI: [10.1103/PhysRevA.94.042112](https://doi.org/10.1103/PhysRevA.94.042112).
- [30] N. Pippenger. The inequalities of quantum information theory. *IEEE Transactions on Information Theory*, 49(4):773–789, 2003. DOI: [10.1109/TIT.2003.809569](https://doi.org/10.1109/TIT.2003.809569).
- [31] J. Pearl. On the testability of causal models with latent and instrumental variables. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 435–443. Morgan Kaufmann Publishers Inc., 1995. DOI: [10.5555/2074158.2074208](https://doi.org/10.5555/2074158.2074208).
- [32] R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, and F. Sciarrino. Quantum violation of an instrumental test. *Nature Physics*, 14(3):291, 2018. DOI: [10.1038/s41567-017-0008-5](https://doi.org/10.1038/s41567-017-0008-5).
- [33] R. Chaves, L. Luft, T. Maciel, D. Gross, D. Janzing, and B. Schölkopf. Inferring latent structures via information inequalities. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, pages 112–121, Corvallis, Oregon, 2014. AUAI Press. DOI: [10.5555/3020751.3020764](https://doi.org/10.5555/3020751.3020764).
- [34] M. Weilenmann. *Quantum causal structure and quantum thermodynamics*. PhD thesis, University of York, 2017. Also available as [arXiv:1807.06345](https://arxiv.org/abs/1807.06345).
- [35] M. Navascues and E. Wolfe. The inflation technique solves completely the classical inference problem. e-print [arXiv:1707.06476](https://arxiv.org/abs/1707.06476), 2017.
- [36] C. Branciard, N. Gisin, and S. Pironio. Characterizing the nonlocal correlations created via entanglement swapping. *Physical Review Letters*, 104(17):170401, 2010. ISSN 1079-7114. DOI: [10.1103/PhysRevLett.104.170401](https://doi.org/10.1103/PhysRevLett.104.170401).
- [37] C. Branciard, D. Rosset, N. Gisin, and S. Pironio. Bilocal versus nonbilocal correlations in entanglement-swapping experiments. *Physical Review A*, 85(3):032119, 2012. ISSN 1050-2947. DOI: [10.1103/PhysRevA.85.032119](https://doi.org/10.1103/PhysRevA.85.032119).
- [38] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899, 1993. ISSN 0031-9007. DOI: [10.1103/PhysRevLett.70.1895](https://doi.org/10.1103/PhysRevLett.70.1895).
- [39] H.-J. Briegel, W. Dür, J. I. Cirac, and P. Zoller. Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication. *Physical Review Letters*, 81(26):5932–5935, 1998. ISSN 0031-9007. DOI: [10.1103/PhysRevLett.81.5932](https://doi.org/10.1103/PhysRevLett.81.5932).
- [40] M. Zukowski, A. Zeilinger, M. A. Horne, and A. Ekert. ”Event-ready-detectors” Bell experiment via entanglement swapping. *Physical Review Letters*, 71(26):4287–4290, 1993. ISSN 1079-7114. DOI: [10.1103/PhysRevLett.71.4287](https://doi.org/10.1103/PhysRevLett.71.4287).
- [41] N. Gisin. The elegant joint quantum measurement and some conjectures about N-

locality in the triangle and other configurations. e-print [arXiv:1708.05556](https://arxiv.org/abs/1708.05556), 2017.

[42] R. Chaves. Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality. *Physical Review A*, 87(2):022102, 2013. ISSN 1050-2947. DOI: [10.1103/PhysRevA.87.022102](https://doi.org/10.1103/PhysRevA.87.022102).

[43] V. Vilasini and R. Colbeck. On the sufficiency of entropic inequalities for detecting

non-classicality in the Bell causal structure. e-print [arXiv:1912.01031](https://arxiv.org/abs/1912.01031), 2019.

[44] O. Klein. Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre. *Zeitschrift für Physik*, 72(11): 767–775, 1931. ISSN 0044-3328. DOI: [10.1007/BF01341997](https://doi.org/10.1007/BF01341997).

A Further details regarding the inequalities for the measurement entropy in GPTs

Before getting to the additional properties, we need a few lemmas. The first is the concavity of H_+ , proven in [17, 23], which follows from the concavity of the Shannon entropy.

Lemma 1. For $s_1, s_2 \in \mathcal{S}_A$ and $0 \leq p \leq 1$, $H_+(A)_{ps_1+(1-p)s_2} \geq pH_+(A)_{s_1} + (1-p)H_+(A)_{s_2}$.

The next lemma says that the infimum in the definition of conditional measurement entropy can be restricted to fine-grained measurements.

Lemma 2. Let $s_{AB} \in \mathcal{S}_{AB}$, then $H_+(A|B) = \inf_{\{f^y\} \in \mathcal{E}_B^*} \sum_y f^y(s_B)H_+(A|_y)$.

Proof. Let $\{f^y\}_{y=1}^n \in \mathcal{E}_B$ and consider the coarse-graining that combines the last two effects in $\{f^y\}_{y=1}^n$ to give a new measurement $\{g^y\}_{y=1}^{n-1}$ where $g^y = f^y$ for $y = 1, \dots, n-2$ and $g^{n-1} = f^{n-1} + f^n$. We have

$$\sum_{y=1}^{n-1} g^y(s_B)H_+(A|_y) = \sum_{y=1}^{n-2} f^y(s_B)H_+(A|_y) + (f^{n-1}(s_B) + f^n(s_B))H_+(\alpha s_{A|n-1} + (1-\alpha)s_{A|n}),$$

where $\alpha = f^{n-1}(s_B)/(f^{n-1}(s_B) + f^n(s_B))$. Using concavity, we have

$$\begin{aligned} \sum_{y=1}^{n-1} g^y(s_B)H_+(A)_{s_{A|y}} &\geq \sum_{y=1}^{n-2} f^y(s_B)H_+(A)_{s_{A|y}} + f^{n-1}(s_B)H_+(s_{A|n-1}) + f^n(s_B)H_+(s_{A|n}) \\ &= \sum_{y=1}^n f^y(s_B)H_+(A)_{s_{A|y}}, \end{aligned}$$

i.e., for this coarse-graining the measurement on Bob cannot decrease the expected measurement entropy on Alice conditioned on the result. Since all coarse-grainings can be formed by a sequence of such combinations, it follows that the infimum on Bob's measurements can be restricted to fine-grained measurements. \square

It is also worth noting the following.

Lemma 3. Let $\{f^y\} \in \mathcal{E}_B$ be a trivial refinement of $\{e^x\} \in \mathcal{E}_B$. Then $\sum_y f^y(s_B)H_+(A|_y) = \sum_x e^x(s_B)H_+(A|_x)$.

Proof. Consider the case in which one of the effects in $\{e^x\}_{x=1}^n$ is split into two to form $\{f^y\}_{y=1}^{n+1}$ with $f^y = e^y$ for $y = 1, \dots, n-1$ and $f^n + f^{n+1} = e^n$. We have $H_+(A|_{n+1}) = H_+(A|_n)$ and hence in this case the claim follows from $f^{n+1} + f^n = e^n$. Since any trivial refinement can be formed by combining such splittings, the result generalizes to all trivial refinements. \square

The following lemma is in essence a restatement of part of the proof of Lemma 2 from [28].

Lemma 4. *Let A be classical and B be a system from an arbitrary GPT. For any measurement $M \in \mathcal{E}_{AB}$ there exists an n -outcome measurement $N_A \in \mathcal{E}_A$ and measurements $N_B^x \in \mathcal{E}_B$ for $x = 1, 2, \dots, n$ such that M is equivalent to performing N_A , then performing N_B^x (where x is the result of N_A).*

Proof. Let $M = \{e^r\}$ and let $\{f^x\}$ be a standard classical measurement on A . We can define a new set of effects on B that act as $e_x^r : s_B \mapsto e^r(s_A^x \otimes s_B)$, where $\{s_A^x\}$ are a set of states on A for which $f^x(s_A^y) = \delta_{x,y}$. For each x , the set $\{e_x^r\}_r$ form a measurement on B :

$$\sum_r e_x^r(s_B) = \sum_r e^r(s_A^x \otimes s_B) = 1 \quad \text{for all } s_B \in \mathcal{S}_B.$$

If we take $N_A = \{f^x\}$ and $N_B^x = \{e_x^r\}_r$, then this is equivalent to measuring M :

$$\begin{aligned} p(r) &= \sum_x p(x)p(r|x) = \sum_x p(x)e_x^r(s_{B|x}) = \sum_x f^x(s_A)e^r(s_A^x \otimes s_{B|x}) = \sum_x e^r(s_A^x \otimes (f^x \otimes I_B)(s_{AB})) \\ &= \sum_x e^r((s_A^x f^x \otimes I_B)(s_{AB})) = e^r(s_{AB}), \end{aligned}$$

where we have used that $\sum_x s_A^x f^x$ is the identity transformation on the classical system A . \square

We will in particular rely on the following corollary of this lemma.

Corollary 2. *Let A be classical and B be a system from an arbitrary GPT. For any fine-grained measurement $M \in \mathcal{E}_{AB}^*$ there exists an n -outcome fine-grained measurement $N_A \in \mathcal{E}_A^*$, and fine-grained measurements $N_B^x \in \mathcal{E}_B^*$ for $x = 1, 2, \dots, n$ such that M is equivalent to performing N_A , then performing N_B^x (where x is the result of N_A).*

Proof. We have already shown that N_A can be taken to be fine-grained. Suppose N_B^1 is not fine grained, and consider \hat{N}_B^1 in which $e_1^{t_1}$ is split into other effects $e_1^{t_1}$ and $e_1^{t_2}$ satisfying $e_1^{t_1} + e_1^{t_2} = e_1^{t_1}$ in a non-trivial way, i.e., with $e_1^{t_1}$ and $e_1^{t_2}$ not proportional to one another or to any other effect e_1^j with $j \neq t$. The measurement that involves measuring N_A and then \hat{N}_B^1 if $x = 1$ and otherwise N_B^x depending on the result is a non-trivial fine-graining of M . Thus, if M is fine-grained, so are N_B^x for all x . \square

Proof of Property 6. For any measurements $\{e^x\} \in \mathcal{E}_A^*$ and $\{f^y\} \in \mathcal{E}_B^*$ and any state $s_{AB} \in \mathcal{S}_{AB}$ we have

$$\begin{aligned} -\sum_x e^x(s_A) \log(e^x(s_A)) - \sum_y f^y(s_B) \log(f^y(s_B)) &\geq -\sum_{xy} (e^x \otimes f^y)(s_{AB}) \log((e^x \otimes f^y)(s_{AB})) \\ &\geq H_+(AB), \end{aligned}$$

where the first inequality uses subadditivity of the Shannon entropy and the last line follows because $\{e^x \otimes f^y\}_{x,y}$ is a measurement on AB and is hence at least as large as the infimum over joint measurements in $H_+(AB)$. Then the fact that $H_+(A)$ and $H_+(B)$ are infima over fine-grained measurements means that $\{e^x\}$ and $\{f^y\}$ can be chosen such that $-\sum_x e^x(s_A) \log(e^x(s_A)) - \sum_y f^y(s_B) \log(f^y(s_B))$ is arbitrarily close to $H_+(A) + H_+(B)$ implies result. \square

Proof of Property 7. Using Corollary 2, the measurement on BC in $H_+(A|BC)$ can be decomposed into a standard measurement on B , yielding y , followed by a fine grained measurement on C depending on the value of y obtained, i.e.,

$$H_+(A|BC) = \sum_y e^y(s_B) \inf_{\{h_y^z\}_z \in \mathcal{E}_C^*} \sum_z h_y^z(s_{C|y}) H(A|_{yz}).$$

For some fixed set $\{h_y^z\}_z \in \mathcal{E}_C^*$, the right hand side (without the inf) takes the form

$$\begin{aligned} \sum_{yz} p(y, z) H(A|_{yz}) &= H(A|YZ) = H(AY|Z) - H(Y|Z) \geq H(AB|Z) - H(Y) \\ &= H(AB|Z) - H_+(B) \geq H_+(AB|C) - H_+(B), \end{aligned}$$

where we have used that there is a standard measurement achieving the infimum for classical systems in $H_+(B)$ and the subadditivity of the Shannon entropy. The result follows because we can choose $\{h_{yz}^z\} \in \mathcal{E}_C^*$ such that the left hand side is arbitrarily close to $H_+(A|BC)$. \square

Proof of Property 8. We start from the definition of $H_+(AB|C)$, and use Corollary 2 to give

$$\begin{aligned}
H_+(AB|C) &= \inf_{\{f^z\} \in \mathcal{E}_C, \{e_z^x\}_x \in \mathcal{E}_{AB}^*} - \sum_z f^z(s_C) \sum_x e_z^x(s_{AB|z}) \log(e_z^x(s_{AB|z})) \\
&= \inf_{\{f^z\} \in \mathcal{E}_C, \{g_{yz}^x\}_x \in \mathcal{E}_A^*} - \sum_z f^z(s_C) \sum_{xy} h^y(s_{B|z}) g_{yz}^x(s_{A|yz}) \log(h^y(s_{B|z}) g_{yz}^x(s_{A|yz})) \\
&= \inf_{\{f^z\} \in \mathcal{E}_C, \{g_{yz}^x\}_x \in \mathcal{E}_A^*} - \sum_z f^z(s_C) \left[\sum_y h^y(s_{B|z}) \log(h^y(s_{B|z})) + \sum_{xy} h^y(s_{B|z}) g_{yz}^x(s_{A|yz}) \log(g_{yz}^x(s_{A|yz})) \right] \\
&\geq \inf_{\{f^z\} \in \mathcal{E}_C} - \sum_z f^z(s_C) \sum_y h^y(s_{B|z}) \log(h^y(s_{B|z})) + \\
&\quad \inf_{\{f^z\} \in \mathcal{E}_C, \{g_{yz}^x\}_x \in \mathcal{E}_A^*} - \sum_z f^z(s_C) \sum_{xy} h^y(s_{B|z}) g_{yz}^x(s_{A|yz}) \log(g_{yz}^x(s_{A|yz})) \\
&\geq H_+(B|C) + H_+(A|BC) .
\end{aligned}$$

In the last inequality we use that a measurement on C followed by a fine-grained measurement on A is a joint measurement on s_{AC} , so the infimum over all joint measurements cannot be larger than this term.

If C is classical then we can drop $\inf_{\{f^z\} \in \mathcal{E}_C}$ and take C to always be measured with a standard classical measurement. This gives equality in both inequalities in the above proof. \square

B Proof of Theorem 1 and Corollary 1

The proof relies on the following Lemmas.

Lemma 5. *For any causal structure C^G where all observed variables share one observed ancestor the topological closure of the set of achievable entropy vectors of the observed variables is convex.*

Proof. We first prove convexity, and then show that the set form a cone.

Let C^G have n observed variables and m unobserved ones. Let \mathbf{H}_1 and \mathbf{H}_2 be two achievable entropy vectors for the n observed variables in C^G . In the following, we show that for any $0 \leq p \leq 1$, there is a sequence of entropy vectors \mathbf{H}'_k within C^G , such that $\lim_{k \rightarrow \infty} \mathbf{H}'_k = p\mathbf{H}_1 + (1-p)\mathbf{H}_2$.

For $i = 1, 2$, suppose that \mathbf{H}_i is generated by using states $\{Y_j^i\}_{j=1}^m$ for the m unobserved nodes and that the observed random variables are $\{X_j^i\}_{j=1}^n$. The strategy for achieving the convex combination is as follows. The common observed variable is taken to be $X_1'(k) = (A, Z)$ where

$$(A, Z) = \begin{cases} (0, 0) & \text{with probability } \frac{k-1}{k} \\ (1, (X_1^1)^k) & \text{with probability } \frac{p}{k} \\ (2, (X_1^2)^k) & \text{with probability } \frac{1-p}{k} \end{cases}$$

and where X^k denotes k i.i.d. copies of a random variable X . Each of the unobserved nodes is prepared in state

$$Y_j' = (Y_j^1)^{\otimes k} \otimes (Y_j^2)^{\otimes k} .$$

Each of the other observed nodes then behaves as follows. The children of X_1 have access to A . If $A = 0$ they output $X_j'(k) = (0, 0)$. If $A = 1$ they perform the operation that would have led to \mathbf{H}_1 k times independently, acting on the first k subsystems of any GPT resources they have access to. They then output $X_j'(k) = (1, (X_j^1)^k)$. If $A = 2$, the procedure is the same except that the operation that would have led to \mathbf{H}_2 is repeated k times by acting on the second k subsystems of any GPT

resources and the output is $X'_j(k) = (2, (X_j^2)^k)$. Note that the first part of the argument is equal to A , so, in this way, the value of A is transferred to all descendants. An analogous strategy is then used for subsequent generations.

For any subset S of the observed random variables $\{X'_j(k)\}_{j=1}^n$ we have

$$H'_k(S) = H'_k(A) + pH^1(S) + (1-p)H^2(S),$$

where H'_k refers to the entropy in the new strategy and H^1 and H^2 refer to the entropies in the original strategies (i.e., according to \mathbf{H}_1 or \mathbf{H}_2). Noting that

$$H'_k(A) = -\frac{k-1}{k} \log \frac{k-1}{k} - \frac{p}{k} \log \frac{p}{k} - \frac{1-p}{k} \log \frac{1-p}{k}$$

tends to 0 as k tends to ∞ , we have $\lim_{k \rightarrow \infty} \mathbf{H}'_k = p\mathbf{H}_1 + (1-p)\mathbf{H}_2$.

If \mathbf{H}_1 and \mathbf{H}_2 are themselves only achievable as limits of entropy vectors the above argument can be followed for each vector in the corresponding sequences tending to \mathbf{H}_1 and \mathbf{H}_2 respectively and thus also holds for \mathbf{H}_1 and \mathbf{H}_2 . This shows that the closure of the set of entropy vectors is convex. \square

The next lemma extends this beyond the case where there is a common observed ancestor.

Lemma 6. *For any causal structure C^G the topological closure of the set of achievable entropy vectors of the observed variables is convex.*

Proof. If all observed variables in C^G have a common observed ancestor, the statement follows by Lemma 5. Otherwise, there are $1 < l \leq n$ observed nodes without any observed ancestors, which we label X_1, \dots, X_l (all other observed nodes (X_{l+1}, \dots, X_n) are descendants of at least one of these nodes). We construct a larger causal structure C' by introducing an observed parent node A_i for each X_i with $i = 1, \dots, l$, where A_i has no direct link to any variable except for X_i . Note that a distribution over the observed variables X_1, \dots, X_n is compatible with C^G if and only if it is the marginal of a distribution over $X_1, \dots, X_n, A_1, \dots, A_l$ that is compatible with C' .

Now let C'' be another causal structure that is constructed from C' by adding a directed link from A_1 to all other A_i with $2 \leq i \leq l$. A distribution over $X_1, \dots, X_n, A_1, \dots, A_l$ is compatible with C' if and only if it is compatible with C'' and, at the same time, obeys $I(A_1 : A_i) = 0$ for all $2 \leq i \leq l$.

The “if” condition follows because any distribution in C' obeys $I(A_1 : A_i) = 0$ for all $2 \leq i \leq l$ and it can be realised in C'' without using the additional causal links $A_1 \rightarrow A_i$.

For the “only if”, we use that $I(A_1 : A_i) = 0$ holds if and only if $p(a_i a_1) = p(a_i)p(a_1)$ ¹⁰, so that any distribution in C'' obeying $I(A_1 : A_i) = 0$ for all $2 \leq i \leq l$ can be written as

$$p(x_1, \dots, x_n, a_1, \dots, a_l) = p(x_{l+1}, \dots, x_n | x_1, \dots, x_l) p(x_1 | a_1) \dots p(x_l | a_l) p(a_1) \dots p(a_l), \quad (25)$$

with the right hand side compatible with C' .

Hence, a distribution over X_1, \dots, X_n is compatible with C^G if and only if it is the marginal of a distribution over $X_1, \dots, X_n, A_1, \dots, A_l$ that is compatible with C'' and obeys $I(A_1 : A_i) = 0$ for all $2 \leq i \leq l$.

The closure of the set of entropy vectors of the observed variables that are compatible with C'' (without any additional constraints) is convex by Lemma 5. The closure of the set of achievable entropy vectors in C^G is the closure of the set of achievable entropy vectors in C'' restricted by the linear equalities $I(A_1 : A_i) = 0$ for all $2 \leq i \leq l$ and projected to the marginals involving only X_1, \dots, X_n . Because these operations preserve convexity, the closure of the set of achievable entropy vectors of C^G is convex. \square

The main theorem of this section now follows as a corollary.

¹⁰This can for instance be seen by writing the mutual information in terms of the relative entropy and using Klein’s inequality [44].

Proof of Theorem 1. Convexity of the set of achievable entropy vectors follows by Lemma 6. That it is a cone follows because if \mathbf{H} is an achievable entropy vector, then $k\mathbf{H}$ for $k \in \mathbb{N}$ is achievable by taking k independent copies of all systems in the strategy achieving \mathbf{H} . Furthermore, in any causal structure C^G , $\mathbf{H} = \mathbf{0}$ is achievable by taking all observed variables to be 0 with probability 1. Hence, by taking convex combinations, if \mathbf{H} is achievable, so is $\lambda\mathbf{H}$ for any $\lambda \in \mathbb{R}_{\geq 0}$. \square

Corollary 1 then follows in a similar way.

Proof of Corollary 1. Consider first postselecting on one of the parentless variables in C^G , and suppose that this variable has k possible values. Let X_1, \dots, X_n be the set of all the observed descendants of the variable that has been postselected on and Y_1, \dots, Y_m be the set of all other observed nodes. In other words, for a fixed distribution, P_{Y_1, \dots, Y_m} , of all other observed nodes we consider k different ways to form X_1, \dots, X_n to give $P_{X_1^1 \dots X_{n-1}^1 X_n^1 Y_1^1, \dots, Y_m^1}, \dots, P_{X_1^k \dots X_{n-1}^k X_n^k Y_1^k, \dots, Y_m^k}$ respectively. For these distributions we define an entropy vector with $k(2^{n+m} - 1)$ components by concatenating the entropy vectors of each of them. Since the marginal distributions obey $P_{Y_1^1 \dots Y_m^1} = \dots = P_{Y_1^k \dots Y_m^k}$, $(k-1)(2^{n+m} - 1)$ components can be removed from the vector. If \mathbf{H}_1 and \mathbf{H}_2 are two such achievable entropy vectors, then for any $0 \leq p \leq 1$, $p\mathbf{H}_1 + (1-p)\mathbf{H}_2$ is also such an achievable entropy vector. This follows by applying the technique used to prove Lemma 6 separately to the causal structure including only one of the k alternatives (i.e., the causal structure formed from the post-selected causal structure by removing all nodes associated with other alternatives). This strategy leads to the same distribution on Y_1, \dots, Y_m for each alternative and hence the overall entropy vector has the same entropies for all subsets of $\{Y_1, \dots, Y_m\}$. We can then postselect on further parentless variables in a similar way. \square

C Proof of Proposition 2

We use X_1, \dots, X_n for the observed variables and Y_1, \dots, Y_m for the unobserved nodes in C . For each unobserved node Y_i we use Y_i^j with $1 \leq j \leq k_i$ for the subsystems associated with the k_i outgoing edges, sometimes using $Y = \{Y_i^j\}_{i=1, j=1}^{m, k_i}$ and $X = X_1, \dots, X_n$ as a shorthand. For any unobserved node Y_i with $1 \leq i \leq m$ and for any $1 \leq j \leq k_i$, we show how to modify Y_i^j to \tilde{Y}_i^j such that, if \tilde{Y}_i^j is shared along the j^{th} outgoing edge instead of Y_i^j , the same distributions among the observed variables are obtained. This construction of \tilde{Y}_i^j will make all conditional entropies of unobserved systems positive.

If $H(Y_i^j | Y \cup X \setminus Y_i^j) \geq 0$, we set $\tilde{Y}_i^j = Y_i^j$. Otherwise, we let $\tilde{Y}_i^j = Y_i^j \otimes \alpha_i^j$, where $\alpha_i^j = \sum_a p_i^j(a) |a\rangle\langle a|$ is a system that is uncorrelated with any other system and obeys $H(\alpha_i^j) = H(Y_i^j)$. Then, $H(\tilde{Y}_i^j) = H(Y_i^j) + H(\alpha_i^j)$ and \tilde{Y}_i^j can be used to produce the same observed distributions as Y_i^j , since α_i^j may be ignored when processing the unobserved systems \tilde{Y}_i^j to obtain observed variables. Furthermore, due to the independence of the α_i^j and by weak monotonicity, for any \tilde{Y}_i^j and any set of variables, S , coexisting with \tilde{Y}_i^j ,

$$H(\tilde{Y}_i^j | S) = H(Y_i^j | S) + H(\alpha_i^j) \geq -H(Y_i^j) + H(\alpha_i^j) = 0, \quad (26)$$

where the last equality follows by construction. (Note that for an observed variable X and a set S coexisting with X , the analogous relation $H(X|S) \geq 0$ already holds.)

We now show that for any two coexisting sets $S, T \subset U$ with $S \cap T = \emptyset$ and where U is a maximal coexisting set, the conditional entropy $H(S|T)$ is positive. First of all, by strong subadditivity,

$$H(S|T) \geq H(S|U \setminus S). \quad (27)$$

Positivity of $H(S|U \setminus S)$ can be shown inductively in the cardinality of S . For cardinality 1 this is implied by (27) and (26). Assuming that this holds for any set with cardinality q , the following shows

that it also holds for any set with cardinality $q + 1$. Let there be a set of variables $S \subseteq U$ of a maximal coexisting set U with cardinality $q + 1$ and a one element subset $S_1 \subseteq S$, Writing $S = S_1 \bar{S}_1$, then

$$\begin{aligned} H(S|U \setminus S) &= H(S_1|U \setminus S) + H(\bar{S}_1|(U \setminus S) \cup S_1) \\ &= H(S_1|U \setminus S) + H(\bar{S}_1|U \setminus \bar{S}_1), \end{aligned}$$

which is at least 0 by the inductive hypothesis. It then follows from (27) that $H(S|T) \geq 0$ for all $T \subseteq (U \setminus S)$.

D Remarks on the quantum entropy vector method

This appendix gives additional information regarding the role of Proposition 2 for the quantum entropy vector method introduced in [9] (see also [22] for a review). For completeness, we first briefly introduce the details of this method.

The *quantum entropy vector method* is based on the von Neumann entropy. For any joint state of coexisting systems associated with some of the nodes (and edges) of a causal structure a joint entropy can be defined, where the notion of coexisting sets is the one discussed for the measurement entropy in the main text. However, the quantum method does not take the conditional entropies as separate variables (these would be redundant because $H(X|Y) = H(XY) - H(Y)$). For all variables within a coexisting set, the following inequalities hold,¹¹

- Strong subadditivity: For any state ρ_{XYZ} , $H(XYZ) + H(Z) \leq H(XZ) + H(YZ)$.
- Weak monotonicity: For any state ρ_{XYZ} , $H(X|Y) + H(X|Z) \geq 0$.

Note that whenever there is no entanglement between two subsystems X and Y of a state ρ_{XY} , the stronger monotonicity statement $H(X|Y) \geq 0$ holds. Since it is always possible to purify an unobserved quantum state $\rho_{A_1 \dots A_n}$, we can impose the following.

- Purification for unobserved systems: For an unobserved system in state $\rho_{A_1 \dots A_n}$ we can take $H(A_1 \dots A_n) = 0$ and for any subsystem $S \subset \{A_1, \dots, A_n\}$ we can take $H(S) = H(\{A_1, \dots, A_n\} \setminus S)$.¹²

Among the variables of different coexisting sets data processing inequalities hold.

- Data Processing: Let ρ_{XY} be the joint state of two sets of coexisting nodes X and Y and let \mathcal{E} be a completely positive trace preserving map taking Y to Z such that $(\mathcal{I} \otimes \mathcal{E})(\rho_{XY}) = \rho_{XZ}$, then $H(X|Y) \leq H(X|Z)$.¹³

In addition, the causal structure will in general imply independence constraints among observed as well as among unobserved systems. These are based on the notion of *d-separation*: for three pairwise disjoint sets of variables X , Y and Z , X and Y are d-separated by Z , if Z blocks any path from any node in X to any node in Y . A path is blocked by Z , if it contains one of the following: $i \rightarrow z \rightarrow j$ or $i \leftarrow z \rightarrow j$ for nodes i, j and a node $z \in Z$ in that path, or if it contains $i \rightarrow k \leftarrow j$, where $k \notin Z$. Note that it is possible that $Z = \emptyset$.

- Independences (following Theorem 22 (i) from [14]): For three pairwise disjoint sets of observed variables, X , Y and Z , if X and Y are d-separated by Z , then $H(X|Z) = H(X|YZ)$. (Note that $Z = \emptyset$ is allowed.)

We show with the following Lemma that weak monotonicity constraints are not relevant in this approach when considering causal structures where none of the unobserved nodes have any parents. These are the scenarios that are usually considered in the literature.

¹¹In both inequalities X, Y, Z are all coexisting and may each be made up of subsystems associated with several nodes.

¹²This can be seen by considering the Schmidt decomposition of the purified state [22].

¹³For a discussion on which data processing inequalities are relevant for computing constraints on entropy vectors we refer to [22].

Lemma 7. For any causal structure C^Q in which the unobserved quantum nodes do not have any parents, all weak monotonicity inequalities are implied by the other inequalities, i.e., for any two coexisting sets S_1, S_2 with $S_1 \cap S_2 \neq \emptyset$,

$$H(S_1 \cap S_2 | S_1 \setminus S_2) + H(S_1 \cap S_2 | S_2 \setminus S_1) \geq 0 \quad (28)$$

is redundant.

Proof. Let A denote the collection of subsystems of all unobserved nodes and S be the maximal coexisting set that includes all unobserved systems A . Then for the coexisting sets S_1, S_2 with $S_1 \cap S_2 \neq \emptyset$ we divide into three cases.

Case 1: $S_1, S_2 \subseteq S$. Let $R_1 = S_1 \cap A$ and $R_2 = S_2 \cap A$. We use the purification of unobserved systems to rewrite,¹⁴

$$\begin{aligned} & H(S_1 \cap S_2 | S_1 \setminus S_2) + H(S_1 \cap S_2 | S_2 \setminus S_1) \\ &= H(R_1) - H(R_1 | R_2) + H(S_1 \setminus R_1) - H((S_1 \setminus R_1) \setminus S_2) + H(R_2) - H(R_2 \setminus R_1) + H(S_2 \setminus R_2) - H((S_2 \setminus R_2) \setminus S_1) \\ &= H(R_1) - H(R_1 \setminus R_2) + H(A \setminus R_2) - H(A \setminus (R_2 \setminus R_1)) + H((S_1 \cap S_2) \setminus A | (S_1 \setminus S_2) \setminus A) + H((S_1 \cap S_2) \setminus A | (S_2 \setminus S_1) \setminus A), \end{aligned} \quad (29)$$

where to obtain the first equality we used that $\rho_{S_1} = \rho_{R_1} \otimes \rho_{S_1 \setminus R_1}$ and $\rho_{S_2} = \rho_{R_2} \otimes \rho_{S_2 \setminus R_2}$ and $R_1 \setminus S_2 = R_1 \setminus R_2$ and $R_2 \setminus S_1 = R_2 \setminus R_1$; in the second line we used the purity of ρ_A . The last two terms in (29) are positive because these are classical conditional entropies (none of the sets contain elements of A). The sum of the remaining four terms is then positive by strong subadditivity.

Case 2: $S_1 \cap S_2 \subseteq S$ and either $S_1 \not\subseteq S$ or $S_2 \not\subseteq S$, or both. In this case we use data-processing to give

$$H(S_1 \cap S_2 | S_1 \setminus S_2) + H(S_1 \cap S_2 | S_2 \setminus S_1) \geq H(S_1 \cap S_2 | T_1) + H(S_1 \cap S_2 | T_2), \quad (30)$$

where $T_1, T_2 \subseteq S$ are the sets of variables that are processed to $S_1 \setminus S_2$ and $S_2 \setminus S_1$ respectively. Since $S_1 \cap S_2 \subseteq S$, $S_1 \cap S_2 = T_1 \cap T_2$ and so positivity of the remaining expression follows using Case 1.

Case 3: $S_1 \cap S_2 \not\subseteq S$. In this case we can find $R_1 \subseteq S$ and R_2 with $R_2 \cap S = \emptyset$ such that $S_1 \cap S_2 = R_1 \cup R_2$, and rewrite

$$H(S_1 \cap S_2 | S_1 \setminus S_2) + H(S_1 \cap S_2 | S_2 \setminus S_1) = H(R_1 | S_1 \setminus S_2) + H(R_2 | S_1 \setminus R_2) + H(R_1 | S_2 \setminus S_1) + H(R_2 | S_2 \setminus R_2). \quad (31)$$

The second and fourth terms are positive since R_2 is classical. The first and third terms correspond to a weak monotonicity inequality like that considered in case Case 2, and so their sum is also positive. \square

By Proposition 2, instead of purifying the unobserved systems and dropping weak monotonicity, we could alternatively replace all weak monotonicity constraints by monotonicity (doing so prevents us from purifying the unobserved quantum systems). The question then arises as to the implications of each for deriving new entropy inequalities for the observed variables. The following lemma shows that the quantum approach outlined in this section (which takes the purification of unobserved systems into account) leads to entropy inequalities that are at least as tight as those obtained by considering monotonicity instead.

Constraints on the observed variables are usually derived starting from:

- (1) The Shannon constraints for the observed variables.
- (2) All independences among observed and unobserved variables that C implies.
- (3) Data processing inequalities.
- (4) Positivity of all entropies.

¹⁴As the unobserved nodes do not have any parents the systems associated with a particular node are independent of those associated with another node as well as independent of other variables coexisting with them.

- (5) Positivity of the conditional mutual information (strong subadditivity).
- (6) Positivity of the conditional entropy (monotonicity) between subsets that cannot be entangled.
- (7) Weak monotonicity between subsets that can be entangled.

Proposition 2 implies that we can add

- (8) Monotonicity between subsets that can be entangled.

and Lemma 7 implies that in the case where all the quantum nodes are parentless, instead of (8) we can add

- (8') The unobserved systems originating at a node are in a pure state, e.g., for a node A with subsystems A_1, A_2 they obey $H(A_1, A_2) = 0$ and $H(A_1) = H(A_2)$.

Lemma 8. *Let C^Q be a causal structure in which each unobserved node is parentless and has at most two children. Consider starting with the constraints (1)–(7) and performing a Fourier-Motzkin elimination to give a set of constraints on the entropies of observed variables. Call the resulting cone Γ . Consider also Γ_2 formed analogously but with (8) in addition to (1)–(7) and likewise Γ'_2 with (8') in addition to (1)–(7). We have $\Gamma \subseteq \Gamma_2$ and $\Gamma \subseteq \Gamma'_2$.*

In other words, including either (8) or (8') does not give a tighter approximation on the set of achievable entropy vectors of the observed variables.

Proof. For any unobserved node A in C , the subsystems A_1 and A_2 only occur jointly in coexisting sets whose state can be written as $\rho_{A_1 A_2} \otimes \rho_R$ where R contains all other systems in that coexisting set. This implies that for any $R_1 \subseteq R$,

$$H(A_1 A_2 R_1) = H(A_1 A_2) + H(R_1) \quad (32)$$

$$H(A_1 R_1) = H(A_1) + H(R_1) \quad (33)$$

$$H(A_2 R_1) = H(A_2) + H(R_1). \quad (34)$$

and hence that $H(A_1 A_2 R_1)$, $H(A_1 R_1)$ and $H(A_2 R_1)$ can be eliminated from all valid inequalities. Strong-subadditivity and monotonicity inequalities that include any of these three are redundant since they decompose into terms that only involve A_1 and A_2 and terms that do not involve those variables, both of which are separately implied by another valid inequality.¹⁵

The remaining types of inequalities involving both A_1 and A_2 are those with the form (up to exchange of A_1 and A_2)

$$H(A_1) + H(A_2) \geq H(A_1 A_2) \quad (35)$$

$$H(T_1 | A_1 S_2 R'_2) \geq H(T_1 | A_1 A_2 R_2) \quad (36)$$

$$H(A_1 T_1 | S_2 R'_2) \geq H(A_1 T_1 | A_2 R_2), \quad (37)$$

where $R_1, R_2, T_1 \subseteq R$, S_2 is obtained by processing A_2 and R_2 , the set T_1 coexists with S_2 , and R'_2 is the subset of all observed variables in R_2 .

By (32) and (34), the inequalities of types (36) and (37) are equivalent to

$$H(T_1 | A_1 S_2 R'_2) \geq H(T_1 | R_2), \quad (38)$$

$$H(A_1 T_1 | S_2 R'_2) \geq H(T_1 | R_2) + H(A_1 A_2) - H(A_2). \quad (39)$$

For (8) we have the additional inequalities

$$H(A_1 A_2) \geq H(A_1) \quad (40)$$

$$H(A_1 A_2) \geq H(A_2) \quad (41)$$

¹⁵Since monotonicity is only included for cq-states, the reduction does not lead to any decompositions that would require positivity of $H(A_1 | A_2)$ or $H(A_2 | A_1)$ to be implied.

and for (8'),

$$H(A_1A_2) = 0 \tag{42}$$

$$H(A_1) = H(A_2). \tag{43}$$

In the following we show that neither (40) and (41) nor (42) and (43) imply any inequalities for the observed variables other than the ones that follow without them.

For (8), the only remaining inequalities containing $H(A_1A_2)$ are (35) and (39), both of which have $H(A_1A_2)$ as a lower bound to other entropies as well as (40) and (41), where $H(A_1A_2)$ is an upper bound. After eliminating $H(A_1A_2)$, we hence obtain $H(A_1) + H(A_2) \geq H(A_1)$ and $H(A_1) + H(A_2) \geq H(A_2)$ as well as $H(A_1T_1|S_2R'_2) \geq H(T_1|R_2)$ and $H(A_1T_1|S_2R'_2) \geq H(T_1|R_2) + H(A_1) - H(A_2)$, where the first two immediately follow from positivity of the entropy and the third is implied by (38) and $H(A_1T_1|S_2R'_2) = H(T_1|A_1S_2R'_2) + H(A_1|S_2R'_2)$.

If we were not to impose the inequalities (40) and (41), the variable elimination would lead to $H(A_1) + H(A_2) \geq 0$ and $H(A_1T_1|S_2R'_2) \geq H(T_1|R_2) - H(A_2)$, the first of which is implied by positivity and the second by (38) and monotonicity for cq-states. We now show that after the elimination of $H(A_2)$ and $H(A_1)$ the additional inequalities we obtained from (40) and (41),

$$H(A_2) \geq -(H(A_1T_1|S_2R'_2) - H(T_1|R_2)) + H(A_1) \tag{44}$$

$$H(A_1) \geq -(H(A_2T_1|S_1R'_1) - H(T_1|R_1)) + H(A_2) \tag{45}$$

become redundant (here we have put back the relation where A_1 and A_2 are interchanged). To see this, assume that in addition the constraint $H(A_2) \geq H(A_1)$ holds (which could always be achieved by adding an independent system in a maximally mixed state to A_2 and which also preserves $H(A_1|A_2) \geq 0$ and $H(A_2|A_1) \geq 0$). This implies all inequalities (44). Since it is then the only inequality where $H(A_2)$ upper bounds other entropies, it is in the elimination of $H(A_2)$ combined with all inequalities that involve $H(A_2)$. It furthermore renders (45) redundant after elimination (since only the inequalities (44) have $H(A_2)$ as an upper bound). Since there is no inequality left with $H(A_1)$ as an upper bound, the subsequent elimination of $H(A_1)$ leads to the same inequalities as we obtain without including $H(A_1|A_2) \geq 0$, $H(A_2|A_1) \geq 0$ and $H(A_2) \geq H(A_1)$.

For (8'), using $H(A_1A_2) = 0$ changes (35) and (39) to $H(A_1) + H(A_2) \geq 0$ and $H(A_1T_1|S_2R'_2) \geq H(T_1|R_2) - H(A_2)$ respectively, which we have seen to be the inequalities that also follow upon elimination of $H(A_1A_2)$ if (42) is not imposed and which we have also seen to be redundant.

Now, in all inequalities where $H(A_1)$ or $H(A_2)$ occur, they are lower bounds (see (38)) or as $H(A_1) \geq 0$ or $H(A_2) \geq 0$. Thus, setting $H(A_2) = H(A_1)$ and then eliminating $H(A_1)$ is equivalent to eliminating them each separately. \square

E Inequalities for the bilocal causal structure with classical variables

In the following we give one representative of each of the 53 types of entropy inequalities for the bilocal causal structure. The remaining inequalities can be generated by using the symmetries under exchange of X_0 and X_1 , Y_0 and Y_1 as well as Z_0 and Z_1 and the exchange of the pairs of variables (X_0, X_1) with (Z_0, Z_1) . We list them in terms of the coefficients of the entropies in the inequalities such that each row understood as a vector \mathbf{v} imposes an inequality $\mathbf{v} \cdot \mathbf{H} \geq 0$ for the entropy vectors \mathbf{H} .

#	$H(X_0)$	$H(X_1)$	$H(Y_0)$	$H(Y_1)$	$H(Z_0)$	$H(Z_1)$	$H(X_0Y_0)$	$H(X_0Y_1)$	$H(X_0Z_0)$	$H(X_0Z_1)$	$H(X_1Y_0)$	$H(X_1Y_1)$	$H(X_1Z_0)$	$H(X_1Z_1)$	$H(Y_0Z_0)$	
1	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	1	0
2	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0
4	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
5	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
6	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
7	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
8	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
9	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
10	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
11	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
12	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	0
13	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-1
14	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-1
15	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
16	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
17	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
18	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-2
19	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0
20	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0
21	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0
22	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
23	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
24	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-1
25	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-1
26	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-1
27	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-1
28	0	-1	0	0	0	-1	0	-1	0	0	0	0	0	0	0	0
29	0	-1	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0
30	0	-1	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0
31	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0
32	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	0
33	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	0
34	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-2
35	0	-1	0	0	0	-1	0	-1	0	0	0	1	0	0	0	-2
36	0	-1	0	0	0	-1	0	-1	0	0	0	0	0	0	0	0
37	0	-1	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0
38	0	-1	-1	0	0	-1	0	0	0	0	0	0	0	-1	0	0
39	0	0	0	-1	0	0	0	0	0	-1	0	0	-1	0	0	0
40	0	0	0	-1	0	0	0	0	0	-1	0	0	-1	0	0	0
41	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
42	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
43	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
44	0	-1	0	0	0	-1	0	0	0	0	0	-1	0	0	0	-1
45	0	-1	0	0	0	-1	0	-1	0	0	0	0	0	0	0	-1
46	0	-1	0	0	0	-1	0	-1	0	0	0	0	0	0	0	-1
47	0	-1	0	0	0	-1	0	-1	0	0	0	0	0	0	0	-1
48	0	-1	0	0	0	-1	0	-1	0	0	0	0	0	0	0	-1
49	0	-1	0	0	0	-1	0	-1	0	0	-1	1	0	0	0	0
50	0	-1	0	0	0	-1	0	-1	0	0	-1	1	0	0	0	0
51	0	-1	0	0	0	-1	0	-1	0	0	-1	1	0	0	0	-1
52	0	-1	0	0	0	-1	0	-1	0	0	-1	1	0	0	0	-1
53	0	-1	0	0	0	-1	-1	-1	0	0	1	0	0	0	0	0

#	$H(Y_0Z_1)$	$H(Y_1Z_0)$	$H(Y_1Z_1)$	$H(X_0Y_0Z_0)$	$H(X_0Y_0Z_1)$	$H(X_0Y_1Z_0)$	$H(X_0Y_1Z_1)$	$H(X_1Y_0Z_0)$	$H(X_1Y_0Z_1)$	$H(X_1Y_1Z_0)$	$H(X_1Y_1Z_1)$
1	0	0	0	0	0	0	0	0	0	0	0
2	0	-1	1	0	0	0	0	0	0	1	0
3	0	0	-1	0	0	-1	1	0	0	1	1
4	0	0	-1	0	0	-1	1	0	0	1	1
5	0	0	-1	0	-1	0	1	0	1	0	1
6	0	0	-1	-1	0	0	1	1	1	0	1
7	0	-1	0	0	-1	1	0	0	1	0	1
8	0	-1	0	0	-1	0	1	0	1	1	0
9	0	-1	0	-1	0	1	0	1	0	0	1
10	0	-1	0	-1	0	0	1	1	0	1	0
11	1	0	0	0	0	0	0	0	0	1	1
12	0	-1	1	0	0	1	0	0	0	0	0
13	0	0	0	1	0	0	0	0	1	0	0
14	0	0	0	0	1	0	0	1	0	0	0
15	0	-2	1	0	0	1	-1	0	0	1	1
16	0	-2	1	0	-1	1	0	0	1	1	0
17	0	-2	1	-1	0	1	0	1	0	1	0
18	1	0	0	0	1	0	0	1	0	0	0
19	0	0	-1	-1	0	0	1	0	1	1	1
20	0	-1	1	0	0	1	-1	0	0	1	1
21	0	-1	0	-1	0	1	0	0	1	1	1
22	1	0	0	1	-1	0	0	0	0	1	1
23	1	0	0	0	0	0	0	1	-1	1	1
24	0	0	0	1	0	1	0	0	1	-1	0
25	0	0	0	1	0	0	1	0	1	0	-1
26	0	0	0	0	1	1	0	1	0	-1	0
27	0	0	0	0	1	0	1	1	0	0	-1
28	0	-1	0	-1	0	1	1	0	1	1	0
29	0	-1	1	1	0	0	-1	0	1	1	0
30	0	-1	1	0	1	0	-1	1	0	1	0
31	0	0	-1	-2	1	0	1	1	0	1	1
32	0	-1	0	-2	1	1	0	1	0	1	1
33	0	-2	1	0	0	1	1	0	0	1	-1
34	1	0	0	1	0	1	0	1	0	-1	0
35	1	0	0	1	0	0	1	1	0	0	-1
36	0	-1	0	-2	1	1	1	1	0	1	0
37	0	-1	1	1	0	1	-2	0	1	0	1
38	0	-1	1	0	1	1	-2	1	0	0	1
39	0	0	0	0	1	-1	1	1	-1	1	1
40	0	0	0	-1	1	1	1	1	-1	1	0
41	1	0	-1	0	-1	0	1	0	1	1	1
42	1	0	-1	-1	0	0	1	1	0	1	1
43	1	-1	0	0	-1	1	0	0	1	1	1
44	1	-1	0	-1	0	1	0	1	0	1	1
45	1	0	-1	0	-1	1	1	0	1	0	1
46	1	0	-1	-1	0	1	1	1	0	0	1
47	1	-1	0	0	-1	1	1	0	1	1	0
48	1	-1	0	-1	0	1	1	1	0	1	0
49	-1	0	0	0	1	0	1	1	1	-1	0
50	-1	-1	1	0	1	0	0	1	1	0	0
51	0	0	0	1	0	0	1	1	1	-1	0
52	0	-1	1	1	0	0	0	1	1	0	0
53	0	0	-1	0	1	1	1	-1	0	0	1