The entropy vector method is unable to certify non-classicality in line-like causal structures

Mirjam Weilenmann[*] and Roger Colbeck[†]

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

(Dated: September 6, 2016)

Bell's theorem shows that our intuitive understanding of causation must be overturned in light of quantum correlations. Nevertheless, quantum mechanics does not permit signalling and hence a notion of cause remains. Understanding this notion is not only important at a fundamental level, but also for technological applications such as key distribution and randomness expansion. It has recently been shown that a useful way to decide which classical causal structures could give rise to a given set of correlations is to use entropy vectors. These are vectors whose components are the entropies of all subsets of the observed variables in the causal structure. The entropy vector method employs causal relationships among the variables to restrict the set of possible entropy vectors.

Here, we consider whether the same approach can lead to useful certificates of non-classicality within a given causal structure. Surprisingly, we find that for a family of causal structures that include the usual bipartite Bell structure they do not. For all members of this family, no function of the entropies of the observed variables gives such a certificate, in spite of the existence of non-classical correlations. It is therefore necessary to look beyond entropy vectors to understand cause from a quantum perspective.

I. INTRODUCTION

Correlation and causation are two different things. They are however connected. Reichenbach's principle [1] says that if two events X and Y are correlated then either X causes Y, Y causes X or they have a common cause. In the standard view of causation, having a common cause corresponds to the existence of a shared random variable from which the observed correlations derive. In other words, if X and Y have a common cause, then there exists a random variable C such that $P_{XY}(x,y) = \sum_c P_C(c) P_{X|c}(x) P_{Y|c}(y)$.

In this work we will take a broader view of causation that allows the common cause to be more general than a shared random variable, a direction that has also been considered in [2–6]. In particular, we will allow shared quantum systems, so that if X and Y have a *quantum* common cause, then there exists a bipartite quantum system ρ and measurements described by POVMs $\{E_x\}_x$ and $\{F_y\}_y$ such that $P_{XY}(x,y) = \mathrm{tr}((E_x \otimes F_y)\rho)$.

In this simple case, there is no separation between the sets of classical and quantum correlations: for any P_{XY} we can find a classical common cause explanation as well as a quantum one. However, for more general causal structures this is not the case. Bell [7] was the first to notice that quantum common causes could allow for stronger correlations than their classical counterparts. Any restriction on the set of correlations that follows under the assumption that any common causes are classical has been termed a Bell inequality, and many such inequalities have been discovered (e.g. [8, 9]). The connection between Bell inequalities and the literature on
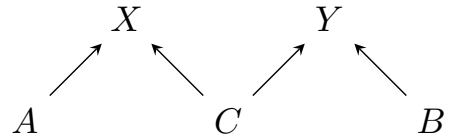
---

*Electronic address: msw518@york.ac.uk
†Electronic address: roger.colbeck@york.ac.uk

FIG. 1: The standard causal structure of a bipartite Bell experiment. Here A, B, X and Y are observed; A and B correspond to input settings and X and Y to outcomes. If the common cause is classical, then the observed correlations satisfy $P_{ABXY} = \sum_C P_A P_B P_C P_{X|AC} P_{Y|BC}$. In the case that A, B, X and Y are binary, the CHSH inequality [8] can be derived. However, if the common cause is quantum this inequality can be violated, but Tsirelson's bound must hold instead [21].

causal structures was elucidated in [10], where a novel take on Bell's theorem was given.

Ruling out classical common causes is important in information theory, and, especially, for device-independent cryptography [11–18]. In particular, it has recently been shown that the ability to demonstrate non-classicality implies the ability to generate secure random numbers [19]. It is therefore important to characterize the set of classical correlations as far as possible. Work in this direction also helps us to understand the meaning of causation in quantum theory.

In the standard Bell scenario, shown in Figure 1, the set of classical correlations is well understood. However, as the scenario is made more complicated, it rapidly becomes difficult to compute all the Bell inequalities [20], and hence to precisely separate the classical region from the non-classical.

An approach to causal structures using entropy has recently been developed [2, 3, 22–27]. The idea behind this approach is to study the entropies of the observed variables that can be realised by correlations within a given

causal structure, rather than the correlations themselves. Note that entropy has been used in (at least) two different ways in the causal structures literature. In this paper we study one of these ways and introduce the term *entropy vector method* for it. When applied to n observed variables, the central object is the vector whose $2^n - 1$ components are the entropies of each subset of the variables (excluding the empty set)[1]. This method is inviting because causal constraints correspond to linear inequalities on entropies, rather than the non-linear relations they imply for the probabilities. This means that entropy vectors are effective at distinguishing whether a set of correlations can be generated within a particular causal structure. Furthermore, the approach does not rely on any assumptions on the size of the alphabet of the involved random variables. In this paper we study the use of this approach as a means of separating classical and quantum versions of a given causal structure, focusing on a family of "line-like" causal structures that include the bipartite Bell structure.

One of the advantages of the entropy vector method is its generality—it applies to any causal structure. However, other ways to use entropy can be useful in this context and inequalities using entropy have been derived for the bipartite Bell scenario [28, 29]. These inequalities do not concern the entropies of the observed variables directly, but rather involve entropies of variables conditioned on particular outcomes of other variables. This technique has recently been generalised to other scenarios [23, 25, 30]. In the discussion we elaborate on this alternative technique and we exemplify its application to line-like causal structures in Appendix B. In contrast to the entropy vector method, this fine-grained technique is not straightforwardly applicable to general causal structures and for many causal structures it is not clear how to motivate entropic inequalities of this type.

II. THE ENTROPY VECTOR METHOD

We first outline the classical case, initially introduced in [31], and its application to causal structures [23]. For a random variable, X, distributed according to P_X we use the Shannon entropy, $H(X) := -\sum_x P_X(x) \log P_X(x)$.[2] The conditional entropy is then defined by $H(X|Y) := H(XY) - H(Y)$ and the conditional mutual information by $I(X:Y|Z) := H(XZ) + H(YZ) - H(XYZ) - H(Z)$.

A distribution over n random variables $X_1, \ldots, X_n$ has an associated entropy vector whose $2^n - 1$ components are the entropies of every subset of variables (excluding the empty set). Because they correspond to entropies of a joint distribution, these components must satisfy certain constraints. For example, they must be positive, obey monotonicity, i.e., $H(S) \leq H(RS)$, and sub-modularity (or strong subadditivity), i.e., $H(RS) + H(ST) \geq H(RST) + H(S)$, where R, S, and T denote disjoint subsets of the n random variables. Monotonicity and sub-modularity are equivalent to the positivity of the conditional entropy and conditional mutual information respectively. This set of linear constraints are called the *Shannon constraints*.

Let $\mathbf{H} : P_{X_1 \ldots X_n} \mapsto \mathbb{R}^{2^n - 1}$ denote the map from a joint distribution to its entropy vector. We will consider the set of entropy vectors that can be formed by applying $\mathbf{H}$ to a probability distribution, i.e., $\Gamma_n^* = \{v \in \mathbb{R}^{2^n - 1} : v = \mathbf{H}(P_{X_1 \ldots X_n})\}$ and its closure $\overline{\Gamma_n^*}$. The latter is known to be convex [31]. It is natural to ask whether any vector $v \in \mathbb{R}^{2^n - 1}$ that obeys the Shannon constraints is also in $\overline{\Gamma_n^*}$. It turns out that this is the case for $n \leq 3$, but does not hold for larger n [32]. Thus, the Shannon constraints are necessary but not sufficient in order for a vector to be the entropy vector of a probability distribution and the set of vectors obeying these constraints is an outer approximation to the set of achievable entropy vectors.

In order to account for the causal structure additional constraints are included. A *causal structure* comprises a set of nodes arranged in a directed acyclic graph (DAG). A subset of these nodes is designated as *observed*. If the causal structure is *classical*, each unobserved node has a corresponding random variable. For a causal structure G, we will use G^{C} to denote its classical version. If all the nodes are observed, a probability distribution is said to be *compatible* with a classical causal structure if it decomposes as

$$P_{X_1 \ldots X_n} = \prod_{i=1}^{n} P_{X_i | X_i^{\downarrow 1}} , \qquad (1)$$

where $X_i^{\downarrow 1}$ denotes the parents of X_i in the DAG. For a classical causal structure G^{C}, we will use $\mathcal{P}(G^{\mathrm{C}})$ to denote the set of compatible distributions. If not all nodes are observed, compatibility is defined by the existence of a joint distribution that is compatible with the equivalent causal structure with all nodes observed and having the correct marginal distribution over the observed nodes (see Figure 1 for an example). We will denote this set $\mathcal{P}_{\mathcal{M}}(G^{\mathrm{C}})$.

A probability distribution decomposes as in (1) if and only if every variable X_i is independent of its non-descendants $X_i^{\uparrow}$ conditioned on its parents $X_i^{\downarrow 1}$ (cf. Theorem 1.2.7 in [33]). Thus, for a DAG with n variables, the compatibility constraints are implied by a minimal set of (at most) n equations. In terms of entropies, these constraints can be concisely written as $I(X_i : X_i^{\uparrow} | X_i^{\downarrow 1}) = 0$, which are linear equalities in the entropies.

---

[1] Note that when we refer to the entropy vector method, we consider entropies of observed variables. In particular, alternative approaches that condition on output values of some of the observed variables are not included in this terminology (see later in the introduction as well as in the discussion and in Appendix B for details on alternative approaches of the latter kind).

[2] In this work lower case letters are used to denote particular instances of upper case random variables, and all random variables are taken to have finite alphabet.

In general, the set of constraints on the underlying causal structure implies additional constraints on the observed variables. These can be found by Fourier-Motzkin elimination [34] (see also [3, 26] for more details on its application to causal structures).

For a causal structure G, we denote the set of achievable entropy vectors by $\Gamma_{\mathcal{M}}^*(G^{\mathrm{C}}) := \{v : \exists P \in \mathcal{P}_{\mathcal{M}}(G^{\mathrm{C}}) \text{ with } v = \mathbf{H}(P)\}$. The closure of this, $\overline{\Gamma_{\mathcal{M}}^*}(G^{\mathrm{C}})$, is convex[3].

The entropy vector approach was generalized to the quantum case in [35], and its application to causal structures detailed in [3], which we now summarize. The relevant generalization of the Shannon entropy is the von Neumann entropy. For a system in state ρ on $\mathcal{H}_A$, it is defined by $H(A) := -\mathrm{tr}(\rho \log \rho)$, and the quantum conditional entropy and conditional mutual information are defined by replacing the Shannon entropy by the von Neumann entropy in the classical definitions. For a quantum system comprising n subsystems, we can again define a vector $v \in \mathbb{R}^{2^n-1}$ whose entries are the corresponding von Neumann entropies. Like the Shannon entropy, the von Neumann entropy is always positive and obeys sub-modularity. However, it does not in general obey monotonicity, but instead satisfies *weak monotonicity*, i.e., $H(R) + H(S) \leq H(RT) + H(ST)$. We call this set of constraints *von Neumann constraints*. Like in the classical case, these constraints are necessary, but not sufficient in order that a given $v \in \mathbb{R}^{2^n-1}$ corresponds to the von Neumann entropies of a joint quantum state [35].

Rather than discuss the quantum version of arbitrary causal structures, we consider here a restricted class that will be sufficient for our purposes. In particular, we will consider causal structures with only two generations, the first of which consists of the unobserved variables and the second of the observed ones. These causal structures are, for example, relevant in the case that spacelike separated observations are made (so that none of the observed variables can be the cause of any other). For convenience, we will use C_i or C, D, E etc. for unobserved nodes, and X_i or W, X, Y etc. for observed ones. In this case, if the causal structure is *quantum*, each edge of the graph has an associated Hilbert space, which can be labelled by the parent and child, e.g., there will be a Hilbert space $\mathcal{H}_{C_X}$ if the DAG contains $C \to X$. For each unobserved node there is an associated quantum state, a density operator on the tensor product of the Hilbert spaces associated with the edges coming from that node. For each observed node there is an associated POVM that acts on the tensor product of the Hilbert spaces associated with the edges that meet at that node. The corresponding correlations are those resulting from performing the specified POVMs on the relevant systems via the Born
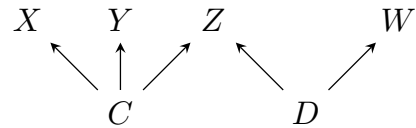

FIG. 2: Example causal structure. If this is quantum, then the realisable correlations are those formed by measuring states of the form $\rho_{C_X C_Y C_Z} \otimes \sigma_{D_Z D_W}$ with separate measurements on $\mathcal{H}_{C_X}$, $\mathcal{H}_{C_Y}$, $\mathcal{H}_{C_Z D_Z}$ and $\mathcal{H}_{D_W}$.

rule. An example is shown in Figure 2. With respect to a causal structure G, we use $\mathcal{P}_{\mathcal{M}}(G^{\mathrm{Q}})$ to denote the set of distributions on the observed nodes that can be realised if the causal structure is quantum.

In the entropic picture, there is an entropy for each observed node and for each edge of the DAG in question (for convenience we will refer to both of these as subsystems in the following[4]). While for n jointly distributed random variables, all the joint entropies make sense, this is no longer the case in a quantum causal structure with n subsystems. In particular, the subsystems corresponding to the edges that meet at an observed node do not coexist with the outcome at that node and hence there is no joint quantum state from which the joint entropy can be derived. For example, if a measurement is performed on $\mathcal{H}_{C_X} \otimes \mathcal{H}_{D_X}$ with outcome X, then $H(C_X D_X X)$ is not well-defined, although $H(C_X D_X)$ is[5]. To avoid this problem, the approach only considers entropies of coexisting sets. Two subsystems are said to *coexist* if neither is a quantum ancestor of the other, and a set of subsystems that pairwise coexist form a *coexisting set*.

Within each coexisting set the von Neumann constraints hold. However, since the observed subsystems are classical, some of the weak monotonicity constraints can be replaced by monotonicity. For example, if either R or S is a set of classical variables, then the monotonicity constraint $H(RS) \geq H(R)$ holds.

The causal constraints are accounted for by the condition that two subsets of a coexisting set are independent (and hence have zero mutual information between them) if they have no shared ancestors. To connect different coexisting sets, data processing inequalities are used. For example, if a measurement is performed on $\mathcal{H}_{C_Y} \otimes \mathcal{H}_{D_Y}$ with outcome Y, then $I(C_Y D_Y : X) \geq I(Y : X)$ (cf. Figure 2).

Like in the classical case, we denote the set of achievable entropy vectors by $\Gamma_{\mathcal{M}}^*(G^{\mathrm{Q}}) := \{v : \exists P \in \mathcal{P}_{\mathcal{M}}(G^{\mathrm{Q}}) \text{ with } v = \mathbf{H}(P)\}$, and its closure $\overline{\Gamma_{\mathcal{M}}^*}(G^{\mathrm{Q}})$ is again convex.

---

[3] This follows from the convexity of $\overline{\Gamma_n^*}$ and the fact that the causal constraints correspond to projections of this.

[4] Note, however, that they are not all subsystems of one joint quantum state.

[5] Note also that in the classical case the analogous argument fails as information can always be copied.
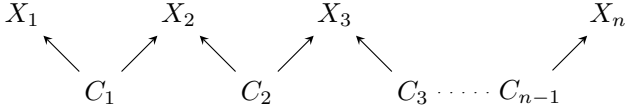
FIG. 3: The causal structure P_n. The nodes X_i represent observed variables, whereas the C_i denote the unobserved classical or quantum systems.

III. LINE-LIKE CAUSAL STRUCTURES

For the remainder of this paper, we consider the family of line-like causal structures shown in Figure 3. The causal structure P_n has observed nodes $X_1, X_2, \ldots, X_n$. Each pair of consecutive observed nodes X_i and X_{i+1} has an unobserved parent C_i.

The case $n = 4$ is in one-to-one correspondence with the bipartite Bell causal structure of Figure 1 [36]. To make the identification, take $X_1 = A$, $X_2 = X$, $X_3 = Y$, $X_4 = B$ and $C_2 = C$. We can assume without loss of generality that $C_1 = A$ and $C_3 = B$: the same set of observed correlations can be generated in either case[6].

In the classical case the node C corresponds to a local hidden variable. Free choice of settings, crucial to the derivation of a Bell inequality, is naturally encoded in the causal structure (e.g., $P_{A|BYC} = P_A$ follows as A has no parents but BYC as its non-descendants), as are the conditions of *local causality*, that $P_{XY|ABC} = P_{X|AC}P_{Y|BC}$. The only difference between P_4^{C} and the quantum case, P_4^{Q}, is the nature of the node C. Bell's original argument then implies that there are non-classical correlations, i.e., there are distributions in $\mathcal{P}_{\mathcal{M}}(P_4^{\mathrm{Q}})$ that are not in $\mathcal{P}_{\mathcal{M}}(P_4^{\mathrm{C}})$.

In the following we will prove that, in spite of this separation, by looking at the entropy vectors no distinction can be made. This is stated more formally as follows.

Theorem 1. $\overline{\Gamma_{\mathcal{M}}^*}(P_n^{\mathrm{Q}}) = \overline{\Gamma_{\mathcal{M}}^*}(P_n^{\mathrm{C}})$ *for all* $n \in \mathbb{N}$.

Note that for $n \leq 3$, $\mathcal{P}_{\mathcal{M}}(P_n^{\mathrm{C}}) = \mathcal{P}_{\mathcal{M}}(P_n^{\mathrm{Q}})$ and hence in these cases the lemma immediately follows [36]. We proceed to give the argument for $n = 4$, deferring the general case to Appendix A.

Note also that the $n = 5$ case is closely related to the so-called bilocality scenario [37, 38], introduced in the context of entanglement swapping. The difference to P_5 is that bilocality also allows an additional observed "input" to the central node. In fact, following an analogous argument to that of Theorem 1 reveals that in the bilo-

cality scenario there is also no separation between the classical and quantum entropy cones.

Proof of Theorem 1 for $n = 4$. The entropy vector of the joint distribution of A, X, Y and B has to obey the Shannon inequalities in both P_4^{C} and P_4^{Q}. In addition, the causal structure directly implies the following independences among the four observed variables[7]:

$$
\begin{aligned}
I(A : YB) &= 0, \\
I(AX : B) &= 0.
\end{aligned}
\tag{2}
$$

In both the classical and the quantum case, if the unobserved subsystems are included, there are further valid (in)equalities implied by the causal structure. The following argument shows, however, that these do not impart any additional constraints on the entropy vector of the observed nodes: in P_4 the Shannon inequalities together with (2) fully characterise the set of achievable entropy vectors of the observed nodes in both the classical and quantum case.

The Shannon inequalities on four variables together with (2) are necessary conditions on a vector $v \in \mathbb{R}^{15}$ in order that there is a distribution P_{AXYB} in $\mathcal{P}_{\mathcal{M}}(P_4^{\mathrm{C}})$ with $\mathbf{H}(P_{AXYB}) = v$. They therefore form an an outer approximation to $\overline{\Gamma_{\mathcal{M}}^*}(P_4^{\mathrm{C}})$. This outer approximation is a convex cone that can equivalently be expressed via its extremal rays. Conversion between these two descriptions can be conveniently done using software such as PORTA [39] or PANDA [40] and results in the following rays, where the components are ordered as

$$
\begin{aligned}
&(H(A), H(X), H(Y), H(B), H(AX), H(AY), \\
&\quad H(AZ), H(XY), H(XB), H(YB), H(AXY), \\
&\quad H(AXB), H(AYB), H(XYB), H(AXYB)),
\end{aligned}
$$

(i)	111122222233333
(ii)	011111122222222
(iii)	111022121122222
(iv)	000100101101111
(v)	001001010110111
(vi)	010010011011011
(vii)	100011100011101
(viii)	001101111111111
(ix)	011011011111111
(x)	110011111011111.

If each of these rays is achievable using a distribution in $\mathcal{P}_{\mathcal{M}}(P_4^{\mathrm{C}})$ then, by convexity of $\overline{\Gamma_{\mathcal{M}}^*}(P_4^{\mathrm{C}})$, the outer approximation is tight. In other words, any vector v that obeys the Shannon constraints and (2) is achievable, i.e.,

---

[6] To see this, note that for any bipartite quantum state ρ_{CD}, if C is measured to generate A, then the post measurement state has the form $\sum_a P_A(a)|a\rangle\langle a| \otimes \rho_D^a$. The same joint state can be generated by sharing $A = a$ with distribution $P_A(a)$ and simulating the statistics of the state ρ_D^a at D conditioned on $A = a$.

[7] Note that A and YB do not share any ancestors (similarly AX and B).

in $\overline{\Gamma_{\mathcal{M}}^{*}}(P_4^{\mathrm{C}})$. We establish this by taking C_1, C_2 and C_3 to be uniform random bits and use the following functions:

- (i): Take $A = C_1$, $X = C_1 \oplus C_2$, $Y = C_2 \oplus C_3$ and $B = C_3$.

- (ii): Let $A = 1$ be deterministic and choose $X = C_2$, $Y = C_2 \oplus C_3$ and $B = C_3$. (iii) can be achieved with an analogous strategy, where $B = 1$ is the deterministic variable.

- (iv): Choose $A = X = Y = 1$ and $B = C_3$. (v), (vi) and (vii) are permutations of this strategy.

- (viii): Let $A = X = 1$ be deterministic and let $Y = B = C_2$. (ix) and (x) are permutations of this.

The outer approximation of the set of entropy vectors that are achievable classically, $\overline{\Gamma_{\mathcal{M}}^{*}}(P_n^{\mathrm{C}})$, given here is also an outer approximation to $\overline{\Gamma_{\mathcal{M}}^{*}}(P_n^{\mathrm{Q}})$. Since the extremal rays are achievable the lemma follows. $\qquad\square$

IV. DISCUSSION

Although for all $n \geq 4$ there are distributions in $\mathcal{P}_{\mathcal{M}}(P_n^{\mathrm{Q}})$ that cannot be achieved in $\mathcal{P}_{\mathcal{M}}(P_n^{\mathrm{C}})$ the entropy vector approach we have outlined is unable to detect this. Even correlations that in other contexts are thought of as strongly non-classical have this masked under the mapping to entropy vectors: no function of the entropy vector acts as a certificate of non-classicality in these causal structures. It is an interesting open question as to whether this is generic: i.e., can entropy vectors ever detect the difference between classical and quantum versions of a given causal structure? We discuss this question in more detail in [41].

Because of the shortcomings of the entropy vector method, other techniques will be needed to separate classical and quantum causal structures. Recently, other approaches to this have been developed, one involving polynomial Bell inequalities [42, 43] and the other drawing on tools from algebraic geometry [44].

As mentioned in the introduction, for certain causal structures (including line-like ones), an alternative entropic technique can be applied, as first introduced by Braunstein and Caves [28]. In our terminology, the inequality of [28] states that in the causal structure P_4^{C}

$$H(Y|X)_{11} + H(X|Y)_{10} + H(X|Y)_{01} - H(X|Y)_{00} \geq 0, \tag{3}$$

where $H(X|Y)_{ab}$ is the conditional entropy of the conditional distribution $P_{XY|A=a,B=b}$.

The crucial idea behind the derivation of inequality (3) is that in the classical case there exists a joint distribution $P'_{X_0 X_1 Y_0 Y_1}$ whose marginals satisfy $P'_{X_a Y_b} = P_{XY|A=a,B=b}$ for all a and b [54, 55]. In the quantum case there is no such distribution in general, and hence (3) does not apply. Such inequalities are not obtained with the entropy vector method because the latter does not consider conditioning on particular outcomes.

It was shown in [25] that every non-local distribution in P_4 can be used to violate such an inequality if one takes an appropriate convex combination with a local distribution. In fact, the inequality (3) and its permutations are the only relevant inequalities for two measurements with dichotomic outcomes for each party [23]. These inequalities can also be generalized to the chained Bell inequalities [28], which allow for A and B to take any number of values [23, 25].

Entropic inequalities of this type (i.e., after conditioning on output values of some of the observed variables) may arise in other classical causal structures [8]. In Appendix B, we show how additional entropic inequalities for P_5^{C} and P_6^{C} may be derived with this technique. It is an open question, however, as to whether any quantum violations of these extra inequalities exist.

It is natural to ask whether the entropy vector method can be used with other entropy measures, the family of Rényi entropies [46] being a natural alternative, as considered in [47]. These do not obey sub-modularity and hence the set of allowed entropy vectors is (using known inequalities) far less constrained than in the von Neumann case. Although Rényi conditional entropies satisfy $H_\alpha(A|BC) \leq H_\alpha(A|B)$ [48–52], because the conditional Rényi entropy cannot be expressed as a difference of unconditional entropies, these relations do not lead to constraints on the Rényi entropy vector. Including conditional entropies as separate elements of the entropy vector would allow use of these relations, but given the expanded length of the vector and the comparatively small number of additional constraints, we don't expect this to be fruitful without further inequalities between Rényi entropies.

One can also look at causal structures that allow post-quantum non-signalling systems, such as non-local boxes [21, 45], to be shared. One approach to this has been presented in [2]. In the case of P_4, this yields the constraints of (2) on the observed variables. Hence, the proof of Theorem 1 can be used to show that functions of the entropy vector of the observed variables cannot detect post-quantum non-locality either. Whether the entropy vectors are ever able to encode information about the physical nature of the involved variables, rather than mere independences, remains an open question.

---

[8] Note, however, that the application of this method to general causal structures is not so straightforward, as the justification of a statement similar to Fine's theorem is for many of them not evident.

---

[1] Reichenbach, H. *The direction of time* (University of California Press, 1956).

[2] Henson, J., Lal, R. & Pusey, M. F. Theory-independent limits on correlations from generalized Bayesian networks. *New Journal of Physics* **16**, 113043 (2014).

[3] Chaves, R., Majenz, C. & Gross, D. Information-theoretic implications of quantum causal structures. *Nature communications* **6**, 5766 (2015).

[4] Pienaar, J. & Brukner, Č. A graph-separation theorem for quantum causal models. *New Journal of Physics* **17**, 073020 (2015).

[5] Fritz, T. Beyond Bell's Theorem II: Scenarios with Arbitrary Causal Structure. *Communications in Mathematical Physics* **341**, 391–434 (2015).

[6] Chaves, R. & Budroni, C. Entropic nonsignalling correlations. e-print `arXiv:1601.07555` (2016).

[7] Bell, J. S. On the Einstein Podolsky Rosen paradox. *Physics* **1**, 195–200 (1964).

[8] Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters* **23**, 880–884 (1969).

[9] Greenberger, D. M., Horne, M. & Zeilinger, A. Going beyond Bell's theorem. In Kafatos, M. (ed.) *Bell's Theorem, Quantum Mechanics and Conceptions of the Universe*, 69–72 (Kluwer Academic, Dordrecht, The Netherlands, 1989).

[10] Wood, C. J. & Spekkens, R. W. The lesson of causal discovery algorithms for quantum correlations: Causal explanations of bell-inequality violations require fine-tuning. *New Journal of Physics* **17**, 033002 (2015).

[11] Ekert, A. K. Quantum cryptography based on Bell's theorem. *Physical Review Letters* **67**, 661–663 (1991).

[12] Mayers, D. & Yao, A. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)*, 503–509 (IEEE Computer Society, Los Alamitos, CA, USA, 1998).

[13] Barrett, J., Hardy, L. & Kent, A. No signaling and quantum key distribution. *Physical Review Letters* **95**, 010503 (2005).

[14] Acín, A., Gisin, N. & Masanes, L. From Bell's theorem to secure quantum key distribution. *Physical Review Letters* **97**, 120405 (2006).

[15] Colbeck, R. *Quantum and Relativistic Protocols For Secure Multi-Party Computation*. Ph.D. thesis, University of Cambridge (2007). Also available as `arXiv:0911.3814`.

[16] Colbeck, R. & Kent, A. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical* **44**, 095305 (2011).

[17] Pironio, S. *et al.* Random numbers certified by Bell's theorem. *Nature* **464**, 1021–4 (2010).

[18] Vazirani, U. & Vidick, T. Fully device-independent quantum key distribution. *Physical Review Letters* **113**, 140501 (2014).

[19] Miller, C. A. & Shi, Y. Universal security for randomness expansion from the spot-checking protocol. e-print `arXiv:1411.6608` (2014).

[20] All the Bell Inequalities - OpenQIProblemsWiki. URL `http://qig.itp.uni-hannover.de/qiproblems/1`.

[21] Tsirelson, B. Some results and problems on quantum Bell-type inequalities. *Hadronic Journal Supplement* **8**, 329–345 (1993).

[22] Steudel, B. & Ay, N. Information-Theoretic Inference of Common Ancestors. *Entropy* **17**, 2304–2327 (2015).

[23] Fritz, T. & Chaves, R. Entropic Inequalities and Marginal Problems. *IEEE Transactions on Information Theory* **59**, 803–817 (2013).

[24] Chaves, R. & Fritz, T. Entropic approach to local realism and noncontextuality. *Physical Review A* **85**, 032113 (2012).

[25] Chaves, R. Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality. *Physical Review A* **87**, 022102 (2013).

[26] Chaves, R., Luft, L. & Gross, D. Causal structures from entropic information: geometry and novel scenarios. *New Journal of Physics* **16**, 043001 (2014).

[27] Chaves, R. *et al.* Inferring latent structures via information inequalities. In *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence*, 112–121 (AUAI Press, Corvallis, Oregon, 2014).

[28] Braunstein, S. L. & Caves, C. M. Information-Theoretic Bell Inequalities. *Physical Review Letters* **61**, 662–665 (1988).

[29] Cerf, N. J. & Adami, C. Entropic Bell inequalities. *Physical Review A* **55**, 3371–3374 (1997).

[30] Pienaar, J. Which causal scenarios are interesting? e-print `arXiv:1606.07798` (2016).

[31] Yeung, R. W. A framework for linear information inequalities. *IEEE Transactions on Information Theory* **43**, 1924–1934 (1997).

[32] Zhang, Z. & Yeung, R. W. A non-Shannon-type conditional inequality of information quantities. *IEEE Transactions on Information Theory* **43**, 1982–1986 (1997).

[33] Pearl, J. *Causality* (Cambridge University Press, 2009).

[34] Williams, H. P. Fourier's method of linear programming and its dual. *The American Mathematical Monthly* **93**, 681–695 (1986).

[35] Pippenger, N. The inequalities of quantum information theory. *IEEE Transactions on Information Theory* **49**, 773–789 (2003).

[36] Fritz, T. Beyond Bell's theorem: correlation scenarios. *New Journal of Physics* **14**, 103001 (2012).

[37] Branciard, C., Gisin, N. & Pironio, S. Characterizing the nonlocal correlations created via entanglement swapping. *Physical Review Letters* **104**, 170401 (2010).

[38] Branciard, C., Rosset, D., Gisin, N. & Pironio, S. Bilocal versus nonbilocal correlations in entanglement-swapping experiments. *Physical Review A* **85**, 032119.

[39] Christof, T. & Loebel, A. POlyhedron Representation

Transformation Algorithm (PORTA) (1997). URL `http://comopt.ifi.uni-heidelberg.de/software/PORTA/index.html`.

[40] Lörwald, S. & Reinelt, G. PANDA: a software for polyhedral transformations. *EURO Journal on Computational Optimization* **3**, 297–308 (2015).

[41] Weilenmann, M. & Colbeck, R. Non-Shannon inequalities in the entropy vector approach to causal structures. e-print `arXiv:1605.02078` (2016).

[42] Chaves, R. Polynomial Bell Inequalities. *Physical Review Letters* **116**, 010402 (2016).

[43] Rosset, D. *et al.* Nonlinear Bell inequalities tailored for quantum networks. *Physical Review Letters* **116**, 010403 (2016).

[44] Lee, C. M. & Spekkens, R. W. Causal inference via algebraic geometry: necessary and sufficient conditions for the feasibility of discrete causal models. e-print `arXiv:1506.03880` (2015).

[45] Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Foundations of Physics* **24**, 379–385 (1994).

[46] Rényi, A. On Measures of Entropy and Information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1: Contributions to the Theory of Statistics*, 547–561 (1960).

[47] Linden, N., Mosonyi, M. & Winter, A. The structure of Renyi entropic inequalities. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **469**, 20120737–20120737 (2013).

[48] Petz, D. Quasi-entropies for finite quantum systems. *Reports on Mathematical Physics* **23**, 57–65 (1986).

[49] Tomamichel, M., Colbeck, R. & Renner, R. A fully quantum asymptotic equipartition property. *IEEE Transactions on information theory* **55**, 5840–5847 (2009).

[50] Müller-Lennert, M., Dupuis, F., Szehr, O., Fehr, S. & Tomamichel, M. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics* **54**, 122203 (2013).

[51] Frank, R. L. & Lieb, E. H. Monotonicity of a relative Rényi entropy. *Journal of Mathematical Physics* **54**, 122201 (2013).

[52] Beigi, S. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics* **54**, 122202 (2013).

[53] Verma, T. & Pearl, J. Causal networks: semantics and expressiveness. *Proceedings of the 4th workshop on uncertainty in artificial intelligence (2013)* 352–359 (1990).

[54] Fine, A. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics* **23**, 1306 (1982).

[55] Fine, A. Hidden Variables, Joint Probability, and the Bell Inequalities. *Physical Review Letters* **48**, 291–295 (1982).

APPENDIX A: PROOF OF THEOREM 1

We rely on the following lemma [31]

Lemma 1. *Consider n variables $X_1, X_2, \ldots, X_n$ and define $\Omega := \{X_1, X_2, \ldots, X_n\}$. Taking positivity of each entropy to be implicit, the Shannon inequalities for these are generated from a minimal set of $n + n(n-1)2^{n-3}$ inequalities:*

$$H(\Omega | \Omega \setminus \{X_i\}) \geq 0, \qquad (A1)$$
$$I(X_i : X_j | X_S) \geq 0, \qquad (A2)$$

where the first is needed for all $X_i \in \Omega$ and the second for all $X_S \subsetneq \Omega$, $X_i, X_j \in \Omega$, $X_i, X_j \notin X_S$, $i < j$.

Now take $X_1, X_2, \ldots, X_n$ to be the observed nodes in P_n and for $i + 1 < j$ define $M_{i,j} := \{X_k\}_{k=i+1}^{j-1}$ as the set of nodes between X_i and X_j. The first part of the proof of Theorem 1 is to show that from these $n + n(n-1)2^{n-3}$ Shannon inequalities at most $\frac{n(n+1)}{2}$ are not implied by the conditional independence constraints and the remaining Shannon inequalities.

To directly read conditional independences off the DAG, we use a condition known as d-separation. For a classical causal structure, if X, Y and Z are disjoint sets of variables then X and Y are said to be *d-separated* by Z if every path from a node in X to a node in Y contains one of (i) $c \to z \to d$ with $z \in Z$, (ii) $c \leftarrow z \to d$ with $z \in Z$ or (iii) $c \to e \leftarrow d$ with $e \notin Z$. As shown by Verma and Pearl [53], if a distribution is compatible with a classical causal structure in which X and Y are d-separated by Z, then $I(X : Y | Z) = 0$.

Lemma 2. *Within the causal structure P_n^{C}, all of the sub-modularity inequalities (A2) with $M_{i,j} \nsubseteq X_S$ are implied by the causal constraints.*

Proof. Let $M_{i,j} \nsubseteq X_S$, then there is at least one node $X_k \notin X_S$ with $i < k < j$. For each such node we can partition $X_S = \{X_S^{k-}, X_S^{k+}\}$, where X_S^{k-} contains all $X_l \in X_S$ with $l < k$ and X_S^{k+} contains the elements with $X_l \in X_S$ with $l > k$ (note that both sets might be empty). Since $\{X_i\} \cup X_S^{k-}$ is d-separated from $\{X_j\} \cup X_S^{k+}$ we have

$$H(\{X_i, X_j\} \cup X_S) = H(\{X_i\} \cup X_S^{k-}) + H(\{X_j\} \cup X_S^{k+}),$$
$$H(\{X_i\} \cup X_S) = H(\{X_i\} \cup X_S^{k-}) + H(X_S^{k+}),$$
$$H(\{X_j\} \cup X_S) = H(X_S^{k-}) + H(\{X_j\} \cup X_S^{k+}),$$
$$H(X_S) = H(X_S^{k-}) + H(X_S^{k+}),$$

and thus (A2) is obeyed with equality. $\square$

Lemma 3. *Within the causal structure P_n^{C}, the $\frac{n(n-1)}{2}$ sub-modularity constraints of the form $I(X_i : X_j | M_{i,j}) \geq 0$ for all X_i, X_j with $i < j$ imply all sub-modularity constraints (A2).*

Proof. Lemma 2 shows this to hold in the case $M_{i,j} \nsubseteq X_S$. Thus, we restrict to the case $M_{i,j} \subseteq X_S$. Let us write $X_S = M_{i,j} \cup X_T$, where $X_T = X_S \setminus M_{i,j}$.

First consider the case where $X_{i-1}, X_{j+1} \notin X_T$. Here $M_{i-1,j+1}$ and X_T are d-separated and hence

$$H(\{X_i, X_j\} \cup M_{i,j} \cup X_T) = H(\{X_i, X_j\} \cup M_{i,j}) + H(X_T)$$
$$H(\{X_i\} \cup X_T) = H(X_i) + H(X_T)$$
$$H(\{X_j\} \cup X_T) = H(X_j) + H(X_T)$$
$$H(M_{i,j} \cup X_T) = H(M_{i,j}) + H(X_T)$$

so that $I(X_i : X_j|M_{i,j} \cup X_T) = I(X_i : X_j|M_{i,j})$.

Next, consider the case where $X_k \in X_T$ for $k = j+1, j+2, \ldots, j+L$, but $X_{i-1}, X_{j+L+1} \notin X_T$. By d-separation, we have $I(X_i : X_j|M_{i,j} \cup X_T) = I(X_i : X_j|M_{i,j} \cup \{X_{j+1}, \ldots, X_{j+L}\})$, and the latter expression can be more concisely written as $I(X_i : X_j|M_{i,j} \cup M_{j,j+L+1})$. Then,

$$I(X_i : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$= I(X_i : M_{j,j+L+1} \cup \{X_j\}|M_{i,j}) - I(X_i : M_{j,j+L+1}|M_{i,j})$$
$$= I(X_i : M_{j,j+L+1} \cup \{X_j\}|M_{i,j})$$
$$= I(X_i : X_j|M_{i,j}) + I(X_i : M_{j,j+L+1}|M_{i,j+1})$$
$$= I(X_i : X_j|M_{i,j}) + I(X_i : M_{j+1,j+L+1} \cup \{X_{j+1}\}|M_{i,j+1}),$$

where we have used $I(X_i : M_{j,j+L+1}|M_{i,j}) = 0$, which follows from d-separation. Noting the relation between the last term in the final line and the third line, we can proceed to recursively decompose the expression into

$$I(X_i : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$= \sum_{l=0}^{L} I(X_i : X_{j+l}|M_{i,j+l}). \quad \text{(A3)}$$

Now suppose $X_k \in X_T$ for $k = i-1, i-2, \ldots i-K$ and $k = j+1, j+2, \ldots, j+L$, but $X_{i-K-1}, X_{j+L+1} \notin X_T$. By d-separation, we have $I(X_i : X_j|M_{i,j} \cup X_T) = I(X_i : X_j|M_{i,j} \cup \{X_{i-K}, \ldots, X_{i-1}\} \cup \{X_{j+1}, \ldots, X_{j+L}\})$, and the latter expression can be more concisely written as $I(X_i : X_j|M_{i,j} \cup M_{i-K-1,i} \cup M_{j,j+L+1})$. Then,

$$I(X_i : X_j|M_{i,j} \cup M_{i-K-1,i} \cup M_{j,j+L+1})$$
$$= I(M_{i-K-1,i} \cup \{X_i\} : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$\quad - I(M_{i-K-1,i} : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$= I(M_{i-K-1,i} \cup \{X_i\} : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$= I(X_i : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$\quad + I(M_{i-K-1,i} : X_j|M_{i-1,j} \cup M_{j,j+L+1})$$
$$= I(X_i : X_j|M_{i,j} \cup M_{j,j+L+1})$$
$$\quad + I(M_{i-K-1,i-1} \cup \{X_{i-1}\} : X_j|M_{i-1,j} \cup M_{j,j+L+1}),$$

where we have used $I(M_{i-K-1,i} : X_j|M_{i,j} \cup M_{j,j+L+1}) = 0$, which follows from d-separation. Noting the relation between the last term in the final line and the third line, we can hence proceed to recursively decompose the expression into

$$I(X_i : X_j|M_{i,j} \cup M_{i-K-1,i} \cup M_{j,j+L+1})$$
$$= \sum_{k=0}^{K} I(X_{i-k} : X_j|M_{i-k,j} \cup M_{j,j+L+1}).$$

The latter can then be decomposed using (A3). $\qquad\square$

Including the n monotonicity constraints, there are at most $\frac{n(n+1)}{2}$ Shannon inequalities that are not implied by the conditional independence relations of P_n^{C}. These inequalities constrain a pointed polyhedral cone with the zero vector as its vertex. They hold for all entropy vectors in P_n^{C} and thus approximate the entropy cone $\overline{\Gamma_{\mathcal{M}}^*}(P_n^{\mathrm{C}})$ from the outside. They are also valid for $\overline{\Gamma_{\mathcal{M}}^*}(P_n^{\mathrm{Q}})$ (recall that two subsets of a coexisting set are independent if they have no shared ancestors). Note that the causal constraints reduce the effective dimensionality of the problem to $\frac{n(n+1)}{2}$, since the entropies of contiguous sequences are sufficient to determine all entropies[9].

The $\frac{n(n+1)}{2}$ inequalities can lead to at most $\frac{n(n+1)}{2}$ extremal rays, which corresponds to the number of ways of choosing $\frac{n(n+1)}{2} - 1$ inequalities to be simultaneously obeyed with equality. In the following we show that this bound is saturated by constructing $\frac{n(n+1)}{2}$ entropy vectors from probability distributions in P_n^{C}, each of which lies on a different extremal ray.

Consider the following set of distributions in P_n^{C} (leading to corresponding entropy vectors). Let $\{C_i\}_{i=1}^{n-1}$ be uniform random bits, and $1 \le i \le j \le n$. For each i, j we define a distribution $D_{i,j}$.

- For $i \le n-1$, $D_{i,i}$ is formed by taking $X_i = C_i$ and $X_k = 1$ for all $k \ne i$, while $D_{n,n}$ has $X_i = C_{i-1}$ and $X_k = 1$ for all $k \ne i$.

- For $i < j$, $D_{i,j}$ is constructed in the following. Note that depending on i and j, each of the parts indexed by k below may also be empty.

  - $X_k = 1$ for $1 \le k \le i-1$,
  - $X_i = C_i$,
  - $X_k = C_{k-1} \oplus C_k$ for $i+1 \le k \le j-1$, where $\oplus$ denotes addition modulo 2,
  - $X_j = C_{j-1}$,
  - $X_k = 1$ for $j+1 \le k \le n$.

Note that the set of distributions $\{D_{i,j}\}_{i,j}$ for $1 \le i \le j \le n$ is in one-to-one correspondence with the contiguous sequences from Ω.

Lemma 4. *The $\frac{n(n+1)}{2}$ entropy vectors of the probability distributions $\{D_{i,j}\}_{i,j}$ with $1 \le i \le j \le n$ are extremal rays of $\overline{\Gamma_{\mathcal{M}}^*}(P_n^{\mathrm{C}})$.*

Proof. It is sufficient to prove the following:

- For each i, $D_{i,i}$ obeys all of the Shannon equalities with equality except the monotonicity relation $H(\Omega) - H(\Omega \setminus \{X_i\}) \ge 0$, which is a strict inequality.

- For $i < j$, $D_{i,j}$ obeys all of the Shannon inequalities with equality except $I(X_i : X_j|M_{i,j}) \ge 0$, which is a strict inequality.

---

[9] There are n contiguous sequences of length 1, $\{H(X_i)\}_{i=1}^{n}$, $n-1$ of length 2, $\{H(X_iX_{i+1})\}_{i=1}^{n-1}$, and so on, leading to $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ in total.

For the n distributions $D_{i,i}$ all variables are independent and thus their entropy vectors automatically satisfy all sub-modularity inequalities with equality. Furthermore, for any $X_S \subsetneq \Omega$ with $X_i \notin X_S$ we have $H(\{X_i\} \cup X_S) = H(X_i)$. Thus, for $j \neq i$ we have

$$H(\Omega) - H(\Omega \setminus \{X_j\}) = 0\,,$$

while for $j = i$

$$H(\Omega) - H(\Omega \setminus \{X_j\}) = H(X_i)$$
$$> 0\,.$$

This establishes the first statement.

Consider now the $(n-1)!$ distributions $D_{i,j}$ with $i < j$. We first deal with the monotonicity constraints. For $k < i$ and $k > j$, we have $H(\Omega) = H(\Omega \setminus X_k) = j - i$. Similarly, since any $j - i - 1$ elements of $M_{i-1,j+1}$ are sufficient to determine the remaining element, we also have $H(\Omega \setminus X_k) = j - i$ for $i \leq k \leq j$. Thus, all the monotonicity constraints hold with equality.

For the sub-modularity constraints, it is useful to note that for any $D_{i,j}$ with $i < j$ we have

$$H(X_k|M_{k,l}) = \begin{cases} 1, & k = i \text{ and } k < l \leq j, \\ 1, & i < k \leq j \text{ and } k < l, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $I(X_k : X_l|M_{k,l}) = H(X_k|M_{k,l}) - H(X_k|M_{k,l+1})$ is zero unless $k = i$ and $l = j$ (in which case it is 1). This establishes the second statement, and hence completes the proof of Lemma 4. □

Note that the entropy vector of each of the $\frac{n(n+1)}{2}$ distributions belongs to a different extremal ray. We have thus shown that for each extremal ray of $\overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{C})$ there is a distribution in $\mathcal{P}_\mathcal{M}(P_n^\mathrm{C})$ whose entropy vector lies on that ray. It follows by convexity that any vector that satisfies all the Shannon constraints and the causal constraints of the marginal scenario in P_n^C is realisable in P_n^C (at least asymptotically). Since the same outer approximation is valid for $\overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{Q})$ and any classical distribution can be realised quantum mechanically, we have $\overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{C}) \subseteq \overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{Q}) \subseteq \overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{C})$ and therefore $\overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{C}) = \overline{\Gamma^*_\mathcal{M}}(P_n^\mathrm{Q})$.

APPENDIX B: REMARKS ON THE BRAUNSTEIN-CAVES TECHNIQUE

The generalization of the Braunstein-Caves technique to other causal structures is difficult, as a restriction on the alphabet size of certain variables is needed. In the case of P_4, we have such a restriction because C_1 and C_3 can be assumed to be equal to the observed A and B, whose alphabets can be determined by observation. In P_n, this can always be done for the outermost nodes, hence, in P_5^C, for example, with observed nodes A, X, Y, Z and B, with A and B binary, there exists a joint distribution $P'_{X_0 X_1 Y Z_0 Z_1}$ that gives the correct marginal distributions, i.e., $P'_{X_a Y Z_b} = P_{XYZ|A=a,B=b}$ for all a and b. This is defined by setting

$$P'_{X_0 X_1 Y Z_0 Z_1}(x,x',y,z,z') = \sum_{C_2 C_3} P_{C_2} P_{C_3} P_{X|A=0,C_2}(x) P_{X|A=1,C_2}(x') P_{Y|C_2 C_3}(y) P_{Z|B=0,C_3}(z) P_{Z|B=1,C_3}(z')\,.$$
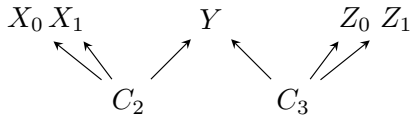

FIG. 4: Causal structure $\tilde{P}_3$ representing the independences of the variables in $P'_{X_0 X_1 Y Z_0 Z_1}$.

For $P'_{X_0 X_1 Y Z_0 Z_1}$, entropic inequalities can be derived with the entropy vector approach applied to the causal structure $\tilde{P}_3$ shown in Figure 4. Note that the P_5 scenario is related to bilocality [37, 38]. The Shannon and conditional independence constraints that restrict the corresponding five variable entropy cone are marginalized to the four triples of variables $\{X_0, Y, Z_0\}$, $\{X_0, Y, Z_1\}$, $\{X_1, Y, Z_0\}$ and $\{X_1, Y, Z_1\}$ and their subsets. As shown in [24], in addition to Shannon inequalities, it yields a further 36 (in)equalities, made up of the 7 families listed below. Note that the first family is a consequence of the Shannon constraints involving all variables, and holds independently of the causal structure.

$$H(X_1|YZ_1) + H(Z_0|X_1Y) + H(Z_1|X_0Y) - H(Z_0|X_0Y) \geq 0,$$
$$H(X_0|YZ_1) + H(Z_0|YX_1) - H(X_0|YZ_0) + H(X_0Y) - H(X_0Z_0) \geq 0,$$
$$H(Y|X_0Z_0) + H(X_1|YZ_1) - H(X_1|YZ_0) \geq 0,$$
$$H(X_1|YZ_0) + H(X_0Y) + H(YZ_1) - H(X_1Y) - H(X_0Z_1) \geq 0, \tag{B1}$$
$$H(X_0YZ_1) + H(X_1YZ_0) - H(X_1YZ_1) - H(X_0Z_0) \geq 0,$$
$$H(X_0|YZ_1) + H(YZ_0) - H(X_0Z_0) \geq 0,$$
$$I(X_0 : Z_0) = 0.$$

These can be expanded to the full set by noting the symmetry between X_0 and X_1, between Z_0 and Z_1 and between X and Z. They form an outer approximation to $\overline{\Gamma^*_{\mathcal{M}}}(\tilde{P}^{\mathrm{C}}_3)$.

The inequalities can be converted to the following extremal rays, with components ordered as
$(H(X_0),\ H(X_1),\ H(Y),\ H(Z_0),\ H(Z_1),\ H(X_0Y),$
$H(X_0Z_0),\ H(X_0Z_1),\ H(X_1Y),\ H(X_1Z_0),\ H(X_1Z_1),$
$H(YZ_0), H(YZ_1), H(X_0YZ_0), H(X_0YZ_1), H(X_1YZ_0),$
$H(X_1YZ_1))$,

(i)	11111222222222222
(ii)	01111111222222222
(iii)	01101101212121222
(iv)	01110110221212122
(v)	10111222111222222
(vi)	10101212101122212
(vii)	10110221110212221
(viii)	11101212212122222
(ix)	11110221221212222
(x)	00001001001010101
(xi)	00010010010101010
(xii)	00100100100111111
(xiii)	01000000111000011
(xiv)	10000111000001100
(xv)	00111111111111111
(xvi)	00110110110111111
(xvii)	00101101101111111
(xviii)	11100111111111111
(xix)	10100111100111111
(xx)	01100100111111111.

These rays can be generated from those of the entropic cone of P_3. To do so, let X, Y and Z be distributed according to one of the six distributions reproducing the extremal rays of the entropy cone of P_3. In the cases where X is a random bit, let either $X_0 = X$ and $X_1 = 1$, or $X_0 = 1$ and $X_1 = X$, or $X_0 = X_1 = X$, and the same for Z. Doing this for all extremal rays of P_3, the above extremal rays (i)–(xx) are recovered (as well as some additional redundant ones). This shows that the above outer approximation to $\overline{\Gamma^*_{\mathcal{M}}}(\tilde{P}^{\mathrm{C}}_3)$ is tight: all entropy vectors that satisfy the Shannon constraints and (B1) are in $\overline{\Gamma^*_{\mathcal{M}}}(\tilde{P}^{\mathrm{C}}_3)$.

The same technique can be applied to P_n via causal structures $\tilde{P}^{\mathrm{C}}_{n-2}$. For P_6 this gives a total of 16 entropic equalities, expressing independences among the involved variables and 153 inequalities (including Shannon inequalities). In the case of P_6, the extremal rays can also be generated starting from those of P_4 and splitting analogously to the treatment for P_3 above. This yields a complete characterization of $\overline{\Gamma^*_{\mathcal{M}}}(\tilde{P}^{\mathrm{C}}_4)$.

All entropic inequalities characterising $\overline{\Gamma^*_{\mathcal{M}}}(\tilde{P}^{\mathrm{C}}_3)$ and $\overline{\Gamma^*_{\mathcal{M}}}(\tilde{P}^{\mathrm{C}}_4)$ can be calculated without considering the unobserved nodes: only the Shannon inequalities and the independences among the observed variables are needed for their derivation. Note that the same independence constraints also hold in the analogous quantum casual structure. However, in the quantum case the observed X_0 and X_1 as well as Z_0 and Z_1 do not coexist and thus do not necessarily allow for a joint distribution. It is thus not justified to analyse the causal structure P^{Q}_5 using the related structure $\tilde{P}^{\mathrm{Q}}_3$, and some of the Shannon constraints among the variables $\{X_0,\ X_1,\ Y,\ Z_0,\ Z_1\}$ may not hold in the quantum case. This treatment does not therefore imply that there are no quantum violations to the classical entropic inequalities in this approach, and, at present, we do not know whether or not violations exist. If we allow post-quantum non-signalling systems, however, such violations have been found [6].