PHYSICAL REVIEW A **100**, 062108 (2019)

Analyzing causal structures using Tsallis entropies

V. Vilasini* and Roger Colbeck†

Department of Mathematics, University of York, Heslington, York YO10 5DD, England, United Kingdom

(Received 9 July 2019; published 5 December 2019)

Understanding cause-effect relationships is a crucial part of the scientific process. As Bell's theorem shows, within a given causal structure, classical and quantum physics impose different constraints on the correlations that are realizable, a fundamental feature that has technological applications. However, in general it is difficult to distinguish the set of classical and quantum correlations within a causal structure. Here we investigate a method to do this based on using entropy vectors for Tsallis entropies. We derive constraints on the Tsallis entropies that are implied by (conditional) independence between classical random variables and apply these to causal structures. We find that the number of independent constraints needed to characterize the causal structure is prohibitively high such that the computations required for the standard entropy vector method cannot be employed even for small causal structures. Instead, without solving the whole problem, we find new Tsallis entropic constraints for the triangle causal structure by generalizing known Shannon constraints. Our results reveal mathematical properties of classical and quantum Tsallis entropies and highlight difficulties of using Tsallis entropies for analyzing causal structures.

DOI: 10.1103/PhysRevA.100.062108

I. INTRODUCTION

Cause-effect relationships between physical systems constrain the correlations that can arise between them. The study of causality allows us to explain observed correlations between different variables in terms of unobserved systems that cause these variables to become correlated. This has found applications in diverse fields of research such as medical testing, socioeconomic surveys and physics. The foundational interest in causal structures stems from the fact that the theory that describes the unobserved systems affects the set of possible correlations over the observed variables. Bell inequalities [1] are constraints on the observed correlations in a classical causal structure [Fig. 1(a)] and can be violated in quantum and generalized probabilistic theories (GPTs). The possibility of such violations leads to applications in device-independent cryptography [2–7].

In the bipartite Bell causal structure [Fig. 1(a)], the set of all joint conditional distributions $P_{XY|AB}$ over the observed nodes $X, Y, A,$ and B that can arise when Λ is classical is relatively well understood. For fixed input and output sizes, it forms a convex polytope and hence membership can be checked using a linear program (although the size of the linear program scales exponentially with the number of inputs and the problem is NP-complete [8]). Because of this, the complete set of Bell inequalities characterizing these polytopes is unknown for $|X|, |Y| > 3$ or $|A|, |B| > 5$ [9–11].

In causal structures with more unobserved common causes [such as the triangle causal structure of Fig. 1(b)], the set of compatible correlations is not well understood. The inflation technique [12] can in principle certify whether or not a given distribution belongs to the classical marginal entropy cone[1] of a causal structure [13]. However, the method does not tell us how to construct a suitable inflation of the causal structure in order to achieve this, or how large this inflation needs to be. Thus, in general, using the inflation technique becomes intractable in practice. The difficulty of solving the general problem in part stems from its nonconvexity. One approach to overcoming this is to analyze the problem in entropy space [14]. This has proven to be useful in a number of cases (see, e.g., [15,16], or [17] for a detailed review), since the problem is convex in entropy space and the entropic inequalities characterizing the relevant sets are independent of the number of measurement outcomes. However, it was shown in [18] that the entropy vector method with Shannon entropies cannot detect the classical-quantum gap for linelike causal structures.[2] Further, even though new Shannon entropic inequalities have been derived using this method, no quantum violation of these has been found for a range of causal structures where nonclassical correlations are known to exist [20,21]. Due to these limitations of Shannon entropies, it is natural to ask whether other entropic quantities could do better.

Here we consider Tsallis entropies in the entropy vector method for analyzing causal structures. One motivation for considering such entropies for the task is that they are a family

*vv577@york.ac.uk

†roger.colbeck@york.ac.uk

[1]The set of possible entropy vectors over the observed nodes of the classical causal structure.

[2]Note that this result holds for the entropic characterization without postselection. Using the postselection technique (see, e.g., [17] for an explanation), one can derive quantum-violatable Shannon entropic inequalities even for linelike causal structures [19]; however, this technique is not generalizable to causal structures that have no parentless observed nodes, such as in Fig. 1(b).

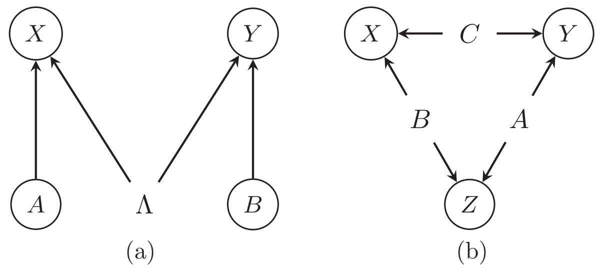

FIG. 1. Some causal structures: Observed nodes are circled and the uncircled ones correspond to unobserved nodes. (a) The bipartite Bell causal structure. The nodes A and B represent the random variables corresponding to Alice's and Bob's choices of input while X and Y represent the random variables corresponding to their outputs. Λ here is the only potentially unobserved node and is the common cause of X and Y. (b) The triangle causal structure. Here, the three observed nodes X, Y, and Z have unobserved, pairwise common causes A, B, and C, but no joint common cause.

with an additional (real) parameter. The set of entropies for all possible values of this parameter conveys more information about the underlying probability distribution than a single member of the family and hence the ability to vary a parameter may give advantages for analyzing causal structures. Tsallis entropies appear to be a good candidate since they satisfy monotonicity, submodularity, and the chain rule, which are desirable properties for their use in the entropy vector method.[3] Tsallis entropies have been considered in the context of causal structures before [23], where they were shown to give an advantage over Shannon entropy in detecting the nonclassicality of certain states in the Bell scenario if one also postselects on the values of observed parentless nodes.[4] Here we consider a systematic treatment that can be applied to an arbitrary causal structure in the absence of postselection. [Note that use of postselection is not possible in causal structures with no observed parentless nodes such as the triangle of Fig. 1(b)].

In Sec. IV, we derive the constraints on the classical Tsallis entropies that are implied by a given causal structure, and in the Appendix we generalize this result to quantum Tsallis entropies for certain cases. In Sec. IV B, we use these constraints in the entropy vector method with Tsallis entropies but find that the computational procedure becomes too time consuming even for simple causal structures such as the bipartite Bell scenario. Despite this limitation, we derive Tsallis entropic inequalities for the triangle causal structure in Sec. V, using known Shannon entropic inequalities of [21] and our Tsallis constraints of Sec. IV. In Sec. VI, we discuss the reasons for the computational difficulty of this method, the drawbacks of using Tsallis entropies for analyzing causal structure, and identify potential future directions.

[3]Other examples of more general entropy measures such as the Rényi entropy [22] do not satisfy one or more of these properties, making it more difficult to get entropic constraints on them using the entropy vector method.

[4]Note that nonclassicality cannot be detected entropically in the Bell causal structure [Fig. 1(a)] without postselection [18].

II. SHANNON ENTROPY AND THE ENTROPY VECTOR METHOD

Given a random variable X distributed according to the discrete probability distribution[5] p_X, the Shannon entropy of X is given by $H(X) = -\sum_x p_X(x)\ln p_X(x)$.[6] Given two random variables X and Y, distributed according to P_{XY}, the conditional Shannon entropy is defined by $H(X|Y) = -\sum_{x,y} p_{XY}(xy)\ln\frac{p_{XY}(xy)}{p_Y(y)}$ and the Shannon mutual information is defined by $I(X:Y) = H(X) - H(X|Y)$. For three random variables X, Y, and Z, we can also define the mutual information between X and Y conditioned on Z, $I(X:Y|Z) = H(X|Z) - H(X|YZ)$.

We will sometimes use the shorthands $p_x = p_X(x) = p(X=x)$ and $p_{x|y} := p_{X|Y}(X=x|Y=y)$, etc., for probability distributions.

We next provide a short overview of the entropy vector method that suffices for the purposes of this paper. For a more detailed overview of the method, see [17]. Consider a joint distribution $p_{X_1,\dots,X_n}$ over n random variables $X_1, X_2, \dots, X_n$. With each such distribution, we associate a vector with $2^n - 1$ components, each of which corresponds to the entropy of an element of the powerset of $\{X_1, X_2, \dots, X_n\}$ (excluding the empty set). This defines the *entropy vector* of $p_{X_1,\dots,X_n}$. Note that this vector encodes the conditional entropies and mutual information via the relations $H(X|Y) = H(XY) - H(Y)$, $I(X:Y) = H(X) + H(Y) - H(XY)$, and $I(X:Y|Z) = H(XZ) + H(YZ) - H(XYZ) - H(Z)$. We use $\mathbf{H}$ to denote the map that takes a probability distribution over n variables to its entropy vector (with $2^n - 1$ components), and Γ_n^* to denote the set of all vectors that are entropy vectors of a probability distribution $p_{X_1,\dots,X_n}$, i.e., $\Gamma_n^* = \{v \in \mathbb{R}^{2^n-1} : \exists p_{X_1,\dots,X_n} \text{ such that } v = \mathbf{H}(p_{X_1,\dots,X_n})\}$. The closure of Γ_n^*, denoted by $\overline{\Gamma_n^*}$ is known to be a convex set for any n [24].

A. Shannon cone

Valid entropy vectors necessarily satisfy certain constraints. These include positivity of the entropies, monotonicity [i.e., $H(R) \leqslant H(RS)$], and submodularity [also known as strong subadditivity; $H(RT) + H(ST) \geqslant H(RST) + H(T)$]. Monotonicity and submodularity are equivalent to the positivity of the conditional entropy $H(S|R)$ and the conditional mutual information $I(R:S|T)$, respectively, and hold for any three disjoint subsets R, S, and T of $\{X_1, \dots, X_n\}$. These linear constraints are together known as the *Shannon constraints* and the set of vectors $u \in \mathbb{R}^{2^n-1}$ obeying all the Shannon constraints form the convex cone known as the *Shannon cone*, Γ_n. Other than positivity (which, following standard practice, we include implicitly), there are a total of $n + n(n-1)2^{n-3}$ independent Shannon constraints for n variables [14]. By definition, the Shannon cone is an outer approximation to $\overline{\Gamma_n^*}$,

[5]We will only be considering random variables defined on a finite set in this paper.

[6]Note that it is common to take logarithms in base 2 and measure entropy in bits; here we use base e corresponding to measuring entropy in nats.

i.e., $\overline{\Gamma_n^*} \subseteq \Gamma_n$.[7] Hence all entropy vectors derived from a probability distribution $p_{X_1,\ldots,X_n}$ obey the Shannon constraints but not all vectors $u \in \mathbb{R}^{2^n-1}$ obeying the Shannon constraints are such that $\mathbf{H}(p_{X_1,\ldots,X_n}) = u$ for some joint distribution $p_{X_1,\ldots,X_n}$.

In the next subsection we discuss how causal structures give additional entropic constraints.

B. Entropy vectors and causal structure

A causal structure can be represented as a directed acyclic graph (DAG) over several nodes, some of which are labeled observed and some of which are labeled unobserved. Each observed node corresponds to a classical random variable,[8] while for each unobserved node there is an associated system the nature of which depends on the theory being considered. A causal structure is called *classical* (denoted $\mathcal{G}^{\mathrm{C}}$), *quantum* (denoted $\mathcal{G}^{\mathrm{Q}}$), or *GPT* (denoted $\mathcal{G}^{\mathrm{GPT}}$) depending on the nature of the unobserved nodes. In the following, we briefly review the framework of classical causal models [25].

A distribution $p_{X_1,\ldots,X_n}$ over n random variables $\{X_1,\ldots,X_n\}$ is said to be *compatible* with a classical causal structure $\mathcal{G}^{\mathrm{C}}$ (with these variables as nodes) if it satisfies the *causal Markov condition*, i.e., the joint distribution decomposes as

$$p_{X_1,\ldots,X_n} = \prod_{i=1}^{n} p_{X_i|X_i^{\downarrow_1}}, \tag{1}$$

where $X_i^{\downarrow_1}$ denotes the set of all parent nodes of the node X_i in the DAG $\mathcal{G}^{\mathrm{C}}$. The Markov condition of Eq. (1) is equivalent to the conditional independence of X_i from its nondescendants, denoted $X_i^{\not\uparrow}$ given its parents $X_i^{\downarrow_1}$ in $\mathcal{G}$, i.e., $\forall i \in \{1,\ldots,n\}$, $p_{X_i X_i^{\not\uparrow}|X_i^{\downarrow_1}} = p_{X_i|X_i^{\downarrow_1}} p_{X_i^{\not\uparrow}|X_i^{\downarrow_1}}$ [25]. All other conditional independences between different subsets of nodes are implied by these n constraints and can be derived from these constraints and standard probability calculus based on Bayes' rule. The concept of *d separation* developed by Geiger [26] and Verma and Pearl [27] provides a method to read off implied conditional independence relations from the graph. In other words, for arbitrary disjoint subsets X, Y, and Z of the nodes, it can be used to determine whether X and Y are conditionally independent given Z.

Definition 1 – Blocked paths. Let $\mathcal{G}$ be a DAG in which X and $Y \neq X$ are nodes and let Z be a set of nodes not containing X or Y. A path from X to Y is said to be *blocked* by Z if it contains either $A \to W \to B$ with $W \in Z$, $A \leftarrow W \to B$ with $W \in Z$, or $A \to V \leftarrow B$ with $V \notin Z$.

Definition 2 – d separation. Let $\mathcal{G}$ be a DAG in which X, Y, and Z are disjoint sets of nodes. X and Y are *d separated* by Z in $\mathcal{G}$ if every path from a variable in X to a variable in Y is *blocked* by Z.

The importance of d separation is that, given a causal structure $\mathcal{G}$, X and Y are d separated by Z in $\mathcal{G}$ if and only if $I(X:Y|Z) = 0$ for all distributions compatible with $\mathcal{G}$ [25].[9] The complete set of d separation conditions gives all the conditional independence relations implied by the DAG. In the case of Shannon entropy for a DAG with n nodes these are all implied by the n constraints:

$$I\big(X_i : X_i^{\not\uparrow}|X_i^{\downarrow_1}\big) = 0 \quad \forall i \in \{1,\ldots,n\}. \tag{2}$$

In other words, a distribution over n variables satisfies Eq. (1) if and only if it satisfies Eq. (2).

Since we wish to contrast classical and quantum versions of causal structures we also define the latter. For the purpose of this paper, it is sufficient to do so for causal structures with at most two generations and in which the first generation can be either observed classical random variables or unobserved quantum nodes, while those of the second generation are only observed classical variables (in the Appendix we also look at a case in which the second generation can be quantum). Each edge emanating from an unobserved node has an associated Hilbert space labeled by the parent and the child. For example, an edge from an unobserved node X to an observed node Y has the associated Hilbert space $\mathcal{H}_{X_Y}$. Each unobserved quantum node corresponds to a density operator in the tensor product of the Hilbert space corresponding to all the edges emanating from that node. For each observed node, there is a positive operator-valued measure (POVM) that acts on the tensor product of the Hilbert spaces associated with the edges that meet at that node. The set of distributions over observed nodes compatible with the quantum causal structure $\mathcal{G}^{Q}$ consists of those distributions that can be obtained by performing the specified POVMs (possibly specified by classical input nodes in the first generation) on the relevant quantum states and using the Born rule. For instance, a distribution P_{ABXY} is compatible with the quantum analog of Fig. 1(a) if there exists a quantum state $\rho \in \mathcal{H}_{\Lambda_X} \otimes \mathcal{H}_{\Lambda_Y}$ and POVMs $\{E_x^a\}_x$ and $\{F_y^b\}_y$ acting on $\mathcal{H}_{\Lambda_X}$ and $\mathcal{H}_{\Lambda_Y}$, respectively, such that $P_{ABXY}(a,b,x,y) = P_A(a)P_B(b)\,\mathrm{Tr}[\rho(E_x^a \otimes F_y^b)]$ for all values of the random variables.

Now, in the case of classical causal structures with unobserved nodes, the compatibility condition requires that there exists a joint distribution $p_{X_1,\ldots,X_n}$ over the n variables satisfying the causal Markov condition and having the correct marginals over the observed nodes. In quantum and more general theories, the existence of a joint state over all the nodes is not guaranteed because there may be sets of systems that do not coexist. (For example, there is no joint quantum state of a system and the outcome of a measurement on it.) Because classical information can be copied, such joint distributions always exist in the classical case. The entropy vector method aims to exploit this difference to certify the nonclassicality of correlations.

The entropic constraints over all the nodes will in general imply constraints on the entropy vector over the observed nodes. These can be obtained by Fourier-Motzkin elimination [28]. The procedure takes the entropy cone over all nodes, that is constrained by the $n + n(n-1)2^{n-3}$ Shannon constraints and the n causal constraints [Eq. (2)] and projects it onto the entropy cone of the observed nodes (eliminating all combinations of entropies involving unobserved nodes). Since nonclassical causal structures do not satisfy the initial assumption of the existence of the joint distribution and entropies, they may give rise to correlations that do not satisfy the

[7] For $n \leqslant 3$, the cones coincide, but for $n \geqslant 4$ they do not [24].

[8] These may represent inputs or outputs of an experiment.

[9] Note that $I(X:Y|Z) = 0$ is equivalent to $P_{XY|Z} = P_{X|Z}P_{Y|Z}$.

marginal constraints on the observed nodes obtained through this procedure. A violation of one of the inequalities certifies the nonclassicality of that causal structure.

For linelike causal structures [of which the bipartite Bell causal structure of Fig. 1(a) is an instance], the classical and quantum Shannon entropy cones coincide and Shannon entropic inequalities cannot certify the nonclassicality of these causal structures even though they support nonclassical correlations [18]. Further, in other scenarios such as the triangle, which is also known to support nonclassical correlations [29], known Shannon entropic inequalities such as those of [20,21] have no known quantum violations. The main question of the current paper is whether using Tsallis entropies can provide tighter, quantum violatable entropic inequalities and hence avoid these limitations.

III. TSALLIS ENTROPIES

For a classical random variable X distributed according to the discrete probability distribution p_X, the order q Tsallis entropy of X for real parameter q is defined as [30]

$$S_q(X) = \begin{cases} -\sum_{\{x:p_x>0\}} p_x^q \ln_q p_x & \text{if } q \neq 1 \\ H(X) & \text{if } q = 1 \end{cases} \tag{3}$$

where we have used the shorthand $\ln_q p_x = \frac{p_x^{1-q}-1}{1-q}$. This q-logarithm function converges to the natural logarithm in the limit $q \to 1$ so that $\lim_{q\to 1} S_q(X) = H(X)$ and the function is continuous in q. For brevity, we will henceforth write $\sum_x$ instead of $\sum_{\{x:p_x>0\}}$, keeping it implicit that probability zero events do not contribute to the sum.[10]

The conditional Tsallis entropy [31] is defined by

$$S_q(X|Y) := \begin{cases} -\sum_{x,y} p_{xy}^q \ln_q p_{x|y} & \text{if } q \neq 1 \\ H(X|Y) & \text{if } q = 1 \end{cases} \tag{4}$$

and converges to the Shannon conditional entropy $H(X|Y)$ in the limit $q \to 1$. Note that there are other ways to define the conditional Tsallis entropy [32] but they do not satisfy the chain rule [Eq. (10)] and will not be considered here.

The unconditional and conditional Tsallis mutual informations are defined analogously to the Shannon case:

$$I_q(X:Y) = S_q(X) - S_q(X|Y), \tag{5}$$

$$I_q(X:Y|Z) = S_q(X|Z) - S_q(X|YZ). \tag{6}$$

Properties of Tsallis entropies

Tsallis entropies satisfy a number of properties that are desirable for their use in the entropy vector method. For any joint distribution over the random variables involved the following properties hold.

(1) Pseudoadditivity [33]: For two independent random variables X and Y, i.e., $p_{XY} = p_X p_Y$, and for all q, the Tsallis entropies satisfy

$$S_q(XY) = S_q(X) + S_q(Y) + (1-q)S_q(X)S_q(Y). \tag{7}$$

[10]Note that this means the Tsallis entropy for $q < 0$ is not robust in the sense that small changes in the probability distribution can lead to large changes in the Tsallis entropy.

Note that in the Shannon case ($q = 1$) we recover additivity for independent random variables.

(2) Upper bound [34]: For $q \geqslant 0$ we have $S_q(X) \leqslant \ln_q d_X$. For $q > 0$ equality is achieved if and only if $P_X(x) = 1/d_X$ for all x (i.e., if the distribution on X is uniform).

(3) Monotonicity [35]: For all q,

$$S_q(X) \leqslant S_q(XY). \tag{8}$$

(4) Strong subadditivity [31]: For $q \geqslant 1$,

$$S_q(XYZ) + S_q(Z) \leqslant S_q(XZ) + S_q(YZ). \tag{9}$$

(5) Chain rule [31]: For all q,

$$S_q(X_1, X_2, \ldots, X_n|Y) = \sum_{i=1}^{n} S_q(X_i|X_{i-1}, \ldots, X_1, Y). \tag{10}$$

The chain rules $S_q(XY) = S_q(X) + S_q(Y|X)$ and $S_q(XY|Z) = S_q(X|Z) + S_q(Y|XZ)$ emerge as particular cases and allow the Tsallis mutual informations of Eqs. (5) and (6) to be written as

$$I_q(X:Y) = S_q(X) + S_q(Y) - S_q(XY), \tag{11}$$

$$I_q(X:Y|Z) = S_q(XZ) + S_q(YZ) - S_q(Z) - S_q(XYZ). \tag{12}$$

Using the chain rule, the monotonicity and strong subadditivity relations [Eqs. (8) and (9)] are equivalent to the non-negativity of the unconditional and conditional Tsallis mutual information. For $q < 1$, strong subadditivity does not hold in general [31], hence we often restrict to the case $q \geqslant 1$ in what follows.

IV. CAUSAL CONSTRAINTS AND TSALLIS ENTROPY VECTORS

In Sec. III, we discussed some of the general properties of Tsallis entropy that hold irrespective of the underlying causal structure over the variables. The causal structure imposes the causal Markov constraints on the joint probability distribution over the variables involved (Sec. II B) and we wish to translate these probabilistic constraints into Tsallis entropic ones in order to use Tsallis entropies in the entropy vector method for analyzing causal structures.

A first observation is that Tsallis entropy vectors do not in general satisfy the causal constraints [Eq. (2)] satisfied by their Shannon counterparts. For a concrete counterexample, consider the simple, three variable causal structure where Z is a common cause of X and Y, and where there are no other causal relations. In terms of Shannon entropies, the only causal constraint in this case is $I(X:Y|Z) = 0$. Taking X, Y, and Z to be binary variables with possible values 0 and 1, the distribution $p_{xyz} = 1/4\ \forall x \in X, y \in Y$ if $z = 0$, and $p_{xyz} = 0$ otherwise, satisfies $p_{xy|z} = p_{x|z} p_{y|z}\ \forall x \in X, y \in Y$ and $z \in Z$ but has a $q = 2$ Tsallis conditional mutual information of $I_2(X:Y|Z) = \frac{1}{4}$. Hence when using Tsallis entropies (and conditional Tsallis entropy as defined in Sec. III) the causal constraint cannot be simply encoded by $I_q(X:Y|Z) = 0$ for $q > 1$.

Given this observation, it is natural to ask whether there are constraints for Tsallis entropies implied by the causal

Markov condition [Eq. (1)]. We answer this question with the following theorems.

Theorem 1. If a joint probability distribution p_{XY} over random variables X and Y with alphabet sizes d_X and d_Y is separable (i.e., $p_{XY} = p_X p_Y$), then for all $q \in [0,\infty)$ the Tsallis mutual information $I_q(X:Y)$ is upper bounded by

$$I_q(X:Y) \leqslant f(q, d_X, d_Y),$$

where the function $f(q, d_X, d_Y)$ is given by

$$\begin{aligned} f(q, d_X, d_Y) &= \frac{1}{(q-1)}\left(1 - \frac{1}{d_X^{q-1}}\right)\left(1 - \frac{1}{d_Y^{q-1}}\right) \\ &= (q-1)\ln_q d_X \ln_q d_Y. \end{aligned}$$

For $q \in (0,\infty)\setminus\{1\}$, the bound is saturated if and only if p_{XY} is the uniform distribution over X and Y.

Proof. The proof follows from the pseudoadditivity of Tsallis entropies (property 1) and the upper bound (property 2). Using these, for all $q \geqslant 0$ and for all separable distributions $p_{XY} = p_X p_Y$, we have

$$\begin{aligned} I_q(X:Y) &= S_q(X) + S_q(Y) - S_q(XY) = (q-1)S_q(X)S_q(Y) \\ &\leqslant \frac{\left(1 - \frac{1}{d_X^{q-1}}\right)\left(1 - \frac{1}{d_Y^{q-1}}\right)}{q-1} = f(q, d_X, d_Y). \end{aligned} \tag{13}$$

Whenever $q \in (0,\infty)\setminus\{1\}$, the bound is saturated if and only if p_{XY} is uniform over X and Y since, for these values of q, $S_q(X)$ and $S_q(Y)$ both attain their maximum values if and only if this is the case. ∎

Theorem 2. If a joint probability distribution p_{XYZ} satisfies the conditional independence $p_{XY|Z} = p_{X|Z}p_{Y|Z}$, then for all $q \geqslant 1$ the Tsallis conditional mutual information $I_q(X:Y|Z)$ is upper bounded by

$$I_q(X:Y|Z) \leqslant f(q, d_X, d_Y).$$

For $q > 1$, the bound is saturated only by distributions in which for some fixed value k the joint probabilities are given by $p_{xyz} = \begin{cases} \frac{1}{d_X d_Y} & \text{if } z=k \\ 0 & \text{otherwise} \end{cases}$ for all x, y, and z.[11]

Proof. Writing out $I_q(X:Y|Z)$ in terms of probabilities we have

$$\begin{aligned} I_q(X:Y|Z) &= \frac{1}{q-1}\left[\sum_{xyz} p^q(xyz) + \sum_z p^q(z) - \sum_{xz} p^q(xz) - \sum_{yz} p^q(yz)\right] \\ &= \frac{1}{q-1}\sum_z p^q(z)\left[\sum_{xy} p^q(xy|z) + 1 - \sum_x p^q(x|z) - \sum_y p^q(y|z)\right] \\ &= \sum_z p^q(z) I_q(X:Y)_{p_{XY|Z=z}}. \end{aligned}$$

[11]These distributions have deterministic Z and there is one such distribution for each value that Z can take.

Using this and Theorem 1, we can bound $I_q(X:Y|Z)$ as

$$\begin{aligned} &\max_{p_{XYZ} = p_Z p_{X|Z} p_{Y|Z}} I_q(X:Y|Z) \\ &= \max_{p_{XYZ} = p_Z p_{X|Z} p_{Y|Z}} \sum_z p_z^q I_q(X:Y)_{p_{XY|Z=z}} \\ &\leqslant \max_{p_Z} \sum_z p_z^q \max_{p_{X|Z}p_{Y|Z}} I_q(X:Y)_{p_{XY|Z=z}} \\ &= \max_{p_Z} \sum_z p_z^q f(q, d_X, d_Y) = f(q, d_X, d_Y). \end{aligned}$$

The last step holds because, for all $q > 1$, $\sum_z p_z^q$ is maximized by deterministic distributions over Z with a maximum value of 1, i.e., only distributions p_{XYZ} that are deterministic over Z saturate the upper bound of $f(q, d_X, d_Y)$. This completes the proof. ∎

Two corollaries of Theorem 2 naturally follow.

Corollary 1. Let X, Y, and Z be random variables with fixed alphabet sizes. Then for all $q \geqslant 1$ we have

$$\max_{\substack{p_{XYZ} \\ p_{XY|Z} = p_{X|Z}p_{Y|Z}}} I_q(X:Y|Z) = \max_{\substack{p_{XY} \\ p_{XY} = p_X p_Y}} I_q(X:Y).$$

Furthermore, for $q > 1$, the maximum on the left-hand side is achieved only by distributions in which for some fixed value k the joint probabilities are given by $p_{xyz} = \begin{cases} \frac{1}{d_X d_Y} & \text{if } z=k \\ 0 & \text{otherwise} \end{cases}$, while the maximum on the right-hand side occurs if and only if P_{XY} is the uniform distribution.

The significance of these relations for causal structures is then given by the following corollary.

Corollary 2. Let $p_{X_1 \ldots X_n}$ be a distribution compatible with the classical causal structure $\mathcal{G}^{\mathrm{C}}$ and let X, Y, and Z be disjoint subsets of $\{X_1, \ldots, X_n\}$ such that X and Y are d separated given Z. Then for all $q \geqslant 1$ we have

$$I_q(X:Y|Z) \leqslant f(q, d_X, d_Y),$$

where d_X is the product of d_{X_i} for all $X_i \in X$, and likewise for d_Y.

Remark 1. The results of this section can be generalized to the quantum case under certain assumptions i.e., as constraints on quantum Tsallis entropies implied by certain quantum causal structures (see the Appendix for details). Note that only constraints on the classical Tsallis entropy vectors derived in this section are required to detect the classical-quantum gap. Hence, the Appendix is not pertinent to the main results of this paper but can be seen as additional results regarding the properties of quantum Tsallis entropies.

A. Number of independent Tsallis entropic causal constraints

We saw previously that in the Shannon case ($q = 1$) the n conditions of the form $I(X_i : X_i^{\uparrow}|X_i^{\downarrow_1}) = 0$ ($i = 1, \ldots, n$) imply all the independence relations that follow from the causal structure. In the Tsallis case, however, the n conditions of the form $I_q(X_i : X_i^{\uparrow}|X_i^{\downarrow_1}) \leqslant f(q, d_{X_i}, d_{X_i^{\uparrow}})$ do not do the same. In the bipartite Bell and triangle causal structures we find that there is no redundancy amongst the 53 and 126 distinct Tsallis entropic inequalities that are implied by the

d separation relations in the corresponding DAGs in the case where the dimension (cardinality) of each individual node is taken to be d. In more detail, we used linear programming to show that each implication of d separation yields a nontrivial entropic causal constraint for all $q > 1$ and $d > 2$ for the bipartite Bell and triangle causal structures. By comparison, in these causal structures five and six independent Shannon entropic constraints imply all others. As an illustration of the difference, in the Shannon case $I(A : BC) = 0$ implies $I(A : B) = I(A : C) = 0$, whereas the analogous implication does not hold in the Tsallis case in general: although $I_q(A : BC) \leqslant f(q, d_A, d_{BC})$ implies $I_q(A : B) \leqslant f(q, d_A, d_{BC})$, it is not the case that $I_q(A : BC) \leqslant f(q, d_A, d_{BC})$ implies $I_q(A : B) \leqslant f(q, d_A, d_B)$.[12]

The number of distinct conditional independences (and hence the number of independent Tsallis constraints that follow from d separation) in a DAG depends on the specific graph; however, for any DAG $\mathcal{G}_n$ with n nodes, the number of such constraints can be upper bounded by that of the n-node DAG where all n nodes are independent, i.e., the n-node DAG with no edges. The number of conditions in this DAG can be thought of as the number of ways of partitioning n objects into four disjoint subsets[13] such that the first two are nonempty and where the ordering of the first two does not matter. Therefore, there are at most $\frac{1}{2}(4^n - 2 \times 3^n + 2^n)$ such conditions.

B. Using Tsallis entropies in the entropy vector method

We used the causal constraints of Corollary 2 in the entropy vector method with the aim of deriving quantum-violatable entropic inequalities for the triangle causal structure [Fig. 1(b)]. To do so, we started with the variables $A, B, C, X, Y,$ and Z of the triangle causal structure, the Shannon constraints, and the causal constraints satisfied by the Tsallis entropy vectors over these variables (Corollary 2) and used a Fourier-Motzkin (FM) elimination algorithm (from PORTA [36]) to eliminate the Tsallis entropy components involving the unobserved variables $A, B,$ and C and obtain the constraints on the observed nodes $X, Y,$ and Z.

The Tsallis entropy vector for the six nodes has $2^6 - 1 = 63$ components. The required marginal scenario with the observed nodes $X, Y,$ and Z has Tsallis entropy vectors with $2^3 - 1 = 7$ components and in this case the Fourier-Motzkin algorithm has to run 56 iterations, each of which eliminates one variable.

Starting with the full set of 126 Tsallis entropic causal constraints for the triangle causal structure as well as the 246 independent Shannon constraints, the Fourier-Motzkin elimination algorithm did not finish within several days on a standard desktop PC and the number of intermediate inequalities generated grew to about 90 000 after 11 steps. Because of this we instead tried starting with a subset composed of 15 of the 126 Tsallis entropic causal constraints,[14] i.e., 261 constraints on 63 dimensional vectors. We considered the case of $q = 2$ and where the six random variables are all binary. Again, in this case the algorithm did not finish after several days. We also tried starting with fewer causal constraints (for example, the six constraints analogous to the Shannon case) as well as using a modified code, optimized to deal with redundancies better but both of these attempts made no significant difference to this outcome.

Such a rapid increase of the number of inequalities in each step is a known problem with Fourier-Motzkin elimination where an elimination step over n inequalities can result in up to $n^2/4$ inequalities in the output and running d successive elimination steps can yield a double exponential complexity of $4(n/4)^{2^d}$ [28]. This rate of increase can be kept under control when the resulting set of inequalities has many redundancies. This happens in the Shannon case where the causal constraints are simple equalities and the system of 246 Shannon constraints plus six Shannon entropic causal constraints reduces to a system of just 91 independent inequalities before the FM elimination. In the Tsallis case, no reduction of the system of inequalities is possible in general due to the nature of the causal constraints. The fact that the Tsallis entropic causal constraints are inequality constraints rather than equalities also contributes to the computational difficulty since each independent equality constraint in effect reduces the dimension of the problem by 1.

We also tried the same procedure on the bipartite Bell causal structure [Fig. 1(a)], again for $q = 2$ and binary variables. Here, starting with the full set of 53 causal constraints again resulted in the program running for over one week without nearing the end, and a similar result was obtained when starting with only eight to ten causal constraints. While starting with fewer causal constraints such as the five conditional independence constraints (one for each node) resulted in a terminating program, no nontrivial entropic inequalities were obtained (i.e., we only obtained constraints corresponding to Shannon constraints or causal constraints that follow directly from d separation).[15]

[12]For an explicit counterexample, consider $p_{ABC} = \{\frac{3}{10}, 0.0, \frac{2}{10}, 0.0, \frac{1}{10}, \frac{1}{10}, \frac{2}{10}, \frac{1}{10}\}$ over binary A, B, and C for which $I_2(A : BC) = 9/25 < 3/8 = f(2, 2, 4)$ but $I_2(A : B) = 13/50 > 1/4 = f(2, 2, 2)$.

[13]The four subsets correspond to the three arguments of the conditional mutual information and a set of "leftovers."

[14]These included the six that follow from "each node N_i is conditionally independent of its descendants given its parents" (denoted as $N_i \perp N_i^{\dagger} | N_i^{\downarrow 1}$) and nine more chosen arbitrarily from the total of 126 independent Tsallis constraints we found for the triangle. The six former constraints for the triangle [Fig. 1(b)] are $A \perp CXB$, $B \perp CYA$, $C \perp BZA$, $X \perp YAZ|CB$, $Y \perp XBZ|AC$, and $Z \perp YCX|AB$. Examples of nine more constraints for which the procedure did not work are $X \perp Y|CB$, $X \perp A|CB$, $X \perp Z|CB$, $Y \perp X|AC$, $Y \perp B|AC$, $Y \perp Z|AC$, $Z \perp Y|AC$, $Z \perp C|AB$, and $Z \perp X|AB$. We also tried some other choices and number of constraints but this did not lead to any improvement.

[15]For example, we were able to obtain $I_2(A : BY) \leqslant \frac{7}{16}$ and $I_2(B : AX) \leqslant \frac{7}{16}$, while, in the case of binary variables and $q = 2$, the independences in the DAG together with Theorem 1 imply $I_2(A : BY) \leqslant \frac{6}{16}$ and $I_2(B : AX) \leqslant \frac{6}{16}$, which are the Tsallis entropic equivalents of the two nonsignaling constraints.

V. NEW TSALLIS ENTROPIC INEQUALITIES FOR THE TRIANGLE CAUSAL STRUCTURE

Despite the limitations encountered in applying the entropy vector method to Tsallis entropies (Sec. IV B), here we find Tsallis entropic inequalities for the triangle causal structure for all $q \geqslant 1$ by using known inequalities for the Shannon entropy [21] and the causal constraints derived in Sec. IV. Using the entropy vector method for Shannon entropies, the following three classes of entropic inequalities were obtained for the triangle causal structure [Fig. 1(b)] in [21].[16] Including all permutations of X, Y, and Z, these yield seven inequalities:

$$-H(X) - H(Y) - H(Z) + H(XY) + H(XZ) \geqslant 0, \tag{14a}$$

$$-5H(X) - 5H(Y) - 5H(Z) + 4H(XY) + 4H(XZ) + 4H(YZ) - 2H(XYZ) \geqslant 0, \tag{14b}$$

$$-3H(X) - 3H(Y) - 3H(Z) + 2H(XY) + 2H(XZ) + 3H(YZ) - H(XYZ) \geqslant 0. \tag{14c}$$

By replacing the Shannon entropy $H()$ with the Tsallis entropy $S_q()$ on the left-hand side of these inequalities and minimizing the resultant expression over our outer approximation to the classical Tsallis entropy cone for the triangle causal structure, one can obtain valid Tsallis entropic inequalities for this causal structure. More precisely, the outer approximation to the classical Tsallis entropy cone for the triangle is characterized by the $6 + 6(6-1)2^{6-3} = 246$ independent Shannon constraints (monotonicity and strong subadditivity constraints) and the 126 causal constraints (one for each conditional independence implied by the causal structure). To perform this minimization we used LPASSUMPTIONS [37], a linear program solver in MATHEMATICA that implements the simplex method allowing for unspecified variables. In our case, we assumed that the dimensions of all the unobserved nodes (A, B, and C) are equal to d_u and those of all the observed nodes (X, Y, and Z) are equal to d_o, and so the unspecified variables are $q \geqslant 1$, $d_u \geqslant 2$, and $d_o \geqslant 2$. We obtained the following Tsallis entropic inequalities for the triangle:

$$-S_q(X) - S_q(Y) - S_q(Z) + S_q(XY) + S_q(XZ) \geqslant B_1(q, d_o, d_u), \tag{15a}$$

$$\begin{aligned}&-5S_q(X) - 5S_q(Y) - 5S_q(Z) + 4S_q(XY) + 4S_q(XZ) + 4S_q(YZ) - 2S_q(XYZ)\\ &\quad \geqslant B_2(q, d_o, d_u) := \max(B_{21}(q, d_o, d_u), B_{22}(q, d_o, d_u)),\end{aligned} \tag{15b}$$

$$-3S_q(X) - 3S_q(Y) - 3S_q(Z) + 2S_q(XY) + 2S_q(XZ) + 3S_q(YZ) - S_q(XYZ) \geqslant B_3(q, d_o, d_u), \tag{15c}$$

where

$$B_1(q, d_o, d_u) = -\frac{1}{q-1}\big(1 - d_o^{1-q}\big)\big(2 - d_o^{1-q} - d_u^{1-q}\big), \tag{16a}$$

$$B_{21}(q, d_o, d_u) = -\frac{1}{q-1}\big(11 + d_u^{3-3q} + 6d_o^{2-2q} + 3d_o^{1-q}d_u^{1-q} - 6d_u^{1-q} - 15d_o^{1-q}\big), \tag{16b}$$

$$B_{22}(q, d_o, d_u) = -\frac{1}{q-1}\big(10 + d_o^{1-q}d_u^{3-3q} + 5d_o^{2-2q} + 2d_o^{1-q}d_u^{1-q} - 5d_u^{1-q} - 13d_o^{1-q}\big), \tag{16c}$$

$$B_3(q, d_o, d_u) = -\frac{1}{q-1}\big(6 + d_o^{1-q}d_u^{2-2q} + 3d_o^{2-2q} + d_o^{1-q}d_u^{1-q} - 3d_u^{1-q} - 8d_o^{1-q}\big). \tag{16d}$$

Note that $\lim_{q\to 1} B_1 = \lim_{q\to 1} B_2 = \lim_{q\to 1} B_3 = 0$ $\forall d_u, d_o \geqslant 2$, recovering the original inequalities for Shannon entropies [Eqs. (14a)–(14c)] as a special case.

In [38], an upper bound on the dimensions of classical unobserved systems needed to reproduce a set of observed correlations is derived in terms of the dimensions of the observed systems. In the case of the triangle causal structure with $d_X = d_Y = d_Z = d_o$ and $d_A = d_B = d_C = d_u$ as considered here, the result of [38] implies that all classical correlations P_{XYZ} can be reproduced by using hidden systems of dimension at most $d_o^3 - d_o$. Since the dimension of the unobserved systems is unknown, it makes sense to take the minimum of the derived bounds over all d_u between 2 and $d_o^3 - d_o$. By taking their derivative, one can verify that for $q > 1$ each of the functions B_1, B_{21}, B_{22}, and B_3 is monotonically decreasing in d_o and d_u, and hence that the minimum is obtained for $d_u = d_o^3 - d_o$ for any given $d_o \geqslant 2$. It follows that for all $q > 1$ and $d_o \geqslant 2$ relations of the same form as Eqs. (15a)–(15c) hold, with the quantities on the right-hand

[16] Note that a tighter entropic characterization was found in [20] based on non-Shannon inequalities, and that the techniques introduced here could also be applied to these.

sides replaced by

$$B_1^*(q,d_o) = B_1\big(q,d_o,d_o^3-d_o\big)$$
$$= -\frac{1}{q-1}\big[2+d_o^{2-2q}-3d_o^{1-q}+d_o\big(-d_o+d_o^3\big)^{-q}-d_o^3\big(-d_o+d_o^3\big)^{-q}-d_o^{2-q}\big(-d_o+d_o^3\big)^{-q}+d_o^{4-q}\big(-d_o+d_o^3\big)^{-q}\big], \tag{17a}$$

$$B_{21}^*(q,d_o) = B_{21}\big(q,d_o,d_o^3-d_o\big)$$
$$= -\frac{1}{q-1}\big[11+6d_o^{2-2q}-15d_o^{1-q}+\big(-d_o+d_o^3\big)^{3-3q}-6\big(-d_o+d_o^3\big)^{1-q}+3d_o^{1-q}\big(-d_o+d_o^3\big)^{1-q}\big], \tag{17b}$$

$$B_{22}^*(q,d_o) = B_{22}\big(q,d_o,d_o^3-d_o\big)$$
$$= -\frac{1}{q-1}\big[10+5d_o^{2-2q}-13d_o^{1-q}+d_o^{1-q}\big(-d_o+d_o^3\big)^{3-3q}-5\big(-d_o+d_o^3\big)^{1-q}+2d_o^{1-q}\big(-d_o+d_o^3\big)^{1-q}\big], \tag{17c}$$

$$B_3^*(q,d_o) = B_3\big(q,d_o,d_o^3-d_o\big)$$
$$= -\frac{1}{q-1}\big[6+3d_o^{2-2q}-8d_o^{1-q}+d_o^{1-q}\big(-d_o+d_o^3\big)^{2-2q}-3\big(-d_o+d_o^3\big)^{1-q}+d_o^{1-q}\big(-d_o+d_o^3\big)^{1-q}\big]. \tag{17d}$$

A quantum violation of any of these bounds would imply that no unobserved classical systems of arbitrary dimension could reproduce those quantum correlations.

Remark 2. Because they are monotonically decreasing, the bounds for $d_u = d_o^3 - d_o$ are not as tight as the d_u-dependent bounds for general $q > 1$. Nevertheless, as $q \to 1$, all the bounds $B^*(q, d_o)$ tend to zero, reproducing the known result of [15] for the Shannon case.

Remark 3. In some cases it may be interesting to show quantum violations of these inequalities for low values of d_u, hence ruling out classical explanations with hidden systems of low dimensions, while possibly leaving open the case of arbitrary classical explanations. This would be interesting if it could be established that using hidden quantum systems allows for much lower dimensions than for hidden classical systems, for example.

Looking for quantum violations

It is known that the triangle causal structure [Fig. 1(b)] admits nonclassical correlations such as Fritz's distribution [29]. The idea behind this distribution is to embed the Clauser-Horne-Shimony-Holt (CHSH) game in the triangle causal structure such that nonlocality for the triangle follows from the nonlocality of the CHSH game. To do so, C is replaced by the sharing of a maximally entangled pair of qubits, and A and B are taken to be uniformly random classical bits. The observed variables X, Y, and Z in Fig. 1(b) are taken to be pairs of the form $X := (\tilde{X}, B)$, $Y := (\tilde{Y}, A)$, and $Z := (A, B)$, where $\tilde{X}$ and $\tilde{Y}$ are generated by measurements on the halves of the entangled pair with B and A used to choose the settings such that the joint distribution $P_{\tilde{X}\tilde{Y}|BA}$ maximally violates a CHSH inequality. By a similar postprocessing of other nonlocal distributions in the bipartite Bell causal structure [Fig. 1(a)] such as the Mermin-Peres magic square game [39,40] and chained Bell inequalities [19], one can obtain other nonlocal distributions in the triangle that cannot be reproduced using classical systems. We explore whether any of these violate any of our inequalities.

Since the values of $B_i(q, d_o, d_u)$ are monotonically decreasing in d_o and d_u, if a distribution realizable in a quantum causal structure does not violate the bounds (15a)–(15c) for all $q \geqslant 1$ and some fixed values of d_o and d_u, then no violations are possible for $d_o' > d_o$, $d_u' > d_u$. We therefore take the smallest possible values of d_o and d_u when showing that a particular distribution cannot violate any of the bounds.

For Fritz's distribution [29], C is a two-qubit maximally entangled state, A and B are binary random variables, while X, Y, and Z are random variables of dimension 4, i.e., the actual observed dimensions are $(d_X, d_Y, d_Z) = (4, 4, 4)$ in this case. Here we see that taking $d_o = 4$ and the smallest possible d_u, which is $d_u = 2$, the left-hand sides of Inequalities (15a)–(15c) evaluated for Fritz's distribution do not violate the corresponding bounds $B_i(q, d_o = 4, d_u = 2)$ for any $q \geqslant 1$. This means that it is not possible to detect any quantum advantage of this distribution (even over the case where the unobserved systems are classical bits) using this method, and automatically implies that it cannot violate the bounds $B_i(q, d_o = 4, d_u)$ for $d_u \geqslant 2$.

We also considered the chained Bell and magic square correlations embedded in the triangle causal structure analogously to the case discussed above. For each of these, we define d^i to be the smallest value of d_o for which the bound $B_i(q, d_o = d^i, d_u = 2)$ cannot be violated for any $q > 1$. The values of d^i are given in Table I for the different cases of the chained Bell correlations and the magic square. Since the values of d^i are always lower than the smallest of the observed dimensions in the problem, and due to the monotonicity of the bounds, it follows that none of these quantum distributions violate any of our inequalities when the observed dimension is set to $d_o^{\min}$.

We further checked for violations of Inequalities (15a)–(15c) by sampling random quantum states for the systems A, B, and C and random quantum measurements the outcomes of which would correspond to the classical variables X, Y, and Z. The value of q was also sampled randomly between 1 and 100. We considered the cases where the shared systems were pairs of qubits with four outcome

TABLE I. Values of d^i for the chained Bell and magic square correlations embedded in the triangle causal structure. The values of N correspond to the number of inputs per party in the chained Bell inequality, which always has two outputs per party (the $N = 2$ case corresponds to Fritz's distribution [29]). When embedded in the triangle, the number of outcomes of the observed nodes is $(d_X, d_Y, d_Z) = (2N, 2N, N^2)$. The last column of the table gives the minimum of the observed node dimensions (d_X, d_Y, d_Z) for each N, which is simply $2N$. For the magic square, the dimensions (d_X, d_Y, d_Z) are (12,12,9). In all cases, the minimum value of d^i such that the Inequalities (15a)–(15c) with bounds $B_i(q, d_o = d^i, d_u = 2)$ are not violated for any $q \geqslant 1$ is less than the minimum observed dimension $d_o^{\min}$, and hence no violations of (15a)–(15c) could be found for the relevant case with $d_o = d_o^{\min}$.

Scenario	d^i			Smallest
	Ineq. (15a) ($i = 1$)	Ineq. (15b) ($i = 2$)	Ineq. (15c) ($i = 3$)	observed dim. $\left(d_o^{\min}\right)$
$N = 2$	2	2	2	4
$N = 3$	3	2	3	6
$N = 4$	4	2	4	8
$N = 5$	5	2	5	10
$N = 6$	6	2	6	12
$N = 7$	7	2	7	14
$N = 8$	8	2	8	16
$N = 9$	9	3	9	18
$N = 10$	10	3	10	20
Magic Sq.	4	2	4	9

measurements ($d_X = d_Y = d_Z = 4$) and qutrits with nine outcome measurements ($d_X = d_Y = d_Z = 9$) but were unable to find violations of any of the inequalities even for the bounds with the $d_o = 4, d_u = 2$ (the two-qubit case) and $d_o = 9, d_u = 2$ (the two-qutrit case), i.e., the bounds obtained when the unobserved systems are classical bits.

Remark 4. In the derivation of Inequalities (15a)–(15c), we set the dimensions of the observed nodes X, Y, and Z to all be equal and those of the unobserved nodes A, B, and C to also all be equal. One could in principle repeat the same procedure taking different dimensions for all six variables but we found the computational procedure too demanding. However, Table I shows that, even when we consider the bounds $B_i(q, d_o, d_u)$ with d_o and d_u much smaller than the actual dimensions, known nonlocal distributions in the triangle considered in Table I do not violate the corresponding Inequalities (15a)–(15c) for any $q \geqslant 1$. Since the bounds are monotonically decreasing in d_u and d_o, even if we obtained the general bounds for arbitrary dimensions of X, Y, Z, A, B, and C, they would be strictly weaker than $B_i(q, d^i, d_u = 2)$ $\forall i \in \{1, 2, 3\}, q \geqslant 1$ and can certainly not be violated by these distributions.

VI. DISCUSSION

We have investigated the use of Tsallis entropies within the entropy vector method to causal structures, showing how causal constraints imply bounds on the Tsallis entropies of the variables involved. Although Tsallis entropies for $q \geqslant 1$ possess many properties that aid their use in the entropy vector method, the nature of the causal constraints makes the problem significantly more computationally challenging than in the case of Shannon entropy. This meant that we were unable to complete the desired computations in the former case, even for some of the simplest causal structures. Nevertheless, we were able to derive classical causal constraints expressed in terms of Tsallis entropy by analogy with known Shannon constraints, but were unable to find cases where these were violated, even using quantum distributions that are known not to be classically realizable. This mirrors an analogous result for Shannon entropies [20].

Tsallis entropies are known to give improvements [23] in cases that involve postselection. While postselection cannot be used for general causal structures (including the triangle), it would be interesting to understand whether using Tsallis entropy helps in other cases for which postselection is applicable.

One could also investigate whether other entropic quantities could be used in a similar way. The Rényi entropies of order α do not satisfy strong subadditivity for $\alpha \neq 0, 1$, while the Rényi as well as the minimum and maximum entropies fail to obey the chain rules for conditional entropies. Thus, use of these in the entropy vector method would require an entropy vector with components for all possible conditional entropies as well as unconditional ones, considerably increasing the dimensionality of the problem, which we would expect to make the computations harder.[17]

Further, one could consider using algorithms other than Fourier-Motzkin elimination to obtain nontrivial Tsallis entropic constraints over observed nodes starting from the Tsallis cone over all the nodes (see, e.g., [42]). These could in principle yield solutions even in cases where FM elimination becomes intractable. However, we found that the FM elimination procedure became intractable even when starting out with only a small subset of the Tsallis entropic causal constraints for a simple causal structure such as the Bell one. This suggests that the difficulty is not only with the number of constraints, but also with their nature (in particular, that they are not equalities and depend nontrivially on the dimensions). Consequently, we bypassed FM elimination and used an alternative technique to obtain Tsallis entropic inequalities for the triangle causal structure (Sec. V).

[17] In some cases, not having a chain rule may not be prohibitive [41].

It is also worth noting that the following alternative definition of the Tsallis conditional entropy was proposed in [32]:

$$\tilde{S}_q(X|Y) = \frac{1}{1-q}\frac{\sum_y p_y^q S_q(X|Y=y)}{\sum_y p_y^q} = \frac{1}{1-q}\left(\frac{\sum_{x,y} p_{xy}^q}{\sum_y p_y^q} - 1\right). \quad (18)$$

Using this definition, Tsallis entropies would satisfy the same causal constraints as the Shannon entropy [Eq. (2)]. However, the conditional entropies defined this way do not satisfy the chain rules of Eq. (10) but instead obey a nonlinear chain rule, $S_q(XY) = S_q(X) + S_q(Y|X) + (1-q)S_q(X)S_q(Y|X)$ [32]. This would again mean that conditional entropies would need to be included in the entropy vector. Furthermore, since Fourier-Motzkin elimination only works for linear constraints, an alternative algorithm would be required to use this chain rule in conjunction with the entropy vector method.

That the inequalities for Tsallis entropy derived in this paper depend on the dimensions of the systems involved could be used to certify that particular observed correlations in a classical causal structure require a certain minimal dimension of unobserved systems to be realizable. To show this would require showing that classically realizable correlations violate one of the inequalities for some d_u. Such bounds would then complement the upper bounds of [38]. However, in some cases we know our bounds are not tight enough to do this. As a simple example, within the triangle causal structure we tried taking $X = (X_B, X_C)$, $Y = (Y_A, Y_C)$, and $Z = (Z_A, Z_B)$ with $X_B = Z_B, X_C = Y_C$, and $Y_A = Z_A$ where each is uniformly distributed with cardinality D, for $D \in \{3, \ldots, 10\}$. In this case it is clear that the correlations cannot be achieved with classical unobserved systems with $d_u = 2$. Taking the bound with $d_u = 2$ and $d_o = D^2$ no violations of Inequalities (15a)–(15c) were seen by plotting the graphs for $q \in [1, 20]$, for the range of D above. Hence, our bounds are too loose to certify lower bounds on d_u in this case.

While our analysis highlights significant drawbacks of using Tsallis entropies for analyzing causal structures, it does not rule out the possibility of Tsallis entropies being able to detect the classical-quantum gap[18] in these causal structures, or others. To overcome the difficulties we encountered we would need either increased computational power or the use of new, alternative techniques for analyzing causal structures (with or without entropies).

ACKNOWLEDGMENTS

We thank Mirjam Weilenmann and Elie Wolfe for useful discussions and also thank Mirjam Weilenmann for sharing some MATHEMATICA code. V.V. acknowledges financial support from the Department of Mathematics, University of York. R.C. is supported by the Engineering and Physical Sciences Research Council (EPSRC) Quantum Communications Hub (Grant No. EP/M013472/1) and by an EPSRC First Grant (Grant No. EP/P016588/1).

APPENDIX: QUANTUM GENERALIZATIONS OF THEOREMS 1 AND 2

In the following, for a (finite-dimensional) Hilbert space $\mathcal{H}$, we use $\mathcal{L}(\mathcal{H})$ to represent the set of linear operators on $\mathcal{H}$, $\mathcal{P}(\mathcal{H})$ to represent the set of positive (semidefinite) operators on $\mathcal{H}$, and $\mathcal{S}(\mathcal{H})$ to denote the set of density operators on $\mathcal{H}$ (positive and trace 1).

Tsallis entropies as defined for classical random variables in Sec. III are easily generalized to the quantum case by replacing the probability distribution by a density matrix [43]. For a quantum system described by the density matrix $\rho \in \mathcal{S}(\mathcal{H})$ on the Hilbert space $\mathcal{H}$ and $q > 0$, the quantum Tsallis entropy is defined by

$$S_q(\rho) = \begin{cases} -\operatorname{Tr} \rho^q \ln_q \rho, & q \neq 1 \\ H(\rho), & q = 1 \end{cases} \quad (A1)$$

where $H(\rho) = -\operatorname{Tr} \rho \ln \rho$ is the von-Neumann entropy of ρ and $\ln_q(x) = \frac{x^{1-q}-1}{1-q}$ as in Sec. III.[19]

Given a density operator $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$, the conditional quantum Tsallis entropy of A given B can then be defined by $S_q(A|B)_\rho = S_q(AB) - S_q(B)$, the mutual information between A and B can be defined by $I_q(A:B)_\rho = S_q(A) + S_q(B) - S_q(AB)$, and for $\rho_{ABC} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ the conditional Tsallis information between A and B given C is defined by $I_q(A:B|C)_\rho = S_q(A|C) + S_q(B|C) - S_q(AB|C)$. In this section we use d_S to represent the dimensions of the Hilbert space $\mathcal{H}_S$.

The following properties of quantum Tsallis entropies will be useful for what follows.

(1) Pseudoadditivity [30]: If $\rho_{AB} = \rho_A \otimes \rho_B$, then

$$S_q(AB) = S_q(A) + S_q(B) + (1-q)S_q(A)S_q(B). \quad (A2)$$

(2) Upper bound [44]: For all $q > 0$, we have $S_q(A) \leqslant \ln_q d_A$ and equality is achieved if and only if $\rho_A = \mathbb{1}_A/d_A$.

(3) Subadditivity [44]: For any density matrix ρ_{AB} with marginals ρ_A and ρ_B, the following holds for all $q \geqslant 1$:

$$S_q(AB) \leqslant S_q(A) + S_q(B). \quad (A3)$$

Using these we can generalize Theorem 1 to the quantum case. This corresponds to the causal structure with two independent quantum nodes and no edges in between them.

Theorem 3. For all separable bipartite density operators, i.e., $\rho_{AB} = \rho_A \otimes \rho_B$ with $\rho_A \in \mathcal{S}(\mathcal{H}_A)$ and $\rho_B \in \mathcal{S}(\mathcal{H}_B)$, the quantum Tsallis mutual information $I_q(A:B)_\rho$ is upper bounded as follows for all $q > 0$:

$$I_q(A:B)_\rho \leqslant f(q, d_A, d_B),$$

[18] Proving that Tsallis entropies are unable to do this would also be difficult. For instance, the proof of [18] that Shannon entropies are unable to detect the gap in linelike causal structures involves first characterizing the marginal polytope through Fourier-Motzkin elimination, which itself proved to be computationally infeasible with Tsallis entropies even for the simplest linelike causal structure, the bipartite Bell scenario.

[19] Analogously to the classical case we keep it implicit that if ρ has any zero eigenvalues these do not contribute to the trace.

where the function $f(q, d_A, d_B)$ is given by

$$f(q, d_A, d_B) = \frac{1}{(q-1)}\left(1 - \frac{1}{d_A^{q-1}}\right)\left(1 - \frac{1}{d_B^{q-1}}\right) = (q-1)\ln_q d_A \ln_q d_B.$$

The bound is saturated if and only if $\rho_{AB} = \frac{\mathbb{1}_A}{d_A} \otimes \frac{\mathbb{1}_B}{d_B}$.

Proof. The proof goes through in the same way as the proof of Theorem 1 for the classical case (properties 1 and 2 are analogous to those needed in the classical proof). ■

Next, we generalize Theorem 2 and Corollaries 1 and 2. This would correspond to the causal constraints on quantum Tsallis entropies implied by the common cause causal structure with C being a complete common cause of A and B (which share no causal relations among themselves). Here, one must be careful in precisely defining the conditional mutual information and interpreting it physically. For example, if the common case C were quantum and the nodes A and B were classical outcomes of measurements on C, then A, B, and C do not coexist and there is no joint state ρ_{ABC} in such a case. This is a significant difference in quantum causal modeling compared to the classical case, and there have been several proposals for how do deal with it [45–48]. In the following we consider two cases.

(1) When C is classical, all three systems coexist and ρ_{ABC} can be described by a classical-quantum state (see Theorem 4).

(2) When C is quantum, one approach is to view ρ_{ABC} not as the joint state of the three systems but as being related to the Choi-Jamiolkowski representations of the quantum channels from C to A and B (see Sec. VI) as done in [47].

The following lemma proven in [49] is required for our generalization of Theorem 2 in the first case.

Lemma 1 [49, Lemma 1]. Let $\mathcal{H}_A$ and $\mathcal{H}_Z$ be two Hilbert spaces and let $\{|z\rangle\}_z$ be an orthonormal basis of $\mathcal{H}_Z$. Let ρ_{AZ} be classical on $\mathcal{H}_Z$ with respect to this basis, i.e.,

$$\rho_{AZ} = \sum_z p_z \rho_A^{(z)} \otimes |z\rangle\langle z|,$$

where $\sum_z p_z = 1$ and $\rho_A^{(z)} \in S(\mathcal{H}_A)\ \forall z$. Then, for all $q > 0$,

$$S_q(AZ)_\rho = \sum_z p_z^q S_q\big(\rho_A^{(z)}\big) + S_q(Z),$$

where $S_q(Z)$ is the classical Tsallis entropy of the variable Z distributed according to P_Z.

Note that the above lemma immediately implies that

$$S_q(A|Z)_\rho = \sum_z p_z^q S_q\big(\rho_A^{(z)}\big). \tag{A4}$$

Theorem 4. Let $\rho_{ABC} = \sum_c p_c \rho_{AB}^{(c)} \otimes |c\rangle\langle c|$, where $\rho_{AB}^{(c)} = \rho_A^{(c)} \otimes \rho_B^{(c)}\ \forall c$; then, for all $q \geqslant 1$,

$$I_q(A:B|C)_{\rho_{ABC}} \leqslant f(q, d_A, d_B).$$

For $q > 1$ the bound is saturated if and only if $\rho_{ABC} = \frac{\mathbb{1}_A}{d_A} \otimes \frac{\mathbb{1}_B}{d_B} \otimes |c\rangle\langle c|_C$.

Proof. Using (A4) we have

$$\begin{aligned} I_q(A:B|C)_{\rho_{ABC}} &= S_q(A|C)_\rho + S_q(B|C)_\rho - S_q(AB|C)_\rho \\ &= \sum_c p_c^q\big[S_q\big(\rho_A^{(c)}\big) + S_q\big(\rho_B^{(c)}\big) - S_q\big(\rho_{AB}^{(c)}\big)\big] \\ &= \sum_c p_c^q I_q(A:B)_{\rho_{AB}^{(c)}}. \end{aligned}$$

The rest of the proof is analogous to Theorem 2, where, using the above, using Theorem 3, and defining the set $\mathcal{R} = \{\rho_{ABC} \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C : \rho_{ABC} = \sum_c p_c \rho_A^{(c)} \otimes \rho_B^{(c)} \otimes |c\rangle\langle c|\}$, we have

$$\begin{aligned} \max_{\mathcal{R}} I_q(A:B|C)_\rho &= \max_{\mathcal{R}} \sum_c p_c^q I_q(A:B)_{\rho_{AB}^{(c)}} \\ &\leqslant \max_{\{p_c\}_c} \sum_c p_c^q(c) \max_{\{\rho_A^{(c)}\}_c, \{\rho_B^{(c)}\}_c} I_q(A:B)_{\rho_{AB}^{(c)}} \\ &= f(q, d_A, d_B), \end{aligned}$$

where the last step follows because, for all $q \geqslant 1$, $\sum_c p_c^q$ is maximized by deterministic distributions over C with a maximum value of 1[20] and $I_q(A:B)_{\rho_{AB}^{(c)}}$ for product states is maximized by the maximally mixed state over A and B for all c (Theorem 3). Thus, for $q > 1$, the bound is saturated if and only if $\rho_{ABC} = \frac{\mathbb{1}_A}{d_A} \otimes \frac{\mathbb{1}_B}{d_B} \otimes |c\rangle\langle c|_C$ for some value c of C. ■

Generalization: When systems do not coexist

There is a fundamental problem with naively generalizing classical conditional independences such as $p_{XY|Z} = p_{X|Z}p_{Y|Z}$ to the quantum case by replacing joint distributions by density matrices: it is not clear what is meant by a conditional quantum state, e.g., $\rho_{A|C}$, since it is not clear what it means to condition on a quantum system, especially when the (joint state of the) system under consideration and the one being conditioned upon do not coexist. There are a number of approaches for tackling this problem, from describing quantum states in space and time on an equal footing [50] to quantum analogs of Bayesian inference [45] and causal modeling [46–48]. In the following, we will focus on one such approach that is motivated by the framework of [47]. Central to this approach is the Choi-Jamiołkowski isomorphism [51,52] from which one can define conditional quantum states.

Definition 3 – Choi state. Let $|\gamma\rangle = \sum_i |i\rangle_R |i\rangle_{R^*} \in \mathcal{H}_R \otimes \mathcal{H}_{R^*}$, where $\mathcal{H}_{R^*}$ is the dual space to $\mathcal{H}_R$ and where $\{|i\rangle_R\}_i$ and $\{|i\rangle_{R^*}\}_i$ are orthonormal bases of $\mathcal{H}_R$ and $\mathcal{H}_{R^*}$, respectively. Given a channel $\mathcal{E}_{R|S} : \mathcal{S}(\mathcal{H}_R) \to \mathcal{S}(\mathcal{H}_S)$, the *Choi state of the channel* is defined by

$$\rho_{S|R} = (\mathcal{E}_{R|S} \otimes \mathcal{I})(|\gamma\rangle\langle\gamma|) = \sum_{ij} \mathcal{E}(|i\rangle\langle j|_R) \otimes |i\rangle\langle j|_{R^*}.$$

Thus, $\rho_{S|R} \in \mathcal{P}(\mathcal{H}_S \otimes \mathcal{H}_{R^*})$.

Now, if a quantum system C evolves through a unitary channel $\mathcal{E}_I(\cdot) = U'(\cdot)U'^\dagger$ to two systems A' and B' where $U' : \mathcal{H}_C \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$, it is reasonable to call the system C a quantum common cause of the systems A' and B'. Further, this would still be reasonable if one were to then perform local completely positive trace preserving (CPTP) maps on the A'

[20]For $q > 1$ such deterministic distributions are the only way to obtain the bound.

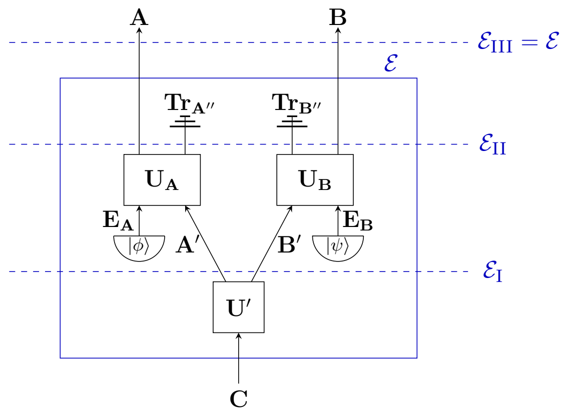

FIG. 2. A circuit decomposition of the channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_{\mathbf{C}}) \to \mathcal{S}(\mathcal{H}_{\mathbf{A}} \otimes \mathcal{H}_{\mathbf{B}})$ when **C** is a *complete common cause* of **A** and **B**: If the map $\mathcal{E}$ from the system C to the systems A and B can be decomposed as shown here, then C is a complete common cause of A and B ([47]). We build up our result step by step considering the channels given by $\mathcal{E}_{\rm i}$ (unitary), $\mathcal{E}_{\rm ii}$ (unitary followed by local isometries), and $\mathcal{E}_{\rm iii} = \mathcal{E}$.

and B' systems. By the Stinespring dilation theorem, these local CPTP maps can be seen as local isometries followed by partial traces, and the local isometries can be seen as the introduction of an ancilla in a pure state followed by a joint unitary on the system and ancilla. This is illustrated in Fig. 2 and is compatible with the definition of quantum common causes presented in [47]. In other words, a system C can be said to be a complete (quantum) common cause of systems A and B if the corresponding channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_C) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ can be decomposed as in Fig. 2 for some choice of unitaries U', U_A, and U_B and pure states $|\phi\rangle_{E_A}$ and $|\psi\rangle_{E_B}$. Note that a more general set of channels than we use here fits the definition of quantum common cause in [47]; whether the theorems here extend to this case we leave as an open question.

In [47] it is shown that whenever a system C is a complete common cause of systems A and B then the Shannon conditional mutual information evaluated on the state $\tau_{ABC^*} = \frac{1}{d_A}\rho_{AB|C}$ satisfies $I(A : B|C^*)_\tau = 0$ where $\rho_{AB|C}$ is the Choi state of the channel from C to A and B. We generalize this result to Tsallis entropies for $q \geqslant 1$ for certain types of channels. We present the result in three cases, each with increasing levels of generality. These are explained in Fig. 2 and correspond to the cases where the map from the complete common cause C to its children A and B is (i) unitary ($\mathcal{E}_{\rm i} = U'$), (ii) unitary followed by local isometries ($\mathcal{E}_{\rm ii}$), and (iii) unitary followed by local isometries followed by partial traces on local systems ($\mathcal{E}_{\rm iii} = \mathcal{E}$).

Lemma 2. Let $\mathcal{E}_{\rm i} : \mathcal{S}(\mathcal{H}_C) \to \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'})$ be a unitary quantum channel, i.e.,

$$\mathcal{E}_{\rm i}(\cdot) = U'(\cdot)U'^\dagger,$$

where $U' : \mathcal{H}_C \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ is an arbitrary unitary operator. If $\rho_{A'B'|C}$ is the corresponding Choi state, then the Tsallis conditional mutual information evaluated on the state $\tau_{A'B'C^*} = \frac{1}{d_C}\rho_{A'B'|C} \in \mathcal{S}(\mathcal{H}_{A'} \otimes \mathcal{H}_{B'} \otimes \mathcal{H}_{C^*})$ satisfies

$$I_q(A' : B'|C^*)_\tau = f(q, d_{A'}, d_{B'}) \quad \forall q > 0.$$

Proof. The conditional mutual information $I_q(A' : B'|C^*)_\tau$ can be written as

$$I_q(A' : B'|C^*)_\tau = \frac{1}{q-1}\big(\mathrm{Tr}_{A'B'C^*}\, \tau^q_{A'B'C^*} + \mathrm{Tr}_{C^*}\, \tau^q_{C^*} - \mathrm{Tr}_{A'C^*}\, \tau^q_{A'C^*} - \mathrm{Tr}_{B'C^*}\, \tau^q_{B'C^*}\big). \tag{A5}$$

We will now evaluate every term in the above expression for the case where the channel that maps the C system to the A' and B' systems is unitary. In this case, $\tau_{A'B'C^*}$ is a pure state and can be written as $\tau_{A'B'C^*} = |\tau\rangle\langle\tau|_{A'B'C^*}$ where

$$|\tau\rangle_{A'B'C^*} = \frac{1}{\sqrt{d_C}}\sum_i U'|i\rangle_C \otimes |i\rangle_{C^*}. \tag{A6}$$

This means that $\mathrm{Tr}_{A'B'C^*}\, \tau^q_{A'B'C^*} = \mathrm{Tr}_{A'B'C^*}\, \tau_{A'B'C^*} \quad \forall q > 0$. Since $\tau_{A'B'C^*}$ is a valid quantum state, it must be a trace-1 operator and we have

$$\mathrm{Tr}_{A'B'C^*}\, \tau^q_{A'B'C^*} = 1 \quad \forall q > 0. \tag{A7}$$

Further, we have $\tau_{C^*} = \mathrm{Tr}_{A'B'}\, \tau_{A'B'C^*} = \frac{\mathbb{1}_{C^*}}{d_C}$ and hence

$$\mathrm{Tr}_{C^*}\, \tau^q_{C^*} = \frac{1}{d_C^{q-1}} = \frac{1}{d_{A'}^{q-1} d_{B'}^{q-1}}. \tag{A8}$$

The second step follows from the fact that $U' : \mathcal{H}_C \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ is unitary so $d_C = d_{A'}d_{B'}$.

Now, the marginals over A' and B' are $\tau_{A'} = \mathrm{Tr}_{B'C^*}\, \tau_{A'B'C^*} = \frac{\mathbb{1}_{A'}}{d_{A'}}$ and $\tau_{B'} = \mathrm{Tr}_{A'C^*}\, \tau_{A'B'C^*} = \frac{\mathbb{1}_{B'}}{d_{B'}}$. By the Schmidt decomposition of $\tau_{A'B'C^*}$, the nonzero eigenvalues of $\tau_{A'}$ are the same as those of $\tau_{B'C^*}$. Since the Tsallis entropy depends only on the nonzero eigenvalues, $S_q(A') = S_q(B'C^*)$ and hence

$$\mathrm{Tr}_{B'C^*}\, \tau^q_{B'C^*} = d_{A'}\left(\frac{1}{d^q_{A'}}\right) = \frac{1}{d^{q-1}_{A'}}. \tag{A9}$$

By the same argument it follows that

$$\mathrm{Tr}_{A'C^*}\, \tau^q_{A'C^*} = d_{B'}\left(\frac{1}{d^q_{B'}}\right) = \frac{1}{d^{q-1}_{B'}}. \tag{A10}$$

Combining Eqs. (A5)–(A10), we have

$$\begin{aligned} I_q(A' : B'|C^*)_\tau &= \frac{1}{q-1}\left(1 + \frac{1}{d^{q-1}_{A'} d^{q-1}_{B'}} - \frac{1}{d^{q-1}_{A'}} - \frac{1}{d^{q-1}_{B'}}\right) \\ &= f(q, d_{A'}, d_{B'}) \quad \forall q > 0. \end{aligned} \tag{A11}$$

■

Lemma 3. Let $\mathcal{E}_{\rm ii} : \mathcal{S}(\mathcal{H}_C) \to \mathcal{S}(\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{\tilde{B}})$ be a quantum channel of the form

$$\begin{aligned} \mathcal{E}_{\rm ii}(\cdot) = (U_A \otimes U_B)[&|\phi\rangle\langle\phi|_{E_A} \otimes U'(\cdot)U'^\dagger \\ &\otimes |\psi\rangle\langle\psi|_{E_B}](U_A \otimes U_B)^\dagger, \end{aligned}$$

where $U' : \mathcal{H}_C \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$, $U_A : \mathcal{H}_{E_A} \otimes \mathcal{H}_{A'} \to \mathcal{H}_{\tilde{A}}$, and $U_B : \mathcal{H}_{B'} \otimes \mathcal{H}_{E_B} \to \mathcal{H}_{\tilde{B}}$ are arbitrary unitaries and $|\phi\rangle_{E_A}$ and $|\psi\rangle_{E_B}$ are arbitrary pure states. If $\rho_{\tilde{A}\tilde{B}|C}$ is the corresponding Choi state, then the Tsallis conditional mutual information evaluated on the state $\tau_{\tilde{A}\tilde{B}C^*} = \frac{1}{d_C}\rho_{\tilde{A}\tilde{B}|C} \in \mathcal{S}(\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_{\tilde{B}} \otimes \mathcal{H}_{C^*})$ satisfies

$$I_q(\tilde{A} : \tilde{B}|C^*)_\tau = f(q, d_{A'}, d_{B'}) \quad \forall q > 0.$$

Proof. Note that the map $\mathcal{E}_{\rm ii}$ is the unitary map $\mathcal{E}_{\rm i}(\cdot) = U'(\cdot)U'^\dagger$ followed by local isometries V_A and V_B on the A' and B' systems, respectively. Since the expression for the conditional mutual information $I_q(\tilde{A} : \tilde{B}|C^*)_\tau$ can be written in terms of entropies, which are functions of the eigenvalues of the relevant reduced density operators, and since the eigenvalues are unchanged by local isometries, this conditional mutual information is invariant under local isometries. The rest of the proof is identical to that of Lemma 2, resulting in

$$I_q(\tilde{A} : \tilde{B}|C^*)_\tau = I_q(A' : B'|C^*)_\tau = f(q, d_{A'}, d_{B'}) \quad \forall q > 0. \tag{A12}$$

■

For the last case where $\mathcal{E}_{\rm iii}(\cdot) = {\rm Tr}_{A''B''}[(U_A \otimes U_B)[|\phi\rangle\langle\phi|_{E_A} \otimes U'(\cdot)U'^\dagger \otimes |\psi\rangle\langle\psi|_{E_B}](U_A \otimes U_B)^\dagger]$, one could intuitively argue that tracing out systems could not increase the mutual information and one would expect that

$$I_q(AA'' : BB''|C^*)_\tau \geqslant I_q(A : B|C^*)_\tau. \tag{A13}$$

Since $I_q(AA'' : BB''|C^*)_\tau = I_q(A : B|C^*)_\tau + I_q(AA'' : B''|BC^*)_\tau + I_q(A'' : B|AC^*)_\tau$, Eq. (A13) would follow from strong subadditivity used twice, i.e., $I_q(AA'' : B''|BC^*)_\tau \geqslant 0$ and $I_q(A'' : B|AC^*)_\tau \geqslant 0$. However, it is known that strong subadditivity does not hold in general for Tsallis entropies for $q > 1$ [53]. Reference [53] also provides a sufficiency condition for strong subadditivity to hold for Tsallis entropies. In the following lemma, we provide another, simple sufficiency condition that also helps bound the Tsallis mutual information $I_q(AA'' : B|C)_\tau$ [or $I_q(A : BB''|C)_\tau$] corresponding to the map $\mathcal{E}_{\rm iii}$ where only one of A'' or B'' is traced out but not both.

Lemma 4 – sufficiency condition for strong subadditivity of Tsallis entropies. If ρ_{ABC} is a pure quantum state, then for all $q \geqslant 1$ we have $I_q(A : B|C)_\rho \geqslant 0$.

Proof. We have

$$I_q(A : B|C) = S_q(AC) + S_q(BC) - S_q(ABC) - S_q(C).$$

Since ρ_{ABC} is pure we have $S_q(ABC) = 0 \ \forall q > 0$ and (from the Schmidt decomposition argument mentioned earlier) $S_q(AC) = S_q(B)$, $S_q(BC) = S_q(A)$, and $S_q(C) = S_q(AB)$. Thus,

$$I_q(A : B|C) = S_q(A) + S_q(B) - S_q(AB) = I_q(A : B) \geqslant 0,$$

which follows from subadditivity of quantum Tsallis entropies for $q \geqslant 1$ [44]. In other words, for pure ρ_{ABC}, strong subadditivity of Tsallis entropies is equivalent to their subadditivity, which holds whenever $q \geqslant 1$. ■

Corollary 3. Let $\mathcal{E}^1_{\rm iii} : \mathcal{S}(\mathcal{H}_C) \to \mathcal{S}(\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_B)$ be a quantum channel of the form

$$\mathcal{E}^1_{\rm iii}(\cdot) = {\rm Tr}_{B''}\{(U_A \otimes U_B)[|\phi\rangle\langle\phi|_{E_A} \otimes U'(\cdot)U'^\dagger \otimes |\psi\rangle\langle\psi|_{E_B}] (U_A \otimes U_B)^\dagger\},$$

where $U' : \mathcal{H}_C \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$, $U_A : \mathcal{H}_{E_A} \otimes \mathcal{H}_{A'} \to \mathcal{H}_{\tilde{A}} \cong \mathcal{H}_A \otimes \mathcal{H}_{A''}$, and $U_B : \mathcal{H}_{B'} \otimes \mathcal{H}_{E_B} \to \mathcal{H}_{\tilde{B}} \cong \mathcal{H}_B \otimes \mathcal{H}_{B''}$ are arbitrary unitaries and $|\phi\rangle_{E_A}$ and $|\psi\rangle_{E_B}$ are arbitrary pure states. If $\rho_{\tilde{A}B|C}$ is the corresponding Choi state, then the Tsallis conditional mutual information evaluated on the state $\tau_{\tilde{A}BC^*} = \frac{1}{d_C}\rho_{\tilde{A}B|C} \in \mathcal{S}(\mathcal{H}_{\tilde{A}} \otimes \mathcal{H}_B \otimes \mathcal{H}_{C^*})$ satisfies

$$I_q(\tilde{A} : B|C^*) := I_q(AA'' : B|C^*)_\tau \leqslant f(q, d_{A'}, d_{B'}) \qquad \forall q \geqslant 1.$$

Proof. Since $I_q(AA'' : BB''|C^*)_\tau = I_q(AA'' : B|C^*)_\tau + I_q(AA'' : B''|BC^*)_\tau$, the purity of $\tau_{\tilde{A}\tilde{B}C^*} = \tau_{AA''BB''C^*}$ and Lemma 4 imply that

$$I_q(AA'' : BB''|C^*)_\tau \geqslant I_q(AA'' : B|C^*)_\tau, \quad \forall q \geqslant 1,$$

or (equivalently) in more concise notation

$$I_q(\tilde{A} : \tilde{B}|C^*)_\tau \geqslant I_q(\tilde{A} : B|C^*)_\tau \quad \forall q \geqslant 1.$$

Finally, using Lemma 3 we obtain the required result. ■

Now, for Eq. (A13) to hold, we do not necessarily need strong subadditivity. Even if $I_q(A'' : B|AC)_\tau \geqslant 0$ does not hold, Eq. (A13) would still hold if $I_q(AA'' : B''|BC)_\tau + I_q(A'' : B|AC)_\tau \geqslant 0$. This motivates the following conjecture.

Conjecture 1. Let $\mathcal{E}_{\rm iii} : \mathcal{S}(\mathcal{H}_C) \to \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ be a quantum channel of the form

$$\mathcal{E}_{\rm iii}(\cdot) = {\rm Tr}_{A''B''}\{(U_A \otimes U_B)[|\phi\rangle\langle\phi|_{E_A} \otimes U'(\cdot)U'^\dagger \otimes |\psi\rangle\langle\psi|_{E_B}] (U_A \otimes U_B)^\dagger\},$$

where $U' : \mathcal{H}_C \to \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$, $U_A : \mathcal{H}_{E_A} \otimes \mathcal{H}_{A'} \to \mathcal{H}_A \otimes \mathcal{H}_{A''}$, and $U_B : \mathcal{H}_{B'} \otimes \mathcal{H}_{E_B} \to \mathcal{H}_B \otimes \mathcal{H}_{B''}$ are arbitrary unitaries and $|\phi\rangle_{E_A}$ and $|\psi\rangle_{E_B}$ are arbitrary pure states. If $\rho_{AB|C}$ is the corresponding Choi state, then the Tsallis conditional mutual information evaluated on the state $\tau_{ABC^*} = \frac{1}{d_C}\rho_{AB|C} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_{C^*})$ satisfies

$$I_q(A : B|C^*)_\tau \leqslant f(q, d_{A'}, d_{B'}) \quad \forall q \geqslant 1.$$

Notice that in Corollary 3 and Conjecture 1 the bounds are functions of $d_{A'}$ and $d_{B'}$ and not of the dimensions of the systems A and B (those in the quantity on the left-hand side). In the case that $d_A \geqslant d_{A'}$ and $d_B \geqslant d_{B'}$, the fact that $f(q, d_A, d_B)$ is a strictly increasing function of d_A and d_B $\forall q \geqslant 0$ allows us to write $I_q(\tilde{A} : B|C^*)_\tau \leqslant f(q, d_{\tilde{A}}, d_B)$ and $I_q(A : B|C^*)_\tau \leqslant f(q, d_A, d_B)$ under the conditions of Corollary 3 and Conjecture 1, respectively. However, if $d_A \leqslant d_{A'}$ and/or $d_B \leqslant d_{B'}$, the bounds $f(q, d_{\tilde{A}}, d_B)$ and $f(q, d_A, d_B)$ are tighter than the bound $f(q, d_{A'}, d_{B'})$ and so not implied. However, based on the several examples that we have checked, we further conjecture the following.

Conjecture 2. Under the same conditions as Conjecture 1,

$$I_q(A : B|C^*)_\tau \leqslant f(q, d_A, d_B) \quad \forall q \geqslant 1.$$

Further, it is shown in [47] that if C is a *complete common cause* of A and B then the corresponding Choi state $\rho_{AB|C}$ decomposes as $\rho_{AB|C} = (\rho_{A|C} \otimes \mathbb{1}_B)(\mathbb{1}_A \otimes \rho_{B|C})$ or $\rho_{AB|C} = \rho_{A|C}\rho_{B|C}$ in analogy with the classical case where if a classical random variable Z is a common cause of the random variables X and Y then the joint distribution over these variables factorizes as $p_{XY|Z} = p_{X|Z}p_{Y|Z}$. Then we have that $\tau_{ABC^*} = \frac{1}{d_C}\rho_{AB|C} = \frac{1}{d_C}\rho_{A|C}\rho_{B|C}$. By further analogy with the classical results of Sec. IV, one may also consider instead a state of the form $\hat{\sigma}_{ABCC^*} = \sigma_C \otimes \frac{1}{d_C}\rho_{A|C}\rho_{B|C} = \sigma_C \otimes \tau_{ABC^*}$, where

$\sigma_C \in \mathcal{S}(\mathcal{H}_C)$.[21] Note that $\hat{\sigma}_{ABCC^*}$ is a valid density operator on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \otimes \mathcal{H}_{C^*}$.

Lemma 5. The state $\hat{\sigma}_{ABCC^*} = \sigma_C \otimes \tau_{ABC^*}$ defined above satisfies

$$I_q(A:B|CC^*)_{\hat{\sigma}} \leqslant f(q, d_A, d_B),$$

whenever $I_q(A:B|C^*)_\tau \leqslant f(q, d_A, d_B)$ holds for the state $\tau_{ABC^*} = \frac{1}{d_A}\rho_{AB|C}$, where $\rho_{AB|C}$ represents the quantum channel from C to A and B and σ_C is the input quantum state to this channel.

Proof. Since $\hat{\sigma}$ is a product state between the C and ABC^* subsystems, by the pseudoadditivity of quantum Tsallis entropies and the chain rule we have

$$\begin{aligned}
I_q(A:B|CC^*)_{\hat{\sigma}} &= S_q(ACC^*) + S_q(BCC^*) \\
&\quad - S_q(ABCC^*) - S_q(CC^*) \\
&= S_q(AC^*) + S_q(BC^*) - S_q(ABC^*) - S_q(C^*) \\
&\quad - (q-1)S_q(C)[S_q(AC^*) + S_q(BC^*) \\
&\quad - S_q(ABC^*) - S_q(C^*)] \\
&= [1-(1-q)S_q(C)]I(A:B|C^*) \\
&= \mathrm{Tr}(\sigma_C^q)I(A:B|C^*).
\end{aligned}$$

Now let p_c be the distribution the entries of which are the eigenvalues of σ_C. We have $\mathrm{Tr}(\sigma_C^q) = \sum_c p_c^q$. Thus, if $q > 1$, $\sum_c p_c^q \leqslant 1$ with equality if and only if $p_c = 1$ for some value of c. It follows that

$$I_q(A:B|CC^*)_{\hat{\sigma}} \leqslant I_q(A:B|C^*)_\tau.$$

Therefore, if $I_q(A:B|C^*)_\tau \leqslant f(q, d_A, d_B)$, we also have $I_q(A:B|CC^*)_{\hat{\sigma}} \leqslant f(q, d_A, d_B)$. ■

[21] This is the analog of the statement $p_{ABC} = p_C p_{A|C} p_{B|C}$ for probability distributions.

[1] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics* (Cambridge University, Cambridge, England, 1987).
[2] A. K. Ekert, Quantum Cryptography Based on Bell's Theorem, Phys. Rev. Lett. **67**, 661 (1991).
[3] D. Mayers and A. Yao, Quantum cryptography with imperfect apparatus, in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)* (IEEE, New York, 1998), pp. 503–509.
[4] J. Barrett, L. Hardy, and A. Kent, No Signalling and Quantum Key Distribution, Phys. Rev. Lett. **95**, 010503 (2005).
[5] R. Colbeck, Quantum and relativistic protocols for secure multi-party computation, Ph.D. dissertation, University of Cambridge, 2007, arXiv:0911.3814.
[6] R. Colbeck and A. Kent, Private randomness expansion with untrusted devices, J. Phys. A **44**, 095305 (2011).
[7] S. Pironio *et al.*, Random numbers certified by Bell's theorem, Nature (London) **464**, 1021 (2010).
[8] I. Pitowski, in *Quantum Probability — Quantum Logic*, Lecture Notes in Physics Vol. 321 (Springer-Verlag, Berlin, Heidelberg, 1989).
[9] L. Masanes, Tight Bell inequality for d-outcome measurements correlations, Quantum Inf. Comput. **3**, 345 (2002).
[10] J.-D. Bancal, N. Gisin, and S. Pironio, Looking for symmetric Bell inequalities, J. Phys. A **43**, 385303 (2010).
[11] T. Cope and R. Colbeck, Bell inequalities from no-signaling distributions, Phys. Rev. A **100**, 022114 (2019).
[12] E. Wolfe, R. W. Spekkens, and T. Fritz, The inflation technique for causal inference with latent variables, J. Causal Inference **7** (2019).
[13] M. Navascues and E. Wolfe, The inflation technique completely solves the causal compatibility problem, arXiv:1707.06476.
[14] R. W. Yeung, A framework for linear information inequalities, IEEE Trans. Inf. Theory **43**, 1924 (1997).
[15] T. Fritz and R. Chaves, Entropic inequalities and marginal problems, IEEE Trans. Inf. Theory **59**, 803 (2013).
[16] R. Chaves, Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality, Phys. Rev. A **87**, 022102 (2013).
[17] M. Weilenmann and R. Colbeck, Analysing causal structures with entropy, Proc. R. Soc. A **473**, 20170483 (2017).
[18] M. Weilenmann and R. Colbeck, Inability of the entropy vector method to certify nonclassicality in linelike causal structures, Phys. Rev. A **94**, 042112 (2016).
[19] S. L. Braunstein and C. M. Caves, Information-Theoretic Bell Inequalities, Phys. Rev. Lett. **61**, 662 (1988).
[20] M. Weilenmann and R. Colbeck, Non-Shannon inequalities in the entropy vector approach to causal structures, Quantum **2**, 57 (2018).
[21] R. Chaves, L. Luft, and D. Gross, Causal structures from entropic information: Geometry and novel scenarios, New J. Phys. **16**, 043001 (2014).
[22] A. Rényi, 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 (University of California Press, Berkeley, CA, 1961), pp. 547–561.
[23] M. Wajs, P. Kurzyński, and D. Kaszlikowski, Information-theoretic Bell inequalities based on Tsallis entropy, Phys. Rev. A **91**, 012114 (2015).
[24] Z. Zhang and R. W. Yeung, A non-Shannon-type conditional inequality of information quantities, IEEE Trans. Information Theory **43**, 1982 (1997).
[25] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University, New York, 2000).
[26] D. Geiger, Towards the formalization of informational dependencies, UCLA Technical Report No. 880053 (1987), http://fmdb.cs.ucla.edu/Treports/880053.pdf.
[27] T. Verma and J. Pearl, Causal networks: Semantics and expressiveness, in *Proceedings of the Fourth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '88

(North-Holland Publishing Co., Amsterdam, The Netherlands, 1990), pp. 69–78.

[28] H. P. Williams, Fourier's method of linear programming and its dual, Am. Math. Mon. **93**, 681 (1986).

[29] T. Fritz, Beyond Bell's theorem: Correlation scenarios, New J. Phys. **14**, 103001 (2012).

[30] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, J. Stat. Phys. **52**, 479 (1988).

[31] S. Furuichi, Information theoretical properties of Tsallis entropies, J. Math. Phys. **47**, 023302 (2006).

[32] S. Abe and A. Rajagopal, Nonadditive conditional entropy and its significance for local realism, Physica A (Amsterdam) **289**, 157 (2001).

[33] E. M. F. Curado and C. Tsallis, Generalized statistical mechanics: Connection with thermodynamics, J. Phys. A: Math. Gen. **24**, L69 (1991).

[34] S. Furuichi, K. Yanagi, and K. Kuriyama, Fundamental properties of Tsallis relative entropy, J. Math. Phys. **45**, 4868 (2004).

[35] Z. Daróczy, Generalized information functions, Information and Control **16**, 36 (1970).

[36] Polyhedral Representation and Transformation Algorithm, http://porta.zib.de/.

[37] R. Colbeck and V. Vilasini, LPASSUMPTIONS, MATHEMATICA package (2019), available at https://github.com/rogercolbeck/LPAssumptions.

[38] D. Rosset, N. Gisin, and E. Wolfe, Universal bound on the cardinality of local hidden variables in networks, Quantum Inf. Comput. **18**, 0910 (2018).

[39] N. D. Mermin, Simple Unified form for the Major No-Hidden-Variables Theorems, Phys. Rev. Lett. **65**, 3373 (1990).

[40] A. Peres, Incompatible results of quantum measurements, Phys. Lett. A **151**, 107 (1990).

[41] M. Weilenmann and R. Colbeck, Analysing causal structures in generalised probabilistic theories, arXiv:1812.04327.

[42] T. Gläßle, D. Gross, and R. Chaves, Computational tools for solving a marginal problem with applications in Bell non-locality and causal modeling, J. Phys. A **51**, 484002 (2018).

[43] X. Hu and Z. Ye, Generalized quantum entropy, J. Math. Phys. **47**, 023502 (2006).

[44] K. Audenaert, Subadditivity of q-entropies for $q > 1$, J. Math. Phys. **48**, 083507 (2007).

[45] M. S. Leifer and R. W. Spekkens, Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference, Phys. Rev. A **88**, 052130 (2013).

[46] F. Costa and S. Shrapnel, Quantum causal modelling, New J. Phys. **18**, 063032 (2016).

[47] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Quantum Common Causes and Quantum Causal Models, Phys. Rev. X **7**, 031021 (2017).

[48] J. Pienaar, A time-reversible quantum causal model, arXiv:1902.00129.

[49] J. S. Kim, Tsallis entropy and general polygamy of multiparty quantum entanglement in arbitrary dimensions, Phys. Rev. A **94**, 062338 (2016).

[50] D. Horsman, C. Heunen, M. F. Pusey, J. Barrett, and R. W. Spekkens, Can a quantum state over time resemble a quantum state at a single time? Proc. R. Soc. A **473**, 20170395 (2017).

[51] A. Jamiołkowski, Linear transformations which preserve trace and positive semidefiniteness of operators, Rep. Math. Phys. **3**, 275 (1972).

[52] M.-D. Choi, Completely positive linear maps on complex matrices, Linear Algebra Appl. **10**, 285 (1975).

[53] D. Petz and D. Virosztek, Some inequalities for quantum Tsallis entropy related to the strong subadditivity, Math. Inequal. Appl. **18**, 555 (2015).