Supplementary Figure 1. The GBN for the triangle.

Supplementary Figure 2. A generalized Bayesian network for $G_{n,m}$.

Supplementary Figure 3. An example of the modified GBN from Lemma 2 for $n = 3$ and $i = 2$. 
a) A test.  

b) A preparation and a measurement test.

c) The unique deterministic test.  
d) Labelling of systems is omitted when not needed explicitly.

Supplementary Figure 4. The graphical calculus for operational probabilistic theories.

| Extremal rays of the marginal triangle scenario |  |
|-----------------------------------------------|--|---|---|---|---|---|---|---|
| type | # of permutations | $H_C$ | $H_B$ | $H_{BC}$ | $H_A$ | $H_{AC}$ | $H_{AB}$ | $H_{ABC}$ |
| 1    | 3               | 0    | 0    | 0    | 1    | 1    | 1    | 1    |
| 2    | 3               | 0    | 1    | 1    | 1    | 1    | 1    | 1    |
| 3    | 1               | 1    | 1    | 2    | 1    | 2    | 2    | 2    |
| 4    | 3               | 3    | 3    | 5    | 2    | 4    | 4    | 6    |

Supplementary Table I. The four kinds of extremal rays defining the marginal entropic cone of the triangle scenario.
Supplementary Note 1: Details about the new IC inequality

In the following we will discuss how to characterize the most general marginal scenario in the information causality game. We will start discussing the purely classical case (i.e. Alice and Bob share classical correlations) and afterwards apply the linear program framework to prove that the tighter information causality inequality (a facet of the classical Shannon cone) is also valid for quantum mechanical correlations.

Without loss of generality, we restrict our attention to the particular case of two input bits. The classical causal structure associated with information causality contains six classical variables \( S = \{ X_1, X_2, Y_1, Y_2, M, \lambda \} \). The variable \( \lambda \) stands here for the classical analog of the quantum state \( \varrho_{AB} \). The most general marginal scenario that is compatible with the information causality game and thus with protocols using more general resources such as nonlocal boxes is given by \( M = \{ X_1, X_2, Y_i, M \} \) (with \( i = 1, 2 \)). The relevant conditional independencies implied by the graph are given by \( I(X_1, X_2 : \lambda) = 0 \) and \( I(X_1, X_2 : Y_1, Y_2 | M, \lambda) = 0 \). CIs like \( I(X_1 : Y_1 | M, \lambda) = 0 \) are implied by the relevant ones together with the polymatroidal axioms for the set \( S \) of variables, and in this sense are thus redundant. Given this inequality description (basic inequalities plus CIs) we need to eliminate from our description, via a FM elimination, all the variable not contained in \( M \).

Our first step was to eliminate from the system of inequalities the variable \( \lambda \). Doing that one obtains a new set of inequalities for the five variables \( S = \{ X_1, X_2, Y_1, Y_2, M \} \). These set of inequalities is simply given by the basic inequalities plus one single non-trivial inequality, implied by the CIs:

\[
H(Y_1, Y_2, M) + H(X_1, X_2) \leq H(M) + H(X_1, X_2, Y_1, Y_2, M)
\] (1)

We then proceed eliminating all variables not contained in \( M \). The final inequality description of the marginal cone of \( M \) can be organized in two groups. The first group contains all inequalities that are valid for the collection of variables in \( M \) independently of the underlying causal relationships between them, that is, they follow from the basic inequalities alone. The second group contains the inequalities that follow from the basic inequalities plus the conditional independencies implied by the causal structure. These are the inequalities capturing the causal relations implied by information causality and there are 54 of them. Among these 54 inequalities, one of particular relevance is the tighter IC inequality (8) given in the main text.

One can prove, using the linear program framework delineated before, that this inequality is also valid for the corresponding quantum causal structure shown in Fig. 1 b) of the main text. Following the discussion in the main text, the sets of jointly existing variables in the quantum case are given by \( S_0 = \{ X_1, X_2, A, B \} \), \( S_1 = \{ X_1, X_2, M, B \} \) and \( S_2 = \{ X_1, X_2, M, Y_i \} \) (with \( i = 1, 2 \)). One can think about these sets of variables in a time ordered manner. At time \( t = 0 \) the joint existing variables are the inputs \( X_1 \) and \( X_2 \) of Alice, together with the shared quantum state \( \varrho_{AB} \). At time \( t = 1 \) Alice encodes the input bits into the message \( M \) also using her correlations with Bob obtained through the shared quantum state. Doing that, Alice disturbs her part \( A \) of the quantum system that therefore does not coexist with the variables defined in \( S_1 \). In the final step of the protocol at time \( t = 2 \), Bob uses the received message \( M \) and its part \( B \) of the quantum state in order to make a guess \( Y_1 \) or \( Y_2 \) about Alice’s inputs. Once more, by doing that \( B \) ceases to coexist with the variables contained in \( S_2 \).

Following the general idea, we write down all the basic inequalities for the sets \( S_0 \) and \( S_1 \) and \( S_2 \), together with the conditional independencies and the data processing inequalities. As discussed before, because the quantum analogous of \( I(X_1, X_2 : Y_1, Y_2 | M, \lambda) = 0 \) has no description in the quantum case, the only CI implied here will be \( I(X_1, X_2 : A, B) = 0 \). The causal relations encoded in the other CIs are taken care by the data processing inequalities.

Below we list all used data processing inequalities:

\[
I(X_1, X_2 : Y_i) \leq I(X_1, X_2 : M, B)
\] (2)

\[
I(X_i : Y_j) \leq I(X_i : M, B)
\] (3)

\[
I(X_i : X_{i\text{\#1}}, Y_j) \leq I(X_i : X_{i\text{\#1}}, B)
\] (4)

\[
I(X_1, X_2 : Y_i, M) \leq I(X_1, X_2 : B, M)
\] (5)

\[
I(X_i : Y_j, M) \leq I(X_i : B, M)
\] (6)

\[
I(X_i : X_{i\text{\#1}}, Y_j, M) \leq I(X_i : X_{i\text{\#1}}, B, M)
\] (7)

Note that some of these DP inequalities may be redundant, that is, they may be implied by other DP inequalities together with the basic inequalities.
We organize all the above constraints into a matrix $M$ and given a certain candidate inequality $I$ we run the linear program discussed in the main text. Doing that one can easily prove that inequality (8) is also valid in the quantum case.

Note that this computational analysis will in general be restricted by the number of variables involved in the causal structure, highlighting the relevance of the analytical proof provided in the main text.

In the main text we have shown that our new inequality (8) can witness, already on the single copy level, the postquantumness of distributions that could not be detected before even in the limit of many copies. In order to generate the plots in Figure (2) of the main text we proceed similarly to the approach in Ref. [14]. To compute the single copy and many copies curves we use the protocol devised in [2].

Supplementary Note 2: Proving the monogamy relations for the triangle scenario

Since 1998 it is known that, for a number of variables $n \geq 4$, there are inequalities valid for Shannon entropies that cannot be derived from the elemental set of polymatroidal axioms (submodularity and monotonicity) [3]. These are the so called non-Shannon type inequalities [4]. More precisely, the existence of these inequalities imply that the true entropic cone (denoted by $\Gamma_n$) is a strict subset of the Shannon cone, that is, the inclusion $\Gamma_n^* \subseteq \Gamma_n$ is strict for $n \geq 4$.

Remember that a convex cone has a dual description, either in terms of its facets or its extremal rays. In terms of its half-space description the strict inclusion $\Gamma_n^* \subset \Gamma_n$ implies in particular that while all Shannon type inequalities are valid for any true entropy vector, they may fail to be tight. In terms of the extremal rays, this implies that some of the extremal rays of the Shannon cone are not populated, that is, there is no well defined probability distribution with an entropy vector corresponding to it.

Sometimes, the projection of the outer approximation $\Gamma_n$ onto a subspace, described by the marginal cone $\Gamma_M$, may lead to the true cone in the marginal space, that is $\Gamma_M = \Gamma_M^*$ [5]. A sufficient condition for that to happen is that all the extremal rays of $\Gamma_M$ are populated. Using this idea, in the following we will prove that all the extremal rays describing the Shannon marginal cone of classical triangle scenario are populated, proving that in this case the Shannon and true marginal cones coincide.

Proceeding with the three steps program delineated in the main text, one can see that the marginal scenario $\{A, B, C\}$ of the triangle scenario is completely characterized by the following non-trivial Shannon type inequalities (and permutations thereof) [6]

$$
I(A : B) + I(A : C) - H(A) \leq 0,
$$

$$
I(A : B) + I(A : C) + I(B : C) - H(A,B) \leq 0,
$$

$$
I(A : B) + I(A : C) + I(B : C) - \frac{1}{2}(H(A) + H(B) + H(C)) \leq 0,
$$

plus the polymatroidal axioms for the three variables $A, B, C$. Given the inequality description of the marginal cone we have used the software PORTA [7] in order to recover the extremal rays of it. There are only 10 extremal rays, that can be organized in the 4 different types listed in Supplementary Table[4].

Below we list the probability distributions reproducing the 4 different types of entropy vectors:

$$
p_1(a, b, c) = \begin{cases} 1/2 & \text{if } a = \{1,2\} \text{ and } b = c = \{1\} \\ 0 & \text{otherwise} \end{cases},
$$

$$
p_2(a, b, c) = \begin{cases} 1/2 & \text{if } a = b = \{1,2\} \text{ and } c = \{1\} \\ 0 & \text{otherwise} \end{cases},
$$

$$
p_3(a, b, c) = \begin{cases} 1/4 & \text{if } a \oplus b \oplus c = 0 \text{ with } a, b, c = \{1,2\} \\ 0 & \text{otherwise} \end{cases},
$$

and
The corresponding output system state is called an event. Physical operations performed, e.g., the preparation of a state, or a measurement. A test has input and output classical. Using the linear programming framework detailed above, one can also prove that the inequality (8) holds if the underlying hidden variables stand for quantum states. The sets of jointly existing variables in the quantum case are given by $S_0 = \{A_1, A_2, B_1, B_2, C_1, C_2\}$, $S_1 = \{A, B_1, B_2, C_1, C_2\}$, $S_2 = \{A_1, A_2, B, C_1, C_2\}$, $S_3 = \{A_1, A_2, B_1, B_2, C\}$, $S_4 = \{A, B, C_1, C_2\}$, $S_5 = \{A, B_1, B_2, C\}$, $S_6 = \{A_1, A_2, B, C\}$ and $S_7 = \{A, B, C\}$. The fact that the quantum states are assumed to be independent is translated in the CI $H(A_1, A_2, B_1, B_2, C_1, C_2) = H(\varrho_{A_1B_1}) + H(\varrho_{A_2C_2}) + H(\varrho_{C_2C_2})$. Classically, the causal constraint that observable variables have no direct influence on each other (all the correlation are mediated by the underlying quantum states) is encoded in CIs like $I(A : B | A_1) = I(A : B | B_1) = 0$ that involve jointly non-coexisting variables in the quantum case. In the quantum case these causal constraints are taken care by the corresponding data processing inequalities. Below we also list all used data processing inequalities:

\[
I(A : B) \leq I(A : B_1, B_2) \tag{15}
\]
\[
I(A : B) \leq I(A_1, A_2 : B) \tag{16}
\]
\[
I(A : B) \leq I(A_1, A_2 : B_1, B_2) \tag{17}
\]
\[
I(A : B, C) \leq I(A : B, C_1, C_2) \tag{18}
\]
\[
I(A : B, C) \leq I(A : B_1, B_2, C) \tag{19}
\]
\[
I(A : B, C) \leq I(A_1 : B_1, B_2, C_1, C_2) \tag{20}
\]
\[
I(A : B, C) \leq I(A_1, A_2 : B, C_1, C_2) \tag{21}
\]
\[
I(A : B, C) \leq I(A_1, A_2 : B_1, B_2, C) \tag{22}
\]
\[
I(A : B, C) \leq I(A_1, B_1, B_2, C_1, C_2) \tag{23}
\]
\[
I(A, B_1 : B_2) \leq I(A_1, A_2, B_1 : B_2) \tag{24}
\]
\[
I(A, C_1 : C_2) \leq I(A_1, A_2, C_1 : C_2) \tag{25}
\]

and similarly for permutations of all variables. Again, note that some of these DP inequalities may be redundant, that is, they may be implied by other DP inequalities together with the basic inequalities.

Supplementary Note 3: Proving the monogamy relation for general non-signalling theories

The goal of this appendix is to prove the inequality

\[
\sum_{j \in \{1, \ldots, n\} \setminus \{i\}} I(V_i : V_j) \leq H(V_i) \tag{26}
\]

for random variables that constitute a generalized Bayesian network [9] with respect to a DAG where each parent correlates at most two of them, i.e., the random variables are results of measurements on a set of arbitrary non-signalling resources shared between two parties. The case of three random variables has been proven in [9], the purpose of this appendix is to prove the generalization to an arbitrary number of random variables. Also we want to prove that for any fixed connectivity number for the parent nodes there are entropic corollaries. To this end we have to introduce a framework to handle generalized probabilistic theories that are non-signaling and have a property called local discriminability that was developed in [9].

An operational probabilistic theory has two basic notions, systems and tests. Tests are the objects that represent any physical operation that is performed, e.g., the preparation of a state, or a measurement. A test has input and output systems and can have a classical random variable as measurement outcome as well. An outcome together with the corresponding output system state is called event. The components are graphically represented by a directed
acyclic graph (DAG) where the nodes represent tests, and the edges represent systems. We use the convention that the diagram is read from bottom to top, i.e. a test's input systems are represented by edges coming from below and its output systems are edges emerging from the top of the node (Supplementary Figure 2(a)). If a system has trivial input or trivial output we omit the edge and represent the node by a half moon shape. Tests with trivial input are called preparations, tests with trivial output are called measurements (Supplementary Figure 2(b)). If we do not need to talk about the systems we omit the labels of the edges, and preparation tests are given greek letter labels, as they have, without loss of generality, only a single event (Supplementary Figure 2(d)). Finally we assume, just as it is done in [8], that there exists a unique way of discarding a system, which is shown in Supplementary Figure 2(c).

We call this the discarding test, and it is shown in [8] that its existence and uniqueness is equivalent to the non-signaling condition. These elements can now be connected by using the output system of one test as input system for another. An arrangement of tests is called a generalized Bayesian network (GBN). We also say that the arrangement forms a GBN with respect to a DAG $\mathcal{G}$, or that a GBN has shape $\mathcal{G}$, if the tests are arranged according to it, analogous to the classical case introduced in the main text.

The main ingredient for the proof of (26) for $n = 3$ in [8] is the following

**Lemma 1** ([8, Thm. 23].) For any probability distribution $p(x, y, z)$ of random variables that are the classical output of a GBN with respect to the DAG in Supplementary Figure 1 there is a probability distribution $p'$ such that

\[
\begin{align*}
    p'(x, z) &= p'(x)p'(z) \\
p'(x, y) &= p(x, y) \\
p'(y, z) &= p(y, z)
\end{align*}
\]

For our purposes we need a generalization of this result. The GBN for the scenario of any parent connecting at most $m$ children can be described as follows. The $n$ random variables $V_1, ..., V_n$ arise from $n$ measurement tests. For any $m$ of these measurement tests there is a preparation test whose output systems are input for exactly these measurements. We denote the preparation test corresponding to a subset $I \subseteq \{1, ..., n\}$, $|I| = m$ by $\sigma_I$. In total there are therefore $\binom{n}{m}$ preparations. This GBN can be found in Supplementary Figure 2 and we denote the corresponding DAG by $\mathcal{G}_{n,m}$.

**Lemma 2.** For any probability distribution arising from a GBN of shape $\mathcal{G}_{n,m}$ and any index $i \in \{1, ..., n\}$ there is a probability distribution $p'$ such that any bivariate marginal involving $V_i$ is equal to the corresponding marginal of $p$ and $p'(v_1, ..., v_{i-1}, v_{i+1}, ..., v_n)$ is compatible with $\mathcal{G}_{n-1,m-1}$.

**Proof:** Analogous to the proof of Lemma 1 in [8] we define a new GBN from the old one as follows:

- Any preparation test $\sigma_I$ with $i \in I$ is left as it is.

- Any preparation test $\sigma_I$ with $i \notin I$ is copied. In one copy the first outgoing edge is discarded and in the second copy all edges except the first are discarded.

The modified GBN is depicted in Supplementary Figure 3 for $n = 3$, $m = 2$ and $i = 2$. It can be seen in an analogous way as in the proof of Theorem 23 in [8] that using the probability distribution that arises from this GBN as $p'$ has the desired properties. This concludes the proof.

We are now ready to prove the inequality (26).

**Theorem 3.** Let $V_1, ..., V_n$ be random variables defined by a GBN of shape $\mathcal{G}_{n,2}$. Then for any $i \in \{1, ..., n\}$

\[
\sum_{j \in \{1, ..., n\} \atop j \neq i} I(V_i : V_j) \leq H(V_i)
\]

**Proof:** Without loss of generality we take $i = 1$. We proceed by induction over $n$. For $n = 2$ the inequality is trivially true. Assume now that the statement is true for $n - 1$. We construct the probability distribution $p'$ according to Lemma 2 and observe that $p'(v_2, ..., v_n)$ arises from $\mathcal{G}_{1,n-1}$, i.e. it is a product distribution. Denote the modified random variables by $V'_1, ..., V'_n$ and calculate
\[ \sum_{j=2}^{n} I(V_1 : V_j) = \sum_{j=2}^{n} I(V'_1 : V'_j) \]

\[ = I(V'_2 : V'_3) - I(V'_2 : V'_3|V'_1) + I(V'_1 : V'_2V'_3) + \sum_{j=4}^{n} I(V'_1 : V'_j) \]

\[ \leq I(V'_1 : V'_2V'_3) + \sum_{j=4}^{n} I(V'_1 : V'_j), \]

where the inequality follows from the independence of \( V'_2 \) and \( V'_3 \) and strong subadditivity. Now observe that with \( X = (V'_2, V'_3) \) the distribution of \( V'_1, X, V'_4, ..., V'_m \) is compatible with \( G_{n-1,2} \) and therefore we have, using the induction hypothesis,

\[ I(V'_1 : V'_2V'_3) + \sum_{j=4}^{n} I(V'_1 : V'_j) \leq H(V'_1). \]

But \( p'(v_1) = p(v_1) \) and therefore (30) is proven.

For general \( m \) the situation is somewhat less simple, but for the special case of \( n = m + 1 \) we can still prove a nontrivial inequality.

**Theorem 4.** Let \( V_1, ..., V_{m+1} \) be random variables corresponding to a GBN of shape \( G_{m+1,m} \). Then

\[ \sum_{k=2}^{m+1} I(V_1 : V_k) \leq \sum_{k=2}^{m+1} I(V'_1 : V'_k) \]

and there is a set of random variables \( X_1, ..., X_{m+1} \) incompatible with \( G_{m+1,m} \) that violates this inequality.

Note that this inequality is, in particular, also true for quantum-classical bayesian networks and, to our knowledge, provides the only known entropic corollaries in this case, too.

**Proof:** (by induction). For \( m = 1 \) the statement is trivially true, as then the two random variables are independent and therefore \( I(V_1 : V_2) = 0 \leq 0 \). Assume now the inequality was proven for \( m - 1 \). Construct random variables \( V'_1, ..., V'_{m+1} \) according to Lemma [3]. Then calculate

\[ \sum_{k=2}^{m+1} I(V_1 : V_k) = \sum_{k=2}^{m+1} I(V'_1 : V'_k) \]

\[ = \frac{1}{m-1} \sum_{k=3}^{m+1} \left[ I(V'_2 : V'_k) - I(V'_2 : V'_3|V'_1) + I(V'_1 : V'_2V'_3) + (m-2)I(V'_1 : V'_k) \right] \]

\[ \leq \frac{1}{m-1} \sum_{k=3}^{m+1} \left[ I(V'_2 : V'_k) + I(V'_1 : V'_2V'_3) + (m-2)I(V'_1 : V'_k) \right] \]

\[ \leq (m-1)H(V'_1) + \frac{1}{m-1} \sum_{k=0}^{m-4} \frac{(m-k-2)(m-k-2)!}{(m-2)!} H(V_{k+2}) \]

\[ = (m-1)H(V'_1) + \sum_{k=1}^{m-3} \frac{(m-k-1)(m-k-1)!}{(m-1)!} H(V_{k+1}) \]

\[ = \sum_{k=0}^{m-3} \frac{(m-k-1)(m-k-1)!}{(m-1)!} H(V_{k+1}), \]

where the first inequality follows from strong subadditivity and the second inequality follows from the induction hypothesis and the trivial bound \( I(X : Y) \leq H(X) \). This completes the proof of the first assertion, i.e. that the inequality is fulfilled by random variables from a GBN of shape \( G_{m+1,m} \). To see that more general random variables violate this inequality, let \( X_i = X, i = 1, ..., m+1 \), where \( X \) is an unbiased coin. In other words, the \( X_i \) are maximally corellated. Then \( H(X_i) = 1 \) and \( I(X_1 : X_i) = 1 \). Therefore we have

\[ \sum_{k=2}^{m+1} I(X_1 : X_k) = m, \]
but
\[
\sum_{k=0}^{m-3} \frac{(m-k-1)(m-k-1)!}{(m-1)!} H(X_{k+1}) = \sum_{k=0}^{m-3} \frac{(m-k-1)(m-k-1)!}{(m-1)!}
\]
\[
= m - \frac{2}{(m-1)!} < m,
\]

hence the inequality \([33]\) is violated. This concludes the proof.

Note that this inequality yields nontrivial constraints for the entropies of random variables resulting from a GBN of any shape \(G_{n,m}, n > m\), as it can be applied to any \(m+1\) of the \(n\) variables.

---

Supplementary References


