Information causality from an entropic and a probabilistic perspective

Sabri W. Al-Safi* and Anthony J. Short†
DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
(Received 18 August 2011; published 11 October 2011)

The information causality principle is a generalization of the no-signaling principle which implies some of the known restrictions on quantum correlations. But despite its clear physical motivation, information causality is formulated in terms of a rather specialized game and figure of merit. We explore different perspectives on information causality, discussing the probability of success as the figure of merit, a relation between information causality and the nonlocal "inner-product game," and the derivation of a quadratic bound for these games. We then examine an entropic formulation of information causality with which one can obtain the same results, arguably in a simpler fashion.

DOI: 10.1103/PhysRevA.84.042323 PACS number(s): 03.67.−a, 03.65.Ud

I. INTRODUCTION

Quantum theory has many strange properties, but perhaps the most surprising is that of *nonlocality*. Some quantum states, known as *entangled* states, cannot be described by giving a separate quantum state for each system, or even by a probabilistic mixture of such states. This is not just an artifact of the mathematical formalism; many entangled states give rise to observable correlations which cannot be explained by any local model [1–3]. However, an important caveat is that these nonlocal correlations cannot be used for superluminal signaling.

Although this area has been extensively studied, we still do not have a good intuition about which nonlocal correlations are achievable in quantum theory, and what they can be used for. They are certainly helpful in some nonlocal tasks [4,5], but it has been shown that even stronger correlations are possible without generating superluminal signaling [6]. Furthermore, there have recently been a number of results describing nonlocal tasks for which quantum entanglement is not helpful at all, while stronger nonlocal correlations give an advantage [7–9]. By gaining a better understanding of quantum nonlocality, we hope to hone our intuitions about its information-theoretic uses, and perhaps learn more about why nature is quantum.

In this paper, we will discuss one particular nonlocal task for which quantum nonlocality is not helpful (at least with the original figure of merit), known as *information causality* [8]. This is an appealing principle which one would reasonably expect to hold, and which quantum theory obeys, yet which can be violated using correlations slightly stronger than quantum theory permits [8,10].

Information causality relates to a particular type of game: A bit string $\mathbf{x}$ of length n is chosen uniformly at random and given to Alice, while Bob is given a random number k, $(1 \leqslant k \leqslant n)$. Alice may then send an m-bit message $\boldsymbol{\alpha}$ to Bob, after which Bob must try to guess x_k, the kth bit of Alice's original bit string. Bob's guess when his input is k is denoted β_k. The parties may decide on a joint strategy and may initially share correlated resources but play from separate locations.

The information causality principle states that

$$I \equiv \sum_{k=1}^{n} I_c(x_k : \beta_k) \leqslant m, \tag{1}$$

where $I_c(X : Y)$ denotes the classical mutual information of variables X and Y [11]. The intuition behind this bound is that the total information that Bob can access about Alice's bits cannot exceed the size of the message she sent. Indeed, the inequality in (1) is saturated if Alice simply sends to Bob the first m bits of $\mathbf{x}$, so that $I_c(x_k : \beta_k) = 1$ if $1 \leqslant k \leqslant m$, and 0 otherwise.

It is proven in [8] that information causality is obeyed in both the quantum and classical world. However, it can be violated in worlds governed by different physical laws (such as "box world" [12,13], which permits all nonsignaling correlations). In what follows, we first discuss probability of success in the information causality game. We then derive a bound which relates information causality to a different nonlocal game, in which Alice and Bob must compute the inner product of two bit strings. Finally, we will explore an alternative formulation and derivation of information causality based on entropy rather than mutual information.

II. PROBABILITY OF SUCCESS FOR INFORMATION CAUSALITY

Although quantum entanglement gives no advantage over a classical strategy in the information causality game when I [defined by (1)] is the figure of merit, it is not true that every quantum strategy can be classically simulated. In fact, if probability of success is used as the figure of merit instead, it can easily be seen that entangled quantum states allow one to do better than in the classical world. For example, in a simple version of the game in which $n = 2$ and $m = 1$, the optimal classical probability of success is $\frac{3}{4}$ (e.g., when Alice sends Bob $\alpha = x_1$ and he guesses $\beta_k = \alpha$, they always win when $k = 1$ and win half the time when $k = 2$). However, by exploiting well-known quantum violations of Bell inequalities, Alice and Bob can achieve a success probability of $\frac{2+\sqrt{2}}{4}$. To do this, Alice and Bob first generate bits a and b satisfying $a \oplus b = (x_1 \oplus x_2)(k-1)$ with probability $\frac{2+\sqrt{2}}{4}$, where $\oplus$ denotes addition modulo 2. This is equivalent to the quantum Tsirelson

*s.w.al-safi@damtp.cam.ac.uk
†a.j.short@damtp.cam.ac.uk

bound for the CHSH inequality [2,14]. Then Alice sends Bob $\alpha = a \oplus x_1$ and Bob outputs $\beta_k = b \oplus \alpha$ [8,15].

It is also possible to obtain very different values of I for strategies with the same probabilities of success. As above, Alice can send Bob her first bit to obtain $I = 1$ and probability of success $\frac{3}{4}$; alternatively, Alice and Bob can randomly "mix" this strategy with one where Alice sends Bob her second bit and he outputs it, so that the overall probability of success is the same but

$$\begin{aligned} I &= I_c(x_1 : \beta_1) + I_c(x_2 : \beta_2) \\ &= 2\left[1 - H\left(\left\{\tfrac{3}{4}, \tfrac{1}{4}\right\}\right)\right] \\ &\approx 0.38. \end{aligned} \tag{2}$$

Furthermore, it is clear that a small amount of noise added to the first strategy will do better than the second strategy in terms of I, but worse in terms of success probability, so these two figures of merit are not monotonically related.

The optimal classical strategy to maximize the probability of success in the case when $m = 1$ has already been derived for general n, in the context of random-access coding [16]. It is attained by using the "majority-vote" strategy, in which Alice simply sends Bob the bit that most frequently occurs in her string. This gives success probability,

$$P^C_{\text{success}} = \frac{1}{2}\left[1 + \frac{1}{2^{n-1}}\binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}\right]. \tag{3}$$

Using Stirling's approximation, one can derive the asymptotic behavior of this probability:

$$P^C_{\text{success}} \approx \frac{1}{2}\left(1 + \sqrt{\frac{2}{\pi n}}\right). \tag{4}$$

We show in the next section that the optimal quantum probability of success for the same situation is

$$P^Q_{\text{success}} = \frac{1}{2}\left(1 + \sqrt{\frac{1}{n}}\right), \tag{5}$$

which is always strictly larger than the classical limit. This extends a result obtained in [17] for particular n [18].

Interestingly, Eq. (5) is also the optimal success probability when Alice is allowed to send a qubit to Bob instead of a classical bit, but Alice and Bob do not share an entangled state [16,17].

The probability of success has a clean operational interpretation as a figure of merit: It is the asymptotic fraction of games one would expect to win over many independent repetitions. Although it sounds appealing, the operational meaning of I is less natural. In particular, suppose Alice and Bob play the game many times, then Alice is told Bob's input k for each round, and she sends him some supplementary classical information which (together with his guesses β_k) he must use to output the correct value of x_k for each round. The average amount of supplementary information per round which Alice must send Bob is $(1 - I/n)$. This follows from a result of [19] that the asymptotic amount of information (using coding over many rounds) required to learn x_k given that you hold β_k is $H(x_k|\beta_k) = H(x_k) - I(x_k : \beta_k)$.

However, although I is a less natural *a priori* figure of merit than success probability, its appeal lies in the simplicity of the bound given by (1). In particular, the maximum value of I is the same for classical or quantum strategies, and can be simply stated for any message length m [by contrast, the maximum success probabilities given by (3)–(5) are complicated, depend on n, and only apply when $m = 1$].

III. INFORMATION CAUSALITY AND THE INNER PRODUCT GAME

Given that the mutual information is a complicated nonlinear function of the associated probabilities, it is surprising that the bound given by information causality can be used to derive the Tsirelson bound, which can be understood as a bound on the quantum success probability for a particular nonlocal game [20]. Even more surprisingly, information causality can be used to generate part of the curved surface of the set of achievable quantum correlations [10].

To investigate this, we note that the proof of the Tsirelson bound given in [8] can be decomposed into several steps. The first is to prove that the information causality principle $I \leqslant m$ implies a bound $\sum_{k=1}^{n}[1 - h(P_k)] \leqslant m$ on the binary entropy [21] of the success probability P_k given a particular input for Bob. This entropic bound can be transformed into a quadratic bound on the *bias* $E_k = (2P_k - 1)$ achieved in the game by noting that $1 - h(P_k) \geqslant \frac{E_k^2}{2\ln 2}$. The information causality principle can therefore be used to generate the bound,

$$\sum_{k=1}^{n} E_k^2 \leqslant 2m \ln 2. \tag{6}$$

Finally, the authors consider a particular strategy for playing the game in which $m = 1$ and n is a power of 2, and show that the ability to generate correlations violating the Tsirelson bound would allow one to violate (6) for sufficiently large n. Hence, given information causality, the Tsirelson bound holds.

As the quadratic bound given by Eq. (6) plays a key role in deriving the Tsirelson bound from information causality, it is interesting to investigate such bounds directly in quantum theory. To facilitate this, we first consider a seemingly unrelated nonlocal game, in which the aim is to produce the inner product of two bit strings. In this inner product game, Alice and Bob are given uniformly random n-bit strings $\mathbf{x}$ and $\mathbf{y}$, respectively. Then without communicating, Alice and Bob must output bits a and b, respectively, such that $a \oplus b = \mathbf{x} \cdot \mathbf{y}$, where $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 \oplus x_2 y_2 \ldots \oplus x_n y_n$.

The ability to win the inner product game perfectly would allow the parties to nonlocally compute any function of their inputs [15], and therefore to solve any communication complexity problem with only a single bit of communication.

We can derive a bound on the inner product game which is very similar to (6). Assume that Alice and Bob share an initial entangled state $|\psi\rangle$, and their outputs are obtained by measuring the operators $\hat{a}_{\mathbf{x}}$ and $\hat{b}_{\mathbf{y}}$, respectively (with eigenvalues 0,1). The bias they achieve in the game when they are given inputs $\mathbf{x}$ and $\mathbf{y}$ is

$$E_{\mathbf{xy}} = \langle\psi|(-1)^{\hat{a}_{\mathbf{x}} + \hat{b}_y + \mathbf{x}\cdot\mathbf{y}}|\psi\rangle, \tag{7}$$

where

$$P^{Q}_{\text{success}} = \frac{1}{2}\left(1 + \frac{1}{2^{2n}}\sum_{\mathbf{xy}} E_{\mathbf{xy}}\right). \tag{8}$$

Similarly, the average bias they achieve when Bob is given $\mathbf{y}$ and we average over Alice's input is given by $E_{\mathbf{y}} = \frac{1}{2^n}\sum_{\mathbf{x}} E_{\mathbf{xy}}$.

To derive a quadratic bound, we adopt a similar approach to [7]. We define the normalized states:

$$|A\rangle = \frac{1}{\sqrt{2^n}}\sum_{\mathbf{x}}(-1)^{\hat{a}_{\mathbf{x}}}|\psi\rangle \otimes |\mathbf{x}\rangle, \tag{9}$$

$$|B_{\mathbf{y}}\rangle = \frac{1}{\sqrt{2^n}}\sum_{\mathbf{x}}(-1)^{\hat{b}_{\mathbf{y}}+\mathbf{x}.\mathbf{y}}|\psi\rangle \otimes |\mathbf{x}\rangle, \tag{10}$$

where the $|B_{\mathbf{y}}\rangle$ states form an orthonormal set satisfying $\langle B_{\mathbf{y}}|B_{\mathbf{y}'}\rangle = \delta\mathbf{y}\mathbf{y}'$.

It follows that

$$\begin{aligned}\sum_{\mathbf{y}} E_{\mathbf{y}}^2 &= \sum_{\mathbf{y}}\langle A|B_{\mathbf{y}}\rangle^2 \\ &= \langle A|\left(\sum_{\mathbf{y}}|B_{\mathbf{y}}\rangle\langle B_{\mathbf{y}}|\right)|A\rangle \\ &\leqslant 1,\end{aligned} \tag{11}$$

where in the last step we have used the fact that $\sum_{\mathbf{y}}|B_{\mathbf{y}}\rangle\langle B_{\mathbf{y}}|$ is a projector and $|A\rangle$ is normalized. A similar result was obtained independently by Pawlowski and Winter, using a different method, and described very recently in Refs. [22,23].

We can also obtain a bound on the probability of success from (11), by taking

$$\begin{aligned}P^{Q}_{\text{success}} &= \frac{1}{2}\left(1 + \frac{1}{2^n}\sum_{\mathbf{y}} E_{\mathbf{y}}\right) \\ &\leqslant \frac{1}{2}\left(1 + \sqrt{\frac{1}{2^n}\sum_{\mathbf{y}} E_{\mathbf{y}}^2}\right) \\ &\leqslant \frac{1}{2}\left(1 + \frac{1}{\sqrt{2^n}}\right).\end{aligned} \tag{12}$$

When $n = 1$, the inner-product game is equivalent to the CHSH game [2], and this bound on the success probability corresponds to the usual Tsirelson bound.

The bound given by Eq. (11) actually holds regardless of the probability distribution over Bob's input $\mathbf{y}$. This generalization allows us to derive a bound on a nonlocal version of the information causality game, in which Alice is given a random n-bit string $\mathbf{x}$, Bob is given a random number k satisfying $1 \leqslant k \leqslant n$, and they attempt to produce bits a and b such that $a \oplus b = x_k$ without communicating. If Bob's bit string in the inner-product game is chosen at random from the set of bit strings containing a single one (i.e., from the bit strings of Hamming weight 1), with k denoting the position of the nonzero bit in $\mathbf{y}$, then $\mathbf{x}\cdot\mathbf{y} = x_k$. In this case, the inner-product game is the same as the nonlocal information causality game and (11) gives

$$\sum_{k} E_k^2 \leqslant 1. \tag{13}$$

Note that this is a stronger bound than (6), which was obtained from information causality. Similarly, an analogous derivation to (12) gives

$$P^{Q}_{\text{success}} \leqslant \frac{1}{2}\left(1 + \sqrt{\frac{1}{n}}\right). \tag{14}$$

We now show that the bound given by (13) can be saturated in quantum theory for any choice of E_k. It was proved in [24] (and used in [22]) that for any set of real vectors $\mathbf{u}_{\mathbf{x}}$ and $\mathbf{v}_{\mathbf{y}}$ of at most unit length, we can find a quantum state $|\psi\rangle$ of a bipartite system, and binary-valued operators $\hat{a}_{\mathbf{x}}$ and $\hat{b}_{\mathbf{x}}$ (which can be measured locally on subsystems A and B), such that

$$\langle\psi|(-1)^{\hat{a}_{\mathbf{x}}+\hat{b}_{\mathbf{x}}}|\psi\rangle = \mathbf{u}_{\mathbf{x}}^T\mathbf{v}_{\mathbf{y}}, \tag{15}$$

and hence

$$E_{\mathbf{xy}} = (-1)^{\mathbf{x}.\mathbf{y}}\mathbf{u}_{\mathbf{x}}^T\mathbf{v}_{\mathbf{y}}. \tag{16}$$

For any desired biases $E_{\mathbf{y}}$ satisfying $\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leqslant 1$ we can consider the vectors,

$$\mathbf{u}_{\mathbf{x}} = \sum_{\mathbf{y}}(-1)^{\mathbf{x}.\mathbf{y}}E_{\mathbf{y}}e_{\mathbf{y}}, \tag{17}$$

$$\mathbf{v}_{\mathbf{y}} = e_{\mathbf{y}}, \tag{18}$$

where $e_{\mathbf{y}}$ denotes an orthonormal basis for a real vector space with dimension equal to the number of different inputs for Bob. This gives $E_{\mathbf{xy}} = E_{\mathbf{y}}$, and hence we can achieve any set of biases satisfying (11). In particular, we could obtain an equal bias for all of Bob's possible inputs in the nonlocal information causality game ($E_k = \frac{1}{\sqrt{n}}$) which would achieve the optimal probability of success $\frac{1}{2}(1 + \frac{1}{\sqrt{n}})$ given by (14).

Although these results apply to the nonlocal version of the information causality game, any strategy can be transferred to the original version of the game with $m = 1$, with the same probability of success. Alice simply sends the message $\alpha = a$ to Bob, and he outputs $\beta = a \oplus b$. This is not the only type of strategy which is possible in the original information causality game (e.g., Bob's measurement could depend on Alice's message). However, in [17] an identical inequality to (14) is derived for the original game, hence the optimal strategy for the nonlocal version of the game is also optimal when transferred to the original game. Note that the strategies used in [8] to derive the Tsirelson bound, and to achieve perfect success given arbitrary nonsignaling resources, are also of this form.

The bound $\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leqslant 1$ for the inner product game seems to capture a great deal about the possible quantum correlations, yet note that this inequality can also be saturated by a classical strategy. In particular, if Alice and Bob output $a = \boldsymbol{\alpha}\cdot\mathbf{x}$ and $b = 0$, they will achieve a bias of 1 when $\mathbf{y} = \boldsymbol{\alpha}$ and 0 in every other case.

IV. INFORMATION CAUSALITY FROM ENTROPY

Given the above, it appears that the particular mathematical form of the mutual information is not central in defining the boundary of the set of quantum correlations (as the proof proceeds via a quadratic bound), and the choice of I rather than probability of success as the figure of merit seems somewhat arbitrary. However, the fact that quantum

theory obeys information causality actually follows from the existence of a natural extension of the classical mutual information to quantum states. Can we focus on this as a defining property of quantum theory?

In general probabilistic theories, the state of a system is characterized by a complete description of the probability of each measurement outcome, for any possible measurement on that system [12,25,26]. A specific probabilistic theory is defined by allowing certain types of systems, and certain states on those systems: For example, classical theory consists of systems specified by a single probability distribution (such as a ball in one of several boxes). In any such theory, it was shown in [8] that the information causality principle $I \leqslant m$ will hold if an analog of the mutual information $I(X:Y)$ can be defined for all systems X and Y (which may be composite) with the following properties:

(i) Consistency: Whenever X and Y are classical systems, I reduces to the classical mutual information, $I(X:Y) = I_c(X:Y)$.

(ii) Data processing: Whenever a transformation is performed on Y alone, $\Delta I(X:Y) \leqslant 0$.

(iii) Chain rule: For all tripartite systems X, Y, Z,

$$I(X:YZ) - I(X:Z) = I(XZ:Y) - I(Z:Y).$$

(iv) Symmetry: $I(X:Y) = I(Y:X)$.

(v) Non-negativity: $I(X:Y) \geqslant 0$.

It is well known that all of these properties are satisfied by I_q and I_c, the quantum and classical versions of the mutual information. The proof of information causality also assumes the validity of some natural operations, in particular, the ability to discard a system, or to prepare a system in a state determined by the value of a classical variable. These transformations can be defined for any theory in the general probabilistic framework of [12]. If we consider discarding both X and Y, we can actually derive non-negativity from the symmetry and data-processing conditions, since (denoting a discarded system by $\varnothing$) $I(X:Y) \geqslant I(X:\varnothing) = I(\varnothing:X) \geqslant I(\varnothing:\varnothing) = 0$, hence condition (v) can easily be eliminated [27].

However, while the other properties seem intuitively reasonable, property (iii) seems like a strange demand. Furthermore, the fact that the mutual information necessarily concerns a pair of systems makes it a somewhat complicated quantity.

In the remainder of this section, we show that the information causality principle follows more simply from the existence of "good" measure of entropy in a general theory. In particular, the entropy only concerns a single system (although this may be composite), and is only required to obey two conditions.

(I) Consistency: If system X is classical, $H(X)$ reduces to the classical entropy, $H(X) = H_c(X)$.

(II) Evolution with an ancilla: For any two systems X and Y, whenever a transformation is performed on Y alone,

$$\Delta H(XY) \geqslant \Delta H(Y). \tag{19}$$

Condition (I) says that H gives the asymptotic compression rate for classical data. Condition (II) can be understood intuitively as saying that a local transformation can generate more uncertainty than its effect on an individual subsystem would suggest, as it can destroy but not create correlations. If we also define a conditional entropy analogously to the quantum and classical quantity, as $H(X|Y) = H(XY) - H(Y)$, we can alternatively re-express (19) as $\Delta H(X|Y) \geqslant 0$. We can also express (II) symmetrically as the requirement that $\Delta H(XY) \geqslant \Delta H(X) + \Delta H(Y)$ under local transformations on X and Y.

Given an entropy function obeying the above conditions, we can define a mutual information analogously to the quantum and classical case as

$$I(X:Y) = H(X) + H(Y) - H(XY). \tag{20}$$

This automatically ensures that conditions (iii) and (iv) are satisfied, removing the awkwardness of having to postulate the chain rule, and (i) and (ii) follow trivially from (I) and (II), respectively.

The existence of an entropy function with properties (I) and (II) is therefore sufficient to derive information causality. Conversely, in any theory in which one can violate Tsirelson's bound, it must be *impossible* to define an entropy which satisfies assumptions (I) and (II). Several entropies which can be applied to any probabilistic theory, and which always obey (I), have been proposed in [28–30]. A different set of entropic conditions which can be used to derive information causality were discussed in [29].

It's not hard to deduce some other standard properties of the entropy from conditions (I) and (II):

Subadditivity. By discarding Y, we find from (II) that the entropy is subadditive:

$$H(XY) \leqslant H(X) + H(Y). \tag{21}$$

When X and Y are independent systems, we can also prepare Y locally, which implies that $H(X,Y) = H(X) + H(Y)$ in this case.

Strong subadditivity. By discarding Z from the composite YZ in the tripartite system XYZ, we obtain strong subadditivity:

$$H(XYZ) + H(Y) \leqslant H(XY) + H(YZ). \tag{22}$$

This inequality is equivalent to subadditivity of the conditional entropy. It can also be iterated for a larger number of systems to give

$$H(X_1 \ldots X_n|Y) \leqslant H(X_1|Y) + \cdots + H(X_n|Y). \tag{23}$$

Positivity of classical entropy. Uncertainty about the state of a classical system X can never be negative, even when one conditions on an arbitrary system Y.

$$\text{System } X \text{ is classical} \Rightarrow H(X|Y) \geqslant 0. \tag{24}$$

We argue this last result in the following way: The state of X is described by a probability distribution on a finite set E of outcomes, and for each outcome $e \in E$ there is a corresponding reduced state $\sigma^{Y|e}$ of Y. We can therefore obtain the joint state of system XY by a local transformation on Y from a classical system that is initially perfectly correlated with (and identical to) X. Before the transformation the conditional entropy is

given by $H(X|X)$, and so is non-negative by (I). Then by (II), after the transformation $H(X|Y)$ must also be non-negative.

Information causality can be proven from the existence of an entropy satisfying (I) and (II) by first constructing the mutual information and then applying the proof of [8]. However, it can also be proved more directly using the properties of the entropy derived above, and this yields a slight generalisation of the information causality principle.

Bob's guess β_k is derived solely from Alice's message $\boldsymbol{\alpha}$ and Bob's system B before that message is sent. Thus whatever the strategy, there is a transformation from $(\boldsymbol{\alpha}, B)$ to β_k for each k:

$$\begin{aligned}\sum_k H_c(x_k|\beta_k) &\geqslant \sum_k H(x_k|\boldsymbol{\alpha}, B)\\ &\geqslant H(\mathbf{x}|\boldsymbol{\alpha}, B)\\ &= H(\mathbf{x},\boldsymbol{\alpha}, B) - H(\boldsymbol{\alpha}, B)\\ &\geqslant H(\mathbf{x},\boldsymbol{\alpha}, B) - H(B) - H(\boldsymbol{\alpha})\\ &= H(\mathbf{x},\boldsymbol{\alpha}, B) - H(\mathbf{x}, B) + H(\mathbf{x}) - H(\boldsymbol{\alpha})\\ &= H(\boldsymbol{\alpha}|\mathbf{x}, B) + H(\mathbf{x}) - H(\boldsymbol{\alpha})\\ &\geqslant H_c(\mathbf{x}) - H_c(\boldsymbol{\alpha}). \end{aligned} \tag{25}$$

This is a generalized form of the information causality principle which makes no assumption on the distribution on Alice's input $\mathbf{x}$. It can be interpreted as saying that the remaining uncertainty that Bob has about Alice's bits after guessing must be more than the original uncertainty about her inputs minus the information gained by the message. In the special case in which Alice's inputs are independent, $H_c(\mathbf{x}) = \sum_k H_c(x_k)$, and we can rearrange (25) to get

$$\sum_k I(x_k : \beta_k) \leqslant H_c(\boldsymbol{\alpha}) \leqslant m, \tag{26}$$

as in [8].

V. CONCLUSIONS

Considering probability of success in the information causality game, we see that quantum theory gives an advantage which is not captured by the figure of merit I which is bounded by (1). Investigating how these probabilities are involved in deriving Tsirelson's bound from information causality [8] leads us to a quadratic quantum bound,

$$\sum_{\mathbf{y}} E_{\mathbf{y}}^2 \leqslant 1, \tag{27}$$

on the biases achieved given different inputs for Bob in the nonlocal inner product game. This applies for an arbitrary distribution over Bob's inputs, and hence to the nonlocal version of the information causality game. This is another example of a bound which quantum and classical correlations can both saturate, but stronger nonlocal correlations can violate. Furthermore, the fact that quantum correlations allow one to achieve any set of biases satisfying this rule means that it captures a significant amount about the set of quantum correlations. Can we construct useful quadratic bounds on quantum performance in other nonlocal tasks?

Instead of considering information causality as a constraint on possible physical theories, it may be helpful to think of it as a consequence of the existence of a "good" measure of entropy in the theory. Indeed, we have shown that information causality can be derived given any extension of the entropy from classical to more general systems which satisfies $\Delta H(XY) \geqslant \Delta H(X) + \Delta H(Y)$ under local transformations. Conversely, any theory which violates information causality (such as box world) cannot have an entropy defined in it which obeys the above evolution law and agrees with the Shannon entropy for classical systems.

Given the above results, as well as those of [28–30], it seems that the existence of a "good" entropy for quantum theory, which shares so many of the properties of the classical entropy, is very special within the class of general probabilistic theories. Are there other theories for which one can define an entropy satisfying (I) and (II), or is this a defining feature of quantum theory [31]? The existence of such an entropy potentially places stronger bounds on quantum theory than information causality alone. It would be interesting to look for other games where quantum theory can do no better than classical when such an entropy exists.

Note added in proof. Very recently, similar results to those in Sec. IV have been obtained independently in [32].

ACKNOWLEDGMENTS

The authors thank Marcin Pawłowski and Sandu Popescu for interesting discussions. A.J.S. also acknowledges the support of the Royal Society.

[1] J. S. Bell, Physics **1**, 195 (1965).
[2] J. Clauser, M. Horne, A. Shimony, and R. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[3] A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **49**, 91 (1982).
[4] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[5] R. Cleve, P. Høyer, B. Toner, and J. Watrous, in *Proceedings of the19th Annual IEEE Conference on Computational Complexity (CCC'04)*, Vol. 236 (IEEE, Piscataway, 2004).
[6] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[7] N. Linden, S. Popescu, A. J. Short, and A. Winter, Phys. Rev. Lett. **99**, 180502 (2007).
[8] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, Nature (London) **461**, 1101 (2009).
[9] M. L. Almeida, J.-D. Bancal, N. Brunner, A. Acín, N. Gisin, and S. Pironio, Phys. Rev. Lett. **104**, 230404 (2010).
[10] J. Allcock, N. Brunner, M. Pawlowski, and V. Scarani, Phys. Rev. A **80**, 040103 (2009).

[11] $I_c(X:Y) = H(X) + H(Y) - H(XY)$, where $H(X)$ is the classical Shannon entropy $H(\{p_i\}) = -\sum_i p_i \log_2 p_i$.
[12] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[13] A. Short and J. Barrett, New J. Phys. **12**, 033034 (2010).
[14] B. Tsirelson, Lett. Math. Phys. **4**, 93 (1980).
[15] W. van Dam, e-print arXiv:quant-ph/0501159 (2005).
[16] A. Ambainis, D. Leung, L. Mancinska, and M. Ozols, e-print arXiv:0810.2937 (2008).
[17] M. Pawlowski and M. Zukowski, Phys. Rev. A **81**, 042326 (2010).
[18] Specifically for $n = 2^k 3^j$ where j and k are integers.
[19] D. Slepian and J. Wolf, IEEE Trans. Inf. Theory **19**, 471 (1973).
[20] In particular, Alice and Bob are given uniformly random bits x and y, and must output bits a and b such that $a \oplus b = xy$. The Tsirelson bound is $P^Q_{\text{success}} \leqslant \frac{2+\sqrt{2}}{4}$, while the CHSH inequality gives $P^C_{\text{success}} \leqslant \frac{3}{4}$.
[21] $h(P_k) = -P_k \log_2 P_k - (1 - P_k)\log_2(1 - P_k)$.
[22] M. Pawlowski and A. Winter, e-print arXiv:1106.2409 (2011).
[23] We can also prove an analog of their more general bound, where Alice's input has probability distribution $p(\mathbf{x})$, by replacing $|A\rangle$ with the unnormalized state $\sqrt{2^n}\sum_{\mathbf{x}} p(\mathbf{x})(-1)^{\hat{a}_{\mathbf{x}}}|\psi\rangle \otimes |\mathbf{x}\rangle$. This gives $\sum_{\mathbf{y}} E^2_{\mathbf{y}} \leqslant 2^n \sum_{\mathbf{x}} p(\mathbf{x})^2$.
[24] B. Tsirelson, J. Sov. Math. **36**, 557 (1987).
[25] L. Hardy, e-print arXiv:quant-ph/0101012 (2001).
[26] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
[27] M. Pawlowski (private communication).
[28] A. J. Short and S. Wehner, New J. Phys. **12**, 033023 (2010).
[29] H. Barnum, J. Barrett, L. O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, New J. Phys. **12**, 033024 (2010).
[30] G. Kimura, K. Nuida, and H. Imai, Rep. Math. Phys. **66**, 175 (2010).
[31] Of course, we could consider a restriction of quantum theory which would share the von-Neumann entropy, but the interesting question is to consider theories which cannot be simulated by quantum theory.
[32] O. C. O. Dahlsten, D. Lercher, and R. Renner, e-print arXiv:1108.4549 (2011).