Entropic inequalities as a necessary and sufficient condition to noncontextuality and locality

Rafael Chaves

Institute for Physics, University of Freiburg, Rheinstrasse 10, D-79104 Freiburg, Germany

The assumption of local realism, in a Bell locality scenario, imposes nontrivial conditions on the Shannon entropies of the associated probability distributions, expressed by linear entropic Bell inequalities. In principle, these entropic inequalities provide necessary but not sufficient criteria for the existence of a local hidden variable model reproducing the correlations, as, for example, the paradigmatic nonlocal Popescu-Rohrlich (PR) box is entropically not different from a classically correlated box. In this paper we show that for the n-cycle scenario, entropic inequalities completely characterize the set of local correlations. In particular, every nonsignaling box which violates the Clauser-Horne-Shimony-Holt (CHSH) inequality—including the PR box—can be locally modified so that it also violates the entropic version of CHSH inequality. As we show, any nonlocal probabilistic model when appropriately mixed with a local model, violates an entropic inequality, thus evidencing a very peculiar kind of nonlocality. As the n-cycle captures equally well both the notion of local realism introduced by Bell and that of noncontextuality presented by the Kochen-Specker theorem, the results are also valid for noncontextuality scenarios.

PACS number(s): 03.65.Ud, 03.67.Mn

I. INTRODUCTION

The quantum nonlocal correlations that may arise in experiments performed by spacelike separated and independent observers, are a key concept in the foundational aspects of quantum mechanics. The expected classical intuition that physical quantities have well-established values previous to any measurement and that signals cannot propagate instantaneously, do not suffice to reproduce the quantum mechanical predictions [1], highlighting a very counterintuitive aspect of quantum theory that has received strong experimental corroboration over the years [2]. From an applied point of view, nonlocality is now recognized as a novel physical resource, which enables protocols such as device-independent quantum key distribution [3], random number generation [4], and the reduction of communication complexity in distributed-computing scenarios [5].

In practice, nonlocality is witnessed through the violation of a *Bell inequality* [1]. Given a certain experimental scenario defined by the number of spatially separated parties, the possible different measurement settings for each party, and the possible outcomes for each setting, local-realistic joint probability distributions form a convex set to which Bell inequalities, a set of linear inequalities of the probabilities, are the nontrivial facets [6]. This geometric approach provides a general framework in which Bell inequalities can be derived, since the task to find the facets of a convex set is a linear program that can be solved efficiently. The problem is that, generally, the size of the linear program grows very fast as the nonlocality scenario becomes less simple, some classes even being known to be a nondeterministic polynomial (NP)-complete problem [6]. In spite of that, some particular characterizations are well known. The bipartite scenario with two dichotomic measurements per party is completely characterized by the Clauser-Horne-Shimony-Holt (CHSH) inequality [7] and a generalization of CHSH to more outcomes is provided by the Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality [8], the CHSH and CGLMP inequalities fully describing the set of local correlations up to

three outcomes [9]. However, for a number of outcomes larger than three, a complete characterization of the inequalities bounding the set of local correlations is still to be found [9,10], highlighting the difficulty and limitations of this approach.

In a conceptually different approach introduced by Braunstein and Caves [11], it was shown that local realism imposes nontrivial conditions already on the level of the Shannon entropies. The Shannon entropies carried by the measurements on two distant systems must satisfy certain inequalities, which can be regarded as *entropic Bell inequalities*. It was recently pointed out that the characterization of the local correlations on the entropic level also defines a linear programming problem [12,13]. One advantage of this entropic approach is that it can readily be applied to quantum systems of arbitrary local dimension and general measurement operators, since the inequalities do not depend on the number of outcomes of the measured observables. That is, while the dimension of the set of local correlations in terms of probabilities grows exponentially with the number of outcomes for each observable, the entropic set of local correlations is independent of that. Another advantage of the entropic approach is that it easily adapts to situations of additional independence requirements, like the bilocality scenarios introduced by a entanglement swapping experiment [14,15] and general correlation and causal model scenarios [16–18]. The independence constraints are nonlinear on the level of probabilities, defining a nonconvex set, while in terms of entropies such constraints are linear and still define a convex set that can be solved by linear programming. In spite of their attractive properties, entropic Bell inequalities are, in principle, sufficient but not necessary conditions to witness nonlocality. That is, there are nonlocal distributions violating a Bell inequality, that, however, do not violate its entropic counterpart [13]. However, as we show in this paper, the situation is more involved than initially thought, as entropic inequalities can, at least in some scenarios, completely characterize the set of local correlations.

In this paper we show that in the n-cycle scenario, any nonlocal distribution when augmented with shared randomness will also violate an entropic Bell inequality. The n-cycle can
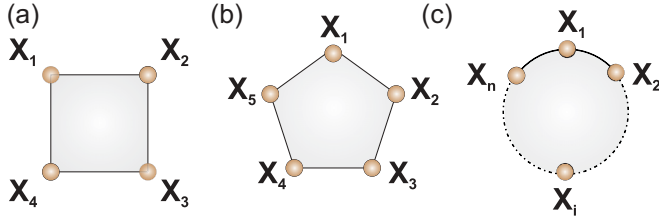
FIG. 1. (Color online) Graphical representation of the n-cycle. The vertices represent different observables and the edges connect observables that are jointly measurable. (a) CHSH scenario, two parties with two measurement settings each. Labeling $X_1 = A_0$, $X_3 = A_1$, $X_2 = B_0$, $X_4 = B_1$, we recover the usual picture where two parties, Alice and Bob, perform spacelike separated measurements. Alice measures one out of two possible measurement settings A_0 or A_1, and similarly for Bob. (b) KCBS scenario, five observables arranged in a cyclic configuration such that each observable is compatible with its neighbors. (c) Generalization of the CHSH/KCBS scenario, with n observables in a cyclic configuration (the n-cycle [22]). For dichotomic observables, the set of local and noncontextual correlations is completely characterized by Eq. (4). For a general number of outcomes, the unique nontrivial entropic inequalities are those given by Eq. (5).

be seen as generalization of the CHSH scenario for an arbitrary number of observables for each party, with only a subset of pairwise observables being jointly measurable (Fig. 1). The n-cycle captures equally well both the notion of local realism introduced by Bell [1] and that of noncontextuality presented by the Kochen-Specker (KS) theorem [19]. A complete characterization of the n-cycle in terms of an exponential number of tight, linear inequalities has recently been found [20]. As we show here, the n-cycle can be equivalently described by a polynomial number of entropic inequalities and a list of local or noncontextual points lying in the facets of the corresponding set of correlations.

II. THE n-CYCLE SCENARIO

The n-cycle scenario is defined for any number $n \geqslant 3$ of observables $X_1, \ldots, X_n$, imposing the restriction that only X_i and X_{i+1} are pairwise jointly measurable for all $i = 1, \ldots, n$ (with $X_{n+k} = X_k$). Any two observables X_i and X_{i+1} are jointly measurable, or compatible, if the result for the measurement of X_i, even if not performed, does not depend on the prior or simultaneous measurement of X_{i+1} and vice versa. This is the notion of noncontextuality captured by the KS theorem [19]. It can readily be turned into the notion of Bell's locality [1], where the compatibility and noncontextuality of the observables is assured by spacelike separation of local measurements of different particles. In particular, for $n = 4$, the n-cycle corresponds to the CHSH Bell scenario [7] [Fig. 1(a)], while for $n = 5$ it is the noncontextuality scenario considered by Klyachko-Can-Binicioglu-Shumovsky (KCBS) [21] [Fig. 1(b)]. For general n, it can be visualized as an n-sided polygon [Fig. 1(c)].

Any correlation that can be reproduced by means of a noncontextual or local hidden variable model is a convex sum

of the deterministic probability distributions, that is,

$$p(x_i x_{i+1} | X_i X_{i+1}) = \sum_\lambda \varrho(\lambda) p(x_i | X_i, \lambda) p(x_{i+1} | X_{i+1}, \lambda), \tag{1}$$

where x_i stands for the outcome $x_i = 0, \ldots, d - 1$ of the corresponding observable X_i. The sum is performed over all the d^n deterministic distributions parameterized by λ [with a distribution $\varrho(\lambda)$], that is, all the distributions that given a certain observable X_i yield with probability 1 a certain outcome x_i. The noncontextual and local deterministic points define a convex set, to which the noncontextuality and Bell inequalities are the nontrivial boundaries.

Similarly, the set of allowed nondisturbing and nonsignaling distributions also define a convex set. Nondisturbing and nonsignaling distributions are defined as the ones for which

$$p(x_i | X_i) = \sum_{x_{i+1}} p(x_i x_{i+1} | X_i X_{i+1}) = \sum_{x_{i-1}} p(x_i x_{i-1} | X_i X_{i-1}), \tag{2}$$

that is, the outcome of X_i cannot be affected by which compatible observable (X_{i-1} or X_{i+1}) it is jointly measured with. For dichotomic observables $x_i = 0, 1$ the vertices of the nondisturbing and nonsignaling set are given by [20]

$$p_{\max}^\gamma(x_i x_{i+1} | X_i X_{i+1}) = \begin{cases} 1/2, & x_i \oplus x_{i+1} = \delta_{-1,\gamma_i} \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $\oplus$ stands for addition modulo 2, $\gamma = \{\gamma_1, \ldots, \gamma_n\}$, $\gamma_i = \{-1, 1\}$ and the total number of $\gamma_i = -1$ is odd. Note that for $n = 4$ and $\gamma = \{1, 1, -1, 1\}$ this represents the maximally nonlocal distribution allowed by nonsignaling, the so-called PR box [23]. Given a certain nondisturbing and nonsignaling probability distribution, any other distribution than can be achieved from it using local reversible transformations is said to be equivalent. Local reversible operations consist of operations of two types: relabeling of the observables that preserve the mutual compatibility of the joint observables $X_i X_{i+1}$, e.g, $X_i \to X_{i+2}$, or relabeling of the outputs (possibly conditioned on the observable), e.g., $x_i \to x_i \oplus \delta_{i,1}$.

Recently, the full characterization of the n-cycle has been found for dichotomic observables [20]. There are 2^{n-1} tight equivalent inequalities given by

$$C_n^\gamma = \sum_i \gamma_i \langle X_i X_{i+1} \rangle \leqslant (n - 2), \tag{4}$$

where $\langle X_i X_j \rangle = p(00 | X_i X_j) + p(11 | X_i X_j) - p(01 | X_i X_j) - p(10 | X_i X_j)$ stands for the expectation value of the observable $X_i X_j$ and again the total number of $\gamma_i = -1$ is odd. A noncontextual or Bell inequality is said to be equivalent to another one, if it can be obtained from it by local reversible operations and/or permutation of observables. Note that $p_{\max}^\gamma$ maximally violates C_n^γ, achieving $C_n^\gamma(p_{\max}^\gamma) = n$, while not violating any other equivalent inequality. In the following we will generally refer to C_n and $p_{\max}$ as the ones with $\gamma_i = 1$ for $i = 1, \ldots, n - 1$ but $\gamma_n = -1$.

Similarly, the complete entropic characterization of the noncontextual and local set of probability distributions has been found for the n-cycle [12,13] (see also Ref. [24] for

the $n = 5$ case). A probability distribution in this scenario is entropically noncontextual or local if and only if the set of n equivalent Braunstein-Caves (BC) entropic inequalities,

$$BC_n^k = H(X_k X_{k+1}) + \sum_{j \neq k,\, k+1} H(X_j)$$
$$- \sum_{j \neq k} H(X_j X_{j+1}) \leqslant 0, \qquad (5)$$

hold for all $k = 1, \ldots, n$ where $H(X_i X_j) = \sum_{x_i, x_j} -p(x_i x_j | X_i X_j) \log_2 p(x_i x_j | X_i X_j)$ is the Shannon entropy of the probability distribution associated with the measurements X_i and X_j. This set of entropic inequalities is said to be maximal in the sense that no other entropic inequality can detect the contextuality and nonlocality not detectable by it, so that this complete set of tight entropic inequalities completely characterizes the region of noncontextual and local probabilistic models in entropy space.

In order to transform an entropic inequality into an equivalent one, the only symmetry operations that can be applied are permutations of the observables. It is a basic feature of Shannon entropy that the entropy of a probability distribution is invariant under permutations of the sample space. Violations of BC inequalities witness then a very peculiar kind of contextuality and nonlocality. If a probabilistic distribution violates the BC inequalities, then so does every other distribution obtained by permuting the outcome probabilities, provided that the permutation leads to the same marginal distributions. This leads to the following phenomenon: From the point of view of the entropic inequality BC_n^k, the maximally contextual and nonlocal distribution $p_{\max}^\gamma$ is not different from a classically correlated distribution,

$$p_C^{\gamma'}(x_i x_{i+1} | X_i X_{i+1}) = \begin{cases} 1/2, & x_i \oplus x_{i+1} = \delta_{-1,\gamma_i'} \\ 0, & \text{otherwise} \end{cases}, \qquad (6)$$

where $\gamma_i' = \gamma_i$ for $i \neq k$ but $\gamma_k' = -\gamma_k$. Note that $p_C^{\gamma'}$ has an even total number of $\gamma_i' = -1$ and does not violate any C_n^γ inequality. Thus, since $p_{\max}^\gamma$ is entropically equivalent to a noncontextual and local distribution, it does not violate any entropic inequality. In this sense, an entropic noncontextual or Bell inequality is a necessary but not sufficient criterion to probe the noncontextual and local behavior a distribution. However, as we show next, entropic inequalities can be turned into a necessary and sufficient condition, since any contextual or nonlocal distribution violating C_n^γ also violates the entropic BC inequalities when properly mixed with a noncontextual and local distribution.

III. ENTROPIC INEQUALITIES COMPLETELY CHARACTERIZE THE *n*-CYCLE SCENARIO WITH DICHOTOMIC OUTCOMES

First, note that the maximum violation of the BC inequality for dichotomic observables is given by $BC_n^k = 1$, since $H(X_j) \leqslant H(X_j X_{j+1})$ and $H(X_k X_{k+1}) \leqslant H(X_k) + H(X_{k+1}) \leqslant H(X_{k+n-1} X_k) + H(X_{k+1}) \leqslant H(X_{k+n-1} X_k) + 1$. The maximal violation of $BC_n^k = 1$ can be achieved by the

probability distribution,

$$p_{E\max}^\gamma = \tfrac{1}{2}\big(p_{\max}^\gamma + p_C^{\gamma'}\big), \qquad (7)$$

for all γ such that $\gamma_k = -1$, since $H(X_j) = H(X_j X_{j+1}) = 1$ for all $j \neq k$ and $H(X_k X_{k+1}) = 2$. That is, a convex combination of two nonviolating distributions may violate an entropic inequality, highlighting the strongly nonlinear character of it. Note that $p_C^{\gamma'}$ can be achieved with 1 bit of shared randomness, that is, $x_i = x_{i+1} = 0$ or $x_i = x_{i+1} = 1$ with the same probability $1/2$ for all $i = 1, \ldots, n$. Similarly the convex combination in (7) also requires 1 bit of shared randomness, outputting $p_{\max}^\gamma$ or $p_C^{\gamma'}$, both with probability $1/2$. When augmented with some shared randomness, it is possible to turn an entropically noncontextual and local distribution into a contextual and nonlocal one. In a noncontextual or Bell scenario, shared randomness is always an available and valid resource. However, for usual noncontextual or Bell inequalities, the mixing with a local point cannot improve the violation of the inequality due to its linearity.

As we show next, the mixing with $p_C^{\gamma'}$ is sufficient to entropically detect any contextual or nonlocal distribution. First we note that in the CHSH scenario ($n = 4$ in the n-cycle), it is known that any probability distribution can be transformed into an isotropic distribution through a local depolarization process, keeping the C_4 value invariant [25]. The isotropic distribution has the property of being invariant under the interchange of the inputs or outputs and being locally unbiased, that is, $p_I(x_i x_{i+1} | X_i X_{i+1}) = p_I(x_{i+1} x_i | X_i X_{i+1})$ and $p_I(x_i | X_i) = 1/2$. For the n-cycle, the isotropic distribution p_I corresponds to a probabilistic mixture,

$$p_I = \epsilon p_{\max} + (1 - \epsilon) p_w, \qquad (8)$$

with $p_w(x_i x_{i+1} | X_i X_{i+1}) = 1/4$ being pure white noise. The corresponding C_n value is given by $C_n(p_I) = n\epsilon$, that is, p_I is contextual and nonlocal for $\epsilon > (n-2)/n$. From the entropic point of view p_I is equivalent to the distribution $\epsilon p_C + (1 - \epsilon) p_w$ and thus no direct violation of entropic BC inequalities is possible.

The first step in our proof is to show that any distribution in the n-cycle scenario can be turned into a isotropic one without changing the values of C_n, that is, a generalization of the depolarization protocol devised in Ref. [25] (see also Ref. [26]) for the $n = 4$ case. Then, for our purposes and to simplify the presentation it is enough to consider isotropic boxes only (however as shown in the Appendix this is not strictly necessary). The depolarization procedure is done in two steps:

(i) $p(x_i x_j | X_i X_j)$ is made locally unbiased by flipping both outputs simultaneously with probability $1/2$, that is, $x_i \to x_i \oplus 1$ and $x_j \to x_j \oplus 1$.

(ii) $p(x_i x_j | X_i X_j) \to (1/n) \sum_{k=0,\ldots,n-1} p(\bar{x}_i^k \bar{x}_j^k | X_{i+k} X_{j+k})$, where $\bar{x}_i^k$ means flipping the output if $i \in \{1, \ldots, n-k\}$. After the second step the initial distribution is in the isotropic form, however, maintaining the value of C_n unchanged.

Computing the BC value for the isotropic distribution mixed with the classically correlated box, that is, $v p_I + (1 - v) p_C$, expanding around $v = 0$ we find that

$$\text{BC}_n = \frac{v}{\ln 4} \{ f(n,\epsilon) - [2 - n(1-\epsilon)] \ln v \}, \qquad (9)$$

with $f(n,\epsilon) = 2 - n(1-\epsilon)(1 + \ln 2) + (n + \epsilon - n\epsilon) \ln(1 - \epsilon) - \epsilon \ln(1 + \epsilon) + \ln[4/(1 - \epsilon^2)]$. For any $2 - n(1 - \epsilon) > 0$ taking a sufficiently small v ensures BC to be positive since $f(n,\epsilon)$ does not depend on v. That is, for any nonlocal isotropic distribution $\epsilon > (n-2)/n$ the BC inequality can be violated. This is the same bound obtained by the direct calculation of $C_n(p_\mathrm{I})$. Given that any distribution can be turned into the isotropic box without changing its C_n value, this means that *any contextual and nonlocal distribution in the n-cycle scenario also violates the entropic* BC_n *inequality when properly mixed with a classically correlated and noncontextual and local distribution.*

We note that we have analyzed the specific case of a distribution violating a specific equivalence of C_n^γ where all $\gamma_i = 1$ but $\gamma_n = -1$. However, the generalization for distributions violating other equivalences of C_n^γ is straightforward. Since there is a one-to-one correspondence between the vertices of the nondisturbing and nonsignaling set and each of the facets of the noncontextual and local set, if a distribution violates C_n^γ, it can be brought to the isotropic form $p_\mathrm{I}^\gamma = p_\mathrm{max}^\gamma + p_\mathrm{w}$ again without changing the value of C_n^γ. Mixing p_I^γ with the corresponding classical correlation $p_\mathrm{C}^{\gamma'}$ will lead to violations of BC_n^k and therefore to the same conclusions as before. It is important to note that the characterization of the noncontextual and local set is achieved with n entropic inequalities, as opposed to the 2^{n-1} linear inequalities (4), however, at the cost of introducing a list of classical points $p_\mathrm{C}^{\gamma'}$.

IV. CONCLUSION

In principle, entropic inequalities only provide a necessary but not sufficient criterion for noncontextuality and local realism. However, we have shown that for the n-cycle with dichotomic outcomes, entropic inequalities turn also to be sufficient, since any contextual and nonlocal probabilistic model will display entropic violations if properly mixed with a classical model. It is quite surprising that a polynomial number of nonlinear and nontight inequalities may completely characterize the set of noncontextual and local correlations, that otherwise would require an exponential number of linear and tight inequalities to do so.

One obvious question is how this result would extend for more complex scenarios, involving more than two outcomes and possibly more parties as in a multipartite Bell test. Even in the bipartite case, $n = 4$ for the n-cycle, the complete characterization of the local correlations is not known for a number of outcomes larger than three [9,10]. Could it be that the BC entropic inequality augmented with shared randomness fully characterizes the n-cycle for a general number of outputs? Another interesting question is to understand the role of entropic inequalities in the bilocality scenario, where independence constraints define a nonconvex set, difficult to characterize in the probability space [15].

Finally, violations of linear Bell inequalities can be understood as a resource, for instance, allowing for higher probability of success in some information tasks [5]. Is there any operational interpretation for the violation of an entropic inequality in terms of a relevant physical task? If that turns out to be the case, an interesting scenario would arise, where

nonlocal but entropically classical correlations could be turned into a useful resource, being activated by the use of shared randomness.

APPENDIX: VIOLATION OF THE ENTROPIC INEQUALITIES WITHOUT THE DEPOLARIZATION PROCEDURE

We show here that given a general nonlocal probability distribution in the CHSH scenario, the mixing with the classical correlation is sufficient to violate the entropic inequality, without the need of the depolarization procedure.

The local set consists of 16 extremal points $p_\mathrm{det}^{\alpha,\beta,\gamma,\delta}$ parametrized as [27]

$$p_\mathrm{det}(ab|xy) = \begin{cases} 1, & a = \alpha x \oplus \beta, \\ & b = \gamma y \oplus \delta \\ 0, & \mathrm{otherwise} \end{cases} \quad (A1)$$

and all the eight nonlocal extremal points $p_\mathrm{PR}^{\alpha,\beta,\gamma}$ of the nonsignaling set can be parametrized as [27]

$$p_\mathrm{PR}(ab|xy) = \begin{cases} 1/2, & a \oplus b = xy + \alpha x + \beta y + \gamma \\ 0, & \mathrm{otherwise} \end{cases}, \quad (A2)$$

here $\alpha, \beta, \gamma, \delta \in \{0,1\}$. To simplify the description we have employed the common notation to the CHSH scenario, that is, X_1 and X_3 corresponding to $x = 0$ and $x = 1$ and X_2 and X_4 corresponding to $y = 0$ and $y = 1$, while a and b label the corresponding outcomes. For a general distribution written as a convex combination of all extreme points,

$$p(ab|xy) = \sum_{\alpha,\beta,\gamma} \varrho_\mathrm{PR}^{\alpha,\beta,\gamma} p_\mathrm{PR}^{\alpha,\beta,\gamma} + \sum_{\alpha,\beta,\gamma,\delta} \varrho_\mathrm{det}^{\alpha,\beta,\gamma,\delta} p_\mathrm{det}^{\alpha,\beta,\gamma,\delta}, \quad (A3)$$

the condition for the violation of the CHSH inequality is that $C_4 = \sum_{\alpha,\beta,\gamma,\delta} (-1)^{(1-\alpha)(\beta+\delta) + \alpha(\beta+\gamma+\delta)} \varrho_\mathrm{det}^{\alpha,\beta,\gamma,\delta} + 2\varrho_\mathrm{PR}^{0,0,0} - 2\varrho_\mathrm{PR}^{0,0,1} > 1$. Mixing the $p(ab|xy)$ with the classical correlated box as before and expanding around $v = 0$, one finds that $\mathrm{BC}_4 = (v/\ln 4)[g + 2 \ln v(1 - C_4)]$, that violates the entropic inequality for $C_4 > 1$ [g is a function of $p(ab|xy)$ but independent of v].

The same argument can be applied to the general n-cycle, and we have tested up to $n = 7$ that the same result holds. For larger n, the procedure becomes unfeasible given that the number of extremal points increases exponentially, but we conjecture that the same result should hold for any n.

[1] J. Bell, Physics **1**, 195 (1964).
[2] S. J. Freedman and J. F. Clauser, Phys. Rev. Lett. **28**, 938 (1972); A. Aspect, P. Grangier, and G. Roger, *ibid.* **49**, 91 (1982); M. A. Rowe *et al.*, Nature (London) **409**, 791 (2001); D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, Phys. Rev. Lett. **100**, 150404 (2008).
[3] S. Pironio *et al.*, New J. Phys. **11**, 045021 (2009).
[4] S. Pironio *et al.*, Nature (London) **464**, 1021 (2010).
[5] C. Brukner, M. Zukowski, J.-W. Pan, and A. Zeilinger, Phys. Rev. Lett. **92**, 127901 (2004); H. Buhrman, R. Cleve, S. Massar, and R. de Wolf, Rev. Mod. Phys. **82**, 665 (2010).
[6] I. Pitowski, *Quantum Probability, Quantum Logic* (Springer, Heidelberg, 1989).
[7] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[8] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, Phys. Rev. Lett. **88**, 040404 (2002).
[9] L. Masanes, Quantum Inf. Comput. **3**, 345 (2002).
[10] J.-D. Bancal, N. Gisin, and S. Pironio, J. Phys. A **43**, 385303 (2010).
[11] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).
[12] T. Fritz and R. Chaves, IEEE Trans. Inform. Theory **59**, 803 (2013).
[13] R. Chaves and T. Fritz, Phys. Rev. A **85**, 032113 (2012).
[14] M. Zukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, Phys. Rev. Lett. **71**, 4287 (1993).
[15] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, Cambridge, 2009).
[16] C. Branciard, N. Gisin, and S. Pironio, Phys. Rev. Lett. **104**, 170401 (2010).
[17] T. Fritz, New J. Phys. **14**, 103001 (2012).
[18] C. J. Wood and R. W. Spekkens, arXiv:1208.4119.
[19] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).
[20] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, arXiv:1206.3212.
[21] A. A. Klyachko, M. A. Can, S. Binicioglu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).
[22] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Phys. Rep. **506**, 1 (2011).
[23] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[24] P. Kurzyński, R. Ramanathan, and D. Kaszlikowski, Phys. Rev. Lett. **109**, 020404 (2012).
[25] L. Masanes, A. Acin, and N. Gisin, Phys. Rev. A **73**, 012112 (2006).
[26] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Klobus, and A. Wojcik, arXiv:1209.3745.
[27] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).