Quantifying Contextuality

A. Grudka,[1] K. Horodecki,[2,3,*] M. Horodecki,[2,4] P. Horodecki,[2,5] R. Horodecki,[2,4] P. Joshi,[4,2] W. Kłobus,[1] and A. Wójcik[1]
[1]*Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*
[2]*National Quantum Information Center of Gdańsk, 81-824 Sopot, Poland*
[3]*Institute of Informatics, University of Gdańsk, 80-952 Gdańsk, Poland*
[4]*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland*
[5]*Faculty of Applied Physics and Mathematics, Technical University of Gdańsk, 80-233 Gdańsk, Poland*
(Received 9 April 2013; revised manuscript received 13 December 2013; published 26 March 2014)

Contextuality is central to both the foundations of quantum theory and to the novel information processing tasks. Despite some recent proposals, it still faces a fundamental problem: how to quantify its presence? In this work, we provide a universal framework for quantifying contextuality. We conduct two complementary approaches: (i) the bottom-up approach, where we introduce a communication game, which grasps the phenomenon of contextuality in a quantitative manner; (ii) the top-down approach, where we just postulate two measures, relative entropy of contextuality and contextuality cost, analogous to existent measures of nonlocality (a special case of contextuality). We then match the two approaches by showing that the measure emerging from the communication scenario turns out to be equal to the relative entropy of contextuality. Our framework allows for the quantitative, resource-type comparison of completely different games. We give analytical formulas for the proposed measures for some contextual systems, showing in particular that the Peres-Mermin game is by order of magnitude more contextual than that of Klyachko *et al.* Furthermore, we explore properties of these measures such as monotonicity or additivity.

DOI: 10.1103/PhysRevLett.112.120401 PACS numbers: 03.65.Ta, 03.65.Ud

Introduction.—Nonlocality is one of the most interesting manifestations of the quantumness of physical systems [1]. It exhibits the strength of correlations that comes out of a quantum state when measured independently by distant parties that share it, which is sometimes higher than that coming from classical resources, and can be even higher for superquantum but nonsignaling resources [2]. Nonlocality has been formulated in terms of "boxes," i.e., families of probability distribution, and has been studied both qualitatively through Bell inequalities as well as quantitatively through measures of nonlocality such as the cost of nonlocality, distillable nonlocality [2–6], or recently as its (anti)robustness [7].

There is, however, another phenomenon known even earlier than Bell's nonlocality, called quantum contextuality [8]. Namely, for certain sets of observables, some of which may be commensurable, their results could not preexist prior to the measurements, or otherwise, one would obtain a logical contradiction sometimes called the Kochen-Specker paradox [9]. In recent years, this phenomenon has been studied in depth. New examples of Kochen-Specker proofs of contextuality have been found [10–12] (see also Refs. [13,14] and references therein for recent results), and the counterparts of Bell inequalities have been introduced, however, in a state-independent fashion [15], i.e., that are violated by any quantum state (see also state-dependent attempts in Refs [16,17] and [18,19] for more recent achievements). The fact that quantum theory is contextual has been also treated experimentally [20–22]; see also Refs. [23–26] and references therein for recent results. In fact, the phenomenon of nonlocality is a special case of contextuality: the commensurability relations are provided by the fact that observables are measured on separate systems. Yet it is not vice versa: the phenomenon of contextuality is more basic, as it can hold in single partite systems.

Since the discovery of quantum contextuality there has been a basic problem: *How to quantify contextuality?* Only recently there were interesting attempts to quantify contextuality in terms of memory cost [27] and the ratio of contextual assignments [28]. There were also some measures of nonlocality, which is a special case of contextuality such as nonlocality cost [2] and relative entropy of nonlocality [29,30]. In this Letter, we propose a universal framework of quantifying contextuality based on two complementary approaches: (i) the bottom-up approach, where we introduce a communication game, which grasps the phenomenon of contextuality in a quantitative manner; (ii) the top-down approach, where we just postulate two measures, contextuality cost and relative entropy of contextuality, analogous to the above mentioned nonlocality measures. We then match the two approaches by showing that the measure emerging from the communication scenario turns out to be equal to the relative entropy of contextuality. We further study properties of the measures such as faithfulness, additivity, or monotonicity, which are analogous to that of entanglement measures. We also compute it for some systems that possess high symmetries.

The presented approach is fully general, as it applies to both the contextuality of the Peres-Mermin description in which the main objects are observables [10,11] (see

Refs. [31–33] for recent progress) as well as the original Kochen-Specker approach [9], described by the so-called orthogonality graphs (see Refs. [19,34–37] and references therein). In fact, our measure gives the natural quantitative way to classify the contextual systems that differ in underling structure and identify possible structural properties needed to obtain high contextuality.

How to quantify contextuality.—Quantum contextuality clearly manifests that quantum mechanical world which cannot be described by a joint probability distribution over a single probability space: there are systems where statistics of observables (some of which are jointly measurable, form a context) cannot be described by a common joint probability distribution. In other words, joint probability distribution that reproduces statistics of some contexts, see Fig. 1(a), at the same time cannot reproduce statistics of other contexts, see Fig. 1(b). For this reason, if we would like to simulate such a system we need at least two common joint probability distributions—see Fig. 1(c) where each of them has to fail in reproducing statistics of some context. Thus, for contextual systems there are inevitable correlations between the contexts and the common joint probability distributions, whereas for noncontextual systems the "which context information" is inaccessible via the joint probability distribution. We will quantify these correlations by means of mutual information since they vanish if and only if the system is noncontextual. This quantity will be called the mutual information of contextuality (MIC). We further show that it equals another quantity that can be viewed as an analogue of relative entropy of entanglement, which we call "relative entropy of contextuality." We study properties of this measure, showing its additivity for some systems, as well as monotonicity under some set of operations. We then compute it for some known systems, developing a technique of symmetrization. Finally, we introduce the measure called "cost of contextuality" and compute it for some systems.

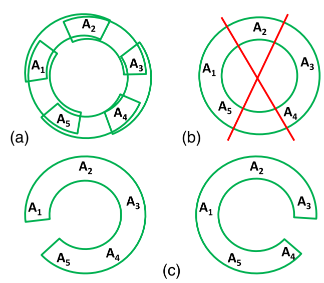

FIG. 1 (color online). Exemplification of contextuality of systems of observables $(A_1, \ldots, A_5)$ (a) Contexts (here neighboring A_i): observables within each context are jointly measurable so that we can ascribe joint probability within context. (b) Ascribing single common joint probability distribution, which has marginals equal to that ascribed in (a) is not possible. (c) Exemplary possible description of the system: by means of two different common joint probability distributions, neither of which reproduces statistics of some context, the left that of A_1, A_5 and the right that of A_3, A_4.

To formalize the above ideas, we consider a set of observables V, some of which are commensurable. Each set of mutually commensurable observables we call a "context" and assign to it a number c. With each context its joint probability distribution over observables that form it is denoted as $g(\lambda_c)$. The set of such contexts $\{g(\lambda_c)\}$ we call a "box." The box is noncontextual if there exists a joint probability distribution $p(\lambda)$ of all observables in V, such that it has marginal distributions on each context c that are equal to $g(\lambda_c)$. Otherwise, we call it contextual.

For illustration, the family of contextual boxes we describe here the so-called chain boxes. The nth chain box, denoted as $\mathrm{CH}_{(n)}$ is based on n dichotomic observables $A_1, A_2, \ldots, A_n$, with the n contexts defined as neighboring pairs of observables $A_i, A_{i+1 \bmod n}$. The distributions of these contexts are fully correlated for all but the last context and fully anticorrelated for the last one, i.e., A_n, A_1 [38]. Note that $\mathrm{CH}_{(4)}$ is the well-known Popescu-Rohrlich (PR) box. The boxes which have only two types of distributions of contexts, equally weighted bit strings with parity 0 and equally weighted bit strings of parity 1 we call XOR boxes. The pair of a set of observables and set of contexts form a hypergraph. The hypergraphs of exemplary XOR boxes [39] that we consider in the Letter are depicted in Fig. 1 in the Supplemental Material [41].

The "which context" game.—To formalize the introduction of the MIC measure, we consider the following game with three persons [Fig. 2]: Alice and Bob (the sender and receiver) and Charlie (adversary). Let the parties preagree on some *a priori* fixed box $B = \{g(\lambda_c)\}$ in hands of Alice. The goal of Alice is to communicate a number of a context c to Bob, through the hands of Charlie. To this end, she chooses the best probability distribution $\{p(c)\}$ and sends c drawn according to it as a challenge to Charlie. Charlie is bounded to do the following: create a distribution $\mathcal{A}_c$ over all variables in V_G, such that it is compatible with $g(\lambda_c)$ on observables that form context c, and send it to Bob. The goal of Charlie is opposite to that of Alice: he wants to diminish the communication of c in this way. Information about context c is given by variable $\mathcal{A}_c$. The amount of correlations between Alice and Bob, given Alice's choice of distribution $\{p(c)\}$ achievable in this game, is

$$I_{\{p(c)\}}(B) \coloneqq \min_{\mathcal{A}_c} I\left(\sum_c p(c)|c\rangle\langle c| \otimes \mathcal{A}_c\right), \tag{1}$$

which is the mutual information of contextuality given *a priori* statistics $\{p(c)\}$ of a box B. We use here Dirac notation only for convenience, meaning a classically correlated system of variables $\mathcal{A}_c$ correlated with register holding value c. Optimizing over strategies of Alice, we obtain the MIC for a box B, i.e., the following quantity:

$$I_{\max}(B) = \sup_{\{p(c)\}} I_{\{p(c)\}}(B). \tag{2}$$

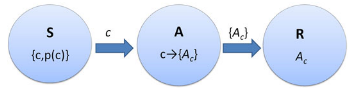

FIG. 2 (color online). The "which context" game. The adversary (A) creates A_c which has context c as that of a chosen box B such that he minimizes communication from sender (S) to receiver (R).

which reports how much correlations Alice and Bob can obtain in this game.

We will argue now that this quantity reports how contextual is box B. Suppose first that B is noncontextual. Then by definition there exists a single joint probability distribution $\mathcal{A}$ over all observables in V_G with marginals $g(\lambda_c)$ on contexts c; hence, $I_{\max}(B) = 0$. However, in the case of contextual box B, by definition Charlie has to use at least two joint probability distributions of all observables in V_G so that on observables of context c, the distribution is $g(\lambda_c)$. Thus, by compactness argument, the value $I_{\max}(B)$ is strictly positive.

(Uniform) relative entropy of contextuality.—We now introduce another measure based directly on the notion of relative entropy distance, in analogy to the measure of nonlocality introduced in Ref. [29]. The first variant, called relative entropy of contextuality, is defined on any box $B = \{g(\lambda_c)\} \in C_G^{(n)}$ as follows:

$$X_{\max}(B) \coloneqq \sup_{p(c)} \min_{\{p(\lambda)\}} \sum_{c\in E_G} p(c) D(g(\lambda_c) \| p(\lambda_c)) \tag{3}$$

where $D(g(\lambda_c)\|p(\lambda_c)) = \sum_i g(\lambda_c)_i \log\{[g(\lambda_c)_i]/[p(\lambda_c)_i]\}$ is the relative entropy distance between distributions $g(\lambda_c)$ and $p(\lambda_c)$ [42,43]. The minimization is taken over all distributions $p(\lambda)$ over $\Omega(A_1) \times \ldots \times \Omega(A_k)$ with marginal distribution on context c equal to $p(\lambda_c)$, and supremum is taken over probability distributions $p(c)$ on the set of numbers of contexts $\{1, \ldots, n\}$.

A natural quantity is also the one which does not distinguish the contexts; i.e., instead of maximization, we set $p(c) = 1/n$ for all c

$$X_u(B) \coloneqq \min_{p(\lambda)} \sum_{c\in E_G} \frac{1}{n} D(g(\lambda_c)\|p(\lambda_c)) \tag{4}$$

where n is number of contexts. We call it the uniform relative entropy of contextuality. By definition, we have $X_{\max} \geq X_u$, but in general these measures are not equal since they differ on direct sum of contextual and noncontextual boxes (see Supplemental Material Section V [41]).

At first, it seems that mutual information of contextuality and relative entropy of contextuality are different, and it is not clear how they are related. Interestingly, one can show that they are equal to each other (see Supplemental Material Theorem 1 [41]), that is,

$$X_{\max} = I_{\max}. \tag{5}$$

We note here that X_u and $X_{\max}$ (and, hence, $I_{\max}$ according to the above result) are faithful.

Analytical formulas.—We calculate now the value of X_u and $X_{\max}$ for the boxes called isotropic XOR boxes. To give an example of isotropic XOR boxes, we consider here the isotropic chain boxes

$$\mathrm{CH}_\alpha^{(n)} = \alpha \mathrm{CH}_{(n)} + (1-\alpha)\mathrm{CH}'_{(n)} \tag{6}$$

where $\mathrm{CH}'_{(n)}$ is the $\mathrm{CH}_{(n)}$ box with correlations and anticorrelations replaced with each other. We just give an idea of how to calculate the (uniform) relative entropy of contextuality for $\mathrm{CH}_{(4)}^{\alpha}$, which is the isotropic Popescu-Rohrlich box denoted as PR_α; the detailed proof for other XOR boxes is shown in the Supplemental Material Sections III and IV [41]. The techniques employed are analogous to those used in entanglement theory, including twirling [44] as well as using symmetries to compute measures based on distance from the set of separable states [45,46], and they were applied in the case of nonlocality, e.g., in Refs. [47,48]. We first compute the value of X_u and then argue that it equals $X_{\max}$ for the isotropic boxes. In order to compute X_u, we observe that for isotropic boxes the minimum in its definition can be taken only over those probability distributions $p(\lambda)$ that give rise to an isotropic box, and $p(\lambda_c)$ is the marginal of $p(\lambda)$ (see the Supplemental Material Theorem 3 [41]).

Let us consider an example of PR_α box (the other examples of isotropic XOR boxes follow similar lines; see the Supplemental Material Section IV [41]), for which

$$X_u(\mathrm{PR}_\alpha) = \min_{p(\lambda)=\mathrm{PR}_{\alpha'}} (1/4) \sum_c D(g(\lambda_c)\|p(\lambda_c)), \tag{7}$$

where $p(\lambda)$ runs over distributions which are from the family of isotropic boxes [47,48] that are noncontextual.

Since any noncontextual box compatible with $G_{\mathrm{CH}}^{(4)}$ has to satisfy the inequality that is equivalent to the CHSH inequality, we have $\frac{1}{4} \leq \alpha' \leq \frac{3}{4}$ (see the Supplemental Material Section IV [41]). The next step is to observe that relative entropy does not change under reversible operations such as in the bit-flip of an output of an observable (see the Supplemental Material Lemma 6 [41]), which gives

$$X_u(\mathrm{PR}_\alpha) = \min_{\frac{1}{4}\leq\alpha'\leq\frac{3}{4}} D(\alpha P_{\mathrm{even}}^{(2)} + (1-\alpha)P_{\mathrm{odd}}^{(2)} \| \alpha' P_{\mathrm{even}}^{(2)} + (1-\alpha')P_{\mathrm{odd}}^{(2)}).$$

Because all isotropic XOR boxes have the above property that $X_u(B_\alpha)$ equals a *single* term of relative entropy no matter how many contexts the box B has, we have that for these boxes $X_{\max} = X_u$ (see Supplemental Material Theorem 7 [41]). It is then easy to show that for $\alpha \geq \frac{3}{4}$ there holds

$$X_{\max}(\mathrm{PR}_\alpha) = X_u(\mathrm{PR}_\alpha) = \log\left(\frac{4}{3^\alpha}\right) - h(\alpha), \qquad (8)$$

where $h(\alpha) = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$ is the binary Shannon entropy. For $\alpha \leq \frac{1}{4}$, $X_u(\mathrm{PR}_\alpha)$ equals the value of $X_u(\mathrm{PR}_{(1-\alpha)})$ according to the above equation. On Fig. 3 we present values of measure X_u for chosen chain boxes $\mathrm{CH}_\alpha^{(n)}$ (quantum ones provided in Ref. [38] and maximally contextual ones).

Comparing contextuality in different scenarios.—Although we have formulated the paradigm of quantifying contextuality using the picture of sets of jointly measurable observables (which is described by a hypergraph) [11], our measures can be applied if additional constrains are imposed, such as mutual exclusiveness, as in the original Kochen-Specker approach. The latter condition means that the observables are binary, and in each context we require that the outcome 1 can occur only for one observable. Quantum mechanically, this is achieved by taking observables to be one-dimensional projectors, and within each context the projectors are mutually orthogonal. In this case, the hypergraph is then equivalently expressed by the so-called orthogonality graph. Our approach allows us now to compare the strength of contextuality coming from those two approaches that at first sight looks quite incomparable. We find that the Peres-Mermin (PM) box [10,11] (belonging to first approach) has $X_{\max} = X_u = \log\frac{6}{5} \approx 0.2630$, while for the Klyachko *et al.* [19] (KCBS) box, that is, based on the original Kohen-Specker approach, the measure turns out to be of order of magnitude smaller $X_{\max} = X_u \approx 0.0467$ (see Supplemental Material subsection V D [41]). This partially reflects the difference of Hilbert dimensionality between the two boxes ($d = 4$ versus $d = 3$) and observables (6 versus 5) but may also suggest the generally more friendly character of Peres-Mermin-type games [10,11] for computational tasks (see also Conclusions below).

Properties of the measures.—One of the most welcome properties of the measure would be its additivity. In Supplemental Material Theorem 9 [41], we show that for families of isotropic XOR boxes X_u and $X_{\max}$ are two-copy additive, i.e., $X_u(B^{\otimes k}) = X_{\max}(B^{\otimes k}) = kX(B)$ for $k = 2$. For boxes that are extremal within the family of isotropic XOR boxes (such as $\mathrm{CH}_{(n)}$, PM, M), X_u and $X_{\max}$ are additive; i.e., the latter statement is true for any natural $k \geq 1$. We conjecture, however, that proposed measures are additive for all isotropic XOR boxes.

Another welcome property would be monotonicity of X_u and $X_{\max}$ under operations which preserve contextuality. We answer partially this question showing in Supplemental Material subsection V A [41] that they are nonincreasing under a natural subclass of contextuality-preserving operations.

The contextuality cost.—Another approach is to base on some known measures of nonlocality and define it properly for all (also one-partite) boxes. This leads us to the *contextuality cost*, which we define as follows:

$$C(B) := \inf\{p \in [0,1] | B = pB_C + (1-p)B_{NC}\} \qquad (9)$$

where infimum is taken over all decompositions of box B into mixture of some noncontextual box B_{NC} and some contextual box B_C. This measure inherits after nonlocality cost the property that it is not increasing under operations that preserve noncontextuality [49]. This holds for the same reason for which the antirobustness of nonlocality is nonincreasing under a class of locality-preserving operations as it is shown in Ref. [7]. We note also that this measure is by definition faithful, and one can easily compute it using linear programming [3]; it is, however, not extensive, i.e., is not proportional to dimension of the system. For the families of isotropic XOR boxes, it can be found analytically, for $\alpha \geq (n-1)/n$, that $C(\mathrm{PM}_\alpha) = 6\alpha - 5$, $C(M_\alpha) = 5\alpha - 4$ and $C(\mathrm{CH}_{(n)}^\alpha) = n\alpha - (n-1)$ (in the same way as shown in Ref. [50] that $C(\mathrm{PR}_\alpha) = 4\alpha - 3$).

Conclusions.—We have proposed a universal framework to quantify contextuality. In particular, we have introduced measures of state-dependent and independent contextuality that are valid for both the single and many-party scenarios. It allows us to compare quantitatively completely different contextual boxes and can be explicitly calculated, showing in particular that the Peres-Mermin box is as a resource significantly more contextual than the Klyachko *et al.* one. Our measure is defined in an information-theoretic manner; hence, it would be interesting to investigate possible relationships between the measure and entropic tests of contextuality put forward in Refs. [51,52] (which have their roots in entropic Bell inequalities [53]). In the context of recent connection of contextuality and quantum speed up [54] our findings may help to quantify the presence of contextual measurements in one-way quantum computation and their relation to quantum speed up in such a model.

Note that our approach can be developed in a few different ways. First, one can define analogous measures to X_u and $X_{\max}$ setting variational distance in place of relative entropy. One can also consider a measure defined as $\min_{\mathcal{A}_c}\sup_{p(c)} I(\sum_c p(c)|c\rangle\langle c| \otimes \mathcal{A}_c)$, i.e., with changed order of min and sup in Eq. (2) which for nonlocal boxes has been studied in Ref. [29]. This measure has more communicational meaning than $X_{\max}$; it is minimal capacity

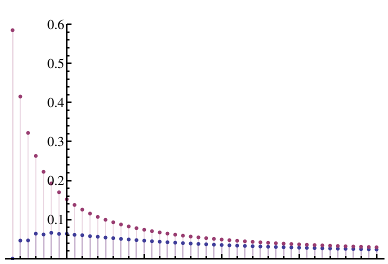

FIG. 3 (color online). Values of measure X_u for $\mathrm{CH}_\alpha^{(n)}$ boxes for $3 \leq n \leq 50$: maximally contextual boxes (upper points, $\alpha = 1$); maximally contextual quantum boxes (lower points) with (i) odd n, $\alpha = [2\cos(\pi/n)]/[1+\cos(\pi/n)]$ and (ii) even n, $\alpha = [1+\cos(\pi/n)]/2$.

of the channel from sender to receiver under adversary's attack. Note that another way of defining relative entropy of contextuality would be to consider a quantity defined on a box B compatible with graph G as $X^*(B) := \inf_{B_{NC} \in NC_G} D(B \| B_{NC})$, where D denotes relative entropy of the boxes B and B_{NC} defined operationally via distinguishability of box B from box B_{NC} in Ref. [55]. It would be interesting to relate such a defined measure with $X_{\max}$ and X_u. Note also that following Ref. [7] it is easy to define and study the notion of (anti)robustness of contextuality.

We thank Marco T. Quintino for helpful comments. A.G. and A.W. thank Paweł Kurzyński for useful discussion. This work is supported by ERC Grant No. QOLAPS and EC Grant No. QUASAR. K.H. also acknowledges Grant No. BMN 538-5300-0975-12. P.J. is supported by Grant No. MPD/2009-3/4 from the Foundation for Polish Science. Part of this work was done in the National Quantum Information Centre of Gdańsk.

*khorodec@inf.ug.edu.pl; writepank@gmail.com

[1] J. S. Bell, Physics **1**, 195 (1964).
[2] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[3] N. Brunner, D. Cavalcanti, A. Salles, and P. Skrzypczyk, Phys. Rev. Lett. **106**, 020402 (2011).
[4] N. Brunner and P. Skrzypczyk, Phys. Rev. Lett. **102**, 160403 (2009).
[5] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vertesi, Phys. Rev. A **80**, 062107 (2009).
[6] M. Forster, Phys. Rev. A **83**, 062114 (2011).
[7] P. Joshi, A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, Quantum Inf. Comput. **13**, 567 (2013).
[8] E. Specker, Dialectica **14**, 239 (1960).
[9] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).
[10] A. Peres, Phys. Lett. A **151**, 107 (1990).
[11] N. D. Mermin, Phys. Rev. Lett. **65**, 3373 (1990).
[12] D. Mermin, Rev. Mod. Phys. **65**, 803 (1993).
[13] S. Yu and C. H. Oh, Phys. Rev. Lett. **108**, 030402 (2012).
[14] A. Cabello, arXiv:1201.0374.
[15] A. Cabello, Phys. Rev. Lett. **101**, 210401 (2008).
[16] A. Cabello, S. Filipp, H. Rauch, and Y. Hasegawa, Phys. Rev. Lett. **100**, 130404 (2008).
[17] Y. Nambu, arXiv:0805.3398.
[18] A. Klyachko, arXiv:quant-ph/0206012.
[19] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).
[20] Y.-F. Huang, C.-F. Li, Y.-S. Zhang, J.-W. Pan, and G.-C. Guo, Phys. Rev. Lett. **90**, 250401 (2003).
[21] G. Kirchmair, F. Zhringer, R. Gerritsma, M. Kleinmann, O. Ghne, A. Cabello, R. Blatt, and C. F. Roos, Nature (London) **460**, 494 (2009).
[22] H. Bartosik, J. Klepp, C. Schmitzer, S. Sponar, A. Cabello, H. Rauch, and Y. Hasegawa, Phys. Rev. Lett. **103**, 040403 (2009).
[23] E. Amselem, L. E. Danielsen, A. J. Lopez-Tarrida, J. R. Portillo, M. Bourennane, and A. Cabello, Phys. Rev. Lett. **108**, 200405 (2012).
[24] C. Zu, Y.-X. Wang, D.-L. Deng, X.-Y. Chang, K. Liu, P.-Y. Hou, H.-X. Yang, and L.-M. Duan, Phys. Rev. Lett. **109**, 150401 (2012).
[25] V. D'Ambrosio, I. Herbauts, E. Amselem, E. Nagali, M. Bourennane, F. Sciarrino, and A. Cabello, Phys. Rev. X **3**, 011012 (2013).
[26] X.-Y. Pan, G.-Q. Liu, Y.-C. Chang, and H. Fan, arXiv:1209.2901.
[27] M. Kleinmann, O. Guhne, J. R. Portillo, J.-A. Larsson, and A. Cabello, New J. Phys. **13**, 113011 (2011).
[28] K. Svozil, arXiv:1103.3980.
[29] W. van Dam, R. Gilles, and P. Grünwald, IEEE Trans. Inf. Theory **51**, 2812 (2005).
[30] F. Brandao (private communication).
[31] M. Waegell and P. K. Aravind, Phys. Rev. A **88**, 012102 (2013).
[32] M. Waegell and P. Aravind, Phys. Lett. A **377**, 546 (2013).
[33] M. Waegell and P. K. Aravind, J. Phys. A **45**, 405301 (2012).
[34] P. Lisonek, P. Badziag, J. R. Portillo, and A. Cabello, arXiv:1308.6012.
[35] A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993).
[36] A. Cabello, S. Severini, and A. Winter, arXiv:1010.2163.
[37] A. Cabello, J. M. Estebaranz, and G. G. Alcaine, Phys. Lett. A **212**, 183 (1996).
[38] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, Phys. Rev. A **88**, 022118 (2013).
[39] Note that XOR boxes are those which win maximally the XOR games, i.e., such games for which payoffs are only the functions of XOR of the output [40].
[40] J. Oppenheim and S. Wehner, Science **330**, 1072 (2010).
[41] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.112.120401 for the proof of relation between Mutual Information of Contextuality and Relative Entropy of Contextuality (REC) and also for numerical calculations of REC for some XOR boxes and derivations of properties of REC.
[42] T. M. Cover and J. A. T., *Elements of Information Theory* (Wiley, New York, 1991).
[43] All logarithms in this Letter are binary.
[44] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[45] V. Vedral and M. B. Plenio, Phys. Rev. A **57**, 1619 (1998).
[46] E. M. Rains, Phys. Rev. A **60**, 179 (1999).
[47] L. Masanes, A. Acin, and N. Gisin, Phys. Rev. A **73**, 012112 (2006).
[48] A. J. Short, Phys. Rev. Lett. **102**, 180502 (2009).
[49] By preserving noncontextuality we mean the operations satisfying the following axioms: (i) transform boxes into boxes, (ii) are linear, (iii) preserve consistency, and (iv) transform noncontextual boxes into noncontextual ones.
[50] K. Horodecki, arXiv:1401.4899.
[51] P. Kurzyński, R. Ramanathan, and D. Kaszlikowski, Phys. Rev. Lett. **109**, 020404 (2012).
[52] R. Chaves and T. Fritz, Phys. Rev. A **85**, 032113 (2012).
[53] S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).
[54] R. Raussendorf, Phys. Rev. A **88**, 022322 (2013).
[55] J. Short and S. Wehner, New J. Phys. **12**, 033023 (2010).