

**Axiomatic approach to contextuality and nonlocality**Karol Horodecki,<sup>1,2</sup> Andrzej Grudka,<sup>3</sup> Pankaj Joshi,<sup>1,4</sup> Waldemar Kłobus,<sup>3</sup> and Justyna Łodyga<sup>3</sup><sup>1</sup>*National Quantum Information Center of Gdańsk, 81-824 Sopot, Poland*<sup>2</sup>*Institute of Informatics, University of Gdańsk, 80-952 Gdańsk, Poland*<sup>3</sup>*Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*<sup>4</sup>*Institute of Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland*

(Received 16 July 2015; published 8 September 2015)

We present a unified axiomatic approach to contextuality and nonlocality based on the fact that both are resource theories. In those theories, the main objects are consistent boxes, which can be transformed by certain operations to achieve certain tasks. The amount of resource is quantified by appropriate measures of the resource. Following a recent paper [J. I. de Vicente, *J. Phys. A: Math. Theor.* **47**, 424017 (2014)], and recent development of abstract approach to resource theories, such as entanglement theory, we propose axioms and welcome properties for operations and measures of resources. As one of the axioms of the measure we propose the asymptotic continuity: the measure should not differ on boxes that are close to each other by more than the distance with a factor depending logarithmically on the dimension of the boxes. We prove that relative entropy of contextuality is asymptotically continuous. Considering another concept from entanglement theory (the convex roof of a measure), we prove that for some nonlocal and contextual polytopes, the relative entropy of a resource is upper bounded up to a constant factor by the cost of the resource. Finally, we prove that providing a measure  $X$  of resource does not increase under allowed class of operations such as, e.g., wirings, the maximal distillable resource which can be obtained by these operations is bounded from above by the value of  $X$  up to a constant factor. We show explicitly which axioms are used in the proofs of presented results, so that analogous results may remain true in other resource theories with analogous axioms. We also make use of the known distillation protocol of bipartite nonlocality to show how contextual resources can be distilled.

DOI: [10.1103/PhysRevA.92.032104](https://doi.org/10.1103/PhysRevA.92.032104)

PACS number(s): 03.65.Ud

**I. INTRODUCTION**

Quantum contextuality stands among the most expressive manifestations of nonclassicality in quantum mechanics [1,2]. In recent years, it has attracted much attention and has been a topic of extensive studies [3–6]. Apart from the interest focused on fundamental concepts, quantum contextuality has been associated with fast computing [7] and quantum information processing [8], which opens a path to possible application of contextual resources in different scenarios.

A particular example of contextuality in the framework where two or more spatially separated parties perform measurements on each subsystem has been termed nonlocality [9]. A lot of effort has been devoted to classifying and quantifying the nonlocality which was identified as a useful resource in the device-independent quantum information processing (see also Sec. IV of Ref. [9]). Although different in nature, nonlocal correlations as well as quantum entanglement proved useful in information processing tasks which cannot be performed with the sole use of classical correlations. This in turn led to the development of the resource theories of entanglement [10] and nonlocality [11].

While approaching to the formulation of a resource theory, three basic ingredients need to be considered. First, the concept of a resource needs to be developed, and showed that it is *useful* regarding some specific tasks, which remain unattainable while having only nonresource objects at disposal. Second, there must be operations by which one may transform resources into one another. Third, one needs to have tools to compare different objects by means of measuring the resource contained by them, namely, a measure of the resource.

Only recently, a theory of resources has been formalized with respect to nonlocal resources [11], steering resources [12],

as well as a general abstract characterization of resource theories has been formulated [13], which captures all needed features and relations that they have in common. In this light, we develop the recent theory of nonlocal resources [11] to include the notion of contextuality, the particular manifestation of which is nonlocality. After identifying the contextual system as useful regarding some computational tasks [14], in this paper we treat contextual systems (“boxes”) as resources in a similar way as it has been done with respect to nonlocal resources. In particular, we describe the notion of contextuality and then, based on the resource theory of entanglement, we formulate a set of axioms for the transformations of contextual resources, as well as for measures intended to quantify the value of given resources.

One of the axioms that we explore is asymptotic continuity, a very desired property from the experimental point of view. Suppose we wish to quantify the amount of resource of a box  $B$  using the measure  $X$ . However, the experimental realization of the box produces an imperfect box  $B'$  which is possibly close to  $B$ :

$$\|B - B'\| \leq \epsilon, \quad (1)$$

where  $\|\dots\|$  denotes the *trace distance between two boxes* [15]. For any measure of a resource, we want it to differ not more than the distance between two considered boxes times a constant logarithmic factor of the dimensionality of the box. Asymptotic continuity of a measure  $X$  therefore means that

$$|X(B) - X(B')| \leq \epsilon \log d + f(\epsilon), \quad (2)$$

where  $d$  is dimensionality of the boxes and  $f$  is a function such that  $f(\epsilon) \xrightarrow{\epsilon \rightarrow 0} 0$ . Throughout this paper logarithms are taken to the base 2.

Only recently, the measures of contextuality have been developed, which enables to quantify the amount of contextuality of boxes [16]. In particular, two measures of contextuality have been introduced: *mutual information of contextuality* (MIC) and *relative entropy of contextuality* (REC). As stated earlier, from an experimental point of view, it is mandatory that any measure of a resource, in particular the measure of contextuality, is asymptotically continuous. With this respect, the main result of the paper is proving that the measure MIC fulfills the axiom of asymptotic continuity. Since the measures MIC and REC are equivalent [16], the asymptotic continuity holds also for the latter, relative entropy distance measure.

The next result of the paper is showing that the relative entropy measure of a resource is upper bounded (up to a constant factor) by another measure which has been defined in Ref. [16], the cost of a resource. We then give examples of applications of this result for a class of bipartite boxes with binary inputs and binary outputs (the most nonlocal of which is *PR* box [17,18]), as well as a class of boxes related to contextual  $n$  cycles [19]. Furthermore, we consider distillation of a resource with the regard that the measure of a resource fulfills the axiom of monotonicity (i.e., that the measure does not increase under the set of allowed operations), and show that distillable contextuality is upper bounded (up to a constant factor) by the value of the measure. We also make use of a distillation protocol as originally devised in [20] to show how contextual resources can be distilled. We then analyze the application of the bound with respect to two relative entropy-based measures of contextuality.

## II. AXIOMATIC APPROACH TO RESOURCE THEORY OF CONTEXTUALITY AND NONLOCALITY

We present a framework for a construction of a general resource theory of contextuality, which can be regarded as a development of resource theory of nonlocality as presented in Ref. [11]. After formalizing the notion of contextual resources with the relation to nonlocal resources, which are to be understood as a specific form of the former, we proceed with formulating the axioms for operations on contextual resources, and axioms for the measures intended to quantify the value of a given resource.

### A. Contextual or nonlocal boxes

In the present setting, the objects of interest are the measurement statistics, without any references to what actual measurements are being performed on actual physical systems. At this point, we do not need to assume that the measurements are performed by spatially separated parties, which, in fact, the notion of nonlocality is all about. In the present view, we consider a set of observables  $\mathcal{M}$  that can be performed on any physical system, where a measurement of each  $M_i \in \mathcal{M}$  gives an outcome  $m_i$  with a probability  $p(m_i|M_i)$ . Let us assume that there exist subsets  $\mathcal{M}_i$  of jointly measurable observables. Each such subset we will call a *context* denoted by  $c$ . The joint probability distribution of obtaining the outcomes  $(m_1, m_2, \dots, m_k)$  while measuring the observables  $M_1, M_2, \dots, M_k$  belonging to a context  $c$  we will denote by

$$p(\lambda_c) := p(m_1, m_2, \dots, m_k | M_1, M_2, \dots, M_k), \quad (3)$$

where  $\lambda_c = (m_1, m_2, \dots, m_k)$  such that  $M_1, M_2, \dots, M_k \in c$ . Note that an observable  $M_i$  may belong to several different contexts. A *box* ( $B$ ) is then a set of joint probability distributions  $B = \{p_B(\lambda_c)\}$  for all contexts in  $\mathcal{M}$  (we will omit the subscript  $B$  of  $p_B(\lambda_c)$  when it is not necessary).

There is, however, a significant constraint which must be obeyed by all boxes to be physically realizable, namely the consistency condition, which states that any marginal distributions for observables that belong to different contexts are independent of a chosen measuring context, i.e.

$$\forall_{i,j} \sum_{\lambda_i \setminus \lambda_j \cap \lambda_j} p(\lambda_i) = \sum_{\lambda_j \setminus \lambda_i \cap \lambda_i} p(\lambda_j). \quad (4)$$

The boxes that fulfill the consistency condition we call *consistent boxes*, and the set of all consistent boxes we denote as  $\mathcal{B}$ . Throughout the paper, while referring to the set of boxes we mean exclusively the set of consistent boxes  $\mathcal{B}$ , without an explicit indication.

At this stage, it should be noticed that the term box is more general than the classical probabilistic model for the whole set of observables  $\mathcal{M}$ , which is given by the joint probability distribution  $p(\lambda|\mathcal{M})$ , such that for all contexts

$$p(\lambda_c) = \sum_{\lambda \setminus \lambda_c} p(\lambda|\mathcal{M}). \quad (5)$$

A set of valuable resources from this perspective are those boxes, which *cannot* be modeled classically, i.e., they cannot be described by a single joint probability distribution for all observables. The set of boxes that constitutes a set of resources we denote as  $\mathcal{B}_v$ , while its elements, i.e., particular valuable boxes, we will denote as  $B_v$ . Otherwise, the set of classical boxes, which are useless from the point of view of valuable resources, we denote as  $\mathcal{B}_{nv}$ , whereas its elements by  $B_{nv}$ . Furthermore, it is known that each box  $B_{nv} \in \mathcal{B}_{nv}$  can be decomposed into a mixture of deterministic boxes which constitute the extremal points of the convex polytope that is identified with  $\mathcal{B}_{nv}$ . The extremal points of the set  $\mathcal{B}_{nv}$  we will denote as  $E_{nv}$ , and each box from  $\mathcal{B}_{nv}$  can be written as  $B_{nv} = \sum_i p_i E_{nv}^i$  for a proper distribution  $\{p_i\}$ . Notice that in general the extremal boxes  $E_{nv}$  are not all the extremal vertices of the set of consistent boxes  $\mathcal{B}$ . The extremal boxes of  $\mathcal{B}$  which do not belong to the set  $\mathcal{B}_{nv}$ , and constitute valuable resources, we will denote as  $E_v$ .

Let us now formalize the resource theory approach regarding the notion of contextuality. The crucial aspect of contextual systems is that one cannot ascribe the values for each observable into the physical system prior measurements such that they will obey the observed statistics. Otherwise, one could, in principle, attribute the observed values of all measurements, which would mean that the construction of a probabilistic model  $p(\lambda|\mathcal{M})$  is possible. Therefore, if we regard a set of contextual boxes  $\mathcal{B}_C$  as a set of resources, then the set of noncontextual boxes  $\mathcal{B}_{NC}$  composes of valueless objects  $\mathcal{B}_{nv} \equiv \mathcal{B}_{NC}$ .

Suppose now that we have instances of physical systems composed of two (or in general more than two) subsystems such that each subsystem is being measured by two spatially separated parties, Alice and Bob. Due to spacelike separation, each pair of measurements performed by the two parties

$x \in \mathcal{M}_A, y \in \mathcal{M}_B$  commute, hence give naturally arising contexts  $c = (x, y)$ . From this perspective, the boxes are given by families of probability distributions  $p(\lambda_c) \equiv p(a, b|x, y)$ , where  $\lambda_c = (a, b)$  is a pair of measurement outcomes of measurements  $(x, y)$ . The comprehensive resource theory of nonlocality was presented in Ref. [11], where nonlocal boxes were identified as valuable resources, whereas local behaviors were considered as the boxes from the set  $\mathcal{B}_{nv}$  (see also Ref. [21]).

With respect to nonlocality, the consistency conditions (4) are the strict analogs of nonsignaling conditions: the marginal distributions for one party are independent of the measurement choice made by spatially separated other party. Hence, the consistency conditions assure nonsignaling in the case when measurement context arises from spacelike separation of different parties and in this case the set of consistent boxes is to be understood simply as a set of nonsignaling boxes.

### B. Axioms of box-resource theories

Once we wish to use the boxes for some specific tasks, we need to specify what are the permitted transformations by which we can process the boxes. Regarding the different sets of boxes, which either constitute valuable resources ( $\mathcal{B}_v$ ), or not ( $\mathcal{B}_{nv}$ ), we formulate two axioms that the operations must obey:

(O1) For *general operations*  $\mathcal{T}$ , given by normalization-preserving and consistency-preserving transformations  $\mathcal{T}$  we must have

$$\forall_{B_1 \in \mathcal{B}_1} \exists_{B_2 \in \mathcal{B}_2} \mathcal{T}(B_1) = B_2. \quad (6)$$

Notice that general transformations  $\mathcal{T}$  transform boxes from one set into another without the assumption of the preservation of the dimensionality of the boxes. In particular, some operations  $\mathcal{T}$  may act as changing the number of contexts of a box by adding or removing some observables from the set  $\mathcal{M}$ . In the case when the dimensionality of boxes (the number of inputs with the respective number of outputs) is preserved, then the general operations can be described by a matrix form [22] (provided that the consistency of a transformed box is preserved)

$$\mathcal{T} \equiv \begin{pmatrix} \alpha_{11} \mathcal{T}_{11} & \dots & \alpha_{1|c|} \mathcal{T}_{1|c|} \\ \vdots & \ddots & \vdots \\ \alpha_{|c|1} \mathcal{T}_{|c|1} & \dots & \alpha_{|c||c|} \mathcal{T}_{|c||c|} \end{pmatrix}, \quad (7)$$

where  $\mathcal{T}_{ij}$  are stochastic matrices acting on the vector of a probability distribution  $p(\lambda_j)$  from a box  $B$ ,  $0 \leq \alpha_{ij} \leq 1$ , and  $\sum_j \alpha_{ij} = 1$ .

(O2) For *free operations*  $\mathcal{L}$ , which constitutes a subset of general operations  $\mathcal{T}$ , we must have

$$\forall_{B_1 \in \mathcal{B}_{nv}} \exists_{B_2 \in \mathcal{B}_{nv}} \mathcal{L}(B_1) = B_2. \quad (8)$$

Similarly, as a set of LOCC (local operations and classical communication) in the context of a resource theory of entanglement [10], and a set of WCCPI (wirings and classical communication prior to inputs) in a resource theory of nonlocality [11,23], the set of free operations  $\mathcal{L}$  is composed of those ‘‘given for free’’ operations in a device-independent

information processing, which by themselves cannot produce a resource from the set  $\mathcal{B}_{nv}$ .

An important class of general operations are reversible operations  $\mathcal{R}$ , for which

$$\exists_{R, R^{-1} \in \mathcal{R}} R^{-1}(R(B)) = B. \quad (9)$$

A particular example of operations from the set  $\mathcal{R}$  are relabelings (see [11] for details).

While considering boxes as resources, we arrive at the question as to whether using general operations we can obtain resources that differ quantitatively from the original box. Similarly as in the entanglement theory, we need to specify what are the relations between different resources, i.e., whether we can state quantitatively that a given resource is more valuable than the other. It is therefore desirable to have a measure to quantify the value of the given resource. In the following, we formulate the axioms that a reliable and usable measure for a given box, call it  $X(B)$ , needs to fulfill.

First of all, we need to know which boxes do not constitute a valuable resource. Thus the basic requirement for the measure  $X$  is as follows:

(M1) *Faithfulness*, which indicates

$$\forall_{B \in \mathcal{B}_{nv}} X(B) = 0. \quad (10)$$

Note that we restricted the axiom of faithfulness solely to the condition given above, and do not set an additional requirement  $X(B_v) > 0$ . After Ref. [11] we recall that some measure defined by usefulness with respect to a given operational task may give  $X(B) = 0$  even for boxes which do not belong to the set  $\mathcal{B}_{nv}$ .

Another property that the measure needs to have must reflect the fact that single-box general operations cannot increase the value of a given resource:

(M2) *Monotonicity*, which states that

$$\forall_{B \in \mathcal{B}} X(\mathcal{T}(B)) \leq X(B). \quad (11)$$

As we have seen earlier, there is a set of reversible operations  $\mathcal{R}$  such that each operation  $R$  from this set has its inverse  $R^{-1}$  for which there holds  $R^{-1}(R(B)) = B$ . Since we already stated that a measure  $X$  should be monotonic under general operations  $\mathcal{T}$ , the measure  $X$  needs to fulfill the following:

(M3) *Partial invariance*, which means that the measure  $X$  is invariant with respect to reversible operations  $\mathcal{R}$  performed on a box

$$\forall_{R \in \mathcal{R}} X(R(B)) = X(B). \quad (12)$$

For the purpose of defining the next axiom, we need to specify the distance measure for a pair of boxes.

*Definition 1.* The trace distance between two boxes  $B$  and  $B'$  is given by [15]

$$\|B - B'\| := \sup_S \|\mathcal{S}(B) - \mathcal{S}(B')\|_D, \quad (13)$$

where  $\|\dots\|_D$  denotes the *norm of difference between two probability distributions* which is given by variational distance between them

$$\|p_B(\lambda_c) - p_{B'}(\lambda_c)\|_D := \sum_{\lambda_c} |p_B(\lambda_c) - p_{B'}(\lambda_c)|, \quad (14)$$

where  $p_B(\lambda_c)$  ( $p_{B'}(\lambda_c)$ ) is the distribution of a measured context  $c$  of a box  $B$  ( $B'$ ) for the outputs  $\lambda_c$ . The supremum in (13) is taken over all operations which transform boxes into a probability distribution according to

$$S(B) := \sum_c \alpha_c \mathcal{T}_c p_B(\lambda_c), \quad (15)$$

where  $\mathcal{T}_c$  are stochastic matrices,  $0 \leq \alpha_c \leq 1$ , and  $\sum_c \alpha_c = 1$ .

To this end, we notice that any experimental realization of boxes involves inevitable distortion from the ideal box we want to realize. Suppose that an experimentally realized imperfect box  $B'$  is close to the target box  $B$ , which constitutes a valuable resource:

$$\|B - B'\| \leq \epsilon, \quad (16)$$

where  $\|\dots\|$  denotes the *trace distance between two boxes* [15]. Now, the desired property of the measure that conforms the requirements raised by the nonideal experimental processes is as follows:

(M4) *Asymptotic continuity*, which means that for two close boxes (16), there holds

$$|X(B) - X(B')| \leq \epsilon \log d + f(\epsilon), \quad (17)$$

where  $d$  is dimensionality of the boxes and  $f$  is a function such that  $f(\epsilon) \xrightarrow{\epsilon \rightarrow 0} 0$ .

Apart from the above axioms, we recall also two welcome properties that the quantifiers of a resource should satisfy:

(P1) *Convexity*, which means that for any box  $B$  which is decomposable into  $B = \sum_i p_i B_i$ , there is

$$X\left(\sum_i p_i B_i\right) \leq \sum_i p_i X(B_i), \quad (18)$$

where  $B_i$  are arbitrary boxes.

If we consider a measure to be extensive we would also require the property termed by the following:

(P2) *Additivity*, for which a measure fulfills

$$\forall_{B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2} X(B_1 \otimes B_2) = X(B_1) + X(B_2), \quad (19)$$

where  $\otimes$  denotes tensor product of boxes, i.e., for boxes  $B_1 = \{p(\lambda_c)\}$  and  $B_2 = \{q(\lambda_{c'})\}$  we have  $B_1 \otimes B_2 = \{p(\lambda_c)q(\lambda_{c'})\}$ .

### III. MEASURES OF CONTEXTUALITY

In this section, we focus on contextuality measures as defined in Ref. [16]. We then introduce the tools needed for proving the asymptotic continuity of the measures. The crucial assumption for the latter is closeness of two respective boxes. As we shall see, the assumption is still valid while considering two close quantum states from which the respective boxes are drawn.

Given a single joint probability distribution of an arbitrary context  $c$  of a box  $B$ , we can trivially define an extended joint probability distribution for all observables in  $\mathcal{M}$ :

$$\mathcal{A}_B(c) := p(\lambda_c)P(\lambda_r|c), \quad (20)$$

where  $P(\lambda_r|c)$  is a joint probability distribution for all the observables from the set  $\mathcal{M} \setminus c$ , and  $\lambda_r$  is the corresponding set of measurement outcomes.

*Definition 2.* For a given box  $B = \{p_B(\lambda_c)\}$  we call its extension a family of distributions:

$$\mathcal{F}(B) := \{\mathcal{A}_B(c)\}, \quad (21)$$

where  $\mathcal{A}_B(c) := p_B(\lambda_c)P(\lambda_r|c)$  is an extension of distribution  $p_B(\lambda_c)$  to all observables of a box  $B$ .

For further purposes, we will write  $\mathcal{E}_{p(c)}(B)$  to denote a distribution, related to the extension  $\mathcal{F}(B)$ , which for the ease of notation we model as a quantum state:

$$\mathcal{E}_{p(c)}(B) := \sum_c p(c)|c\rangle\langle c| \otimes \mathcal{A}_B(c), \quad (22)$$

for an arbitrary probability distribution  $p(c)$  of contexts.

We will now recall the definitions of two measures of contextuality as presented in [16]. The first measure, mutual information of contextuality, captures the idea that a contextual system cannot be described by a single joint probability distribution for all observables that can be measured on the system. It quantifies the correlations between the different joint probability distributions consistent with each of the measured context and the number of a chosen context. The second measure, relative entropy of contextuality, is defined in terms of a statistical distance between a set of probability distributions describing a contextual system to the closest single noncontextual joint probability distribution. The relative entropy of contextuality is a natural extension to contextual systems, an analogous measure of nonlocality presented in [24], called *statistical strength of nonlocality proofs*.

*Definition 3.* Mutual information of contextuality of a given box  $B$  we call

$$I_{\max}(B) := \sup_{\{p(c)\}} \inf_{\{\mathcal{A}_B(c)\}} I\left(\sum_c p(c)|c\rangle\langle c| \otimes \mathcal{A}_B(c)\right), \quad (23)$$

where  $I$  is mutual information between probability distributions.

We also have the following:

*Definition 4.* Relative entropy of contextuality of a given box  $B$  we call

$$X_{\max}(B) := \sup_{\{p(c)\}} \inf_{\{p(\lambda)\}} \sum_c p(c)D(p_B(\lambda_c)||p(\lambda_c)), \quad (24)$$

where  $D$  is relative entropy, and infimum is taken over all joint probability distributions for all observables in  $\mathcal{M}$ ,  $p(\lambda)$ , such that  $p(\lambda_c)$  is a proper marginal distribution for a given context:

$$p(\lambda_c) = \sum_{\mathcal{M} \setminus c} p(\lambda). \quad (25)$$

Note that given a box  $B$ , we have

$$I_{\max}(B) = X_{\max}(B), \quad (26)$$

hence we can use  $I_{\max}(B)$  and  $X_{\max}(B)$  interchangeably. This equality will prove useful, as the arguments given in [25] that REC is asymptotic continuous are not sufficient for a proof of the latter fact. Instead, will prove the asymptotic continuity of MIC.

A specific measure of contextuality also utilizing the concept of relative entropy is given in the following.



*Definition 5.* Uniform relative entropy of contextuality of a given box  $B$  we call

$$X_u(B) := \inf_{\{p(\lambda)\}} \sum_c \frac{1}{n} D(p_B(\lambda_c) || p(\lambda_c)), \quad (27)$$

where  $n$  is a number of different contexts of a box  $B$ .

We now observe the following fact, namely, when the trace distance of two boxes is small then for any extension of one box there exists an extension of the other, which is close to the former.

*Observation 1.* If two boxes  $B = \{p_B(\lambda_c)\}$  and  $B' = \{p_{B'}(\lambda_c)\}$  satisfy  $\|B - B'\| \leq \delta$ , then for  $\delta > 0$  and for any fixed  $\{p(c)\}$  we have

$$\forall_{\mathcal{E}_{p(c)}(B)} \exists_{\mathcal{E}_{p(c)}(B')} \|\mathcal{E}_{p(c)}(B) - \mathcal{E}_{p(c)}(B')\|_D \leq \delta. \quad (28)$$

*Proof.* The left-hand side of Eq. (28) equals

$$\sum_c p(c) \|\mathcal{A}_B(c) - \mathcal{A}_{B'}(c)\|_D, \quad (29)$$

where

$$\mathcal{A}_B(c) := p_B(\lambda_c) P(\lambda_r | c), \quad (30)$$

$$\mathcal{A}_{B'}(c) := p_{B'}(\lambda_c) P(\lambda_r | c). \quad (31)$$

Consider the difference between the two distributions  $\mathcal{A}_B(c)$  and  $\mathcal{A}_{B'}(c)$ :

$$\begin{aligned} & \|\mathcal{A}_B(c) - \mathcal{A}_{B'}(c)\|_D \\ &= \sum_{\lambda_c, \lambda_r} |p_B(\lambda_c) P(\lambda_r | c) - p_{B'}(\lambda_c) P(\lambda_r | c)| \\ &= \sum_{\lambda_c} \left\{ |p_B(\lambda_c) - p_{B'}(\lambda_c)| \left[ \sum_{\lambda_r} P(\lambda_r | c) \right] \right\} \\ &= \sum_{\lambda_c} |p_B(\lambda_c) - p_{B'}(\lambda_c)| \\ &= \|\mathcal{S}^*(B) - \mathcal{S}^*(B')\|_D \\ &\leq \sup_{\mathcal{S}} \|\mathcal{S}(B) - \mathcal{S}(B')\|_D \\ &\leq \delta, \end{aligned} \quad (32)$$

where the first inequality comes from the fact that  $P(\lambda_r | c)$  is a probability distribution for each  $c$ ,  $\mathcal{S}^*$  is such that  $\alpha_c = 1$  and  $\mathcal{T}_c$  is the identity matrix for a chosen context  $c$ , while the last inequality follows from the assumption of the observation. Using the last inequality back in (29) we obtain (28), which ends the proof. ■

We need to introduce the notation for the corresponding extensions given by Eq. (31), as it is described in the following.

*Definition 6.* Consider two boxes  $B = \{p_B(\lambda_c)\}$  and  $B' = \{p_{B'}(\lambda_c)\}$ . For any extension  $\mathcal{F}(B)$ , the extension  $\mathcal{F}(B')$  given by Eq. (31), which satisfy  $\|\mathcal{E}_{p(c)}(B) - \mathcal{E}_{p(c)}(B')\|_D \leq \delta$ , we will denote as  $\mathcal{F}(B'|B)$ .

In the next section, we will derive asymptotic continuity of mutual information of contextuality  $I_{\max}$ . As we shall see in the following, the closeness of quantum states  $\rho$  and  $\rho'$  implies the closeness of the respective boxes  $B$  and  $B'$ , while assuming perfect measurements on quantum states. Consider two boxes

$B$  and  $B'$ , each of them drawn by the same set of measurements  $M_c$ , but on a different quantum state  $\rho$  and  $\rho'$ , respectively, such that  $\|\rho - \rho'\| \leq \delta$ , where in case of matrices  $\|A\| := \sqrt{AA^\dagger}$ . Then,  $I_{\max}$  can be proved to be asymptotic continuous with respect to trace distance for quantum states  $\|\rho - \rho'\| < \delta$ .

*Observation 2.* Consider any two states  $\rho, \rho'$  such that  $\|\rho - \rho'\| < \delta$ . Let  $\mathcal{O} = \{M_c\}$  be a set of operators, which generates two respective boxes  $B$  and  $B'$  on states  $\rho$  and  $\rho'$ , respectively, in such a way that each distribution of the box  $B$  ( $B'$ ) is given by  $p_B(\lambda_c) \equiv \{\text{Tr} M_c \rho\}$  ( $p_{B'}(\lambda_c) \equiv \{\text{Tr} M_c \rho'\}$ ). Then, we have

$$\|B - B'\| < 2\delta. \quad (33)$$

*Proof.* Consider a distribution  $\alpha_c^*$  and stochastic matrices  $\mathcal{T}_c^*$  that realize supremum in definition of  $\|B - B'\|$ . Let  $M^*$  be a measurement operator defined as

$$M^* := \sum_c \alpha_c^* \mathcal{T}_c^* M_c. \quad (34)$$

From the assumption we have that

$$\begin{aligned} \delta &> \|\rho - \rho'\| \\ &= \sup_{\tilde{M}} \text{Tr} \tilde{M} (\rho - \rho') \\ &\geq \text{Tr} P^+ M^* (\rho - \rho'), \end{aligned} \quad (35)$$

where  $P^+$  is a projector onto the positive subspace of  $M^*(\rho - \rho')$ . The right-hand side of the last inequality above is equal to

$$S^+ \equiv \sum_{\{+\}} (r_\rho - r_{\rho'}), \quad (36)$$

where  $r_\rho = \text{Tr} M^* \rho$  ( $r_{\rho'} = \text{Tr} M^* \rho'$ ) and the sum is over all terms  $r_\rho - r_{\rho'} > 0$ . Changing the roles of  $\rho$  and  $\rho'$ , we obtain analogously

$$\begin{aligned} \delta &> \|\rho' - \rho\| \\ &= \sup_{\tilde{M}} \text{Tr} \tilde{M} (\rho' - \rho) \\ &\geq \sum_{\{+\}} (r_{\rho'} - r_\rho) \\ &= \left| \sum_{\{-\}} (r_\rho - r_{\rho'}) \right| \\ &\equiv S^-. \end{aligned} \quad (37)$$

From the above inequality, and (36), we have that  $\max(S^+, S^-) < \delta$ , and since the distributions  $r_\rho$  and  $r_{\rho'}$  are obtained with the optimal  $\alpha_c^*$  and  $\mathcal{T}_c^*$ , we obtain

$$\|B - B'\| = \|r_\rho - r_{\rho'}\|_D = S^+ + S^- \leq 2 \max(S^+, S^-) < 2\delta, \quad (38)$$

which ends the proof. ■

#### IV. ASYMPTOTIC CONTINUITY OF $I_{\max}$

In this section, we will prove the asymptotic continuity for the measure of contextuality given by mutual information of contextuality (23). We will focus on MIC rather than *relative entropy of contextuality*, and we will introduce a

method of proving the asymptotic continuity. Nevertheless, the asymptotic continuity of the former (MIC) implies the same for the latter (REC) since the two measures are equal to one another [16].

At this stage, let us also recall the property of asymptotic continuity for von Neumann entropy. Consider two quantum states  $\rho_1$  and  $\rho_2$  of dimension  $d$ , for which

$$\|\rho_1 - \rho_2\| \leq 1/2, \quad (39)$$

where  $\|\dots\|$  is trace norm. Von Neumann entropy is asymptotically continuous since

$$|S(\rho_1) - S(\rho_2)| \leq \|\rho_1 - \rho_2\| \log d + \eta(\|\rho_1 - \rho_2\|), \quad (40)$$

where  $\eta(x) = -x \log x$ .

Furthermore, quantum conditional entropy is also asymptotically continuous [26], i.e., for  $\|\rho_1 - \rho_2\| \leq \epsilon$  we have

$$|S_{X|Y}(\rho_1) - S_{X|Y}(\rho_2)| \leq 4\|\rho_1 - \rho_2\| \log d + 2\eta(\|\rho_1 - \rho_2\|) + 2\eta(1 - \|\rho_1 - \rho_2\|), \quad (41)$$

where  $X, Y$  denote two subsystems of the states  $\rho_1$  and  $\rho_2$ .

Now, Observation 1 allows us to write the following.

*Observation 3.* If for some  $\delta > 0$  two boxes  $B = \{p_B(\lambda_c)\}$  and  $B' = \{p_{B'}(\lambda_c)\}$  satisfy

$$\|B - B'\| \leq \delta, \quad (42)$$

then for any fixed  $\{p(c)\}$  we have

$$\forall_{\mathcal{E}_{p(c)}(B)} \exists_{\mathcal{E}_{p(c)}(B'|B)} |I(\mathcal{E}_{p(c)}(B)) - I(\mathcal{E}_{p(c)}(B'|B))| \leq g(\delta), \quad (43)$$

with

$$g(\delta) = 5\delta \log d + 2\eta(1 - \delta) + 3\eta(\delta), \quad (44)$$

where  $I$  is mutual information,  $d = \min(\prod_{i=1}^k a_i, |E_G|)$ , where  $\prod_{i=1}^k a_i$  and  $|E_G|$  are the dimensions of the two subsystems of the box  $B$ , respectively.

*Proof.* Due to Observation 1 for every  $\mathcal{E}_{p(c)}(B)$  there exists  $\mathcal{E}_{p(c)}(B'|B)$  such that  $\|\mathcal{E}_{p(c)}(B) - \mathcal{E}_{p(c)}(B'|B)\|_D \leq \delta$ . Using the definition of mutual information, the left-hand side of (43) can be written as

$$\begin{aligned} & |I(\mathcal{E}_{p(c)}(B)) - I(\mathcal{E}_{p(c)}(B'|B))| \\ &= |S_X(\mathcal{E}_{p(c)}(B)) - S_{X|Y}(\mathcal{E}_{p(c)}(B)) \\ &\quad - S_X(\mathcal{E}_{p(c)}(B'|B)) + S_{X|Y}(\mathcal{E}_{p(c)}(B'|B))|. \end{aligned} \quad (45)$$

Then, making use of a triangle inequality, we can bound this expression by

$$\begin{aligned} & |S_X(\mathcal{E}_{p(c)}(B)) - S_X(\mathcal{E}_{p(c)}(B'|B))| \\ &\quad + |S_{X|Y}(\mathcal{E}_{p(c)}(B)) - S_{X|Y}(\mathcal{E}_{p(c)}(B'|B))| \\ &= 5\delta \log d + 2\eta(1 - \delta) + 3\eta(\delta), \end{aligned} \quad (46)$$

where in the last step we used asymptotic continuity of von Neumann entropy (40) as well as quantum conditional entropy (41). ■

Consider the following lemma:

*Lemma 1.* For any real-valued function  $f$ , and for any two sets  $T$  and  $T'$ , if there exists a real-valued function  $g$ , such that for any positive  $\delta$  and any  $\rho \in T$  there exists  $\sigma_\rho \in T'$ , such

that

$$|f(\rho) - f(\sigma_\rho)| \leq g(\delta), \quad (47)$$

and for any  $\sigma \in T'$  there exists  $\rho_\sigma \in T$ , such that

$$|f(\rho_\sigma) - f(\sigma)| \leq g(\delta), \quad (48)$$

then there holds

$$\left| \inf_{\rho \in T} f(\rho) - \inf_{\sigma \in T'} f(\sigma) \right| \leq g(\delta) + \delta \quad (49)$$

and

$$\left| \sup_{\rho \in T} f(\rho) - \sup_{\sigma \in T'} f(\sigma) \right| \leq g(\delta) + \delta, \quad (50)$$

provided that  $\inf_{\rho \in T} f(\rho)$  and  $\inf_{\sigma \in T'} f(\sigma)$ , as well as  $\sup_{\rho \in T} f(\rho)$  and  $\sup_{\sigma \in T'} f(\sigma)$  are bounded.

We prove the lemma in Appendix A. We can now state the following theorem:

*Theorem 1.* For any fixed  $p(c)$ , the function

$$I_{p(c)}(B) = \inf_{\{\mathcal{A}_B(c)\}} I \left( \sum_c p(c) |c\rangle\langle c| \otimes \mathcal{A}_B(c) \right) \quad (51)$$

is asymptotically continuous, i.e., for any  $\|B - B'\| \leq \delta$  there is

$$|I_{p(c)}(B) - I_{p(c)}(B')| \leq g(\delta) + \delta, \quad (52)$$

with  $g(\delta)$  given in the right-hand side of (44).

*Proof.* Assume that  $\|B - B'\| \leq \delta$  for two boxes  $B = \{g_B(c)\}$  and  $B' = \{g_{B'}(c)\}$ . Let us first consider  $I_{p(c)}(B)$ . By definition of infimum, there exist a sequence of extensions of distributions  $\mathcal{F}^n(B) = \{A^n_B(c)\}$ , such that

$$\lim_{n \rightarrow \infty} I(\mathcal{E}^n_{p(c)}(B)) = I_{p(c)}(B), \quad (53)$$

and similarly there exist  $\mathcal{F}^n(B') = \{A^n_{B'}(c)\}$ , such that

$$\lim_{n \rightarrow \infty} I(\mathcal{E}^n_{p(c)}(B')) = I_{p(c)}(B'). \quad (54)$$

Let us now specify the following two sets:

$$T := \bigcup_{n=1}^{\infty} \{I(\mathcal{E}^n_{p(c)}(B)), I(\mathcal{E}^n_{p(c)}(B|B'))\}, \quad (55)$$

$$T' := \bigcup_{n=1}^{\infty} \{I(\mathcal{E}^n_{p(c)}(B'|B)), I(\mathcal{E}^n_{p(c)}(B'))\}, \quad (56)$$

and also denote  $V$  ( $V'$ ) as the set of  $I(\mathcal{E}_{p(c)}(B))$  ( $I(\mathcal{E}_{p(c)}(B'))$ ) for all distributions  $\mathcal{A}_B(c)$  ( $\mathcal{A}_{B'}(c)$ ). Notice that  $\{I(\mathcal{E}^n_{p(c)}(B))\} \subset T$ , and therefore we have that

$$\inf T \leq \inf \{I(\mathcal{E}^n_{p(c)}(B))\} = I_{p(c)}(B). \quad (57)$$

On the other hand,  $T \subset V$ , and therefore

$$I_{p(c)}(B) = \inf V \leq \inf T, \quad (58)$$

where the equality is by definition of  $I_{p(c)}(B)$ . The inequalities (57) and (58) lead to

$$\inf T = I_{p(c)}(B), \quad (59)$$

and the same reasoning also gives

$$\inf T' = I_{p(c)}(B'). \quad (60)$$

Now, due to Observation 3, there is

$$|I(\mathcal{E}_{p(c)}^n(B)) - I(\mathcal{E}_{p(c)}^n(B'|B))| \leq g(\delta), \quad (61)$$

$$|I(\mathcal{E}_{p(c)}^n(B')) - I(\mathcal{E}_{p(c)}^n(B|B'))| \leq g(\delta). \quad (62)$$

Let us take in the assumption of Lemma 1 the function  $f : \mathbf{R} \rightarrow \mathbf{R}$  to be simply the identity function  $f(x) = x$  (it is easy to see that other assumptions are satisfied by the construction and the fact that the boxes  $B$  and  $B'$  are close to each other and by inequalities (61) and (62)). Then, by Lemma 1 we obtain

$$|\inf_{t \in T} t - \inf_{t' \in T'} t'| \leq g(\delta) + \delta, \quad (63)$$

which is exactly what we want to prove.  $\blacksquare$

We can now state the main theorem of this section. The result is crucial from the experimental point of view. Namely, it states that for two close boxes (one of which is the ideal box we aim to obtain, while the second is its imperfect experimental realization), the measure of the amount of contextuality of the experimentally obtained box cannot differ from the measure of contextuality of the target box by more than the distance of the two boxes with a factor depending only logarithmically on the dimension of them.

*Theorem 2.* The measure of contextuality  $I_{\max}$  is asymptotically continuous, i.e., for any two boxes  $B$  and  $B'$  fulfilling  $\|B - B'\| \leq \delta$ , there is

$$|I_{\max}(B) - I_{\max}(B')| \leq 6g(\delta) + 3\delta, \quad (64)$$

with

$$g(\delta) = 5\delta \log d + 2\eta(1 - \delta) + 3\eta(\delta). \quad (65)$$

*Proof.* The proof follows the similar lines as the proof of Theorem 1. Let us first consider  $I_{\max}(B)$ . By definition of supremum, there exist a sequence of distributions  $\{p_n(c)\}$ , such that

$$\lim_{n \rightarrow \infty} I_{p_n(c)}(B) = I_{\max}(B), \quad (66)$$

and similarly there exist  $\{p'_n(c)\}$ , such that

$$\lim_{n \rightarrow \infty} I_{p'_n(c)}(B') = I_{\max}(B'). \quad (67)$$

Let us now specify the following two sets:

$$T := \bigcup_{n=1}^{\infty} \{I_{p_n(c)}(B), I_{p'_n(c)}(B)\}, \quad (68)$$

$$T' := \bigcup_{n=1}^{\infty} \{I_{p_n(c)}(B'), I_{p'_n(c)}(B')\}, \quad (69)$$

and also denote  $V$  ( $V'$ ) as the set of  $I_{p_n(c)}(B)$  ( $I_{p'_n(c)}(B')$ ) for all distributions  $\{p_n(c)\}$  ( $\{p'_n(c)\}$ ). Applying the similar reasoning as in the proof of Theorem 1, but this time concerning the suprema, we arrive at

$$\sup T = I_{\max}(B), \quad (70)$$

and the same reasoning gives also

$$\sup T' = I_{\max}(B'). \quad (71)$$

Now, due to Theorem 1, there is

$$|I_{p_n(c)}(B) - I_{p_n(c)}(B')| \leq g(\delta) + \delta, \quad (72)$$

$$|I_{p'_n(c)}(B') - I_{p'_n(c)}(B)| \leq g(\delta) + \delta. \quad (73)$$

Let us take in the assumption of Lemma 1 the function  $f : \mathbf{R} \rightarrow \mathbf{R}$  to be simply the identity function  $f(x) = x$  (it is easy to see that other assumptions are satisfied by the construction and the fact that the boxes  $B$  and  $B'$  are close to each other and by inequalities (72) and (73)). Then, by Lemma 1 we obtain

$$|\sup_{t \in T} t - \sup_{t' \in T'} t'| \leq g(\delta) + 2\delta, \quad (74)$$

which is exactly what we want to prove.  $\blacksquare$

From the two main theorems presented here, we have an immediate corollary, which states that the relative entropy of contextuality is continuous with respect to quantum states, provided ideal measurements.

*Corollary 1.* For two quantum states  $\rho$  and  $\sigma$  such that  $\|\rho - \sigma\| \leq \delta$ , for any set of quantum measurements  $\mathcal{M} = \{M_c\}$  which generates on the states respective boxes  $B = \{\text{Tr} M_c \rho\}$  and  $B' = \{\text{Tr} M_c \sigma\}$ , we have

$$|I_{p(c)}(B) - I_{p(c)}(B')| \leq 2g(\delta) + 2\delta, \quad (75)$$

as well as

$$|I_{\max}(B) - I_{\max}(B')| \leq 12g(\delta) + 6\delta, \quad (76)$$

with  $g(\delta) = 5\delta \log d + 2\eta(1 - \delta) + 3\eta(\delta)$ .

*Proof.* This corollary follows directly from Observation 2, and Theorems 1 and 2.

## V. COMPUTABLE UPPER BOUND ON THE MEASURE OF A RESOURCE

In this section, we will connect two measures of box resources: the one which is based on relative entropy, with the other which reports how much the creation of a box costs. In case of the resource which is contextuality, we will show that the relative entropy of contextuality for chain boxes is upper bounded (up to a normalization factor) by the cost of contextuality. In derivation of this result, we will use the property of (P1) convexity of the relative entropy of contextuality, which was observed in [27], but without a formal proof, which we provide in Appendix B.

To achieve this, we will show a general result, which holds for a measure satisfying certain axioms. First, we assume that sets  $\mathcal{B}$  and  $\mathcal{B}_{nv}$  are convex polytopes and  $\mathcal{B}_{nv} \subset \mathcal{B}$ . We will now need the notion of a *convex roof of a measure* which is defined for the measure  $X$  as

$$\hat{X}(B) := \inf_{\{p_i, E_i\}} \sum_i p_i X(E_i), \quad (77)$$

where  $E_i$  are extremal boxes of the polytope of all consistent boxes, and infimum is taken over all ensembles of the box  $B$  with all extremal boxes, so that  $\sum_i p_i E_i = B$ .

We begin with noticing the following observation:

*Observation 4.* For any convex measure  $X$  and  $B \in \mathcal{B}$  we have

$$X(B) \leq \hat{X}(B). \quad (78)$$

*Proof.* It follows from the definition of  $\hat{X}$ . ■

Next, we note that some of the extremal boxes of the polytope of consistent boxes are valuable ( $E'_v$ ), while the others are not ( $E''_v$ ). Such a polytope  $\mathcal{B}$  in some cases can satisfy the following property, which we call *vertex-equivalence property*:

$$\forall_{E_v, E'_v \in \mathcal{B}} \exists_{R \in \mathcal{R}} R(E_v) = E'_v, \quad (79)$$

where  $E_v, E'_v$  are extremal, valuable boxes. In other words, any valuable box can be transformed into any other valuable box by means of reversible operations.

We will now show that if the polytope  $\mathcal{B}$  has the vertex-equivalence property, a convex measure  $X$  satisfying axioms (M1) and (M3) is upper bounded (up to a normalization factor  $X(E_v)$  for some extremal valuable box  $E_v$ ) by another measure which we will call the *cost of the resource*. We first formalize the latter measure, as a generalization of the well-known measure of the *cost of nonlocality*:

*Definition 7.* For a box  $B \in \mathcal{B}$  the cost of the resource for this box is

$$C(B) = \inf_p \{p : B = pB_v + (1 - p)B_{nv}, B_v \in \mathcal{B}, B_{nv} \in \mathcal{B}_{nv}\}. \quad (80)$$

We are ready to present the main result of this section:

*Proposition 1.* Let  $\mathcal{B}$  be a polytope of consistent boxes satisfying the vertex equivalence property. Also, let  $X$  be a measure of resource acting on boxes  $B \in \mathcal{B}$ , which satisfies the axioms of (M1) faithfulness, (M3) local invariance, and (P1) convexity. We have then

$$X(B) \leq C(B)X(E_v), \quad (81)$$

where  $E_v$  is arbitrarily fixed, extremal, valuable box in  $\mathcal{B}$ , and  $C(\dots)$  is the cost of the resource.

*Proof.* Let us fix  $B \in \mathcal{B}$  arbitrarily. Since  $X$  is convex, by Observation 4, we have

$$X(B) \leq \inf_{\{p_i, E_i\}} \sum_i p_i X(E_i). \quad (82)$$

Now, for all  $E_i \in \mathcal{B}_{nv}$  we have by axiom (M1)  $X(E_i) = 0$ . Thus, we have

$$X(B) \leq \inf_{\{p_i, E_i\}} \sum_{i \in \mathcal{I}} p_i X(E_i), \quad (83)$$

where  $\mathcal{I}$  is the set of indices for an ensemble  $\{p_i, E_i\}$  such that if  $i \in \mathcal{I}$  then  $E_i$  is valuable box. On the other hand, by the vertex equivalence property of the polytope  $\mathcal{B}$  there is  $X(E_i) = X(E_v)$  for all  $i \in \mathcal{I}$  and  $E_v$  being some extremal valuable box fixed arbitrarily. Indeed, all the valuable boxes have the same value of the measure  $X$  since they are reversibly transformable one into another, and the measure  $X$  satisfies the axiom (M3), i.e., does not change under such transformations. Hence,

$$X(B) \leq X(E_v) \inf_{\{p_i, E_i\}} \sum_{i \in \mathcal{I}} p_i, \quad (84)$$

and furthermore

$$\hat{X}(B) = X(E_v) \inf_{\{p_i, E_i\}} \sum_{i \in \mathcal{I}} p_i. \quad (85)$$

We will now prove that the minimal  $\sum_{i \in \mathcal{I}} p_i$  equals exactly  $C(B)$ , which makes the thesis. Consider again any pure ensemble into boxes  $E_i$  with probabilities  $p_i$ . We can perform a valid decomposition of a box  $B$  into

$$B = pB_v + (1 - p)B_{nv}, \quad (86)$$

where  $p = \sum_{i \in \mathcal{I}} p_i$  and, by definition of  $\mathcal{I}$ ,  $B_v = \sum_{i \in \mathcal{I}} p_i E_i / (\sum_{i \in \mathcal{I}} p_i)$ , and  $B_{nv} = \sum_{i \notin \mathcal{I}} p_i E_i / (\sum_{i \notin \mathcal{I}} p_i)$ . Thus, there is  $C(B) \leq p$ , and since the ensemble was arbitrary, we have that

$$C(B) \leq \inf_{\{p_i, B_i\}} \sum_{i \in \mathcal{I}} p_i. \quad (87)$$

This, via (85), implies that

$$C(B)X(B_v) \leq \hat{X}(B). \quad (88)$$

We now prove converse inequality. To begin with, consider any decomposition which achieves  $C(B)$  (the case when infimum in its definition is not attained, is then obvious):

$$B = qB_v + (1 - q)B_{nv}. \quad (89)$$

Now, consider a decomposition of a box  $B_v$  into extremal boxes. Since  $q = C(B)$ , there cannot be any deterministic box in the decomposition of  $B_v$ , or else we would have smaller value of  $C(B)$  by subtracting this box from  $B_v$  and thereby lowering  $q$ . Thus, we can write  $B_v = \sum_j r_j E_j$ , with  $E_j$  extremal valuable boxes. Let us also decompose a box  $B_{nv}$  into extremal nonvaluable boxes (it is possible since this box is by definition not valuable, and the set  $\mathcal{B}_{nv}$  is a convex polytope by assumption):  $B_{nv} = \sum_{k=1}^K s_k D_k$ . We then have an ensemble of the box  $B$  of the form  $\chi = \{qr_1, \dots, qr_N, (1 - q)s_1, \dots, (1 - q)s_K, E_1, \dots, E_N, D_1, \dots, D_K\}$  for some natural numbers  $N, K \geq 1$ , i.e., by construction:

$$B = q \sum_{j=1}^N r_j E_j + (1 - q) \sum_{k=1}^K s_k D_k. \quad (90)$$

Now, we have  $X(D_k) = 0$  for  $k \in \{1, \dots, K\}$ . Consider then the set of indices  $\mathcal{I} = \{1, \dots, N\}$ . Since  $\hat{X}$  is defined as a function which is minimized over all ensembles of the box  $B$ , it is upper bounded by the value of the function on the ensemble  $\chi$ . Thus, we have

$$\begin{aligned} \hat{X}(B) &= \inf_{\{p_i, B_i\}} \sum_i p_i X(B_i) \\ &\leq \sum_{i \in \mathcal{I}} q r_i X(E_i) \\ &= X(B_v) \sum_i q r_i \\ &= X(B_v)q, \end{aligned} \quad (91)$$

where the second last equality is due to the fact that for all  $i \in \mathcal{I}$   $X(E_i) = X(B_v)$  since the polytope of consistent boxes satisfies the vertex equivalence property. This proves that

$$\hat{X}(B) \leq C(B)X(B_v), \quad (92)$$

which together with the opposite inequality (88) proves the thesis. ■



Let us note here that in case when  $\mathcal{B}_{nv}$  is a polytope, the cost of the resource can be computed by linear programming, hence, the above proposition provides a computable upper bound on a convex measure satisfying (M1) and (M3).

In the next section, we present two examples of polytopes, which satisfy the vertex-equivalence property, for which the above proposition applies. We present now more rough bound, which holds for all polytopes, including those which do not satisfy the latter property.

*Proposition 2.* Let  $\mathcal{B}$  be a polytope of consistent boxes. Also let  $X$  be a measure of a resource acting on boxes  $B \in \mathcal{B}$ , which satisfies axioms of (M1) faithfulness and (P1) convexity. We then have

$$X(B) \leq C(B) \max_{E_v \in \mathcal{B}} X(E_v), \quad (93)$$

where maximum is taken over all extremal, valuable boxes  $E_v$  from  $\mathcal{B}$ , and  $C(\dots)$  is the cost of the resource.

*Proof.* The proof goes along similar lines as the second part of the proof of Proposition 1, namely, starting right after the inequality (88). We then repeat the proof until the sequence of (in)equalities (91) and modify them while observing that for  $s = \sum_{i \in \mathcal{I}} q r_i$ ,

$$\sum_{i \in \mathcal{I}} q r_i X(E_i) \leq \max_{i \in \mathcal{I}} X(E_i) \times s \leq \max_{E_v \in \mathcal{B}} X(E_v) \times s, \quad (94)$$

where in the first inequality maximum is taken over extremal valuable boxes  $E_i$  from the particular ensemble of the box  $B$ , while in the second, over all extremal valuable boxes from the polytope  $\mathcal{B}$ . According to the remaining part of the proof of Proposition 2, we obtain

$$\hat{X}(B) \leq C(B) \times \max_{E_v \in \mathcal{B}} X(E_v), \quad (95)$$

which is a less-tight version of the inequality (92) as required. ■

#### Examples of the upper bounds on the $X_{\max}$ via contextuality (nonlocality) cost

We can now apply the Proposition 1 in two cases. The first is the case of the set  $\mathcal{B}$  equal to the set all bipartite nonsignaling boxes with two binary inputs and two binary outputs (we will denote it as  $\mathcal{B}(2 \times 2)$ ). In this case, it is known that the only extremal valuable boxes of  $\mathcal{B}$  have a form

$$B_{rst}(a, b|x, y) = \begin{cases} 1/2 & \text{if } a \oplus b = xy \oplus rx \oplus sy \oplus t, \\ 0 & \text{else,} \end{cases} \quad (96)$$

where  $a, b, x, y, r, s, t$  are binary.

It is then clear from the above form that all extremal valuable boxes can be transformed reversibly into the  $PR$  box. Moreover, the set of all nonvaluable boxes is called the set of *local* boxes, and it is a convex polytope  $\mathcal{B}_{nv}(2 \times 2) \subset \mathcal{B}(2 \times 2)$ . We have then the following corollary:

*Corollary 2.* For any box  $B \in \mathcal{B}(2 \times 2)$  there holds

$$X_{\max}(B) \leq \inf_{\{p_i, B_i\}} \sum_i p_i X_{\max}(B_i) = C(B) \times \log \frac{4}{3}, \quad (97)$$

where  $B_i$  are extremal boxes and  $C(B)$  is the cost of nonlocality.

*Proof.* It is straightforward to check that  $X_{\max}$  satisfies the axioms (M1) and (M3), and, as it was mentioned in [27] and is proved in Appendix B,  $X_{\max}$  is convex. Moreover, as we have mentioned before, all the nonlocal boxes with two binary inputs and outputs can be transformed by local reversible operations into the  $PR$  box, i.e.,  $B_{000}$ , for which  $X_{\max}(B_{000}) = \log \frac{4}{3}$  [16]. Hence, by Proposition 1 we have the thesis. ■

We now generalize the above result to the case of resources of contextuality. We consider a class of boxes corresponding to  $n$ -cycle hypergraph (a class of  $CH^{(n)}$  boxes) scenarios as presented in Refs. [16,19,28,29]. Any box  $B \in CH^{(n)}$  is given by  $B = \{p_B(\lambda_c)\}$ , where a probability distribution for each context  $c$  can be written as

$$p_B(\lambda_{c_i}) \equiv p(m_i m_{i+1} | M_i M_{i+1}) = \frac{1}{2}(1 + \langle m_i m_{i+1} \rangle), \quad (98)$$

with  $m_j = \pm 1$ , where we use the convention  $m_n m_{n+1} = m_1 m_n$ .

The corresponding polytope of boxes compatible with this hypergraph will be called  $\mathcal{B}^n$ . Let us note here that  $\mathcal{B}^4 \equiv \mathcal{B}(2 \times 2)$ . It is known [16] that the contextuality measure of any extremal valuable box  $E_v$  for arbitrary  $n$  is given by

$$X_{\max}(E_v) = \log \frac{n}{n-1}. \quad (99)$$

*Corollary 3.* For a box  $B \in \mathcal{B}^n$  we have

$$X_{\max}(B) \leq \hat{X}_{\max}(B) = C(B) \log \frac{n}{n-1}. \quad (100)$$

*Proof.* According to the description of a box  $B \in CH^{(n)}$  given in Eq. (98), it can be uniquely described by a collection of  $n$  correlators

$$B \equiv (\langle m_1 m_2 \rangle, \langle m_2 m_3 \rangle, \dots, \langle m_n m_1 \rangle). \quad (101)$$

Now, for any given  $n$  we have  $2^{n-1}$  extremal contextual boxes of the form (101), where  $\langle m_i m_{i+1} \rangle = \pm 1$ , such that the number of negative components is odd [19]. It now suffices to observe that we can obtain all the other extremal contextual boxes from, e.g.,  $(-1, 1, 1, \dots, 1, 1)$  simply by bit-flipping the chosen outputs, which is a contextuality-preserving operation (a particular form of relabeling the outputs). Indeed, by performing a bit-flip  $m_j \rightarrow -m_j$  we change the sign of any pair of neighboring correlators. For instance, performing a bit-flip  $m_1 \rightarrow -m_1$  on the box  $(-1, 1, 1, \dots, 1, 1)$  produces another extremal contextual box  $(1, -1, 1, \dots, 1, 1)$ . Consequently, performing a bit-flip on each consecutive  $m_j$ , we can generate all extremal contextual boxes with exactly one correlator equal to  $-1$ . Now, given a box  $(1, 1, 1, \dots, 1, -1)$  we again perform a bit-flip  $m_1 \rightarrow -m_1$  which produces a box  $(-1, -1, 1, \dots, 1, -1)$  from which we can generate all the boxes with three correlators equal to  $-1$ . In doing so, we can generate all  $2^{n-1}$  extremal contextual boxes (all those with odd number of correlators equal to  $-1$ ) by a contextuality-preserving operation. We see then that  $\mathcal{B}^n$  satisfies vertex-equivalence property. Moreover,  $X_{\max}$  does not change under bit-flip of outputs of some of the observables, hence assumptions of Proposition 2 are satisfied. In consequence, the bound (100) is true. ■

## VI. A BOUND ON DISTILLABLE RESOURCE IN BOX THEORIES

In this section, we consider a scenario analogous to the scenario of distillation of entanglement in the entanglement theory. Namely, we assume that  $n$  copies of a box  $B$  are provided to a party (or parties in case of nonlocality). We distinguish a target box  $B_v^T$  which is valuable, approximation of which one wants to “distill” out of  $n$  copies of  $B$ . We demand that the distilling operations satisfy the axiom (O2), i.e., do not create a valuable box out of nonvaluable ones. We define distillable resource  $D(B_v^T|B)$  for a box  $B$ , in analogous manner to definition of distillable entanglement [30,31] as the highest ratio of number  $k$  which is such that the output of distillation protocol approximates  $[B_v^T]^{\otimes k}$ , to the number of used boxes  $B$  which is  $n$ , in asymptotic limit of large  $n$ . The main result of this section states that the so-called *regularized* measure of resource  $X$ , for which  $X$  satisfies the axioms (M2) of monotonicity and (M4) asymptotic continuity, up to a constant factor  $X(B_v^T)$ , is an upper bound to distillable resource  $D(B_v^T|B)$ :

$$X^\infty(B) \geq X(B_v^T)D(B_v^T|B), \quad (102)$$

where  $X^\infty = \lim_{n \rightarrow \infty} \frac{X(B^{\otimes n})}{n}$  is the regularized measure  $X$ .

### A. Proof of the upper bound

In this section, we prove inequality (102) under some assumptions on measure  $X$  and a target box  $B_v^T$ . We begin with definition of *rate of distillability of a box  $B_v^T$  from a box  $B$* , denoted as  $D(B_v^T|B)$ .

*Definition 8.* For a box  $B \in \mathcal{B}$ , consider a sequence  $\Lambda_n$  of operations satisfying the axioms (O1) and (O2), such that  $\Lambda_n(B^{\otimes n}) = B_n$ . The set  $\mathcal{D} = \{\Lambda_n\}$  is called a protocol distilling a target box  $B_v^T$  from  $B$ , if

$$\lim_{n \rightarrow \infty} \|B_n - [B_v^T]^{\otimes k_n}\| = 0. \quad (103)$$

For a given distillation protocol  $\mathcal{D}$ , its rate is given by

$$r(\mathcal{D}) = \limsup_{n \rightarrow \infty} \frac{k_n}{n}. \quad (104)$$

The rate of distillability of the box  $B_v^T$  from a box  $B$  is given by

$$D(B_v^T|B) = \sup_{\mathcal{D}} r(\mathcal{D}). \quad (105)$$

We can proceed to show the main result of this section, which states that asymptotically continuous and monotonous measure of resource is, up to a constant factor, an upper bound on  $D(B_v^T|B)$ , as it is stated in the following proposition.

*Proposition 3.* Let  $B_v^T \in \mathcal{B}_v$  be some target box. Let  $X$  be a measure which satisfies the axioms of (M2) monotonicity, (M4) asymptotic continuity, and also let  $X$  be superadditive on  $B_v^T$ . Then, we have

$$D(B_v^T|B)X(B_v^T) \leq X^\infty(B). \quad (106)$$

*Proof.* Consider  $n$  copies of a box  $B$ . The purpose is to distill the largest number of (approximate) copies of target boxes  $B_v^T$ . Let us fix  $\delta > 0$ . Then, there exists a protocol  $\mathcal{D} = \{\Lambda_n\}$  such

that  $r(\mathcal{D}) > D(B_v^T|B) - \delta$ . It follows also that for sufficiently large  $n$ ,

$$\Lambda_n(B^{\otimes n}) = B_n, \quad (107)$$

such that there holds  $\|B_n - B_v^T{}^{\otimes k_n}\|_D \leq \epsilon_n$ , where  $0 < \epsilon_n \rightarrow 0$  with  $n \rightarrow \infty$ . Then, we have the following chain of inequalities:

$$\begin{aligned} X(B^{\otimes n}) &\geq X(\Lambda_n(B^{\otimes n})) \\ &= X(B_n) \\ &\geq X([B_v^T]^{\otimes k_n}) - f(\epsilon_n) \\ &\geq k_n X(B_v^T) - f(\epsilon_n), \end{aligned} \quad (108)$$

where the first inequality holds by the axiom of (M2) monotonicity of  $X$  under operations  $\Lambda_n$  satisfying the axiom (O2). The first equality is by (107), the second inequality is by asymptotic continuity of  $X$ , where  $f(\dots)$  is some continuous function. The last inequality is by the assumption of superadditivity of  $X$  on a box  $[B_v^T]^{\otimes k_n}$  (see in this context Theorem 9 of [16]).

If we now divide the first and the last terms of the above chain of (in)equalities by  $n$ , we have

$$\frac{X(B^{\otimes n})}{n} \geq X(B_v^T) \frac{k_n}{n} - \frac{f(\epsilon_n)}{n}. \quad (109)$$

In fact, the left-hand side of the above inequality approaches  $X^\infty$  in the limit  $n \rightarrow \infty$  and, by continuity of  $f$ , the right-hand side approaches  $X(B_v^T)[D(B_v^T|B) - \delta]$ , as it was expected. Since  $\delta$  was arbitrary, taking the limit of  $\delta \rightarrow 0$  proves inequality (106). ■

We have now the following remark:

*Remark 1.* If additionally a measure  $X$  is subadditive, there is  $X \geq X^\infty$ , hence, by the above proposition, we obtain

$$X(B) \geq X(B_v^T)D(B_v^T|B). \quad (110)$$

### B. Protocol of distillation of contextuality

Following, we consider a particular example of a distillation protocol of contextual resources. We will consider a distillation protocol which uses two copies of weakly contextual resource and transform them into a single copy of more contextual resource. We will show this by a specific indicator of contextuality  $\beta$  (which quantifies the violation of a specific contextual inequality) rather than a measure of contextuality  $X_{\max}$  which is uneasy to calculate for boxes used in the distillation process. The distillation protocol itself is easily implementable in the experiment, as it involves only postprocessing of the data (measurement outcomes).

First, let us describe in detail contextual resources from which we distill more valuable resource. An *XOR box* is a consistent box  $\{p(a_1, a_2, \dots, a_m | c_i)\}$  such that for each input  $c_i$  and binary outputs  $a_1, \dots, a_m \in \{0, 1\}$ , either ( $p^{od}$ : odd context)

$$p^{od}(a_1, \dots, a_m | c_i) = \begin{cases} \frac{1}{2^{m-1}} & \forall \oplus_j^m a_j = 1, \\ 0 & \text{otherwise} \end{cases} \quad (111)$$

or ( $p^e$ : even context)

$$p^e(a_1, \dots, a_m | c_i) = \begin{cases} \frac{1}{2^{m-1}} & \forall \bigoplus_j^m a_j = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (112)$$

Note that an XOR box can in principle be contextual or noncontextual. For the rest of this section, we assume that as an XOR box we mean only contextual box, of which the examples are presented in Ref. [16] under the name of PM box, M box, CH box.

A *correlated box* is a box such that for each input  $c_i$  and binary outputs  $a_1, \dots, a_m \in \{0, 1\}$  there is

$$p(a_1, \dots, a_m | c_i) = \begin{cases} \frac{1}{2^{m-1}} & \forall \bigoplus_j^m a_j = 0, \\ 0 & \text{otherwise.} \end{cases} \quad (113)$$

All correlated boxes which correspond to a hypergraph of a certain XOR box are noncontextual: a correlated box can be obtained from a single joint probability distribution  $p(\mathcal{M})$  which is decomposable into probabilistic points with the property  $\bigoplus_j^m a_j = 0$  for all  $c_i$ .

The protocol of distillation, that we will employ, was originally used in [20] for distilling nonlocal resources, whereas in terms of contextual resources it works as follows: on inputs  $c_i$  for two copies of a resource one inputs the same number and then receives outputs  $(a_1, \dots, a_m)$  and  $(a'_1, \dots, a'_m)$ , respectively. Then, one computes the final output as  $(a_1 \oplus a'_1, \dots, a_m \oplus a'_m)$ . This procedure we will call a node-wise XOR operation.

Furthermore, we will use the parameter  $\beta$  (for a definition see Ref. [16]) as a contextuality indicator of a given box, by means of violating the contextuality inequality: if we denote  $B^*$  as a reference extremal isotropic XOR box, then for any contextual box  $B$  the contextuality inequality

$$\beta_{B^*}(B) \leq n - 1 \quad (114)$$

( $n$  is the number of contexts) is violated. In the following theorem, we will show that by performing node-wise XOR operation on two copies of a contextual box, one can concentrate the contextuality content in terms of increasing the value  $\beta$ .

*Theorem 3.* Let  $\Lambda : \mathcal{B}_v \otimes \mathcal{B}_v \rightarrow \mathcal{B}_v$  be a linear and noncontextuality-preserving node-wise operating map. There exists a map  $\Lambda$  such that

$$\beta_{B_x}(\Lambda(B_v^{\otimes 2})) > \beta_{B_x}(B_v), \quad (115)$$

where  $B_v \in \mathcal{B}_v$ , and  $\beta_{B_x}(B_v) = 2^{m-1} \langle B_x | B_v \rangle$ , where  $B_x$  is an extremal isotropic XOR box.

The theorem says that there exists a noncontextuality-preserving map which can be used for distillation of contextuality. Indeed, this result holds for any even number of copies of a box  $B_v$ . In the following, we will prove the theorem for  $\Lambda$  being a node-wise XOR operation. Then, we will show that a node-wise XOR operation is noncontextuality preserving, i.e., it cannot distill a contextual resource from noncontextual boxes  $B \in \mathcal{B}_{nv}$ .

*Proof.* Let  $\Lambda_{\text{XOR}}$  be the node-wise XOR operation, i.e., for any context  $c_i$  it acts as:

$$p(a_1, \dots, a_m | c_i) \otimes p(a'_1, \dots, a'_m | c_i) \xrightarrow{\Lambda_{\text{XOR}}} p(a_1 \oplus a'_1, \dots, a_m \oplus a'_m | c_i). \quad (116)$$

Consider probability distributions  $p^{od}, p^e$ , which constitute an XOR box. Let us check how  $\Lambda_{\text{XOR}}$  acts on different compositions of  $p^{od}$  and  $p^e$ :

$$\begin{aligned} p^e \otimes p^e &= p(\bigoplus_j^m a_j = 0 | c_i) \otimes p(\bigoplus_j^m a'_j = 0 | c_i) \\ &\xrightarrow{\Lambda_{\text{XOR}}} p(\bigoplus_j^m (a_j \oplus a'_j) = 0 | c_i) \\ &= p^e, \end{aligned} \quad (117)$$

where we used a simple identity  $(\bigoplus_j^m a_j) \oplus (\bigoplus_j^m a'_j) = \bigoplus_j^m (a_j \oplus a'_j)$ . Similarly, one can show the following:

$$\begin{aligned} p^e \otimes p^{od} &\xrightarrow{\Lambda_{\text{XOR}}} p^{od}, \\ p^{od} \otimes p^e &\xrightarrow{\Lambda_{\text{XOR}}} p^{od}, \\ p^{od} \otimes p^{od} &\xrightarrow{\Lambda_{\text{XOR}}} p^e. \end{aligned} \quad (118)$$

Consider now a box  $B$  defined as a linear combination of the extremal isotropic XOR box  $B_x$  and a correlated box  $B_c$ :

$$B = \alpha B_x + (1 - \alpha) B_c. \quad (119)$$

It is easy to verify that

$$\begin{aligned} \beta_{B_x}(B) &= 2^{m-1} \langle B_x | B \rangle \\ &= 2^{m-1} [\alpha \langle B_x | B_x \rangle + (1 - \alpha) \langle B_x | B_c \rangle] \\ &= (n - 1 + \alpha) \end{aligned} \quad (120)$$

since  $\langle B_x | B_x \rangle = n/2^{m-1}$  and  $\langle B_x | B_c \rangle = (n - 1)/2^{m-1}$ . We see that the contextuality inequality (114) is violated for any  $\alpha \in (0, 1]$ , therefore,  $B \in \mathcal{B}_v$  except  $\alpha = 0$ .

For two copies of the box  $B$ , we have

$$\begin{aligned} B^{\otimes 2} &= \alpha^2 B_x^{\otimes 2} + (1 - \alpha)^2 B_c^{\otimes 2} \\ &\quad + \alpha(1 - \alpha)(B_x \otimes B_c + B_c \otimes B_x), \end{aligned} \quad (121)$$

and after node-wise XOR operation

$$\begin{aligned} \beta_{B_x}(\Lambda_{\text{XOR}}(B^{\otimes 2})) &= 2^{m-1} \langle B_x | \Lambda_{\text{XOR}}(B^{\otimes 2}) \rangle \\ &= 2^{m-1} [\alpha^2 \langle B_x | \Lambda_{\text{XOR}}(B_x^{\otimes 2}) \rangle + (1 - \alpha)^2 \langle B_x | \Lambda_{\text{XOR}}(B_c^{\otimes 2}) \rangle \\ &\quad + \alpha(1 - \alpha) \langle B_x | \Lambda_{\text{XOR}}(B_x \otimes B_c + B_c \otimes B_x) \rangle]. \end{aligned} \quad (122)$$

Taking into account (118) one can show that

$$\begin{aligned} \langle B_x | \Lambda_{\text{XOR}}(B_x^{\otimes 2}) \rangle &= (n - 1) \langle p^e | p^e \rangle + \langle p^{od} | p^e \rangle \\ &= (n - 1) / 2^{m-1}, \end{aligned} \quad (123)$$

$$\begin{aligned} \langle B_x | \Lambda_{\text{XOR}}(B_c^{\otimes 2}) \rangle &= (n - 1) \langle p^e | p^e \rangle + \langle p^{od} | p^e \rangle \\ &= (n - 1) / 2^{m-1}, \end{aligned} \quad (124)$$

and

$$\begin{aligned} \langle B_x | \Lambda_{\text{XOR}}(B_x \otimes B_c + B_c \otimes B_x) \rangle &= 2(n - 1) \langle p^e | p^e \rangle + 2 \langle p^{od} | p^{od} \rangle = 2n / 2^{m-1} \end{aligned} \quad (125)$$

since for single-context probability vectors  $\langle p^e | p^e \rangle = 1/2^{m-1}$  and  $\langle p^{od} | p^e \rangle = 0$ .

Inserting these values into Eq. (122), we get

$$\beta_{B_x}(\Lambda_{\text{XOR}}(B^{\otimes 2})) = [\alpha^2 + (1 - \alpha)^2](n - 1) + 2\alpha(1 - \alpha)n. \quad (126)$$

Then, by comparing (120) with (126) one can see that for  $0 < \alpha < 1/2$  we obtain

$$\beta_{B_x}(\Lambda_{\text{XOR}}(B^{\otimes 2})) > \beta_{B_x}(B). \quad (127)$$

■

*Note:* We conjecture that node-wise XOR operation is the only operation which results in distillation.

We will now show that node-wise XOR operation is indeed noncontextuality preserving, i.e., it satisfies the axiom (O2). Suppose that we aim to distill contextuality from noncontextual boxes  $B$  and  $B'$ . We need to show that the box  $B'' = \Lambda_{\text{XOR}}(B \otimes B')$  is also noncontextual. Now, since the box  $B = \{p(a_1, \dots, a_m | c_i)\}$  ( $B' = \{p'(a'_1, \dots, a'_m | c_i)\}$ ) is noncontextual, then there exists a joint probability distribution  $p(a_1, a_2, \dots) (p'(a'_1, a'_2, \dots))$  for all observables in  $\mathcal{M}$ . Denote  $\lambda$  ( $\lambda'$ ) as a string of  $2^{|\mathcal{M}|}$  outputs  $(a_1, a_2, \dots)$  ( $(a'_1, a'_2, \dots)$ ), so that  $p(a_1, a_2, \dots) (p'(a'_1, a'_2, \dots))$  is a linear combination of deterministic points indexed by  $\lambda$  ( $\lambda'$ ), each with probability  $p(\lambda)$  ( $p'(\lambda')$ ). Note that a node-wise XOR operation is a map  $\Lambda_{\text{XOR}} : \{\lambda\} \times \{\lambda'\} \rightarrow \{\lambda''\}$ , where the string of outputs  $\lambda''$  are defined by

$$(a''_1, a''_2, \dots) = (a_1 \oplus a'_1, a_2 \oplus a'_2, \dots). \quad (128)$$

The box  $B''$  is then a linear combination of deterministic points indexed by  $\lambda''$ , each with probability

$$p''(a''_1, a''_2, \dots) = \sum_{\{\lambda\} \times \{\lambda'\} \oplus} p(a_1, a_2, \dots) p'(a'_1, a'_2, \dots), \quad (129)$$

where the above sum is over all composition of strings  $\lambda$  and  $\lambda'$  such that (128) holds. For example, in case of  $|\mathcal{M}| = 2$  we would have, e.g., for a string  $(a''_1 = 0, a''_2 = 1)$

$$p''(01) = p(00)p'(01) + p(01)p'(00) + p(10)p'(11) + p(11)p'(10). \quad (130)$$

Note also that  $p''(a''_1, a''_2, \dots)$  forms a well-defined probability distribution because summing all probabilities

$$\begin{aligned} \sum_{\{\lambda''\}} p''(a''_1, a''_2, \dots) &= \sum_{\{\lambda\} \times \{\lambda'\}} p(a_1, a_2, \dots) p'(a'_1, a'_2, \dots) \\ &= \sum_{\{\lambda\}} \sum_{\{\lambda'\}} p(a_1, a_2, \dots) p'(a'_1, a'_2, \dots) \\ &= 1. \end{aligned} \quad (131)$$

The first equality is based on the observation that the inverse image of the map  $\Lambda_{\text{XOR}}$  for all elements in  $\{\lambda''\}$  results in disjoint partitions of the entire product set  $\{\lambda\} \times \{\lambda'\}$ . Thus, we have shown that the box  $B''$  is noncontextual since there exists a joint probability distribution  $p''(a''_1, a''_2, \dots)$  which defines  $B''$ . Similarly, one can show that any node-wise operation is also noncontextuality preserving.

### C. Towards application of Proposition 3

We can pass now to consider for which resources and measures the assumptions of the above proposition are satisfied. We consider a bipartite scenario with nonlocal correlations as a resource, as for that contextuality follows the same lines, and faces the same problems.

(1) *Possible bound via  $X_{\text{max}}$ .* Consider  $B_v^T$  to be  $PR$  box, and a measure  $X$  to be  $X_{\text{max}}$ . Then,  $X_{\text{max}}$  is additive on  $B_v^T$  (see [16]). By Theorem 2,  $X_{\text{max}}$  is also asymptotically continuous. We can consider distillation protocol via restricted set of operations, namely, the *wirings* [23]. These operations transform suitably defined nonvaluable boxes to nonvaluable ones, i.e., local boxes into local ones [32] (see easier formulation [33]), hence satisfy the axiom (O2) (see Appendix C). However, one would need a proof that  $X_{\text{max}}$  does not increase under wirings, which we leave as an open question. It is easy to check that for isotropic boxes  $PR_\alpha = \alpha B_{000} + (1 - \alpha) B_{001}$  (for a formal definition see [16]) this bound would be nontrivial in the whole range of  $\alpha \in (3/4, 1]$ .

(2) *On bound via  $X_u$ .* Similarly, as for  $X_{\text{max}}$ , the measure  $X_u$  is additive on  $PR$  boxes, and is asymptotically continuous via a proof analogous to that of Theorem 2. However, it is definitely not monotonous under general operations which satisfy the axiom (O2), i.e., those that transform local boxes into local ones. This is because it can increase under partial trace. Indeed, consider a box with a hypergraph  $G$  equal to a direct sum of two hypergraphs  $G_1 \oplus G_2$ , such that  $G_2$  has two vertices connected by a single edge, and the context corresponding to this edge is local with Alice. Let also  $G_1$  be a hypergraph of a nonlocal box. Let now the parties have a box  $B$  corresponding to  $G$ , which is  $PR$  box, and a local box with Alice called  $L$  so that the box  $B$  equals  $PR \oplus L$ . By Theorem 8 of [16] there is

$$X_u(PR \oplus L) = \frac{4}{5} X_u(PR) + \frac{1}{5} X_u(L). \quad (132)$$

Now, since  $X_u(L) = 0$  as this box is local, we have that  $X_u(PR \oplus L) < X_u(PR)$ , hence, by removing or adding  $L$  one can increase or decrease the value of  $X_u$ . Let us note here that  $X_{\text{max}}$  does not suffer from the same problem, as

$$\begin{aligned} X_{\text{max}}(PR \oplus L) &= \max\{X_{\text{max}}(PR), X_{\text{max}}(L)\} \\ &= X_{\text{max}}(PR), \end{aligned} \quad (133)$$

so that adding or removing a local box does not change the value of  $X_{\text{max}}$ . Despite the fact that  $X_u$  is not monotonous under locality-preserving operations, monotonicity under wirings is still possible for it, which we also leave as an open problem.

It is worth mentioning that while considering the measure of contextuality  $X_u$ , we observe that it is a normalized version of nonlocality quantifier as referred in Ref. [11]. Notice, however, that although the unnormalized statistical distance measure of nonlocality, given by infimum over local distributions may increase under local transformations (in particular enlarging the number of inputs of a box [11]), it is not necessarily so when the number of added new inputs are properly accounted. Thus, a normalized measure of contextuality ( $X_u$  as well as  $X_{\text{max}}$ ) prevents the increase of relative entropy while trivial expansion of the number of contexts takes place.

## VII. CONCLUSIONS

Using an axiomatic approach common to resource theories, we have developed the theories of contextuality, and its most celebrated example, which is nonlocality. Crucially from the experimental point of view, we have studied the axiom of asymptotic continuity, and proved that recently established



measure of contextuality: the *relative entropy of contextuality* [16] obeys that axiom. We thereby have showed that for an experimental setup which produces an imperfect box  $B'$ , close to the intended box  $B$ , the amount of contextuality measured by the relative entropy of contextuality  $X(B')$  cannot differ from  $X(B)$  by more than the distance  $\|B - B'\|$  with a factor depending only logarithmically on the dimension of the boxes.

We have also considered a general measure of resource  $X$ , with properties satisfying three proposed axioms: faithfulness, local invariance, and convexity. We have focused on boxes  $B$  from the polytope satisfying vertex equivalence property, i.e., which is such that all its contextual vertices are reversibly exchangeable into each other. We have shown that in such polytopes the measure  $X$  is upper bounded by the measure called the *cost of the resource*  $C(B)$  with a multiplication factor  $X(E_v)$  for some extremal valuable box  $E_v$ . Interestingly, due to this factor, we were able to bound an *extensive measure* (which grows linearly with number of copies), by a *nonextensive one* (which takes values in  $[0, 1]$  on any box irrespective of its dimension). The mentioned bound is linear function of the box. It would be interesting to find a nonlinear one, which is more tight and still easily computable. We have supported the latter results by two examples of its application: for bipartite boxes with binary inputs and outputs, as well as for the boxes related to contextual chain box. Analogous, but weaker, upper bound holds in the case of the polytopes  $\mathcal{B}$ , which do not satisfy the vertex equivalence property of  $\mathcal{B}$ .

We have studied a distillation protocol of a valuable target box  $B_v^T$  from many copies of some input boxes  $B$ , and in full analogy with theory of entanglement measures, we have provided an upper bound on the rate of distillability of the resource  $D(B_v^T | B)$ . It is expressed by a measure of resource  $X$  which satisfies another two proposed axioms: monotonicity under allowed class of operations, asymptotic continuity and superadditivity on target boxes:  $X([B_v^T]^{\otimes k}) \geq kX(B_v^T)$ . From our investigation, we can conclude that the relative entropy of contextuality for bipartite boxes with two binary inputs and outputs may be an upper bound on distillable nonlocality in the form of the Popescu-Rohrlich boxes. The only fact which needs to hold for the latter to be true is the nonincreasing of this measure under wirings. We leave this remaining question as an open problem.

Finally, checking whether other measures of contextuality or nonlocality such as, e.g., [36,37] satisfy proposed axioms, would be vital for their further use, and it would be also interesting to find new ones which satisfy the axioms by definition.

*Note added in proof.* Recently, we became aware of the results of [38]. It seems that our results can be set in a more general framework of general resource theories formulated there in more abstract language. We also note that in [39] a general formalism for contextuality scenarios based on the combinatorics of hypergraphs has been developed.

#### ACKNOWLEDGMENTS

K.H. thanks J. I. de Vicente for discussion. This work is supported by ERC Advanced Grant QOLAPS, and the Polish Ministry of Science and Higher Education Grant No. IdP2011 000361. P.J. is supported by Grant No. MPD/2009-3/4 from

Foundation for Polish Science. Part of this work was done in National Quantum Information Centre of Gdańsk.

#### APPENDIX A: FORMAL PROOF OF LEMMA 1

Here, we give the formal proof of Lemma 1.

*Proof.* By definition of infimum, for any  $\delta_n > 0$  there exists a sequence  $\rho_n$  such that

$$f_T^* := \inf_{\rho \in T} f(\rho) \leq f(\rho_n) \leq f_T^* + \delta_n. \quad (\text{A1})$$

Moreover, by the assumption (47) there exists a sequence  $\sigma_n$ , such that

$$f(\rho_n) - g(\delta) \leq f(\sigma_n) \leq f(\rho_n) + g(\delta). \quad (\text{A2})$$

Combining the above two sequences of inequalities we obtain

$$f_T^* - g(\delta) \leq f(\sigma_n) \leq f_T^* + \delta_n + g(\delta). \quad (\text{A3})$$

Now, there exists  $n_0$  such that for every  $n \geq n_0$  there holds  $\delta > \delta_n$ , and hence

$$f_T^* - g(\delta) - \delta \leq f(\sigma_n) \leq f_T^* + \delta + g(\delta). \quad (\text{A4})$$

This means that we obtained a sequence  $f(\sigma_n)$  which is bounded (we use here the fact that the infima are bounded), and by Bolzano-Weierstrass theorem there exists a subsequence  $n_k$ , such that  $f(\sigma_{n_k})$  has a limit. Thus, we have in particular

$$\lim_{n \rightarrow \infty} f(\sigma_{n_k}) - f_T^* \leq g(\delta) + \delta. \quad (\text{A5})$$

Since by definition  $f_{T'}^* := \inf_{T'} f(\sigma) = \lim_{n \rightarrow \infty} f(\sigma_{n_i})$  for some sequence  $\{\sigma_{n_i}\}$ , we have from the above inequality that  $\{\sigma_{n_k}\}$  may be suboptimal (the infimum over a set is the infimum of the set of limits of sequences from this set), hence,

$$f_{T'}^* - f_T^* \leq g(\delta) + \delta. \quad (\text{A6})$$

Analogously, exchanging  $T$  and  $T'$  we can arrive at

$$f_T^* - f_{T'}^* \leq g(\delta) + \delta, \quad (\text{A7})$$

which proves the thesis for infima. The proof for supremum goes analogously, with only a change of inequalities to opposite and signs in front of  $\delta_n$  in (A1), which leads us exactly to the expression (A4), but for the supremum. The rest of the proof goes symmetrically, hence we skip it. ■

#### APPENDIX B: CONVEXITY OF $I_{\max}$

In this section, we will present an explicit proof of another property of the measure of contextuality, which is its convexity. This property was used in Ref. [27] (see Eq. (2)), but without formal proof. We will first prove convexity of  $I_{p(c)}$  and then using the definition of supremum we will show convexity of the measure  $I_{\max}$ . Note that the convexity of the mutual information of contextuality  $I_{\max}$  means that the relative entropy of contextuality  $X_{\max}$  is also convex because of equivalence of the two measures.

Let us denote  $B_{\text{mix}}$  as a convex combination of boxes:

$$B_{\text{mix}} = \sum_i p_i B_i, \quad (\text{B1})$$

where  $B_i$  are not necessarily extremal (or deterministic) boxes. Then, by definition of a box we have

$$p_{B_{\text{mix}}}(\lambda_c) = \sum_i p_i p_{B_i}(\lambda_c). \quad (\text{B2})$$

We now have the following:

$$\begin{aligned} I_{p(c)}\left(\sum_i p_i B_i\right) &= I_{p(c)}(B_{\text{mix}}) \\ &= \min_{p(\lambda)} \sum_c p(c) D(p_{B_{\text{mix}}}(\lambda_c) \| p(\lambda_c)) \\ &\leq \sum_c p(c) D\left(p_{B_{\text{mix}}}(\lambda_c) \left\| \sum_i p_i p^{i*}(\lambda_c)\right.\right), \end{aligned} \quad (\text{B3})$$

where  $p^{i*}(\lambda_c)$  is a marginal distribution obtained from a joint probability distribution  $p^{i*}(\lambda)$  optimal for a particular box  $B_i$ . The above inequality comes from the fact that the distribution  $p^*(\lambda) = \sum_i p_i p^{i*}(\lambda)$  does not necessarily give a desired minimum over all distributions  $p(\lambda)$ . Furthermore, we have

$$\begin{aligned} &\sum_c p(c) D\left(\sum_i p_i p_{B_i}(\lambda_c) \left\| \sum_i p_i p^{i*}(\lambda_c)\right.\right) \\ &\leq \sum_i p_i \sum_c p(c) D(p_{B_i}(\lambda_c) \| p^{i*}(\lambda_c)) \\ &= \sum_i p_i I_{p(c)}(B_i), \end{aligned} \quad (\text{B4})$$

where the inequality comes from joint convexity of relative entropy distance for each  $c$ , while for the last equality we utilized the optimality of  $p^{i*}(\lambda)$  for each box  $B_i$ . Using the results (B3) and (B4) we arrive at

$$I_{p(c)}\left(\sum_i p_i B_i\right) \leq \sum_i p_i I_{p(c)}(B_i). \quad (\text{B5})$$

Now, by the definition of supremum, for any  $\delta_n > 0$  there exists a distribution  $p_n(c)$ , such that

$$I_{\max}\left(\sum_i p_i B_i\right) \leq I_{p_n(c)}\left(\sum_i p_i B_i\right) + \delta_n, \quad (\text{B6})$$

hence, by convexity of  $I_{p_n(c)}$  we have

$$I_{\max}\left(\sum_i p_i B_i\right) \leq \sum_i p_i I_{p_n(c)}(B_i) + \delta_n. \quad (\text{B7})$$

Notice that for each  $i$  the definition of  $I_{\max}$  assures that  $I_{p_n(c)}(B_i) \leq I_{\max}(B_i)$ . Thus,

$$I_{\max}\left(\sum_i p_i B_i\right) \leq \sum_i p_i I_{\max}(B_i) + \delta_n, \quad (\text{B8})$$

and because  $\delta_n$  can be arbitrarily small, we obtain the desired convexity of  $I_{\max}$ .

### APPENDIX C: LOCALLY PERFORMED WIRINGS SATISFY THE AXIOM (O2)

In this section, we present a formal proof of the fact that if Alice and Bob have access to  $n$  boxes, such that the collection of the latter admits a local hidden variable model, then by means of locally performed wirings [23] one cannot transform the collection of boxes into a valuable (nonlocal) resource shared by the two parties.

Consider then a collection of  $n$  boxes shared by Alice and Bob which admits a local hidden variable model with respect to both parties:

$$B_L^n = \sum_{\lambda} p_{\lambda} p^{(\lambda)}(\mathbf{a}|\mathbf{x}) \otimes p^{(\lambda)}(\mathbf{b}|\mathbf{y}), \quad (\text{C1})$$

for some probability distribution  $\{p_{\lambda}\}$ , where  $\mathbf{a} = (a_1, \dots, a_m)$  ( $\mathbf{b} = (b_1, \dots, b_m)$ ) is the vector of Alice's (Bob's) outputs when one of the input from  $\mathbf{x} = (x_1, \dots, x_n)$  ( $\mathbf{y} = (y_1, \dots, y_n)$ ) is chosen.

We will assume that *locally* the distribution of the collection of boxes as seen by one party (e.g., Alice)  $p(\mathbf{a}|\mathbf{x})$  is nonsignaling [33], i.e., the following conditions are satisfied:

$$\forall_{1 \leq i \leq m, \mathbf{a}^{\neq i}, \mathbf{x}^{\neq i}, x_i, x'_i} \sum_{a_i} p(\mathbf{a}|\mathbf{x}^{\neq i}, x_i) = \sum_{a_i} p(\mathbf{a}|\mathbf{x}^{\neq i}, x'_i), \quad (\text{C2})$$

and analogously for Bob. Note that the nonsignaling conditions given above imply nonsignaling with respect to all subsets of inputs, i.e., marginal distribution of the outputs  $\mathbf{a}^{\neq i, j, \dots}$  does not depend on changing the inputs  $\mathbf{x}^{\neq i, j, \dots}$  [35].

Consider now the partition of constituent boxes  $A_1 : A_2 : B$ , where  $A_1 \equiv \{x_1, \dots, x_k\}$ ,  $A_2 \equiv \{x_{k+1}, \dots, x_n\}$ ,  $B \equiv \{y_1, \dots, y_n\}$  for an arbitrary  $1 \leq k \leq n-1$ . As it was shown in Ref. [34], the locality in the partition  $A_1, A_2 : B$  may not be preserved when the subsystems  $A_1$  and  $A_2$  cooperate, i.e., when they perform a suitable wiring. This happens when no constraints are imposed on the distribution  $p^{(\lambda)}(\mathbf{a}|\mathbf{x})$  in the decomposition (C1). However, when the local distribution admits nonsignaling conditions (C2), then the operation of wiring of the subsystems  $A_1$  and  $A_2$  will not lead to emergence of signaling for the one-partite distribution  $p(\mathbf{a}|\mathbf{x})$ . Since such nonsignaling bilocal distributions (NSBL) constitute a closed set under wirings [34], then we see that locally performed wirings will not produce a valuable (nonlocal) resource from useless (local) objects (see Supplemental Material of Ref. [34], where the nonsignaling conditions (C2) need to be assumed).

[1] S. Kochen and E. P. Specker, J. Math. Mech. **17**, 59 (1967).

[2] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).

[3] A. Cabello, Phys. Rev. Lett. **101**, 210401 (2008).

[4] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Phys. Rev. Lett. **101**, 020403 (2008).

- [5] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger, *Nature (London)* **474**, 490 (2011).
- [6] Kirchmair *et al.*, *Nature (London)* **460**, 494 (2009).
- [7] R. Raussendorf, *Phys. Rev. A* **88**, 022322 (2013).
- [8] G. Cañas, M. Arias, S. Etcheverry, E. S. Gómez, A. Cabello, G. B. Xavier, and G. Lima, *Phys. Rev. Lett.* **113**, 090404 (2014).
- [9] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, *Rev. Mod. Phys.* **86**, 419 (2014).
- [10] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [11] J. I. de Vicente, *J. Phys. A: Math. Theor.* **47**, 424017 (2014).
- [12] R. Gallego and L. Aolita, [arXiv:1409.5804](https://arxiv.org/abs/1409.5804).
- [13] B. Coecke, T. Fritz, and R. W. Spekkens, [arXiv:1409.5531](https://arxiv.org/abs/1409.5531).
- [14] M. Howard, J. Wallman, V. Veitch, and J. Emerson, *Nature (London)* **510**, 351 (2014).
- [15] M. Christandl and B. Toner, *J. Math. Phys.* **50**, 042104 (2009).
- [16] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, *Phys. Rev. Lett.* **112**, 120401 (2014).
- [17] P. Rastall, *Found. Phys.* **15**, 963 (1985).
- [18] S. Popescu and D. Rohrlich, *Found. Phys.* **24**, 379 (1994).
- [19] M. Araújo, M. T. Quintino, C. Budroni, M. T. Cunha, and A. Cabello, *Phys. Rev. A* **88**, 022118 (2013).
- [20] M. Forster, S. Winkler, and S. Wolf, *Phys. Rev. Lett.* **102**, 120401 (2009).
- [21] A. Fine, *Phys. Rev. Lett.* **48**, 291 (1982).
- [22] J. Barrett, *Phys. Rev. A* **75**, 032304 (2007).
- [23] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vertesi, *Phys. Rev. A* **80**, 062107 (2009).
- [24] W. van Dam, R. D. Gill, and P. D. Grünwald, *IEEE Trans. Inf. Theory* **51**, 2812 (2005).
- [25] B. Synak-Radtke and M. Horodecki, *J. Phys. A: Math. Gen.* **39**, L423 (2006).
- [26] R. Alicki and M. Fannes, *J. Phys. A: Math. Gen* **37**, L55 (2004).
- [27] R. Ramanathan and P. Horodecki, *Phys. Rev. Lett.* **112**, 040404 (2014).
- [28] S. L. Braunstein and C. M. Caves, *Ann. Phys. (NY)* **202**, 22 (1990).
- [29] S. Wehner, *Phys. Rev. A* **73**, 022110 (2006).
- [30] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, *Phys. Rev. A* **53**, 2046 (1996).
- [31] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wothers, *Phys. Rev. A* **54**, 3824 (1996).
- [32] P. Joshi, A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, *Quant. Inf. Comp.* **13**, 567 (2013).
- [33] K. Horodecki, [arXiv:1401.4899](https://arxiv.org/abs/1401.4899).
- [34] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, *Phys. Rev. Lett.* **109**, 070401 (2012).
- [35] E. Hänggi, R. Renner, and S. Wolf, *Eurocrypt* **2010**, 216 (2010).
- [36] K. Svozil, *Nat. Comp.* **11**, 261 (2012).
- [37] M. Kleinmann, O. Guehne, J. R. Portillo, J. Larsson, and A. Cabello, *New J. Phys.* **13**, 113011 (2011).
- [38] B. Coecke, T. Fritz, and Robert W. Spekkens, [arXiv:1409.5531](https://arxiv.org/abs/1409.5531).
- [39] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, *Commun. Math. Phys.* **334**, 533 (2015).