

# Measures of Contextuality and Noncontextuality

Janne V. Kujala<sup>1</sup> and Ehtibar N. Dzhafarov<sup>2</sup>

<sup>1</sup>University of Turku, janne.kujala@utu.fi

<sup>2</sup>Purdue University, ehtibar@purdue.edu

## Abstract

We discuss three measures of the degree of contextuality in contextual systems of dichotomous random variables. These measures are developed within the framework of the Contextuality-by-Default (CbD) theory, and apply to inconsistently connected systems (those with “disturbance” allowed). For one of these measures of contextuality, presented here for the first time, we construct a corresponding measure of the degree of noncontextuality in noncontextual systems. The other two CbD-based measures do not suggest ways in which degree of noncontextuality of a noncontextual system can be quantified. We find the same to be true for the contextual fraction measure developed by Abramsky, Barbosa, and Mansfield. This measure of contextuality is confined to consistently connected systems, but CbD allows one to generalize it to arbitrary systems.

KEYWORDS: contextuality, connectedness, disturbance, measures of contextuality, measures of noncontextuality

## 1 Introduction

**1.1** We will consider certain *measures of contextuality* (degree of contextuality in a contextual system) and see if they can be naturally extended into *measures of noncontextuality* (degree of noncontextuality in a noncontextual system). What we mean by an extension being “natural” is that it uses essentially the same mathematical construction as the measure of contextuality being extended. Let us illustrate this by an example. Let  $\mathcal{R}$  be a system of random variables, and let  $F(\mathcal{R})$  be a real-valued continuous functional, in the sense that small changes in the distributions of  $\mathcal{R}$  result in small changes of  $F(\mathcal{R})$ . Let the following Bell-type inequality be accepted as a definition, or derived as a theorem: the system  $\mathcal{R}$  is noncontextual if and only if  $F(\mathcal{R}) \leq 0$ . Suppose that in the universe of possible systems  $\mathcal{R}$  the value of  $F(\mathcal{R})$  varies on the interval  $(a, b)$ , with  $a < 0 < b$ . It is natural then to consider a positive value of  $F(\mathcal{R})$  as the degree of contextuality of  $\mathcal{R}$ , increasing as  $F(\mathcal{R})$  increases from 0 to  $b$ :

$$F(\mathcal{R}) > 0 \implies \text{CNT} = F(\mathcal{R}). \quad (1)$$

Equally naturally, this measure can be extended to a measure of noncontextuality, increasing as  $F(\mathcal{R})$  decreases from 0 to  $a$ :

$$F(\mathcal{R}) \leq 0 \implies \text{NCNT} = -F(\mathcal{R}). \quad (2)$$

By contrast, if the functional  $F(\mathcal{R})$  varied on an interval  $[0, b)$ , the degree of contextuality would be defined as before, but it would not naturally extend to a measure of noncontextuality: all noncontextual system would be mapped into zero, so any extension would require ideas and principles other than those used in the construction of the functional  $F$ .

**1.2** We will consider three contextuality measures, all based on the Contextuality-by-Default (CbD) theory and applicable to arbitrary systems of dichotomous random variables.<sup>1</sup> Two of our measures,  $\text{CNT}_1$  and  $\text{CNT}_2$ , are, in a well-defined sense, mirror images of each other, but we will see that only one of them,  $\text{CNT}_2$ , is naturally extendable to a measure of noncontextuality.  $\text{CNT}_2$  for a contextual system of random variables is defined as the  $L_1$ -distance between the surface of a certain polytope and an external point representing the system. The points lying on or inside the polytope represent noncontextual systems, and it is natural to define the extension of  $\text{CNT}_2$  into a measure of noncontextuality,  $\text{NCNT}_2$ , as the  $L_1$ -distance from an internal point of the polytope to its surface.  $\text{CNT}_1$ , too, can be defined as the  $L_1$ -distance between a certain polytope and an external point representing a contextual system. However, all noncontextual systems

---

<sup>1</sup>In the contemporary version of CbD [18, 20], any system of random variables is to be presented in a *canonical form*, one in which each original random variable is replaced with a set of jointly distributed dichotomous variables.

in this case are represented by points lying on the surface of the polytope, as points of zero contextuality. As a result, any extension of  $\text{CNT}_1$  into a measure of noncontextuality would require that one go beyond the construction underlying  $\text{CNT}_1$ .

**1.3** The third CbD-based measure,  $\text{CNT}_3$ , is of a different kind. Here, one maps the system into a certain distribution of quasiprobabilities, numbers that sum to unity but are allowed to be negative.  $\text{CNT}_3$  is measured by how small the negative part of the quasiprobability distribution can be made: the larger this minimal negative mass the more contextual the system. This measure is not naturally extendable to a measure of noncontextuality because all noncontextual systems are identically characterized by this negative mass being zero.

**1.4** We also consider the measure of contextuality called *contextual fraction*, proposed in Refs. [1,6,24] and developed in Ref. [2]. The logic of this measure is similar to that of  $\text{CNT}_3$ : contextuality is measured by how close certain quasiprobabilities (in this case, nonnegative numbers allowed to sum to less than unity) can be made to a proper probability distribution. The measure has been only formulated under the constraint that random variables measuring the same property in different contexts are identically distributed. This constraint, called *consistent connectedness* in CbD, is more generally known as the *no-disturbance* principle (or “no-signaling”, in the case of spatially distributed systems). We provide a CbD-based generalization of contextual fraction to arbitrary systems, and show that this measure, too, does not have a natural noncontextuality counterpart.

**1.5** It is important here to dispel a possible confusion. Any measure of the degree of contextuality in a contextual system can be associated with some complementary measure that can be interpreted as the degree of noncontextuality in this contextual system. Thus, in the opening example of this paper, if  $F(\mathcal{R}) > 0$ , one could define its contextuality degree as  $F(\mathcal{R})/b$ , and consider  $1 - F(\mathcal{R})/b$  the “degree of noncontextuality” in this contextual system. In Abramsky, Barbosa, and Mansfield’s construction [2] a generalized version of this construction is introduced explicitly: 1 minus contextual fraction is *noncontextual fraction*, for any contextual system. Our usage of the term “measure of noncontextuality” is different — it refers to a *degree of noncontextuality in a noncontextual system*. The noncontextual fraction of Abramsky and colleagues is not a measure of noncontextuality in this sense because it identically equals 1 for all noncontextual systems. Returning to our opening example,  $1 - F(\mathcal{R})/b$  is not a measure of noncontextuality in our sense, because it is predicated on contextuality,  $F(\mathcal{R}) > 0$ . If  $F(\mathcal{R}) \leq 0$ , something like  $F(\mathcal{R})/a$ , assuming  $a < 0$ , would be an appropriate measure of noncontextuality (and, if one so wishes, one could define  $1 - F(\mathcal{R})/a$  as the “contextual fraction” of this noncontextual system).

**1.6** One’s interest to measures of noncontextuality can be justified in essentially the same way as one’s interest to measures of contextuality, except that one knows much more about the latter. Thus, one should be interested if a system, be it contextual or noncontextual, is stably so: whether a small perturbation of the random variables it is comprised of will change its (non)contextuality status. Larger values of (non)contextuality mean more stable (non)contextuality. A closely related reason is statistical. If contextuality or noncontextuality of a system is established on a sample level, one should be interested in whether this finding is reliable: e.g., whether a high-level confidence interval for its contextuality or noncontextuality degree lies entirely in the range of contextuality or noncontextuality values, respectively. This is an especially important task in fields outside quantum physics, e.g., in the contextuality analysis of human behavior [8,12,23]. Non-physical applications provide additional reasons for one’s interest in measures of noncontextuality: e.g., some models of decision making can predict both contextual and noncontextual systems of random variables, and noncontextual systems may be linked to features of decision making that are, if anything, of greater interest than those in contextual systems [8]. In quantum physics, there is a growing interest to the question of whether certain classical systems could exhibit contextuality similar to that found in quantum systems [25,29,32]. Here, it might be useful to quantify the “classicality” and “non-classicality” of systems by measures of, respectively, noncontextuality and contextuality, preferably chosen so that they form each other’s natural extensions. One can argue that degree of contextuality has been linked to quantum advantage in computation, communication complexity, and other matters of intrinsic or practical interest [9,10,26], while nothing like this is currently known about noncontextuality. However, this may very well be due to the simple fact that no measures of noncontextuality have so far been proposed and studied.

**1.7** The three measures of contextuality considered in this paper are CbD-based, which means that they are not constrained by the assumption of consistent connectedness. We had to leave out a large number of interesting contextuality measures discussed in the literature under the consistent connectedness constraint (with the exception of the contextual fraction measure that we generalize to apply to arbitrary systems). Thus, most of the measures of nonlocality (as a special case of contextuality) reviewed in Ref. [11] cannot be naturally extended to measures of locality (noncontextuality). The

measures of contextuality constructed in relation to Bell-type criteria of noncontextuality for consistently connected systems, as, e.g., in Ref. [4], usually can be extended to measures of noncontextuality along the lines of our opening example. The fact that we do not discuss these measures in detail is a reflection of the focus of this paper rather than our view of their relative importance.

## 2 Basics of the Contextuality-by-Default approach

**2.1** A *system*  $\mathcal{R}$  of random variables is a set whose elements are random variables  $R_q^c$  labeled in two ways: by their *contents*  $q \in Q$  (that which the random variable measures or responds to) and their *contexts*  $c \in C$  (the conditions under which this random variable is recorded):

$$\mathcal{R} = \{R_q^c : c \in C, q \in Q, q \prec c\}, \quad (3)$$

where  $q \prec c$  indicates that content  $q$  is measured (or responded to) in context  $c$ . Throughout this paper, the set of contents  $Q$  and the set of contexts  $C$  are finite, and all random variables in the system are Bernoulli, with values 0/1.

**2.2** In CbD, with no loss of generality, one can make  $q \prec c$  hold true for all  $q$  and  $c$ , by placing in every “empty”  $(q, c)$ -cell a dummy variable with a single possible value [16]. We will not be using this construction in this paper, as it is convenient to think of the relation  $\prec$  as the *format* of the system  $\mathcal{R}$ , the arrangement of the random variables without information of their distributions.

**2.3** In each context  $c$ , the subset of random variables

$$R^c = \{R_q^c : q \in Q, q \prec c\} \quad (4)$$

is *jointly distributed*, i.e., it is a random variable in its own right. It is referred to as the *bunch* for (or corresponding to) context  $c$ . For each content  $q$ , the subset of random variables

$$\mathcal{R}_q = \{R_q^c : c \in C, q \prec c\} \quad (5)$$

is referred to as the *connection* for (or corresponding to) content  $q$ . The elements of a connection are not jointly distributed, they are *stochastically unrelated*. (This is reflected in the notation,  $R$  vs  $\mathcal{R}$ : the bunch  $R^c$  is a random variable in its own right, while the connection  $\mathcal{R}_q$  is not.) More generally, any  $R_q^c$  and  $R_{q'}^{c'}$  are stochastically unrelated unless  $c = c'$ . The terminology above is illustrated in Fig. 1.<sup>2</sup>

**2.4** A system is *consistently connected* (satisfies the “no-disturbance” requirement) if the distribution of each random variable in it depends on its content only. If this is not the case, the system is *inconsistently connected*. The latter term can also be used for arbitrary systems, that may but need not be consistently connected. Consistent connectedness (non-disturbance) can sometimes be understood in the *strong sense*, as in Refs. [1, 15]: if several contents  $q_1, \dots, q_k$  are measured in two contexts  $c, c'$ , then the joint distributions of  $\{R_{q_1}^c, \dots, R_{q_k}^c\}$  and  $\{R_{q_1}^{c'}, \dots, R_{q_k}^{c'}\}$  coincide. Nothing will change in this paper if consistent connectedness is understood in this strong sense, because we generally assume systems are not consistently connected even in the weaker, more general sense.

**2.5** The general definition of a *coupling* for an indexed set  $\mathcal{X}$  of random variables is that it is a jointly distributed and identically indexed set  $Y$  of random variables such that, for any subset  $X$  of  $\mathcal{X}$  possessing a joint distribution, the corresponding subset of  $Y$  is identically distributed. In particular, every element of  $\mathcal{X}$  is distributed as the corresponding element of  $Y$ . In accordance with this general definition, a coupling of a system  $\mathcal{R}$  is a set of jointly distributed random variables

$$S = \{S_q^c : c \in C, q \in Q, q \prec c\}, \quad (6)$$

such that, for any context  $c$ , the bunches  $S^c$  and  $R^c$  are identically distributed.

<sup>2</sup>One might protest that given the notion of a context and a content, the corresponding notions of a bunch and a connection are unnecessary. It is indeed possible, as we have done in some of our publications, to avoid the use of the latter two terms by speaking instead of context-sharing and content-sharing variables. However, in discussing measures of (non)contextuality and algorithms computing them, the use of the terms in question, e.g., when speaking of “bunch probabilities” and “connection probabilities”, is convenient.

(a)	$R_1^1$	$R_2^1$			$c^1$
		$R_2^2$	$R_3^2$	$R_4^2$	$c^2$
	$R_1^3$		$R_3^3$		$c^3$
	$R_1^4$			$R_4^4$	$c^4$
	$R_1^5$	$R_2^5$	$R_3^5$		$c^5$
	$q_1$	$q_2$	$q_3$	$q_4$	$\mathcal{R}$

(b)	*	*			$c^1$
		*	*	*	$c^2$
	*		*		$c^3$
	*			*	$c^4$
	*	*	*		$c^5$
	$q_1$	$q_2$	$q_3$	$q_4$	$\mathcal{R}$

(c)	$R^1$	*	*		$c^1$
	$R^2$		*	*	$c^2$
	$R^3$	*		*	$c^3$
	$R^4$	*		*	$c^4$
	$R^5$	*	*	*	$c^5$
	$q_1$	$q_2$	$q_3$	$q_4$	$\mathcal{R}$

(d)	$\mathcal{R}_1$	$\mathcal{R}_2$	$\mathcal{R}_3$	$\mathcal{R}_4$	$c^1$
	*	*			$c^2$
	*	*	*	*	$c^3$
	*		*		$c^4$
	*	*	*	*	$c^5$
	$q_1$	$q_2$	$q_3$	$q_4$	$\mathcal{R}$

Figure 1: Illustration for the basic terms of CbD. (a) A system  $\mathcal{R}$  of random variables, with 4 contents measured in 5 contexts. Each variable in the system is uniquely identified by its content and its context. All random variables are dichotomous, 0/1. (b) The format of the system  $\mathcal{R}$ , showing which content is measured in which context. It can also be used as a simplified representation of  $\mathcal{R}$ , since the identification  $R_q^c$  is uniquely reconstructed from the position of the corresponding star. (c) Shows 5 bunches of the system. The random variables within a bunch are jointly distributed, i.e., each bunch is a random variable. (d) Shows 4 connections of the system. The random variables within a connection are stochastically unrelated.

**2.6** In CbD, we are interested not in just any coupling of  $\mathcal{R}$  but in those with a certain property. It can be introduced as follows. For each separately taken connection  $\mathcal{R}_q$  one can find its *multimaximal coupling*

$$T_q = \{T_q^c : c \in C, q \prec c\}, \quad (7)$$

defined as a coupling in which, for every  $c, c'$  in which  $q$  is measured, the probability of  $T_q^c = T_q^{c'}$  is maximal possible (among all possible couplings of  $\mathcal{R}_q$ , or, equivalently, for given marginal distributions of  $R_q^c$  and  $R_q^{c'}$ ). With only dichotomous variables in play, every connection has a unique multimaximal coupling, and for any subset  $\{R_q^c, R_q^{c'}, \dots, R_q^{c'' \dots'}\}$  of a connection  $\mathcal{R}_q$ , the probability of  $T_q^c = T_q^{c'} = \dots = T_q^{c'' \dots'}$  is maximal possible [20, 21]. It should also be noted that maximization of  $T_q^c = T_q^{c'} = \dots = T_q^{c'' \dots'}$  means that both

$$T_q^c = T_q^{c'} = \dots = T_q^{c'' \dots'} = 1 \quad (8)$$

and

$$T_q^c = T_q^{c'} = \dots = T_q^{c'' \dots'} = 0 \quad (9)$$

are maximized (recall that the values of all variables are encoded 0/1).

**2.7** A system  $\mathcal{R}$  is defined as *noncontextual* if it has a coupling  $S$  whose restrictions to all connections are multimaximal couplings of these connections. The system is *contextual* if no such coupling  $S$  exists. Equivalently, a system  $\mathcal{R}$  is noncontextual if it has a coupling  $S$  which is also a coupling for the multimaximal couplings  $T_q$  of connections  $\mathcal{R}_q$  ( $q \in Q$ ). Yet another way of saying this is to define  $\mathcal{R}$  as contextual if its bunches are incompatible with the multimaximal couplings of its connections (cannot be “sewn together” within a single overall distribution).

### 3 Vectorial representation of systems

**3.1** Any system  $\mathcal{R}$  can be described by a *vector of bunch probabilities*. Abramsky and Brandenburger [1] call it an *empirical model*. For any context  $c$ , assuming the  $n_c$  random variables in its bunch were enumerated  $R_1^c, \dots, R_{n_c}^c$ , we

define

$$\mathbf{p}^{(c)} = \begin{bmatrix} p_1^{(c)} \\ \vdots \\ p_i^{(c)} \\ \vdots \\ p_{2^{n_c}}^{(c)} \end{bmatrix} = \begin{bmatrix} \Pr [R_1^c = 0, \dots, R_j^c = 0, \dots, R_{n_c}^c = 0] \\ \vdots \\ \Pr [R_1^c = r_1, \dots, R_j^c = r_j, \dots, R_{n_c}^c = r_{n_c}] \\ \vdots \\ \Pr [R_1^c = 1, \dots, R_j^c = 1, \dots, R_{n_c}^c = 1] \end{bmatrix}, \quad (10)$$

where  $\Pr$  stands for probability, and  $r_1, \dots, r_{n_c}$  run through all  $2^{n_c}$  combinations of 0/1's. The vector of bunch probabilities is defined as

$$\mathbf{p}^{(b)} = \begin{bmatrix} \mathbf{p}^{c_1} \\ \vdots \\ \mathbf{p}^{c_j} \\ \vdots \\ \mathbf{p}^{c_{|C|}} \end{bmatrix} \quad (11)$$

(the boldface index  $\mathbf{b}$  stands for “bunches”).

**3.2** For any content  $q$ , assuming the  $m_q$  elements of the corresponding connection were enumerated  $R_q^1, \dots, R_q^{m_q}$ , any coupling  $(T_q^1, \dots, T_q^{m_q})$  (not necessarily multimaximal) of this connection is defined by

$$\mathbf{p}_{(q)} = \begin{bmatrix} p_{(q),1} \\ \vdots \\ p_{(q),i} \\ \vdots \\ p_{(q),2^{m_q}} \end{bmatrix} = \begin{bmatrix} \Pr [T_q^1 = 0, \dots, T_q^j = 0, \dots, T_q^{m_q} = 0] \\ \vdots \\ \Pr [T_q^1 = s_1, \dots, T_q^j = s_j, \dots, T_q^{m_q} = s_{m_q}] \\ \vdots \\ \Pr [T_q^1 = 1, \dots, T_q^j = 1, \dots, T_q^{m_q} = 1] \end{bmatrix}, \quad (12)$$

with the same meaning of the terms as in (10). The *vector of connection probabilities* is defined as

$$\mathbf{p}^{(c)} = \begin{bmatrix} \mathbf{p}_{q_1} \\ \vdots \\ \mathbf{p}_{q_j} \\ \vdots \\ \mathbf{p}_{q_{|Q|}} \end{bmatrix} \quad (13)$$

(the boldface index  $\mathbf{c}$  stands for “connections”).

**3.3** Finally we stack up the two vectors, for bunches and for connections, to obtain the *complete vector of probabilities*.

$$\mathbf{p}^{(\cdot)} = \begin{bmatrix} \mathbf{p}^{(b)} \\ \mathbf{p}^{(c)} \end{bmatrix}. \quad (14)$$

We include connection couplings in this representation of the system even though they are computed rather than observed. This means that one and the same system can be represented by multiple probability vectors, depending on the couplings we choose for connections.

**3.4** Complete vectors of probabilities will be used in Section 7, when we discuss a generalized version of a measure proposed in Ref. [2]. However, for purposes of computing CbD-based measures of contextuality,  $\mathbf{p}^{(\cdot)}$  is not convenient because of its redundancy: one cannot change any component of  $\mathbf{p}^{(c)}$  or  $\mathbf{p}_{(q)}$  without changing some of its other components. We will deal therefore with one of the numerous versions of a *reduced vector of probabilities* in which components can be changed independently. The variant we choose is introduced in Ref. [19]. It is based on the idea of replacing  $\mathbf{p}^{(c)}$  in (10) with probabilities

$$\Pr [R_{q_i}^c = 1 : i \in I] = \left\langle \prod_{i \in I} R_{q_i}^c \right\rangle \quad (15)$$

for various subsets  $I$  of  $\{1, \dots, n_c\}$ . Analogously,  $\mathbf{p}_{(q)}$  in (12) is replaced with probabilities

$$\Pr [T_q^{c_j} = 1 : j \in J] = \left\langle \prod_{j \in J} T_q^{c_j} \right\rangle \quad (16)$$

for various subsets  $J$  of  $\{1, \dots, m_q\}$ . These probabilities (and also the events whose probabilities they are, when this cannot cause confusion) are referred to as  $k$ -marginals, where  $k = 0, 1, 2, \dots$  is the *order of the marginals* (the number of the random variables involved). The 0-marginal is a constant  $\langle \rangle$  taken to be 1, and the 1-marginals

$$\langle R_q^c \rangle = \Pr [R_q^c = 1] = \Pr [T_q^c = 1] = \langle T_q^c \rangle \quad (17)$$

are shared by the bunches and the connections. Because of this, to avoid redundancy, we put the 0-marginal and all 1-marginals in one group, all higher-order marginals for bunches into a second group, and all higher-order marginals for (couplings of) connections into a third group.

**3.5** Let us order in some way all random variables in the system:  $R_{q_1}^{c_1}, \dots, R_{q_N}^{c_N}$ . Define

$$\mathbf{p}_1 = \begin{bmatrix} 1 \\ \Pr [R_{q_1}^{c_1} = 1] \\ \vdots \\ \Pr [R_{q_i}^{c_i} = 1] \\ \vdots \\ \Pr [R_{q_N}^{c_N} = 1] \end{bmatrix} = \begin{bmatrix} \langle \rangle \\ \langle R_{q_1}^{c_1} \rangle \\ \vdots \\ \langle R_{q_i}^{c_i} \rangle \\ \vdots \\ \langle R_{q_N}^{c_N} \rangle \end{bmatrix}, \quad (18)$$

where the boldface index  $\mathbf{1}$  stands for “low-order marginals”.

**3.6** For a given context  $c$ , let us enumerate  $1, \dots, 2^{n_c} - n_c - 1$  all nonempty and non-singleton subsets of the corresponding bunch:  $\binom{n_c}{2}$  2-marginals followed by  $\binom{n_c}{3}$  3-marginals etc. Define

$$\mathbf{p}^c = \begin{bmatrix} p_1^c \\ \vdots \\ p_i^c \\ \vdots \\ p_{2^{n_c} - n_c - 1}^c \end{bmatrix} = \begin{bmatrix} \Pr [R_1^c = 1, R_2^c = 1] \\ \vdots \\ \Pr [R_{i,1}^c = 1, \dots, R_{i,j}^c = 1, \dots, R_{i,n_{i,c}}^c = 1] \\ \vdots \\ \Pr [R_1^c = 1, \dots, R_j^c = 1, \dots, R_{n_c}^c = 1] \end{bmatrix} = \begin{bmatrix} \langle R_1^c R_2^c \rangle \\ \vdots \\ \langle R_{i,1}^c \dots R_{i,j}^c \dots R_{i,n_{i,c}}^c \rangle \\ \vdots \\ \langle R_1^c \dots R_j^c \dots R_{n_c}^c \rangle \end{bmatrix}, \quad (19)$$

and the *reduced vector of bunch probabilities*

$$\mathbf{p}_b = \begin{bmatrix} \mathbf{p}^{c_1} \\ \vdots \\ \mathbf{p}^{c_j} \\ \vdots \\ \mathbf{p}^{c_{|C|}} \end{bmatrix}. \quad (20)$$

**3.7** We analogously define, having imposed some couplings (not necessarily multimaximal) on the connections,

$$\mathbf{p}_q = \begin{bmatrix} p_{q,1} \\ \vdots \\ p_{q,i} \\ \vdots \\ p_{q,2^{m_q} - m_q - 1} \end{bmatrix} = \begin{bmatrix} \Pr [T_q^1 = 1, T_q^2 = 1] \\ \vdots \\ \Pr [T_q^{i,1} = 1, \dots, T_q^{i,j} = 1, \dots, T_q^{i,m_{i,q}} = 1] \\ \vdots \\ \Pr [T_q^1 = 1, \dots, T_q^j = 1, \dots, T_q^{m_q} = 1] \end{bmatrix} = \begin{bmatrix} \langle T_q^1 T_q^2 \rangle \\ \vdots \\ \langle T_q^{i,1} \dots T_q^{i,j} \dots T_q^{i,m_{i,q}} \rangle \\ \vdots \\ \langle T_q^1 \dots T_q^j \dots T_q^{m_q} \rangle \end{bmatrix}. \quad (21)$$

The *reduced vector of connection probabilities* is

$$\mathbf{p}_c = \begin{bmatrix} \mathbf{p}_{q_1} \\ \vdots \\ \mathbf{p}_{q_j} \\ \vdots \\ \mathbf{p}_{q_{|Q|}} \end{bmatrix}. \quad (22)$$

As mentioned in Section 2.6, with reference to [20,21], if the couplings of connections are chosen to be multimaximal, then all these probabilities are maximal possible, given the values of the corresponding 1-marginal probabilities.

**3.8** Without loss of generality, one can delete from  $\mathbf{p}_c$  all  $k$ -marginals with  $k > 2$ . As shown in Ref. [20], the 1-marginals and 2-marginals define multimaximal couplings of the connections uniquely, and this makes them sufficient for all CbD-based measures of contextuality.

**3.9** Finally,

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_l \\ \mathbf{p}_b \\ \mathbf{p}_c \end{bmatrix} \quad (23)$$

is the *reduced vector of probabilities* representing system  $\mathcal{R}$ .

**3.10** Any component of  $\mathbf{p}^c$  or of  $\mathbf{p}_q$  can generally change its value while other components remain fixed (which is impossible in  $\mathbf{p}^{(c)}$  and  $\mathbf{p}_{(q)}$ ). However, the range of possible changes is limited: every  $k$ -marginal probability is limited from above by any  $k - 1$ -marginal it contains, and from below by any  $(k + 1)$ -marginal containing it.

## 4 Contextuality in vectorial representation

**4.1** Consider a system  $\mathcal{R}$  with  $N$  dichotomous random variables, and let  $\mathbf{v}$  be the  $2^N$ -component vector of possible values of a(ny) coupling  $S$  of the entire system. An element of  $\mathbf{v}$  can be viewed as a conjunction of events

$$\{S_q^c = r_q^c : c \in C, q \in Q, q \prec c\}, \quad (24)$$

with  $r_q^c = 0/1$ . Then any given  $S$  is specified by a  $2^N$ -vector  $\mathbf{x}$  of the probabilities with which the corresponding elements of  $\mathbf{v}$  occur. Clearly,

$$\mathbf{x} \geq 0, \|\mathbf{x}\| = 1, \quad (25)$$

where the inequality is componentwise, and the norm is  $L_1$ . We call  $\mathbf{x}$  a *coupling vector* for  $\mathcal{R}$ .

**4.2** Let  $\mathbf{p}$  be a (reduced) vector of probabilities. Then the  $i$ th component of  $\mathbf{p}$  is the joint probability

$$\Pr [S_q^c = 1 : (c, q) \in D_i] = \left\langle \prod_{(c,q) \in D_i} S_q^c \right\rangle \quad (26)$$

for some  $D_i$ . The latter can be a low-marginal event, in which case it is empty or a singleton; or  $D_i$  can be a bunch event, in which case it consists of a fixed  $c$  paired with two or more  $q$ 's; or else it can be a connection event, in which case it has a fixed  $q$  paired with two or more  $c$ 's (or with precisely two  $c$ 's, in view of Section 3.8).

**4.3** We now construct a Boolean matrix  $\mathbf{M}$  having  $2^N$  columns, with the  $j$ th column being labeled by the  $j$ th value of  $\mathbf{v}$  ( $j = 1, \dots, 2^N$ ). This matrix is the same for all systems in the format of  $\mathcal{R}$ . The  $i$ th row of this matrix is labeled by the event (26) whose probability is the  $i$ th element of  $\mathbf{p}$ . If all the random variables in this event equal to 1 in the  $j$ th value of  $\mathbf{v}$ , then we put 1 in the cell  $(i, j)$  of  $\mathbf{M}$ . All other cells of  $\mathbf{M}$  are filled with zeros.

**4.4** The matrix can be presented as

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_b \\ \mathbf{M}_c \end{pmatrix}, \quad (27)$$

with  $\mathbf{M}_1$ ,  $\mathbf{M}_b$ , and  $\mathbf{M}_c$  corresponding to the  $\mathbf{p}_1$ -part (low-marginal probabilities),  $\mathbf{p}_b$ -part (bunch probabilities), and  $\mathbf{p}_c$ -part (connection probabilities) of  $\mathbf{p}$ , respectively. In particular, the first row of  $\mathbf{M}$  corresponds to the zero-marginal  $\mathbf{1}$ , and this row contains 1 in all cells.

**4.5** Let

$$\mathbf{p}^* = \begin{bmatrix} \mathbf{p}_1^* \\ \mathbf{p}_b^* \\ \mathbf{p}_c^* \end{bmatrix} \quad (28)$$

be a (reduced) vector of probabilities whose  $\mathbf{p}_1$ -part and  $\mathbf{p}_b$ -part consist of *empirical probabilities* (estimated from an experiment or predicted by a model), and  $\mathbf{p}_c$ -part consists of the connection probabilities for multimaximal couplings. Then the system  $\mathcal{R}$  represented by  $\mathbf{p}^*$  is noncontextual if and only if

$$\mathbf{M}\mathbf{x} = \mathbf{p}^* \quad (29)$$

for some nonnegative coupling vector  $\mathbf{x}$ . If no such nonnegative  $\mathbf{x}$  exists, then  $\mathcal{R}$  is contextual. In reference to (25), note that  $\|\mathbf{x}\| = 1$  is guaranteed by the first row of  $\mathbf{M}$  (consisting of 1's only) and first element of  $\mathbf{p}^*$  ( $\langle \rangle = 1$ ).

**4.6** As a step towards measures of contextuality, consider the convex polytope

$$\mathbb{P} = \{\mathbf{p} : \mathbf{M}\mathbf{x} = \mathbf{p}, \text{ for some } \mathbf{x} \geq 0\}. \quad (30)$$

It corresponds to the set of all possible couplings of all systems having the same format as  $\mathcal{R}$  (because matrices  $\mathbf{M}$  are in a one-to-one correspondence with system formats).

**4.7** A specific system  $\mathcal{R}$  is defined by specifying the vectors  $\mathbf{p}_1 = \mathbf{p}_1^*$  and  $\mathbf{p}_b = \mathbf{p}_b^*$ . This defines a convex polytope which is a cross-section of the polytope  $\mathbb{P}$ ,

$$\mathbb{P}_c = \{\mathbf{p}_c : \mathbf{M}_c\mathbf{x} = \mathbf{p}_c, \text{ for some } \mathbf{x} \geq 0, \mathbf{M}_1\mathbf{x} = \mathbf{p}_1^*, \mathbf{M}_b\mathbf{x} = \mathbf{p}_b^*\}. \quad (31)$$

We refer to it as the *feasibility polytope* (for system  $\mathcal{R}$ ). It corresponds to the set of all possible couplings of system  $\mathcal{R}$ .

**4.8** A symmetrically opposite construction is the convex polytope

$$\mathbb{P}_b = \{\mathbf{p}_b : \mathbf{M}_b\mathbf{x} = \mathbf{p}_b, \text{ for some } \mathbf{x} \geq 0, \mathbf{M}_1\mathbf{x} = \mathbf{p}_1^*, \mathbf{M}_c\mathbf{x} = \mathbf{p}_c^*\}. \quad (32)$$

We can call it the *noncontextuality polytope* (for system  $\mathcal{R}$ ), as it corresponds to all noncontextual systems with the same 1-marginals as  $\mathcal{R}$ . Another way of describing  $\mathbb{P}_b$ , to emphasize its symmetry with  $\mathbb{P}_c$ , is that  $\mathbb{P}_b$  corresponds to all possible couplings of the multimaximal couplings of the system's connections.

**4.9** Clearly, system  $\mathcal{R}$  is noncontextual if and only if  $\mathbf{p}_c^* \in \mathbb{P}_c$  and  $\mathbf{p}_b^* \in \mathbb{P}_b$ , with the two statements implying each other,

$$\mathbf{p}_c^* \in \mathbb{P}_c \iff \mathbf{p}_b^* \in \mathbb{P}_b. \quad (33)$$

## 5 Measures of contextuality

**5.1** For a contextual system  $\mathcal{R}$ ,  $\mathbf{p}_c^*$  is outside  $\mathbb{P}_c$ , and the  $L_1$ -distance between them is a natural measure of contextuality,

$$\text{CNT}_1 = \min_{\mathbf{p}_c \in \mathbb{P}_c} \|\mathbf{p}_c^* - \mathbf{p}_c\| = \mathbf{1} \cdot \mathbf{p}_c^* - \max_{\mathbf{p}_c \in \mathbb{P}_c} (\mathbf{1} \cdot \mathbf{p}_c), \quad (34)$$

where the equality follows from  $\mathbf{p}_c^* \geq \mathbf{p}_c$  (componentwise). We can also write

$$\text{CNT}_1 = \mathbf{1} \cdot \mathbf{p}_c^* - \max_{\mathbf{x} \geq 0, \mathbf{M}_1\mathbf{x} = \mathbf{p}_1^*, \mathbf{M}_b\mathbf{x} = \mathbf{p}_b^*} (\mathbf{1} \cdot \mathbf{M}_c\mathbf{x}). \quad (35)$$



To interpret, since the system  $\mathcal{R}$  is contextual, its bunches are incompatible with the multimaximal couplings of its connections.  $\text{CNT}_1$  measures how close the couplings of these connections that are compatible with the system's bunches can be made to the multimaximal ones.

**5.2** This measure was the first one proposed within the framework of CbD [27, 28].<sup>3</sup> Its linear programming implementation is

$$\boxed{\begin{array}{lll} \text{find} & \text{maximizing} & \text{subject to} \\ \mathbf{x} & \mathbf{1} \cdot \mathbf{M}_c \mathbf{x} & \mathbf{x} \geq 0 \\ & & \mathbf{M}_1 \mathbf{x} = \mathbf{p}_1^* \\ & & \mathbf{M}_b \mathbf{x} = \mathbf{p}_b^* \end{array}}. \quad (36)$$

A solution  $\mathbf{x}^*$  must exist, and any such  $\mathbf{x}^*$  yields

$$\text{CNT}_1 = \mathbf{1} \cdot (\mathbf{p}_c^* - \mathbf{M}_c \mathbf{x}^*) = \mathbf{1} \cdot (\mathbf{p}^* - \mathbf{M} \mathbf{x}^*). \quad (37)$$

**5.3** The ‘‘symmetrically opposite’’ measure is the  $L_1$ -distance of  $\mathbf{p}_b^*$  from the contextuality polytope  $\mathbb{P}_b$ ,

$$\text{CNT}_2 = \min_{\mathbf{p}_b \in \mathbb{P}_b} \|\mathbf{p}_b^* - \mathbf{p}_b\| = \min_{\mathbf{x} \geq 0, \mathbf{M}_1 \mathbf{x} = \mathbf{p}_1^*, \mathbf{M}_c \mathbf{x} = \mathbf{p}_c^*} \|\mathbf{p}_b^* - \mathbf{M}_b \mathbf{x}\|. \quad (38)$$

The interpretation should be clear. Since  $\mathcal{R}$  is contextual, the multimaximal couplings of its connections are not compatible with its bunches.  $\text{CNT}_2$  measures how close the bunches that are compatible with these multimaximal couplings can be made to the observed bunches.

**5.4** This measure is introduced here for the first time. Its linear programming implementation is

$$\boxed{\begin{array}{lll} \text{find} & \text{minimizing} & \text{subject to} \\ \mathbf{x} & \mathbf{1} \cdot \mathbf{d} & -\mathbf{d} \leq \mathbf{p}_b^* - \mathbf{M}_b \mathbf{x} \leq \mathbf{d} \\ & & \mathbf{x}, \mathbf{d} \geq 0 \\ & & \mathbf{M}_1 \mathbf{x} = \mathbf{p}_1^* \\ & & \mathbf{M}_c \mathbf{x} = \mathbf{p}_c^* \end{array}}. \quad (39)$$

Again, for any solution  $\mathbf{x}^*$ ,

$$\text{CNT}_2 = \|\mathbf{p}_b^* - \mathbf{M}_b \mathbf{x}^*\| = \|\mathbf{p}^* - \mathbf{M} \mathbf{x}^*\|. \quad (40)$$

**5.5** The third measure to consider,  $\text{CNT}_3$ , has been proposed in Ref. [19], and brought to its present form in Refs. [18, 20]. However, the ideas on which it is based date back to Paul Dirac, with contemporary elaborations, including relating it to contextuality, found in Refs. [1, 3, 13, 14, 30]. The measure is based on the observation [19] that if one drops the nonnegativity constraint in (25), replacing thereby probability distributions with signed-measure distributions, then the set

$$\mathbb{Y} = \{\mathbf{y} : \mathbf{M} \mathbf{y} = \mathbf{p}^*\} \quad (41)$$

is nonempty. Clearly,

$$\mathbf{y} \in \mathbb{Y} \implies \mathbf{1} \cdot |\mathbf{y}| \geq 1, \quad (42)$$

were the absolute value is computed componentwise. The system then is noncontextual if and only if there is a  $\mathbf{y} \in \mathbb{Y}$  with

$$\mathbf{1} \cdot |\mathbf{y}| = \mathbf{1} \cdot \mathbf{y} = 1. \quad (43)$$

It follows that

$$\text{CNT}_3 = \min_{\mathbf{y} \in \mathbb{Y}} (\mathbf{1} \cdot |\mathbf{y}|) - 1 \quad (44)$$

is a natural measure of contextuality. As shown in Ref. [19], this minimum is always attained. The quantity  $\mathbf{1} \cdot |\mathbf{y}|$  is called *total variation* (of the signed measure), so  $\text{CNT}_3$  can be referred as the minimum total variation measure (minus 1).

<sup>3</sup>More precisely, the measure proposed in Refs. [27, 28] is twice  $\text{CNT}_1$ .

**5.6** The linear programming implementation of  $\text{CNT}_3$  is

$$\boxed{\begin{array}{lll} \text{find} & \text{minimizing} & \text{subject to:} \\ \mathbf{y}_+, \mathbf{y}_- & \mathbf{1} \cdot \mathbf{y}_- & \mathbf{M}(\mathbf{y}_+ - \mathbf{y}_-) = \mathbf{p}^* \\ & & \mathbf{y}_+, \mathbf{y}_- \geq 0 \end{array}}. \quad (45)$$

With any solution  $\mathbf{y}_+^*, \mathbf{y}_-^*$ ,

$$\text{CNT}_3 = \mathbf{1} \cdot |\mathbf{y}_+^* - \mathbf{y}_-^*| - 1. \quad (46)$$

## 6 Noncontextuality

**6.1** Consider now the situation when  $\mathcal{R}$  is noncontextual. With respect to the  $\text{CNT}_1$  measure, this means that  $\mathbf{p}_c^* \in \mathbb{P}_c$ . The question we pose is whether  $\text{CNT}_1$  can be extended into a noncontextuality measure by computing

$$\text{NCNT}_1 = \inf_{\mathbf{p}_c \notin \mathbb{P}_c} \|\mathbf{p}_c^* - \mathbf{p}_c\| = \min_{\mathbf{p}_c \in \partial \mathbb{P}_c} \|\mathbf{p}_c^* - \mathbf{p}_c\|, \quad (47)$$

where  $\partial(\text{polytope})$  indicates the boundary of the polytope. The answer to this question turns out to be negative: while this distance is well-defined, it is zero for any  $\mathbf{p}_c^*$ . Indeed, if  $\mathbf{p}_c^*$  were an interior point of  $\mathbb{P}_c$ , one could increase some of the probabilities in (21) by a small amount and still remain within  $\mathbb{P}_c$ . But this is impossible, since all  $k$ -marginal probabilities with  $k > 1$  have maximal possible values.  $\text{CNT}_1$  does not have a noncontextual counterpart.

**6.2** The situation is different with  $\text{CNT}_2$ . The measure

$$\text{NCNT}_2 = \inf_{\mathbf{p}_b \notin \mathbb{P}_b} \|\mathbf{p}_b^* - \mathbf{p}_b\| = \min_{\mathbf{p}_b \in \partial \mathbb{P}_b} \|\mathbf{p}_b^* - \mathbf{p}_b\| \quad (48)$$

is well-defined and varies as  $\mathbf{p}_b^*$  varies within  $\mathbb{P}_b$ . To interpret, since  $\mathcal{R}$  is noncontextual, the multimaximal couplings of its connections are compatible with its bunches.  $\text{NCNT}_2$  measures how far these bunches are from those that are not compatible with these multimaximal couplings. To compute  $\text{NCNT}_2$  we can make use of the following theorem [31]: a point on the boundary of a convex polytope  $L_1$ -closest to an interior point differs from the latter in a single coordinate. This means that all we have to do is to increase or decrease the probabilities in (19) one by one as far as possible without leaving the polytope, and to choose the smallest increase or decrease at the end.

**6.3** The linear programming implementation of this procedure is as follows. Let all elements of  $\mathbf{p}_b^*$  be enumerated  $1, \dots, K$ . Then, for every  $i = 1, \dots, K$ ,

$$\boxed{\begin{array}{lll} \text{find} & \text{maximizing} & \text{subject to} \\ d_i^+, \mathbf{x} & d_i^+ & \mathbf{p}_b^* + d_i^+ \mathbf{e}_i = \mathbf{M}_b \mathbf{x} \\ & & d_i^+, \mathbf{x} \geq 0 \\ & & \mathbf{M}_1 \mathbf{x} = \mathbf{p}_1^* \\ & & \mathbf{M}_c \mathbf{x} = \mathbf{p}_c^* \end{array}} \quad (49)$$

and

$$\boxed{\begin{array}{lll} \text{find} & \text{maximizing} & \text{subject to} \\ d_i^-, \mathbf{x} & d_i^- & \mathbf{p}_b^* - d_i^- \mathbf{e}_i = \mathbf{M}_b \mathbf{x} \\ & & d_i^-, \mathbf{x} \geq 0 \\ & & \mathbf{M}_1 \mathbf{x} = \mathbf{p}_1^* \\ & & \mathbf{M}_c \mathbf{x} = \mathbf{p}_c^* \end{array}}, \quad (50)$$

where  $\mathbf{e}_i$  is the unit vector with the  $i$ th component equal to 1. Once the solutions  $d_i^{*+}, d_i^{*-}$  for  $i = 1, \dots, K$  are determined,

$$\text{NCNT}_2 = \min_{i=1, \dots, K} \{\min(d_i^{*+}, d_i^{*-})\}. \quad (51)$$

**6.4** Consider now  $\text{CNT}_3$ . If  $\mathcal{R}$  is noncontextual, there is a nonnegative  $\mathbf{x}$  such that  $\mathbf{M}\mathbf{x} = \mathbf{p}^*$ , and

$$\text{CNT}_3 = \mathbf{1} \cdot \mathbf{x} - 1 = 0.$$

There seems to be no way to extend it to a noncontextuality measure without modifying its logic.

## 7 Contextual fraction

**7.1** We discuss next the measure called *contextual fraction* (CNTF) [1, 2] (see also Refs. [6, 24]). It has been formulated for consistently connected systems only, and its Cbd-based generalization is not unique. We will consider one such generalization, obtained by treating multimaximal couplings of connections as if they were additional bunches. This can in fact be used to formally redefine every system into a consistently connected one, as proposed by Amaral, Duarte, and Oliveira in Ref. [5].<sup>4</sup> We prefer, however, to simply complement bunch probabilities with connection probabilities rather than redefining the system itself.

**7.2** Unlike the three measures considered above, however, CNTF requires that we deal with complete vectors of probability  $\mathbf{p}_{(\cdot)}$ , as defined in (14), rather than reduced vectors. We need to accordingly replace the Boolean matrix  $\mathbf{M}$  with the Boolean matrix  $\mathbf{M}_{(\cdot)}$  such that the system  $\mathcal{R}$  represented by  $\mathbf{p}_{(\cdot)}^*$  is noncontextual if and only if

$$\mathbf{M}_{(\cdot)}\mathbf{x} = \mathbf{p}_{(\cdot)}^*, \quad (52)$$

with  $\mathbf{x}$  defined as above. The structure of  $\mathbf{M}_{(\cdot)}$  is described in Ref. [19], and its summary is as follows. Recall the definition of vector  $\mathbf{v}$  in Section 4. The columns of matrix  $\mathbf{M}_{(\cdot)}$  are labeled by the elements of  $\mathbf{v}$  in the same way as in matrix  $\mathbf{M}$ . The  $i$ th row of this matrix is labeled by the event whose probability is the  $i$ th element of  $\mathbf{p}_{(\cdot)}^*$  in 10 and 12. If all the random variables in this event have the same values in the  $j$ th value of  $\mathbf{v}$ , then we put 1 in the cell  $(i, j)$  of  $\mathbf{M}_{(\cdot)}$ . All other cells of  $\mathbf{M}_{(\cdot)}$  are filled with zeros.

**7.3** Consider the convex polytope

$$\mathbb{Z} = \{\mathbf{z} : \mathbf{M}_{(\cdot)}\mathbf{z} \leq \mathbf{p}_{(\cdot)}^*, \mathbf{z} \geq 0, \mathbf{1} \cdot \mathbf{z} \leq 1\}. \quad (53)$$

This polytope is nonempty, because, e.g., it contains  $\mathbf{z} = \mathbf{0}$ . If  $\mathbb{Z}$  contains a  $\mathbf{z}$  with  $\mathbf{1} \cdot \mathbf{z} = 1$ , then  $\mathbf{M}_{(\cdot)}\mathbf{z} = \mathbf{p}_{(\cdot)}^*$  because the elements of every  $\mathbf{p}^{(c)}$  in (10) and every  $\mathbf{p}_{(q)}$  and (12) sum to 1, and the corresponding rows of  $\mathbf{M}_{(\cdot)}$  sum to a row consisting of 1's only. Such a system therefore is noncontextual. If  $\mathbf{1} \cdot \mathbf{z} < 1$  for all  $\mathbf{z} \in \mathbb{Z}$ , the system is contextual, and its degree of contextuality can be measured by the difference between 1 and the maximal total mass  $\mathbf{1} \cdot \mathbf{z}$  achievable in  $\mathbb{Z}$ :

$$\text{CNTF} = 1 - \max_{\mathbf{z} \in \mathbb{Z}} (\mathbf{1} \cdot \mathbf{z}). \quad (54)$$

The linear programming formulation of this measure is

$$\boxed{\begin{array}{lll} \text{find} & \text{maximizing} & \text{subject to:} \\ \mathbf{z} & \mathbf{1} \cdot \mathbf{z} & \mathbf{M}_{(\cdot)}\mathbf{z} \leq \mathbf{p}_{(\cdot)}^* \\ & & \mathbf{z} \geq 0 \\ & & \mathbf{1} \cdot \mathbf{z} \leq 1 \end{array}}. \quad (55)$$

Can this measure be naturally extended to a measure of noncontextuality? The answer is negative, for the same reason as in the case of  $\text{CNT}_3$ . If  $\mathbb{Z}$  contains a vector  $\mathbf{z}$  such that  $\mathbf{1} \cdot \mathbf{z} = 1$ , then  $\mathbf{M}_{(\cdot)}\mathbf{z} = \mathbf{p}_{(\cdot)}^*$  and  $\text{CNTF} = 0$ . We do not have conceptual means here to distinguish different noncontextual systems.

## 8 Conclusion

**8.1** We have provided an overview of three Cbd-based measures of contextuality. Two of them,  $\text{CNT}_1$  and  $\text{CNT}_2$  are  $L_1$ -distances between a probability vector representing a system and a convex polytope. For either of the measures, if the probability vector is not outside the polytope, a natural way of extending this measure to a noncontextuality measure is to compute the  $L_1$ -distance from the point to the surface of the polytope. We have seen, however, that in the case of  $\text{CNT}_1$ , the probability vector never gets inside the feasibility polytope (31): as  $\text{CNT}_1$  decreases to zero, and the vector

<sup>4</sup>Amaral and coauthors use maximal couplings rather than multimaximal ones, and they allow for multivalued variables (as we did in the older version of Cbd, e.g., in Ref. [19]). The difference between the two couplings of a set  $\{X_1, \dots, X_n\}$  is that in the multimaximal coupling  $\{Y_1, \dots, Y_n\}$  we maximize probabilities of all equalities  $Y_i = Y_j$  (whence it follows that we also maximize the probability of  $Y_{i_1} = Y_{i_2} = \dots = Y_{i_k}$  for any subset of  $\{Y_1, \dots, Y_n\}$ ), whereas a maximal coupling  $\{Z_1, \dots, Z_n\}$  only maximizes the single chain equality  $Z_1 = Z_2 = \dots = Z_n$ . The reasons we adhere to multimaximal couplings and canonical systems, with all variables dichotomized, were laid out in Refs. [18, 20, 21]. Here, it will suffice to say that our multimaximal couplings are unique, whereas maximal couplings generally are not, even for dichotomous variables (if there are more than two of them). Since measures of (non)contextuality generally depend on what couplings are being used, the approach advocated in Ref. [5] faces the problem of choice.

becomes noncontextual, it sticks to the polytope's surface. By contrast, for  $\text{CNT}_2$ , as its value decreases to zero, the vector of probabilities continues to move inside the contextuality polytope (32).  $\text{CNT}_2$  therefore is naturally extended to  $\text{NCNT}_2$ , the distance from an interior point to the polytope's surface.

**8.2** The third CbD-based measure,  $\text{CNT}_3$ , is of a different kind. It measures the degree of contextuality by dropping the nonnegativity constraint in (25) for coupling vectors  $\mathbf{x}$ , and determining how close the vectors  $\mathbf{y}$  thus obtained can be made to a proper probability distribution while satisfying  $\mathbf{M}\mathbf{y} = \mathbf{p}^*$ . The logical structure of this measure is close to Abramsky, Barbosa, and Mansfield's CNTF measure, CbD-generalized to apply to arbitrary systems. Here the degree of contextuality is measured by replacing the summing-to-unity constraint in (25) with  $\mathbf{1} \cdot \mathbf{z} \leq 1$ , and determining how close the vectors  $\mathbf{z}$  can be made to a proper probability distribution while satisfying  $\mathbf{M}_{(\cdot)}\mathbf{z} \leq \mathbf{p}_{(\cdot)}^*$ . Both these measures do not lend themselves to natural noncontextuality extensions. The values of these two measures, total variation less 1 in  $\text{CNT}_3$  and 1 minus total mass in CNTF, vary with probability vectors representing contextual systems but freeze at zero values for all noncontextual systems.

**8.3** The relationship between the four measures of contextuality discussed in this paper is far from being clear. One nice feature of  $\text{CNT}_1$ , the oldest CbD-based measure, is that it is proportional to the violation of the generalized Bell inequalities in the case of cyclic systems [22, 27, 28]:

$$\text{CNT}_1 = \frac{1}{4} \max_{(\iota_1, \dots, \iota_k) \in \{-1, 1\}^n: \prod_{i=1}^n \iota_i = -1} \sum_{i=1}^n \iota_i \langle R_i^i R_{i \oplus 1}^i \rangle - n + 2 - \sum_{i=1}^n |\langle R_i^i \rangle - \langle R_i^{i \oplus 1} \rangle|, \quad (56)$$

where the dichotomous variables are assumed to be  $\pm 1$ -valued rather than Bernoulli.<sup>5</sup> Here, the bunch for context  $c_i$  ( $i = 1, \dots, n$ ) consists of two random variables  $R_i^i, R_{i \oplus 1}^i$ , where  $i \oplus 1 = i + 1$  for  $i < n$  and  $n \oplus 1 = 1$ . (This inequality generalizes to arbitrary systems the inequality proved in Ref. [7], in a very different way, for consistently connected systems.) Our analysis [17] shows that in the case of cyclic systems

$$\text{CNT}_1 = \text{CNT}_2. \quad (57)$$

It has been conjectured, based on numerical computations conducted with the help of Víctor Cervantes, that in the case of cyclic systems

$$\text{CNT}_3 = 2\text{CNT}_1/(n-1) = 2\text{CNT}_2/(n-1). \quad (58)$$

Beyond cyclic systems, however, we know that  $\text{CNT}_1$  and  $\text{CNT}_3$  are not generally related to each other by any function [13]. The relations between the three CbD-based measures and CNTF is yet to be investigated.

**8.4** It is worth mentioning that all measures of contextuality (and noncontextuality) involve a certain degree of arbitrariness. For instance, both  $\text{CNT}_1$  and  $\text{CNT}_2$  could be constructed with another  $L_p$  or  $L_\infty$  replacing  $L_1$ , and there seem to be no unchallengeable principles to guide one's choice (although  $L_1$  may be argued to be preferable because of the additivity of probabilities). The choice of a reduced vector of probabilities adds another dimension of arbitrariness: although all reduced representations are linear transformations of the complete one, min  $L_1$ -distance values for them may not be related to each other in a simple way. It seems therefore that one could profitably use several measures to characterize a given system. At the same time, the fact that only some measures of contextuality naturally extend into measures of noncontextuality may provide principled guidance in constraining the multitude of possibilities.

**Acknowledgments.** We are grateful to Víctor Cervantes for numerical experimentation with the measures described in this paper, as well as for critically reading the manuscript, and to Bárbara Amaral for sharing with us her work. We thank the participants of the Purdue Winer Memorial Lectures 2018, especially Adán Cabello, for many fruitful discussions. We also thank two anonymous reviewers for their insightful and constructive comments.

## References

- [1] Abramsky S, Brandenburger A 2011. The sheaf-theoretic structure of non-locality and contextuality. *New J. Phys.* 13, 113036-13075.
- [2] Abramsky S., Barbosa R. S., Mansfield S 2017 The contextual fraction as a measure of contextuality. *Phys. Rev. Lett.* 119, 050504.

---

<sup>5</sup>See footnote 3.

- [3] Al-Safi S, Short A 2013 Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities. *Phys. Rev. Lett.* 111, 170403.
- [4] Amaral B, Cunha MT, Cabello A 2015 Quantum theory allows for absolute maximal contextuality. *Phys. Rev. A* 92, 062125.
- [5] Amaral B, Duarte C, Oliveira RI 2018 Necessary conditions for extended noncontextuality in general sets of random variables. *J. Math. Phys.* 59, 072202.
- [6] Amselem E, Danielsen LE, López-Tarrida AJ, Portillo JR, Bourennane M, Cabello A 2012 Experimental fully contextual correlations. *Phys. Rev. Lett.* 108, 200405
- [7] Araújo M, Quintino MT, Budroni C, Cunha MT, Cabello A 2013 All noncontextuality inequalities for the n-cycle scenario, *Phys. Rev. A* 88, 022118.
- [8] Basieva I, Cervantes VH, Dzhafarov EN, Khrennikov A (in press) True contextuality beats direct influences in human decision making. *J. Exp. Psych. Gen.*
- [9] Bermejo-Vega J, Delfosse N, Browne DE, Okay C, Raussendorf R 2017 Contextuality as a resource for models of quantum computation with qubits. *Phys. Rev. Lett.* 119, 120505.
- [10] Brukner C, Zukowski M, Pan J-W, Zeilinger A, 2004 Bell's inequalities and quantum communication complexity. *Phys. Rev. Lett.* 92, 127901.
- [11] Brunner N, Cavalcanti D, Pironio S, Scarani V, Wehner S 2014 Bell nonlocality. *Rev. Mod. Phys.* 86, 419.
- [12] Cervantes VH, Dzhafarov EN 2018 Snow Queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. *Decision* 5, 193-204.
- [13] de Barros JA, Dzhafarov EN, Kujala JV, Oas G 2016. Measuring Observable Quantum Contextuality. In H. Atmanspacher, T. Filk, E. Pothos (Eds.) *Lect. Not. Comp. Sci.* 9535, 36-47.
- [14] de Barros JA, Oas G, Suppes P (2015) Negative probabilities and counterfactual reasoning on the double-slit experiment. In J.-Y. Beziau, D. Krause, & J. B. Arenhart (Eds.), *Conceptual Clarification: Tributes to Patrick Suppes (1992-2014)*. London: College Publications.
- [15] Dzhafarov EN 2003 Selective influence through conditional independence. *Psychom.* 68, 7–25.
- [16] Dzhafarov EN 2017 Replacing nothing with something special: Contextuality-by-Default and dummy measurements. In A. Khrennikov & T. Bourama (Eds.), *Quantum Foundations, Probability and Information*. Springer.
- [17] Dzhafarov EN 2019 Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems. [arXiv:1907.03328](https://arxiv.org/abs/1907.03328).
- [18] Dzhafarov EN, Cervantes VH, Kujala JV 2017 Contextuality in canonical systems of random variables. *Phil. Trans. Roy.Soc. A* 375, 20160389.
- [19] Dzhafarov EN, Kujala JV 2016 Context-content systems of random variables: The Contextuality-by-Default theory. *J. Math. Psych.* 74, 11-33.
- [20] Dzhafarov EN, Kujala JV 2017 Contextuality-by-Default 2.0: Systems with binary random variables. In J.A. de Barros, B. Coecke, E. Pothos (Eds.) *Lect. Not. Comp. Sci.* 10106, 16-32.
- [21] Dzhafarov EN, Kujala JV 2017 Probabilistic foundations of contextuality. *Fort. Phys.* 65, 1-11.
- [22] Dzhafarov EN, Kujala JV, Larsson, J-Å 2015 Contextuality in three types of quantum-mechanical systems. *Found. Phys.* 7, 762-782.
- [23] Dzhafarov EN, Zhang R, Kujala JV 2016 Is there contextuality in behavioral and social systems? *Phil. Trans. Roy. Soc. A* 374, 20150099.
- [24] Eiltsur AC, Popescu S, Rohrlich D 1992 Quantum nonlocality for each pair in an ensemble. *Phys. Lett. A* 162, 25.
- [25] Frustaglia D, Baltanás JP, Velázquez-Ahumada MC, Fernández-Prieto A, Lujambio A, Losada V, Freire MJ, Cabello A 2016 Classical Physics and the Bounds of Quantum Correlations. *Phys. Rev. Lett.* 116, 250404

- [26] Howard M, Wallman J, Veitch V, Emerson J 2014 Contextuality supplies the ‘magic’ for quantum computation. *Nature* 510, 351–355.
- [27] Kujala JV, Dzhafarov EN 2016 Proof of a conjecture on contextuality in cyclic systems with binary variables. *Found. Phys.* 46, 282-299.
- [28] Kujala JV, Dzhafarov EN, Larsson J-Å 2015 Necessary and sufficient conditions for extended noncontextuality in a broad class of quantum mechanical systems. *Phys. Rev. Lett.* 115, 150401.
- [29] Markiewicz M, Kaszlikowski D, Kurzyński P, Wójcik A 2019 From contextuality of a single photon to realism of an electromagnetic wave. *Quant. Inf.* 5, 1-5.
- [30] Spekkens RW 2008 Negativity and contextuality are equivalent notions of nonclassicality. *Phys. Rev. Lett.* 101, 020401.
- [31] Tuentler HJH 2006 Minimum L1-distance projection onto the boundary of a convex set: Simple characterization. *J. Optim. Theor. Appl.*, 112, 441-445.
- [32] Zhang A, Xu H, Xie J, Zhang H, Smith BJ, Kim MS, Zhang L 2019 Experimental test of contextuality in quantum and classical systems. *Phys. Rev. Lett.* 122, 080401.