Contextuality-by-Default 2.0: Systems with Binary Random Variables

Ehtibar N. Dzhafarov[1], Janne V. Kujala[2]

[1]Purdue University
ehtibar@purdue.edu

[2]University of Jyväskylä
jvk@iki.fi

Abstract. The paper outlines a new development in the Contextuality-by-Default theory as applied to finite systems of binary random variables. The logic and principles of the original theory remain unchanged, but the definition of contextuality of a system of random variables is now based on multimaximal rather than maximal couplings of the variables that measure the same property in different contexts: a system is considered noncontextual if these multimaximal couplings are compatible with the distributions of the random variables sharing contexts. A multimaximal coupling is one that is a maximal coupling of any subset (equivalently, of any pair) of the random variables being coupled. Arguments are presented for why this modified theory is a superior generalization of the traditional understanding of contextuality in quantum mechanics. The modified theory coincides with the previous version in the important case of cyclic systems, which include the systems whose contextuality was most intensively studied in quantum physics and behavioral sciences.

Keywords: contextuality, connection, consistent connectedness, cyclic system, inconsistent connectedness, maximal coupling, multimaximal coupling.

1 Introduction: From maximality to multimaximality

The Contextuality-by-Default (CbD) theory [7–12, 14, 16, 23–25] was proposed as a generalization of the traditional contextuality analysis in quantum physics [2, 3, 5, 18, 19, 21, 26, 32]. The latter has been largely confined to *consistently connected* systems of random variables, those adhering to the *"no-disturbance"* principle [27, 31]: the distributions of measurement outcomes remain unchanged under different measurement conditions (*contexts*). CbD allows for *inconsistently connected* systems, those in which context may influence the distribution of measurement outcomes for one and the same property [11, 14–17, 23, 25]. In accordance with the CbD interpretation of the traditional contextuality analysis, this generalization is achieved by replacing the *identity couplings* used in dealing with consistently connected systems by *maximal couplings*.

Recall that, given a set of random variables $X, Y, \ldots, Z$, a coupling of this set is any set of jointly distributed random variables, $(X', Y', \ldots, Z')$, with

$$X \sim X',\ Y \sim Y',\ \ldots,\ Z \sim Z',$$

where $\sim$ stands for "has the same distribution as." The coupling $(X', Y', \ldots, Z')$ is maximal if (using Pr as a symbol for probability) the value of

$$p_{eq} = \Pr\left[X' = Y' = \ldots = Z'\right]$$

is maximal possible among all possible couplings of $X, Y, \ldots, Z$. The identity coupling is a special case of a maximal coupling, when $p_{eq} = 1$. The latter is possible if and only if all random variables $X, Y, \ldots, Z$ (hence also $X', Y', \ldots, Z'$) are identically distributed:

$$X \sim Y \sim \ldots \sim Z.$$

The notion of a maximal coupling, however, is not the only possible generalization of the identity couplings. And it has recently become apparent that it is not the best possible generalization either. The maximal-couplings-based definition of (non)contextual systems adopted in CbD does not have a certain intuitively plausible property that is enjoyed by the identity-couplings-based definition of consistently connected (non)contextual systems. This property is that *any subsystem of a consistently connected noncontextual system is noncontextual.* A subsystem is obtained by dropping from a system some of the random variables. An inconsistently connected noncontextual system in the previously published version of CbD ("CbD 1.0") does not generally have this property: by dropping some of its components one may be able to make it contextual.

In the new version, "CbD 2.0," preservation of noncontextuality for subsystems is achieved by replacing the notion of a maximal coupling in the definition of (non)contextual systems by the notion of a *multimaximal coupling.* This term designates a coupling every subcoupling whereof is a maximal coupling for the corresponding subset of the random variables being coupled (see Definition 1 below).

The remainder of the paper is a systematic presentation of this idea and of how it works in the analysis of contextuality. CbD 1.0 and CbD 2.0 coincide when dealing with consistently connected systems (as they must, because they both generalize this special case). They also coincide when dealing with the important class of cyclic systems [16, 24, 25] (see Section 4). None of the principles upon which CbD is based changes in version 2.0 (Section 2). The recently proposed logic of constructing a universal measure of contextuality [12] also transfers to version 2.0 without changes (Section 5).

2 Contextuality-by-Default theory: Basics

We briefly recapitulate here the main aspects of the Contextuality-by-Default theory. We recommend, however, that the reader look through some of the recent accounts of CbD 1.0, e.g., Refs. [11, 14], or (especially) Ref. [12].

Each random variable in CbD is double-indexed, R_q^c, where q is referred to as the *content* of the random variables, that which R_q^c measures or responds to, and c is referred to as its *context*, the conditions under which R_q^c measures or responds to q.

Remark 1. Following Ref. [12] we will write "conteXt" and "conteNt" to prevent their confusion in reading. The conteXt and conteNt of a random variable uniquely identify it within a given system of random variables.

Two random variables R_q^c and $R_{q'}^{c'}$ are jointly distributed if and only if they share a conteXt: $c = c'$. Otherwise they are *stochastically unrelated*. All random variables sharing a conteXt form a jointly distributed *bunch* of random variables. All random variables sharing a conteNt form a *connection*, the elements of which are pairwise stochastically unrelated. It is necessary that all random variables in a connection have the same set of possible values (more generally, the same set and sigma-algebra).

The present paper is primarily about systems in which all random variables are binary. It is immaterial for contextuality analysis how these values are named, insofar as they are identically named and identically interpreted within each connection. For instance, if $R_q^c = 1$ means "spin-up along axis z in particle 1" and $R_q^c = 2$ means "spin-down along axis z in particle 1," then all random variables $R_q^{c'}$ ($c' \neq c$) should have the same possible values, 1 and 2, with the same meanings. Note that for another conteNt q', the values of $R_{q'}^c$ need not be denoted in the same way even if they have analogous interpretations: e.g., we may have $R_{q'}^c = 3 =$ "spin-up along axis z in particle 2" and $R_{q'}^c = 4 =$ "spin-down along axis z in particle 2".

The matrix below provides an example of a *conteXt-conteNt system* (c-c system) of random variables:

$$
\begin{array}{|c|c|c|c|c}
\hline
R_1^1 & R_2^1 & \cdot & R_4^1 & c_1 \\
\hline
R_1^2 & \cdot & R_3^2 & \cdot & c_2 \\
\hline
R_1^3 & R_2^3 & R_3^3 & R_4^3 & c_3 \\
\hline
\multicolumn{1}{c}{q_1} & \multicolumn{1}{c}{q_2} & \multicolumn{1}{c}{q_3} & \multicolumn{1}{c}{q_4} & \boxed{\boxed{\mathcal{R}_{ex}}}
\end{array}
.
$$

Each row here is a bunch of jointly distributed random variables, each column is a connection ("between bunches"). Note that not every conteNt should be measured in a given conteXt.

The system $\mathcal{R}_{ex}$ can be conveniently used to illustrate the logic of contextuality analysis. We first consider the connections separately, and for each of them find all couplings that satisfy a certain property C. Let's call them C-*couplings*. Then we determine if these C-couplings are *compatible* with a coupling of the bunches of the c-c system (equivalently put, with a coupling of the entire c-c system).

The compatibility in question means the following. A coupling of (the bunches of) the c-c system is a set of jointly distributed random variables

$$\begin{array}{|c|c|c|c|c}
\hline
S_1^1 & S_2^1 & & S_4^1 & c_1 \\
\hline
S_1^2 & \cdot & S_3^2 & \cdot & c_2 \\
\hline
S_1^3 & S_2^3 & S_3^3 & S_4^3 & c_3 \\
\hline
\multicolumn{1}{c}{q_1} & \multicolumn{1}{c}{q_2} & \multicolumn{1}{c}{q_3} & \multicolumn{1}{c}{q_4} & \boxed{\boxed{S_{ex}}}
\end{array},$$

such that

$$\left(S_1^1, S_2^1, S_4^1\right) \sim \left(R_1^1, R_2^1, R_4^1\right),$$

$$\left(S_1^2, S_3^2\right) \sim \left(R_1^2, R_3^2\right),$$

$$\left(S_1^3, S_2^3, S_3^3, S_4^3\right) \sim \left(R_1^3, R_2^3, R_3^3, R_4^3\right).$$

Since the elements of S_{ex} are jointly distributed, the marginal distributions of the columns corresponding to the connections of $\mathcal{R}_{ex}$ are well-defined:

$$\begin{array}{lll}
\left(S_1^1, S_1^2, , S_1^3\right) & \text{is a coupling of connection} & R_1^1, R_1^2, R_1^3, \\
\left(S_2^1, S_2^3\right) & \text{is a coupling of connection} & R_2^1, R_2^3, \\
\left(S_3^2, S_3^3\right) & \text{is a coupling of connection} & R_3^2, R_3^3, \\
\left(S_4^1, S_4^3\right) & \text{is a coupling of connection} & R_4^1, R_4^3.
\end{array}$$

In CbD we pose the following question: is there a coupling S_{ex} such that the subcouplings corresponding to the connections are C-couplings? If the answer is affirmative, then we say that the bunches of $\mathcal{R}_{ex}$ are compatible with at least some of the combinations of the C-couplings for its connections — and the c-c system is considered *partially* C-*noncontextual*. Otherwise, if no such a coupling S_{ex} exists, we say that the bunches of $\mathcal{R}_{ex}$ are incompatible with any of the C-couplings for its connections — and the c-c system is considered *completely* C-*contextual*. The intuition is that in a completely C-contextual c-c system the conteXts "interfere" with one's ability to couple the measurements of each conteNt in a specified (by C) way — while the connections can be coupled in this way if they are considered separately, ignoring the conteXts.

The adjectives "partially" and "completely" do not belong to the original theory. They are added here because one can also consider a stronger (more restrictive) notion of noncontextual c-c systems and, correspondingly, a weaker (less restrictive) notion of contextual c-c systems. We say that a c-c system is *completely* C-*noncontextual* if the bunches of $\mathcal{R}_{ex}$ are compatible with any combinations of the C-couplings for its connections; and it is *partially* C-*contextual* if the bunches of $\mathcal{R}_{ex}$ are incompatible with at least some of these combinations.

The intuition is that in a completely C-noncontextual c-c system the conteXts "do not interfere" in any way with C-couplings of the measurements of any given conteNt (as if the connections were taken separately, ignoring conteXts).

In CbD 1.0 the C-couplings are maximal couplings, as defined in the opening paragraph of the paper. In CbD 2.0 C-couplings are multimaximal couplings, as defined below. We will see that if all random variables in a system are binary and C is multimaximality, then every connection has a unique C-coupling (Theorem 1-Corollary 1). In this case the notions of partial and complete (non)contextuality coincide, allowing us to drop these adjectives when speaking of (non)contextual c-c systems.

Remark 2. It is important to accept that noncontextuality of a c-c system (even if complete) does not mean that the conteXts are irrelevant and can be ignored. On the contrary, they are relevant "by defaults because, e.g., R_2^1 and R_2^3 in the second connection of $\mathcal{R}_{ex}$ are distinct and stochastically unrelated random variables. Moreover, the distributions of R_2^1 and R_2^3 may very well be different (i.e., the c-c system may be inconsistently connected), and this does not necessarily mean that the system is contextual (even if only partially) in the sense of our definitions. The measurements of the conteNt q_2 in conteXt c_3 can be "directly" influenced by the jointly-made measurements of q_3 (in which case we can speak of "signaling" or "disturbance"), while in context c_1 this influence is absent [1,22]. It is also possible that the experimental set-up in context c_3 is different from that in context c_1, in which case we can speak of conteXt-dependent biases [28, 29]. All of this may account for the different distributions of R_2^1 and R_2^3, and none of this by itself makes the system contextual. See Refs. [11,12,17] for argumentation against confusing signaling and contextual biases with contextuality. (Of course, if one so wishes, they can be called forms of contextuality, but in a different sense from how contextulaity is understood in quantum physics and in CbD.)

3 Multimaximal couplings for binary variables

Definition 1. Let $R_q^1, \ldots, R_q^k$ $(k > 1)$ be a connection of a system. A coupling $\left(T_q^1, \ldots, T_q^k\right)$ of $R_q^1, \ldots, R_q^k$ is a *multimaximal coupling* if, for any $m > 1$ and any subset $\left(T_q^{i_1}, \ldots, T_q^{i_m}\right)$ of $\left(T_q^1, \ldots, T_q^k\right)$, the value of

$$\Pr\left[T_q^{i_1} = \ldots = T_q^{i_m}\right]$$

is largest possible among all couplings of $R_q^{i_1}, \ldots, R_q^{i_m}$.

The multimaximality plays the role of the constraint C in the definition of C-couplings given in the previous section. One finds multimaximal couplings for each of the connections and then investigates their compatibility with the c-c system's bunches.

It is known that a maximal coupling exists for any connection [33]. This is not true for multimaximal couplings in general: such a coupling need not exist if the number of possible values for the random variables in a connection exceeds 2.

Example 1. Consider a connection consisting of random variables R_q^1, R_q^2, R_q^3 each having values $1, 2, 3$ with the following probabilities

	1	2	3
R_q^1	0	1/2	1/2
R_q^2	1/2	0	1/2
R_q^3	1/2	1/2	0

.

If a multimaximal coupling $\left(T_q^1, T_q^2, T_q^3\right)$ exists, we should have (see Ref. [33], or Theorem 3.3 in Ref. [12])

$$\begin{array}{ccc} \Pr[T_q^1 = T_q^2 = 1] = 0 & \Pr[T_q^1 = T_q^2 = 2] = 0 & \Pr[T_q^1 = T_q^2 = 3] = 0.5 \\ \Pr[T_q^2 = T_q^3 = 1] = 0.5 & \Pr[T_q^2 = T_q^3 = 2] = 0 & \Pr[T_q^2 = T_q^3 = 3] = 0 \\ \Pr[T_q^1 = T_q^3 = 1] = 0 & \Pr[T_q^1 = T_q^3 = 2] = 0.5 & \Pr[T_q^1 = T_q^3 = 3] = 0 \end{array}$$

from which we have in particular

$$\Pr[T_q^1 = T_q^2 = 3] = \Pr[T_q^1 = T_q^3 = 2] = \Pr[T_q^2 = T_q^3 = 1] = 0.5.$$

But these three events are pairwise mutually exclusive, so the sum of their probabilities cannot exceed 1. □

It can also be shown that, in the case of random variables with more than two possible values, a multimaximal coupling, if it exists, is not generally unique.

Example 2. Consider a connection consisting of random variables R_q^1, R_q^2, R_q^3 each having one of six values (denoted $1, 1', 2, 2', 3, 3'$) with the following probabilities

	1	1′	2	2′	3	3′
R_q^1	0	0	0	1/2	0	1/2
R_q^2	0	1/2	0	0	1/2	0
R_q^3	1/2	0	1/2	0	0	0

.

Then the distinct couplings whose distributions are shown below,

$\left(\dot{T}_q^1, \dot{T}_q^2, \dot{T}_q^3\right) =$	$(2', 1', 1)$	$(3', 3, 2)$	otherwise
prob. mass	$1/2$	$1/2$	0

and

$\left(\ddot{T}_q^1, \ddot{T}_q^2, \ddot{T}_q^3\right) =$	$(2', 3, 2)$	$(3', 1', 1)$	otherwise
prob. mass	$1/2$	$1/2$	0

,

are both multumaximal couplings. □

However, the situation is different if the random variables in a connection are all binary: multimaximal couplings in this case always exist and are unique. In the theorem to follow we denote the values of all variables R_q^i by $1, 2$, and we will write values of $\left(T_q^1, \ldots, T_q^k\right)$ as strings of 1's and 2's, without commas.

Theorem 1. *Let $R_q^1, \ldots, R_q^k$ be a connection with binary random variables arranged so that the values of $p_i = \Pr\left[R_q^i = 1\right]$ are sorted $p_1 \leq \ldots \leq p_k$. Then $\left(T_q^1, \ldots, T_q^k\right)$ is a multimaximal coupling of $R_q^1, \ldots, R_q^k$ if and only if all values of $\left(T_q^1, \ldots, T_q^k\right)$ are assigned zero probability mass, except for*

$$\left[\begin{array}{c|c} \text{value of } \left(T_q^1, \ldots, T_q^k\right) & \text{probability mass} \\ 11\ldots1 & p_1 \\ 21\ldots1 & p_2 - p_1 \\ 22\ldots1 & p_3 - p_2 \\ \vdots & \vdots \\ \overbrace{2\ldots2}^{l}\underbrace{1\ldots1}_{k-l} & p_{l+1} - p_l \\ \vdots & \vdots \\ 22\ldots2 & 1 - p_k \end{array}\right]. \tag{†}$$

Proof. Note that the distribution of $\left(T_q^1, \ldots, T_q^k\right)$ in the theorem's statement is well-defined, and that $\left(T_q^1, \ldots, T_q^k\right)$ is indeed a coupling of $R_q^1, \ldots, R_q^k$: for any $1 \leq l \leq k$,

$$\Pr\left[T_q^l = 1\right] = \sum_{m=0}^{l-1} \Pr\left[\overbrace{2\ldots2}^{m}\underbrace{1\ldots1}_{k-m}\right] = \sum_{m=0}^{l-1} (p_{m+1} - p_m) = p_l = \Pr\left[R_q^l = 1\right].$$

Sufficiency. The "if" part is checked directly: for any $1 \leq i_1 < \ldots < i_m \leq k$,

$$\Pr\left[T_q^{i_1} = \ldots = T^{i_m} = 1\right] = \textstyle\sum_{m=0}^{i_1-1} \Pr\left[\overbrace{2\ldots2}^{m}\underbrace{1\ldots1}_{k-m}\right]$$

$$= \textstyle\sum_{m=0}^{i_1-1} (p_{m+1} - p_m) = p_{i_1} = \Pr\left[T_q^{i_1} = 1\right],$$

which is the maximal possible value for the leftmost probability. Analogously,

$$\Pr\left[T_q^{i_1} = \ldots = T^{i_m} = 2\right] = \sum\nolimits_{m=i_m}^{k} \Pr\left[\overbrace{2\ldots2}^{m}\underbrace{1\ldots1}_{k-m}\right]$$

$$= \sum\nolimits_{m=i_m}^{k} (p_{m+1} - p_m) = 1 - p_{i_m} = \Pr\left[T_q^{i_m} = 2\right],$$

which is also the maximal possible probability. This establishes that $\left(T_q^{i_1}, \ldots, T^{i_m}\right)$ is a maximal coupling for $\left(R_q^{i_1}, \ldots, R^{i_m}\right)$.

Necessity. The "only if" part of the statement is proved by (i) observing that $\Pr[22\ldots2] = 1 - p_k$, and (ii) proving that if l is the ordinal position of the first 1 in the value of $\left(T_q^1, \ldots, T_q^k\right)$, then

$$\Pr\left[\overbrace{2\ldots2}^{l-1}\underbrace{1\ldots1}_{k-l+1}\right] = p_l - p_{l-1},$$

and for all other strings with the first 1 in the lth position the probabilities are zero. We prove (ii) by induction on l. For $l = 1$, we have

$$p_1 = \Pr[11\ldots1].$$

Since

$$p_1 = \Pr\left[T_q^1 = 1\right] = \Pr[11\ldots1] + \sum \Pr\left[1\underbrace{\ldots}_{\text{not all 1's}}\right],$$

all the summands under the summation operator must be zero. Let the statement be proved up to and including $l < k$. We have

$$p_{l+1} = \Pr\left[T_q^{l+1} = \ldots = T_q^k = 1\right] = \Pr\left[\overbrace{2\ldots2}^{l}\underbrace{1\ldots1}_{k-l}\right] + \sum \Pr\left[\overbrace{\ldots}^{\text{not all 2's}}\underbrace{1\ldots1}_{k-l}\right].$$

By the induction hypothesis, all summands under the summation operator are zero, except for

$$\left[\begin{array}{c|c} \text{value of } \left(T_q^1, \ldots, T_q^k\right) & \text{probability mass} \\ 11\ldots1 & p_1 \\ 21\ldots1 & p_2 - p_1 \\ 22\ldots1 & p_3 - p_2 \\ \vdots & \vdots \\ \overbrace{2\ldots2}^{l-1}\underbrace{1\ldots1}_{k-l+1} & p_l - p_{l-1} \end{array}\right].$$

These values sum to p_l. Hence

$$\Pr\left[\overbrace{2\ldots2}^{l}\underbrace{1\ldots1}_{k-l}\right] = p_{l+1} - p_l.$$

We also have

$$\begin{aligned} p_{l+1} = \Pr\left[T_q^{l+1} = 1\right] &= \Pr\left[\overbrace{2\ldots2}^{l}\underbrace{1\ldots1}_{k-l}\right] + \sum \Pr\left[\overbrace{\ldots}^{\text{not all 2's}}\ \underbrace{1\ldots1}_{k-l}\right] \\ &\quad + \sum \Pr\left[\overbrace{\ldots1}^{l+1}\ \underbrace{\ldots}_{\text{not all 1's}}\right] \\ &= (p_{l+1} - p_l) + p_l + \sum \Pr\left[\overbrace{\ldots1}^{l+1}\ \underbrace{\ldots}_{\text{not all 1's}}\right] \end{aligned}$$

whence the summands under the last summation operator must all be zero. □

Corollary 1. *A multimaximal coupling* $\left(T_q^1, \ldots, T_q^k\right)$ *exists and is unique for any connection* $R_q^1, \ldots, R_q^k$ *with binary random variables.*

The significance of this result is that insofar as we confine our analysis to c-c systems of binary random variables, every bunch (a row in a c-c matrix) has a known distribution and every connection (a column in the c-c matrix) has a uniquely imposed on it distribution. The only question is whether the distributions along the rows and along the columns of a c-c matrix are mutually compatible, i.e., can be viewed as marginals of an overall coupling of the entire c-c system.

We can now formulate the CbD 2.0 definition of (non)contextuality in systems with binary random variables.

Definition 2. A coupling of a c-c system is called *multimaximally connected* if every subcoupling of this coupling corresponding to a connection of the system is a multimaximal coupling of this connection.

Definition 3. A c-c system of binary random variables is *noncontextual* if it has a multimaximally connected coupling. Otherwise it is *contextual.*

Remark 3. As explained in the next section, any (non)contextual system of binary random variables is completely (non)contextual. Because of this it is unnecessary to use the qualification "completely" in the definition above. Note that this definition applies only to systems of binary random variables. The extension of this definition to arbitrary random variables is not unique, and we leave this topic outside the scope of this paper (but will discuss it briefly in Section 6).

4 Properties of contextuality

Contextuality analysis of the systems of binary random variables is simplified by the following theorem, proved in Ref. [13].

Theorem 2. *Let $R_q^1, \ldots, R_q^k$ be a connection with binary random variables arranged so that the values of $p_i = \Pr\left[R_q^i = 1\right]$ are sorted $p_1 \leq \ldots \leq p_k$. Then $\left(T_q^1, \ldots, T_q^k\right)$ is a multimaximal coupling of $R_q^1, \ldots, R_q^k$ if and only if $\left(T_q^i, T_q^{i+1}\right)$ is a maximal coupling of $\left\{R_q^i, R_q^{i+1}\right\}$ for $i = 1, \ldots, k-1$.*

In other words, in the case of binary random variables, multimaximality can be defined in terms of certain pairs of random variables rather than all possible subsets thereof, as it was done in Definition 1. As shown in Section 6 below, a pairwise formulation can also be used in the general case, for arbitrary random variables.

The main motivation for switching from the maximal couplings of CbD 1.0 to multimaximal couplings is to be able to prove the following theorem.

Theorem 3. *In a noncontextual c-c system of binary random variables every subsystem (obtained from the system by removing from it some of the random variables) is noncontextual.*

Proof. Let S be a multimaximally connected coupling of a system $\mathcal{R}$. Let $\mathcal{R}'$ be a system obtained by deleting a random variable R_q^c from $\mathcal{R}$; and let S' be the set of random variables obtained by deleting from S the corresponding random variable S_q^c. Then S' is a multimaximally connected coupling of $\mathcal{R}'$. Indeed, S' is jointly distributed, its subcouplings corresponding to the system's bunches have the same distributions as these bunches (including the bunch for conteXt c), and its subcouplings corresponding to the system's connections are multimaximal couplings (including the connection for context q, by the definition of a multimaximal coupling). $\square$

There are other desirable properties of the revised definition of contextuality.

First of all we should mention a property shared by CbD 1.0 and 2.0, one that should hold for any reasonable definition of contextuality. If a c-c system is consistently connected (i.e., $R_q^c \sim R_q^{c'}$ for all q, c, c' such that q is measured in both c and c'), then the system is (non)contextual if and only if it is (non)contextual in the traditional sense (as interpreted in CbD): the multimaximal couplings for connections consisting of identically distributed random variables are identity couplings.

Another property worth mentioning is that, using the terminology introduced at the end of Section 2, whether a c-c system of binary random variables is contextual or noncontextual, it is always completely contextual (respectively, completely noncontextual). This follows from the fact that multimaximal couplings for connections consisting of binary random variables are unique, whence if the combination of these unique couplings is (in)compatible with the system's bunches then it is *all* combinations of the couplings that are (in)compatible with the system's bunches.

A third property we find important follows from the fact that if a connection contains just two random variables, then their maximal coupling is their multimaximal coupling. As a result, the theory of contextuality for cyclic c-c systems [16, 17, 20, 25] remains unchanged. Recall that a cyclic c-c system of binary random variables is one in which (1) any bunch consists of two random variables, and (2) any connection consists of two random variables (and, without loss of generality, the c-c system cannot be decomposed into two disjoint cyclic c-c systems). The conteXt-conteNt matrix below shows a cyclic system with 3 conteNts and 3 conteXts (their numbers in a cyclic system are always the same, and called the *rank* of the c-c system):

$$
\begin{array}{|c|c|c|c}
\hline
R_1^1 & R_2^1 & \cdot & c_1 \\
\hline
\cdot & R_2^2 & R_3^2 & c_2 \\
\hline
R_5^1 & \cdot & R_3^3 & c_3 \\
\hline
\multicolumn{1}{c}{q_1} & \multicolumn{1}{c}{q_2} & \multicolumn{1}{c}{q_3} & \boxed{\boxed{\mathcal{CYC}_3}}
\end{array}
.
$$

A prominent example of a noncyclic c-c system each of whose connections consist of two binary random variables is one derived from the Cabello-Estebaranz-Alcaine proof [4] of the Kochen-Specker theorem in 4D space: the system there consists of 36 random variables arranged into 9 bunches (shown below by columns) containing 4 random variables each, and 18 connections (shown by rows) containing two random variables each:

$$
\begin{array}{c}
\\
q_{0001} \\ q_{0010} \\ q_{1100} \\ q_{1200} \\ q_{0100} \\ q_{1010} \\ q_{1020} \\ q_{1212} \\ q_{1221} \\ q_{0011} \\ q_{1111} \\ q_{0102} \\ q_{1001} \\ q_{1002} \\ q_{0120} \\ q_{1121} \\ q_{1112} \\ q_{2111}
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc} c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 \end{array} \\
\left[\begin{array}{ccccccccc}
\star & \star & & & & & & & \\
\star & & & & \star & & & & \\
\star & & \star & & & & & & \\
\star & & & & & & \star & & \\
 & \star & & & \star & & & & \\
 & \star & & & & & & \star & \\
 & \star & & \star & & & & & \\
 & & \star & \star & & & & & \\
 & & \star & & & \star & & & \\
 & & \star & & & & \star & & \\
 & & & \star & & \star & & & \\
 & & & \star & & & & \star & \\
 & & & & \star & & & & \star \\
 & & & & \star & \star & & & \\
 & & & & & \star & & & \star \\
 & & & & & & \star & \star & \\
 & & & & & & \star & & \star \\
 & & & & & & & \star & \star
\end{array}\right]
\end{array}
.
$$

Here, the star symbol in the cell defined by conteXt c_i and conteNt q_j designates a binary random variable R_j^i (the quadruple index at q represents a ray in a 4D real Hilbert space, as labeled in Ref. [4]). The contextual analysis of such systems generalizes the 4D version of the Kochen-Specker theorem in the same way (although computationally more demanding) in which cyclic c-c systems of rank 3,4,5 generalize the treatment of, respectively, the Suppes-Zanotti-Leggett-Garg [19,32], EPR-Bohm-Bell [2,5,18], and Klyachko-Can-Binicoglu-Shumovsky systems [20]. More general proofs of the Kochen-Specker theorem (e.g., by Peres [30]) translate into systems with more than two binary random variables per connection. The multimaximal-couplings-based analysis here will yield different results from the maximal-couplings-based one.

5 A measure of contextuality

In accordance with the linear consistency theorem proved in Ref. [12], a c-c system of random variables always has a *quasi-coupling* that agrees with a given set of couplings imposed on its connections. Let us clarify this.

A *quasi-random variable* X is defined by assigning to its possible values real numbers (not necessarily nonnegative) that sum to 1. These numbers are called *quasi-probability masses*, or simply *quasi-probabilities*. For instance, a variable X with values 1 and 2 to which we assign quasi-probabilities $\mathsf{qPr}\left[X=1\right]=-5$, $\mathsf{qPr}\left[X=2\right]=6$ is a quasi-random variable. A quasi-random variable is a proper random variable if and only if the quasi-probabilities assigned to its values are nonnegative. If a quasi-random variable X is a vector, $(X_1,\ldots,X_n)$, it can be referred to as a vector of jointly distributed quasi-random variables, even if each X_i is a proper random variable. A vector of jointly distributed quasi-random variables may very well have marginals (subvectors) that are proper random vectors.

A quasi-coupling of a c-c system $\mathcal{R}$ is a vector S of jointly distributed quasi-random variables in a one-to-one correspondence with the elements of $\mathcal{R}$, such that every subcoupling of S that corresponds to a bunch of the system has a (proper) distribution that coincides with that of the bunch. Finally, the quasi-coupling S agrees with a set of multimaximal couplings of the system's connections if any subcoupling of S that corresponds to a connection has the same (proper) distribution as this connection's multimaximal coupling.

As an example, consider again our c-c system $\mathcal{R}_{ex}$:

$$
\begin{array}{|c|c|c|c|c}
\cline{1-4}
R_1^1 & R_2^1 & \cdot & R_4^1 & c_1 \\
\cline{1-4}
R_1^2 & \cdot & R_3^2 & \cdot & c_2 \\
\cline{1-4}
R_1^3 & R_2^3 & R_3^3 & R_4^3 & c_3 \\
\cline{1-4}
\multicolumn{1}{c}{q_1} & \multicolumn{1}{c}{q_2} & \multicolumn{1}{c}{q_3} & \multicolumn{1}{c}{q_4} & \boxed{\boxed{\mathcal{R}_{ex}}}
\end{array}
.
$$

Let all random variables be binary. Then, as we know, each connection has a unique multimaximal coupling. Let us denote these couplings (going from the leftmost column to the rightmost one in the matrix)

$$\left(T_1^1, T_1^2, T_1^3\right), \left(T_2^1, T_2^3\right), \left(T_3^2, T_3^3\right), \left(T_4^1, T_4^3\right).$$

The theorem mentioned in the opening line of this section says that one can always find a quasi-coupling S for $\mathcal{R}_{ex}$,

$$\begin{array}{|c|c|c|c|c}
\hline
S_1^1 & S_2^1 & & S_4^1 & c_1 \\
\hline
S_1^2 & \cdot & S_3^2 & \cdot & c_2 \\
\hline
S_1^3 & S_2^3 & S_3^3 & S_4^3 & c_3 \\
\hline
\multicolumn{1}{c}{q_1} & \multicolumn{1}{c}{q_2} & \multicolumn{1}{c}{q_3} & \multicolumn{1}{c}{q_4} & \boxed{\boxed{S_{ex}}}
\end{array},$$

such that

$$\left(S_1^1, S_1^2, S_1^3\right) \sim \left(T_1^1, T_1^2, T_1^3\right),$$

$$\left(S_2^1, S_2^3\right) \sim \left(T_2^1, T_2^3\right),$$

$$\left(S_3^2, S_3^3\right) \sim \left(T_3^2, T_3^3\right),$$

$$\left(S_4^1, S_4^3\right) \sim \left(T_4^1, T_4^3\right).$$

Clearly, the system $\mathcal{R}_{ex}$ is noncontextual if and only if among all such quasi-couplings S_{ex} there is at least one proper coupling.

It is convenient for our purposes to look at this in the following way (introduced in Ref. [12] but derived from an idea proposed in Ref. [6]). For each quasi-coupling S_{ex} one can compute its *total variation*. The latter is defined as the sum of the absolute values of all quasi-probabilities assigned to the values of S_{ex} (i.e., to all 2^9 combinations of values of $S_1^1, S_2^1, \ldots, S_4^3$). If S_{ex} is a proper coupling, this total variation equals 1, otherwise it is greater than 1. Therefore, if the system $\mathcal{R}_{ex}$ is contextual, then the total variation of its quasi-couplings is always greater than 1. As shown in Ref. [12], one can always find a quasi-coupling S_{ex}^* of $\mathcal{R}_{ex}$ that has the smallest possible value of the total variation. This value (perhaps, less 1, if one wants zero rather than 1 to be the smallest value) can be taken to be a *measure of contextuality*.

Generalizing, we have the following statement.

Theorem 4. *Any c-c system of binary random variables has a quasi-coupling whose subcouplings corresponding to the system's connections are their multimaximal couplings. Among all such quasi-couplings there is at least one with the smallest possible value of total variation (which value is then considered a measure of contextuality for the system).*

6 Conclusion: How to generalize

For c-c systems with binary random variables multimaximal couplings are definitely a better way of generalizing identity couplings of the traditional contextuality analysis than maximal couplings. A system that is deemed noncontextual in terms of multimaximal couplings has noncontextual subsystems. The contextuality of a contextual system and noncontextuality of a noncontextual system are both complete if one uses multimaximal couplings to define them. And the theory specializes to the previous version (CbD 1.0) when applied to cyclic systems and to other systems whose connections consist of pairs of random variables.

The question to pose now is what one should do with non-binary random variables. The most straightforward way to construct a general theory is to simply drop the qualification "binary" in Definition 3. There are, however, some complications associated with this approach. Connections involving non-binary variables may not have multimaximal couplings (Section 3) One has to decide whether such systems are contextual, and how to measure the degree of contextuality in them if they are. Another complication, shared with the CbD 1.0, is that multimaximal couplings are not unique if the random variables are not all binary, because of which one no longer can ignore the difference between complete and partial forms of (non)contextuality. Conceptual and computational adjustments have to be made.

At the same time, some of the properties mentioned in Section 4 hold for arbitrary random variables, at least for categorical ones (those with finite number of values). Theorem 3 obviously holds for arbitrary random variables if noncontextuality is taken to be partial. The definition of the (non)contextuality of a system of random variables reduces to the traditional one when a system is consistently connected. Theorem 2 also generalizes to arbitrary random variables, although in a somewhat weaker form due to the loss of the linear ordering of the distributions within a connection.

Theorem 5. *Let $R_q^1, \ldots, R_q^k$ be a connection. Then $\left(T_q^1, \ldots, T_q^k\right)$ is a multimaximal coupling of $R_q^1, \ldots, R_q^k$ if and only if $\left(T_q^c, T_q^{c'}\right)$ is a maximal coupling of $\left\{R_q^c, R_q^{c'}\right\}$ for all $c < c'$ in $\{1, \ldots, k\}$.*

Proof. The "only if" part is true because pairs are subsets. To prove the "if" part, assume the contrary: there is a subset of the connection (without loss of generality, the connection itself, $R_q^1, \ldots, R_q^k$) such that its coupling $\left(T_q^1, \ldots, T_q^k\right)$ is not maximal while $\Pr\left[T_q^c = T_q^{c'}\right]$ is maximal possible for all c, c'. Then, by the theorem on maximal couplings (see Ref. [33] or Ref. [12], Theorem 3.3) there is a value v in the common set of values for all random variables T_q^c such that

$$\Pr\left[T_q^1 = T_q^2 = \ldots = T_q^k = v\right] < \min_{c \in \{1, \ldots, n\}} \left(\Pr\left[T_q^c = v\right]\right),$$

while, for any $c, c' \in \{1, \ldots, k\}$,

$$\Pr\left[T_q^c = T_q^{c'} = v\right] = \min\left(\Pr\left[T_q^c = v\right], \Pr\left[T_q^{c'} = v\right]\right).$$

Then, by replacing each T_q^c with

$$\widetilde{T}_q^c = \begin{cases} 1\, if & T_q^c = v \\ 2\, if \; otherwise \end{cases},$$

and considering $\left(\widetilde{T}_q^1, \ldots, \widetilde{T}_q^k\right)$ a coupling for some connection consisting of binary random variables, we come to a contradiction with Theorem 2. □

There is a complication, however, that seems especially serious for simply dropping the qualification "binary" in Definition 3: this approach allows a noncontextual system of random variables to become contextual under *coarse-graining*. The latter means lumping together some of the values of the variables constituting some of the connections. Thus, if R_q^c has values $1, 2, 3, 4$, one could lump together 1 and 2 and obtain a random variables with three values (and do the same for all other random variables in the connection for conteNt q). It is natural to expect that a system should preserve its noncontextuality under such course-graining, but this is not the case generally.

Example 3. The system consisting of the single connection with six values $(1, 1', 2, 2', 3, 3')$ in Example 2 is noncontextual, because it does have multimaximal couplings. However, if one lumps together i and i' and denotes the lumped value i $(= 1, 2, 3)$, one obtains the system considered in Example 1, which is contextual because it does not have a multimaximal coupling. □

A radical solution for all the problems mentioned is to deal with binary random variables only. This can be achieved by replacing each non-binary random variable R_q^c in a system with a bunch of jointly distributed dichotomizations thereof (that thereby becomes a sub-bunch of the bunch representing conteXt c). For instance, if R_q^c has values $1, 2, 3, 4$, then it could be represented by $2^{4-1} - 1 = 7$ jointly distributed binary random variables. The joint distribution is very simple: of the 2^7 values of this bunch all but 4 have zero probability masses. Of course, every other random variable with conteNt q should be dichotomized in the same way, replacing thereby the corresponding connection with 7 new connections. Coarse-graining in this approach becomes a special case of extracting from a system a subsystem. The price one pays for the conceptual simplicity thus achieved is a great increase of the numbers of random variables in each bunch (becoming infinite if the original system involves non-categorical random variables), although the cardinality of the supports of the bunches remains unchanged. It is to be seen if this dichotomization approach proves feasible.

Acknowledgments. This research has been supported by NSF grant SES-1155956 and AFOSR grant FA9550-14-1-0318. We are grateful to Victor H. Cervantes for his critical comments on the manuscript.

References

1. Bacciagaluppi, G. (2015). Leggett-Garg inequalities, pilot waves and contextuality. International Journal of Quantum Foundations 1, 1-17.

2. Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. Physics 1: 195-200.
3. Bell, J. (1966). On the problem of hidden variables in quantum mechanics. Review of Modern Physics 38, 447-453.
4. Cabello, A., Estebaranz, J. M., & Alcaine, G. G. (1996). Bell-Kochen-Specker theorem: A proof with 18 vectors. Physics Letters A 212:183
5. Clauser, J.F., Horne, M.A., Shimony, A., & Holt, R.A. (1969). Proposed experiment to test local hidden-variable theories. Physical Review Letters 23:880–884.
6. de Barros, J.A., Oas, G. (2014). Negative probabilities and counter-factual reasoning in quantum cognition. Physica Scripta T163:014008.
7. de Barros, J.A., Dzhafarov, E.N., Kujala, J.V., Oas, G. (2015). Measuring Observable Quantum Contextuality. Lecture Notes in Computer Science 9535, 36-47.
8. Dzhafarov, E.N., & Kujala, J.V. (2014). A qualified Kolmogorovian account of probabilistic contextuality. Lecture Notes in Computer Science 8369, 201-212.
9. Dzhafarov, E.N., & Kujala, J.V. (2014). Contextuality is about identity of random variables. Physica Scripta T163, 014009.
10. Dzhafarov, E.N., & Kujala, J.V. (2015). Random variables recorded under mutually exclusive conditions: Contextuality-by-Default. Advances in Cognitive Neurodynamics IV, 405-410.
11. Dzhafarov, E.N., & Kujala, J.V. (2016). Conversations on contextuality. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology, pp. 1-22. New Jersey: World Scientific.
12. Dzhafarov, E.N., & Kujala, J.V. (2016). Context-content systems of random variables: The Contextuality-by-Default theory. To appear in Journal of Mathematical Psychology [arXiv:1511.03516].
13. Dzhafarov, E.N., & Kujala, J.V. (2016). Probabilistic foundations of contextuality. To be published in Fortschritte der Physik - Progress of Physics. [arXiv:1604.08412].
14. Dzhafarov, E.N., Kujala, J.V., Cervantes, V.H. (2016). Contextuality-by-Default: A brief overview of ideas, concepts, and terminology. Lecture Notes in Computer Science 9535, 12-23.
15. Dzhafarov, E.N., Zhang, & R., Kujala, J.V. (2015). Is there contextuality in behavioral and social systems? Philosophical Transactions of the Royal Society A 374: 20150099.
16. Dzhafarov, E.N., Kujala, J.V., & Larsson, J.-A. (2105). Contextuality in three types of quantum-mechanical systems. Foundations of Physics 7, 762-782.
17. Dzhafarov, E.N., Kujala, J.V., Cervantes, V.H., Zhang, R., & Jones, M. (2016). On contextuality in behavioral data. Philosophical Transactions of the Royal Society A 374: 20150234.
18. Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. Physical Review Letters 48: 291–295.
19. Leggett, A.J., & Garg A. (1985). Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? Physical Review Letters 54: 857–860.
20. Klyachko, A.A., Can, M.A., Binicioglu, S., & Shumovsky,A.S. (2008). A simple test for hidden variables in spin-1 system. Physical Review Letters 101:020403.
21. Kochen, S., & Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. Journal of Mathematics and Mechanics, 17:59–87.
22. Kofler, J., & Brukner, C. (2013). Condition for macroscopic realism beyond the Leggett-Garg inequalities, Physical Review A 87:052115.
23. Kujala, J.V., & Dzhafarov, E.N. (2016). Probabilistic Contextuality in EPR/Bohm-type systems with signaling allowed. In E.N. Dzhafarov, S. Jordan,

R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology, pp. 287-308. New Jersey: World Scientific.
24. Kujala, J.V., Dzhafarov, E.N. (2016). Proof of a conjecture on contextuality in cyclic systems with binary variables. Foundations of Physics, 46, 282-299.
25. Kujala, J.V. , & Dzhafarov, E.N., & Larsson, J.-A. (2015). Necessary and sufficient conditions for maximal contextuality in a broad class of quantum mechanical systems. Physical Review Letters 115:150401.
26. Kurzynski, P., Ramanathan, R., & Kaszlikowski, D. (2012). Entropic test of quantum contextuality, Physical Review Letters 109:020404.
27. Kurzynski, P., Cabello, A., & Kaszlikowski, D. (2014). Fundamental monogamy relation between contextuality and nonlocality. Physical Review Letters 112:100401.
28. Lapkiewicz, R., Li, P., Schaeff, C., Langford, N.K., Ramelow, S., Wiesniak, M., & Zeilinger, A. (2011). Experimental non-classicality of an indivisible quantum system. Nature 474: 490.
29. (2013). Lapkiewicz, R., Li, P., Schaeff, C., Langford, N.K., Ramelow, S., Wiesniak, M., & Zeilinger, A.Comment on "Two fundamental experimental tests of nonclassicality with qutrits". arXiv:1305.5529.
30. Peres, A. (1995). Quantum Theory: Concepts and Methods. Dordrecht: Kluwer.
31. Ramanathan, R., Soeda, A., Kurzynski, P., & Kaszlikowski, D. (2012). Generalized monogamy of contextual inequalities from the no-disturbance principle. Physical Review Letters 109:050404.
32. Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? Synthese 48:191–199.
33. Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration.* New York: Springer.