Contextuality and dichotomizations of random variables

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

[1]University of Turku, Finland, jvk@iki.fi
[2]Purdue University, USA, ehtibar@purdue.edu

Abstract

The Contextuality-by-Default approach to determining and measuring the (non)contextuality of a system of random variables requires that every random variable in the system be represented by an equivalent set of dichotomous random variables. In this paper we present general principles that justify the use of dichotomizations and determine their choice. The main idea in choosing dichotomizations is that if the set of possible values of a random variable is endowed with a pre-topology (V-space), then the allowable dichotomizations split the space of possible values into two linked subsets ("linkedness" being a weak form of pre-topological connectedness). We primarily focus on two types of random variables most often encountered in practice: categorical and real-valued ones (including continuous random variables, greatly underrepresented in the contextuality literature). A categorical variable (one with a finite number of unordered values) is represented by all of its possible dichotomizations. If the values of a random variable are real numbers, then they are dichotomized by intervals above and below a variable cut point.

1 Introduction

This paper deals with *systems of random variables*

$$\mathcal{R} = \left\{ R_q^c : c \in C, q \in Q, q \prec c \right\}, \tag{1}$$

where Q denotes the set of properties q being measured (generically referred to as *contents* of the random variables), C denotes the set of conditions under which the measurements are made (referred to as *contexts* of the random variables), and the relation $\prec$, a subset of $Q \times C$, indicates which content is measured in which context. As an example, consider the system of random variables describing Bohm's version of the Einstein-Podolsky-Rosen experiment (EPR/B) [1], the one for which Bell derived his celebrated inequalities [2–4]:

R_1^1	R_2^1			$c=1$
	R_2^2	R_3^2		$c=2$
		R_3^3	R_4^3	$c=3$
R_1^4			R_4^4	$c=4$
$q=1$	$q=2$	$q=3$	$q=4$	$\mathcal{R}$

. (2)

Here, the random variables represent measurements of spins of two entangled spin-$1/2$ particles, one measured by Alice along the axis $q = 1$ or 3, and the other measured by Bob along the axis $q = 2$ or 4. The contexts c here are defined by the four combinations of the Alice-Bob choices of axes, and we have $(q = 1) \prec (c = 1)$, $(q = 2) \prec (c = 1)$, $(q = 2) \prec (c = 2)$, etc. More generally, contexts c may be defined by any systematically varied conditions under which measurements are made. Thus, in the system

R_1^1	R_2^1	$c=1$
R_1^2	R_2^2	$c=2$
$q=1$	$q=2$	$\mathcal{R}$

(3)

the contexts $c=1$ and $c=2$ may represent two orders in which two measurements, of $q=1$ and of $q=2$, are performed: $1\to 2$ and $2\to 1$. In this case, every q is measured in every c.

The sets of contexts and contents, C and Q, can be infinite and even uncountable, although in all practical applications known to us C is finite. The systems $\mathcal{R}$ are classified into *contextual* and *noncontextual* ones. The traditional approaches to contextuality are confined to systems without disturbance, i.e., those in which any two R_q^c and $R_q^{c'}$ have identical distributions. We call such systems *consistently connected* [5]. In fact, in the traditional analysis one usually assumes that consistent connectedness holds in the *strong form*: for any pair of contexts c, $c\prime$,

$$\left\{R_q^c : q\in Q, q\prec c, c'\right\} \overset{d}{=} \left\{R_q^{c'} : q\in Q, q\prec c, c'\right\}, \tag{4}$$

where $\overset{d}{=}$ means "have the same distribution." That is, the joint distributions for identically subscripted random variables are identical [6,7]. For instance, in system (3), strong consistent connectedness means that the two distributions, in $c=1$ and $c=2$, are identical: $\left\{R_1^1, R_2^1\right\} \overset{d}{=} \left\{R_1^2, R_2^2\right\}$.

Our approach, however, called Contextuality-by-Default (CbD) [8–15], also applies to systems with disturbance (e.g., signaling ones), generically referred to as *inconsistently connected* systems. The reason for this is that, both in quantum physics and in non-physical applications, inconsistently connected systems are abundant. Declaring them all contextual or denying the applicability to them of the notion of contextuality seems unreasonably restrictive, as this leaves important empirical phenomena outside contextuality analysis. For instance, system (2) describes not only the EPR/B experiment, but also a single photon two-slit experiment. In this application $q=1$ and 3 stand for the left slit open and closed, respectively, and $q=2$ and 4 stand for the right slit open and closed, respectively. The random variables R_q^c are binary, indicating whether the photon in a given trial hits a localized detector having passed through q when the two slits are in a particular closed-open arrangement c. For example, $R_{q=2}^{c=2}=1$ if the particle passes through the open right slit and hits the detector when the left slit is closed. This system is inconsistently connected, and its CbD analysis shows that it is noncontextual [16]. By contrast, a three-slit single particle experiment, as shown in the same paper, is described by an inconsistently connected system that can be contextual or noncontextual depending on specific distributions of the random variables.

To give another example, system (3) can describe two sequential projective measurements performed on a single particle, and then the system can be easily shown to be inconsistently connected (and CbD analysis shows it is noncontextual [17]). Outside quantum mechanics, system (3) describes an important behavioral phenomenon called "question order effect" [18]. Mathematically, this phenomenon is precisely the inconsistent connectedness. In this application, $q=1$ and 2 are two questions that can be asked of a responder in one of two possible orders.

A practical benefit offered by CbD compared to traditional approaches to contextuality is the ability to analyze real experiments, in which inconsistent connectedness is present either due to the nature of the experimental object, or due to unavoidable or inadvertent design biases. Thus, an important quantum-mechanical experiment [19] aimed at testing the contextuality inequalities for cyclic systems of rank 5 [20] (cyclic systems will be defined below) exhibits two violations of consistent connectedness, one of them expected, the other inadvertent. This makes the traditional theory of contextuality inapplicable without elaborate work-arounds. The CbD analysis of this experiment [8] faces no such difficulty, and demonstrates contextuality in these data with no "corrections" thereof involved. Several other applications of CbD to quantum-mechanical experiments can be found in the literature, e.g. [21–26]. Bacciagaluppi [27,28] used CbD to study the Leggett-Garg paradigm [29] where "signaling in time" typically leads to inconsistent connectedness [30, 31].

In human behavior (including decision making and psychophysical judgments) inconsistent connectedness is universal. Numerous attempts to demonstrate contextuality in behavioral and social systems (reviewed in [17]) have failed because they overlooked or could not properly handle this fact. Contextuality in some systems of random variables describing human behavior was, however, unambiguously demonstrated in recent experiments [32–34].

This paper focuses on a particular aspect of CbD, one that has not been sufficiently elaborated previously. Namely, CbD requires that in contextuality analysis of arbitrary systems every non-binary random variable R_q^c should be dichotomized, replaced with a set of jointly distributed binary variables. We explain in this paper why this should be done and how one is to choose the set of such dichotomizations. We do this by systematically introducing the basics of CbD and relating them to several principles or desiderata for an acceptable theory of contextuality. In the process, we also explain other features of CbD, such as the use of multimaximally connected couplings.

Let us explain the terminology. The distribution of each random variables R_q^c shows the probabilities of various measurable subsets of the set E_q of possible values of R_q^c. The types of the sets E_q endowed with measurable subsets are virtually unlimited: the elements of E_q can be numbers, functions, sets, etc. It is, however, always possible to present R_q^c by a set of jointly distributed *binary* random variables, those attaining values 0 and 1. Indeed, for every measurable subset A of E_q one can form a random variable

$$R_{q,A}^c = \left[R_q^c \in A\right] := \begin{cases} 1 & \text{if } R_q^c \in A \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

The joint distribution of $\left\{R_{q,A}^c : A \in \Sigma_q\right\}$, where Σ_q is the sigma-algebra on E_q, is uniquely determined by and uniquely determines the distribution of R_q^c. The binary variables $R_{q,A}^c$ are called *dichotomizations* of R_q^c, and $\{A, E_q - A\}$ is called a dichotomization of E_q. We can agree not to distinguish $R_{q,A}^c$ and R_{q,E_q-A}^c, and also exclude $A = \emptyset$ and $A = E_q$, for obvious reasons.

The problem of choice arises because in most cases $\left\{R_{q,A}^c : A \in \Sigma_q\right\}$ is too large a set of dichotomizations, and one can equivalently represent R_q^c by much smaller sets $\left\{R_{q,A}^c : A \in \varUpsilon_q\right\}$, with $\varUpsilon_q$ a proper subset of Σ_q. For instance, if a random variable R_q^c is absolutely continuous with respect to the usual Lebesgue measure, the set of possible dichotomizations includes $R_{q,A}^c$ for all Borel-measurable A (or "one half" of them, as we do not distinguish A and $E_q - A$). However, as shown in [10], using this set would lead to the disappointing conclusion that all inconsistently connected systems comprising such random variables are contextual (contravening thereby the Analyticity principle formulated in Section 3). One can do much better by observing that the distribution of such a variable is uniquely described by its distribution function

$$F_q^c(x) = \Pr\left[R_q^c \leq x\right], \tag{6}$$

whence it follows that R_q^c can be equivalently represented by a much smaller set of the variables

$$R_{q,(-\infty,x]}^c = \left[R_q^c \leq x\right]. \tag{7}$$

We will see that this choice of dichotomizations is dictated by the general principles formulated in Section 4. The theory of continuous and other real-valued random variables is discussed in Section 6.

Another class of random variables that plays an important role in contextuality analysis is the class of *categorical* variables, those with a finite set of values that are arbitrary labels, with no ordering. Let, e.g., R_q^c have the probability mass function

$$\begin{array}{rcccc} \textit{value}: & 1 & 2 & 3 & 4 \\ \textit{probability}: & p_1 & p_2 & p_3 & p_4 \end{array}. \tag{8}$$

It has 7 distinct dichotomizations,

$$R_{q,\{1\}}^c, R_{q,\{2\}}^c, R_{q,\{3\}}^c, R_{q,\{4\}}^c, R_{q,\{1,2\}}^c, R_{q,\{2,3\}}^c, R_{q,\{1,3\}}^c, \tag{9}$$

but R_q^c can also be presented by a subset of them, say,

$$R_{q,\{1\}}^c, R_{q,\{2\}}^c, R_{q,\{3\}}^c, \tag{10}$$

with the joint distribution

$R^c_{q,\{1\}}$	$R^c_{q,\{2\}}$	$R^c_{q,\{3\}}$	*probabilty*
1	0	0	p_1
0	1	0	p_2
0	0	1	p_3
0	0	0	p_4
$\vdots$	$\vdots$	$\vdots$	0

. (11)

The theory of categorical random variables is discussed in Section 7. The application of the general principles here leads one to choose the complete set (9) of dichotomizations over its subsets.

2 Systems and their couplings

The random variables R^c_q in (1) are measurable functions

$$R^c_q : \Omega_c \to E_q, \tag{12}$$

which implies that for any given context $c \in C$, the random variables in

$$R^c := \{R^c_q : q \in Q,\, q \prec c\} \tag{13}$$

are jointly distributed (defined on the same sample space Ω_c).[1] The set of random variables R^c is called a *bunch* (intuitively, the variables measured jointly).

For a given content $q \in Q$, the set

$$\mathcal{R}_q := \{R^c_q : c \in C,\, q \prec c\} \tag{14}$$

is called a *connection*; the random variables in a connection take their values in the same space E_q, endowed with the same sigma-algebra Σ_q. The random variables in a connection are *stochastically unrelated* (defined on different sample spaces Ω_c).

A powerful way to investigate relations between stochastically unrelated variables is to construct their probabilistic copies and make them jointly distributed. For instance, to find out how different are two stochastically unrelated variables X_1 and X_2 one can consider all jointly distributed variables (Y_1, Y_2) such that $Y_1 \overset{d}{=} X_1$ and $Y_2 \overset{d}{=} X_2$, and ask what is the minimal probability with which Y_1 and Y_2 can differ. Note that this question is meaningless if posed for the $\{X_1, X_2\}$ themselves, but the minimal probability of $Y_1 \neq Y_2$ can be viewed as a degree of difference between X_1 and X_2. The pair (Y_1, Y_2) in this example is a special case of the notion defined next.

Definition 1. A *coupling* of a set $\{X_i : i \in I\}$ of random variables (generally stochastically unrelated) is a *jointly distributed* set $\{Y_i : i \in I\}$ of correspondingly indexed random variables where each Y_i has the same distribution as X_i :

$$Y_i \overset{d}{=} X_i, \tag{15}$$

for all $i \in I$. Two couplings of the same set of random variables are considered indistinguishable if they have the same distribution.[2]

[1] We use script letters, $\mathcal{R}, \mathcal{R}_q$, etc., for a set of random variables if they are not necessarily jointly distributed. If, as in R^c, all elements of a set are jointly distributed, then R^c is a random variable in its own right, and we can use ordinary italics.

[2] This means that the choice of a domain space for $\{Y_i : i \in I\}$ is irrelevant. There is a canonical way of constructing this space. Let the set of values of X_i be E_i, with the induced sigma-algebra Σ_q. Then the domain space for $\{Y_i : i \in I\}$ can be chosen as the set $\prod_{i \in I} E_i$ endowed with $\bigotimes_{i \in I} \Sigma_i$. With this choice, every Y_i is a coordinate projection function.

For the wealth of uses of this notion in probability theory, see e.g. [35]. The following special types of couplings are important in analyzing systems of random variables. An *independent coupling* is a coupling $\{Y_i : i \in I\}$ such that all random variables in it are independent. This coupling exists and is unique for any $\{X_i : i \in I\}$. A *maximal coupling* $\{Y_1, Y_2\}$ of a pair $\{X_1, X_2\}$ of random variables is a coupling that maximizes the *coupling probability* $\Pr[Y_1 = Y_2]$ among all couplings of $\{X_1, X_2\}$. This coupling also always exist, but is not generally unique unless X_1, X_2 are binary variables.[3]

Definition 2. A *multimaximal* coupling $\{Y_i : i \in I\}$ of $\{X_i : i \in I\}$ is a coupling such that $\{Y_i, Y_j\}$ is a maximal coupling of $\{X_i, X_j\}$ for all $i, j \in I$.

Multimaximal couplings play a central role in contextuality analysis. As a special case, the *identity coupling* of $\{X_i : i \in I\}$ is a coupling $\{Y_i : i \in I\}$ such that all random variables in it are identical: $\Pr[Y_1 = Y_2] = 1$ for all $i, j \in I$. Such a coupling only exists if variables in $\{X_i : i \in I\}$ are identically distributed, and then $\{Y_i : i \in I\}$ is the unique multimaximal coupling of $\{X_i : i \in I\}$.

The following theorem characterizes the multimaximal couplings for binary (0/1) random variables.

Theorem 3. *Let $\{X_i : i \in I\}$ be a set of binary random variables, and let $\{Y_i : i \in I\}$ be a coupling of it. Then, the following statements are equivalent:*

1. *The coupling $\{Y_i : i \in I\}$ is multimaximal.*
2. *Given any finite subset $\{i_1, \ldots, i_n\} \subset I$ of indices such that*

$$\Pr[X_{i_1} = 1] \geq \cdots \geq \Pr[X_{i_n} = 1],$$

the distribution of $Y_{i_1}, \ldots, Y_{i_n}$ is given by the probability mass function $p(y_{i_1}, y_{i_2}, y_{i_3}, \ldots, y_{i_{n-1}}, y_{i_n})$ determined by the $n+1$ probabilities (all other probabilities being zero)

$$\begin{aligned}
p(0,0,0,\ldots,0,0) &= 1 - \Pr[X_{i_1} = 1],\\
p(1,0,0,\ldots,0,0) &= \Pr[X_{i_1} = 1] - \Pr[X_{i_2} = 1],\\
p(1,1,0,\ldots,0,0) &= \Pr[X_{i_2} = 1] - \Pr[X_{i_3} = 1],\\
&\vdots\\
p(1,1,1,\ldots,1,0) &= \Pr[X_{i_{n-1}} = 1] - \Pr[X_{i_n} = 1],\\
p(1,1,1,\ldots,1,1) &= \Pr[X_{i_n} = 1].
\end{aligned}$$

3. *For any finite subset $\{i_1, \ldots, i_n\} \subset I$ of indices, both*

$$\Pr[Y_{i_1} = Y_{i_2} = \cdots = Y_{i_n} = 0] \text{ and } \Pr[Y_{i_1} = Y_{i_2} = \cdots = Y_{i_n} = 1]$$

are maximal possible probabilities among all couplings $\{Y_i : i \in I\}$ of $\{X_i : i \in I\}$.

4. *For any pair of indices $i, j \in I$ such that $\Pr[X_i = 1] \geq \Pr[X_j = 1]$, any one of the following statements:*
   - *(a)* $\Pr[Y_i = 1, Y_j = 1] = \Pr[X_j = 1]$*, the maximal value among all couplings of $\{X_i, X_j\}$.*
   - *(b)* $\Pr[Y_i = 0, Y_j = 0] = \Pr[X_i = 0]$*, the maximal value among all couplings of $\{X_i, X_j\}$.*
   - *(c)* $\Pr[Y_i = 0, Y_j = 1] = 0$.

Proof. With reference to [10], 1⇒ 2 ⇒ 3 ⇒1. That any of 4a, 4b, 4c implies the other two is established by direct computation, any of them is implied by 2, on putting $n = 2$, and 4a with 4b imply 1. □

[3] Denoting the measurable spaces in which X_1, X_2 are taking their values by (E_1, Σ_1) and (E_2, Σ_2), the definition of maximal coupling is predicated on the assumption that the diagonal set $\{(x, y) \in E_1 \times E_2 : x = y\}$ is measurable in $\Sigma_1 \otimes \Sigma_2$. For dichotomous random variables this condition is satisfied trivially.

Theorem 4. *The multimaximal coupling* $\{Y_i : i \in I\}$ *of a set of binary random variables* $\{X_i : i \in I\}$ *is unique.*

Proof. Consider two couplings $\{Y_i : i \in I\}$ and $\{Y_i' : i \in I\}$ of $\{X_i : i \in I\}$, and let E_i denote the set of values of X_i, for $i \in I$. It follows from Theorem 3 (statement 2) that their distributions agree on all *cylinder subsets* of E^I (i.e., Cartesian products of measurable $S_i \subset E_i$, $i \in I$, with $S_i = E_i$ for all but a finite number of $i \in I$). The cylinder sets form a π-*system* (a nonempty collection of sets closed under finite intersections). Since the two distributions agree on a π-system (because cylinder sets correspond to finite subsets $J \subset I$), it follows [36] that they must agree on the σ-algebra generated by the π-system, in this case the product σ-algebra of E^I. □

Theorem 5. *Given any set of binary random variables* $\{X_i : i \in I\}$*, the set* $\{Y_i : i \in I\}$ *defined by*

$$Y_i = [U \leq \Pr(X_i = 1)] := \begin{cases} 1, & U \leq \Pr(X_i = 1), \\ 0, & \textit{otherwise}, \end{cases}$$

where U *is a uniform random variable on* $[0,1]$*, is the (unique) multimaximal coupling of* $\{X_i : i \in I\}$*.*

Proof. That $\{Y_i : i \in I\}$ is a coupling of $\{X_i : i \in I\}$ follows from the fact that (i) all Y_i are functions of the same random variable U (hence they are jointly distributed), and (ii) $\Pr[Y_i = 1] = \Pr[U \leq \Pr(X_i = 1)] = \Pr(X_i = 1)$. For any $i, j \in I$,

$$\Pr[Y_i = Y_j = 1] = \Pr[U \leq \min[\Pr(X_i = 1), \Pr(X_j = 1)]] = \min[\Pr(X_i = 1), \Pr(X_j = 1)],$$

which implies multimaximality by Theorem 3 (statement 4a). □

3 Contextuality

Definition 6. A *coupling* of a system $\mathcal{R} = \{R_q^c : c \in C, q \in Q, q \prec c\}$ is an identically indexed set $S = \{S_q^c : c \in C, q \in Q, q \prec c\}$ of random variables whose set of bunches $\{S^c : c \in C\}$ is a coupling of the set of bunches $\{R^c : c \in C\}$ of $\mathcal{R}$.

In other words, S is a coupling of $\mathcal{R}$ if (as suggested by the notation) the elements of S are jointly distributed and, for any $c \in C$,

$$S^c \overset{d}{=} R^c. \tag{16}$$

Clearly, S contains as its marginals the couplings $S_q = \left\{S_q^c : c \in C, q \prec c\right\}$ for each of the connections $\mathcal{R}_q = \left\{R_q^c : c \in C, q \prec c\right\}$ of $\mathcal{R}$. We can view the coupling S as a system in its own right, and its marginals S_q as connections of this system. Then we can equivalently say either that, within S, connections $\mathcal{R}_q$ have couplings S_q with some property P (e.g., multimaximal couplings) or that $\mathcal{R}$ has a coupling S whose connections S_q have the property P (e.g., multimaximal connections).

The traditional approach [6, 7, 31, 37–42] is that a system is noncontextual if it has a coupling whose connection S_q is the identity coupling of $\mathcal{R}_q$, for every $q \in Q$. Recall that in the identity coupling, $\Pr\left[S_q^c = S_q^{c'}\right] = 1$ for all components of the connection S_q.[4] Thus, for a system $\mathcal{R}$ to be noncontextual in the traditional sense, the system must be consistently connected, and all random variables R_q^c, $R_q^{c'}$, $R_q^{c''}$, ... in every connection $\mathcal{R}_q$ must correspond to one and the same random variable $T_q := S_q^c = S_q^{c'} = S_q^{c''} = \ldots$ in some coupling S of $\mathcal{R}$. If a system is consistently connected but such a coupling does not exist, the system is considered contextual. For an inconsistently connected system, identity couplings of connections do not exist, because $\Pr\left[S_q^c = S_q^{c'}\right]$ cannot reach 1 unless

[4] The notion of a coupling in the traditional approach is not used explicitly (see [12, 13] for difficulties this creates). To our knowledge, Thorisson [35] (Ch. 1, Sec. 10.4, p. 29) was first to use couplings in contextuality analysis of a system. In CbD, they play a central role.

$R_q^c \overset{d}{=} R_q^{c'}$. All inconsistently connected systems therefore, if one follows the logic of the traditional approach, have to be treated as (trivially) contextual, or else as systems whose contextuality status is undefined. This violates a general principle that we will now formulate.

Let us define the format of the system $\mathcal{R}$ in (1) as

$$\mathsf{f} := (\prec, \{(E_q, \Sigma_q) : q \in Q\}). \tag{17}$$

A format therefore specifies which content q is measured in which context c (the sets Q and C are then effectively determined as projections of $\prec$), and it also specifies the type of the random variables involved, i.e. their sets E_q of possible values and the associated sets Σ_q of events. Let us agree to exclude the trivial formats in which every connection consists of a single random variable. The principle in question is as follows.

Analyticity For any given format, among all inconsistently connected systems of this format there are noncontextual systems.

In other words, contextuality status of an inconsistently connected system should depend on its bunch distributions rather than be predetermined by its format. The importance of this principle is that it rules out trivial extensions of the traditional contextuality theory, including the one that declares all inconsistently connected systems contextual.

In CbD, the concept of (non-)contextuality is extended to inconsistently connected systems by replacing identity couplings with multimaximal couplings: the general idea is that a system is noncontextual if, for all $q \prec c, c'$ simultaneously, the value of $\Pr\left[S_q^c = S_q^{c'}\right]$ in some coupling S reaches its maximum (which is 1 if and only if $R_q^c \overset{d}{=} R_q^{c'}$). However, the established definition in CbD requires that the variables in the systems be dichotomized prior to being subjected to contextuality analysis.

Definition 7. A *split representation* of a system $\mathcal{R} = \{R_q^c : c \in C, q \in Q, q \prec c\}$ is a system

$$\mathcal{D} = \{R_{q,A}^c : c \in C, q \in Q, A \in \Upsilon_q, (q, A) \prec c\},$$

where

(i) $R_{q,A}^c$ are binary variables defined by (5),
(ii) the values of $R_{q,A}^c$ for all $A \in \Upsilon_q$ uniquely determine the value of R_q^c,
(iii) $\Upsilon_q \subseteq \Sigma_q$, and Σ_q is the minimal sigma-algebra containing Υ_q,
(iv) $(q, A) \prec c$ if and only if $q \prec c$ and $A \in \Upsilon_q$.[5]

For any $q \in Q$, the subsystem

$$\mathcal{D}_q = \{R_{q,A}^c : c \in C, A \in \Upsilon_q, (q, A) \prec c\}$$

is a split-representation of the system consisting of the single connection $\mathcal{R}_q$.

The indexation of Υ_q implies that the same set of dichotomizations is applied to all random variables in a given connection $\mathcal{R}_q$. All these variables, we remind, have the same set of values E_q and the same sigma-algebra Σ_q. We also remind the convention (to avoid trivial redundancy) that if $A \in \Upsilon_q$ then $E_q - A \notin \Upsilon_q$, and A is a proper, nonempty subset of E_q.

The split representation $\mathcal{D}$ retains the same set of contexts C as in $\mathcal{R}$ but splits each "old" content q into a set of "new" contents $\{(q, A) : A \in \Upsilon_q\}$. For example, suppose that in the system

R_1^1	R_2^1		$c = 1$
R_1^2		R_3^2	$c = 2$
	R_2^3	R_3^3	$c = 3$
$q = 1$	$q = 2$	$q = 3$	$\mathcal{R}$

(18)

[5] There is a slight abuse of notation here: we use the same symbol $\prec$ to indicate the format relation of both $\mathcal{R}$ and $\mathcal{D}$.

the random variables in connection $\mathcal{R}_{q=1}$ have values $E_1 = \{1,2,3,4\}$, the variables in connection $\mathcal{R}_{q=2}$ have values $E_2 = \{a,b,c\}$, and the variables in connection $\mathcal{R}_{q=3}$ are binary. Then, we could represent the original system as

$[R_1^1 \in \{1\}]$	$[R_1^1 \in \{1,2\}]$	$[R_1^1 \in \{1,2,3\}]$	$[R_2^1 \in \{a\}]$	$[R_2^1 \in \{b\}]$		$c=1$
$[R_1^2 \in \{1\}]$	$[R_1^2 \in \{1,2\}]$	$[R_1^2 \in \{1,2,3\}]$			R_3^2	$c=2$
			$[R_2^3 \in \{a\}]$	$[R_2^3 \in \{b\}]$	R_3^3	$c=3$
$q=(1,\{1\})$	$q=(1,\{1,2\})$	$q=(1,\{1,2,3\})$	$q=(2,\{a\})$	$q=(2,\{b\})$	$q=3$	$\mathcal{D}$

. (19)

Since the choice of a split representation of a given $\mathcal{R}$ is not unique without additional constraining principles (to be discussed later), we denote

$$\Upsilon := \{\Upsilon_q : q \in Q\}, \tag{20}$$

and refer to $\mathcal{D}$ in Definition 7 as the Υ-split representation of $\mathcal{R}$.

Definition 8. A system $\mathcal{R}$ is Υ-*noncontextual* if its Υ-split representation $\mathcal{D}$ has a coupling whose connections are multimaximal (such a coupling is called *multimaximally connected*). Otherwise $\mathcal{R}$ is Υ-*contextual.*

Recall that the connections of a coupling of $\mathcal{D}$ are multimaximal if and only if they are multimaximal couplings of the connections of $\mathcal{D}$. Obviously, Υ-split representation of a system is a system of binary random variables, and it is its own and only split representation (up to relabeling of values).

We will see in the following that at least in some cases the set Υ need not be mentioned because its choice is determined uniquely by certain principles, to be formulated in Section 4. Even without these principles, however, Υ obviously need not be mentioned if the variables in $\mathcal{R}$ are binary to begin with. If the sets of contents and contexts in such a system are finite, the contextuality status of the system (as well as measures of (non)contextuality, not discussed in this paper) can be computed by linear programming. For the important special case of cyclic systems, the contextuality status and measures of (non)contextuality can be determined analytically based on formulas derived in [9,14,15]. A cyclic system of rank $n \in \{2,3,\dots\}$ has $2n$ binary random variables arranged in bunches $\{R_i^i, R_{i\oplus 1}^i\}$ for $i = 1,\dots,n$, where $i \oplus 1 = i+1$ for $i < n$ and $n \oplus 1 = 1$. Thus, (2) and (3) represent cyclic systems of ranks 4 and 2, respectively. Cyclic systems cover a large part of traditionally considered systems in physics and behavioral psychology. CbD allows one to analyze these systems with any amount of inconsistent connectedness present.

Definition 8 can be equivalently stated as follows:

Definition 9. A system $\mathcal{R}$ is Υ-*noncontextual* if it has a coupling S whose Υ-split representation is multimaximally connected.

This version is often preferable, because of the following observation.

Lemma 10. *A coupling S of a system $\mathcal{R}$, and the Υ-split representation of S uniquely determine each other.*

Proof. Given a coupling S (which is a system in its own right), a construction of its Υ-split representation is given by Definition 7(i). Conversely, given a Υ-split representation D of a coupling S, Definition 7(ii) implies that each random variable S_c^q of S is fully determined by its representation as $\{R_{q,A}^c : A \in \Upsilon_q\}$ in D, and so the joint distribution of all S_q^c is also determined as all random variables in D are jointly distributed. □

It is clear from the proof that even though the split representation $\mathcal{D}$ may be very large, the support of its coupling has the same cardinality as the support of a coupling S of the original system $\mathcal{R}$.

We conclude this section with the following simple observation.

Theorem 11. *The definition of contextuality in CbD satisfies Analyticity with respect to any Υ-split representation.*

Proof. Given any format, choose, e.g., $\mathcal{R}$ of this format in which all random variables are deterministic (making sure their values vary within a connection so that the system is inconsistently connected). A unique coupling S then trivially exist for $\mathcal{R}$, and any connection in this coupling is multimaximal. Any split representation of S will retain the multimaximality of connections, proving that $\mathcal{R}$ is Υ-noncontextual with respect to any Υ. □

4 Why multimaximality, and why dichotomizations?

CbD is an extension of the traditional understanding of contextuality effected by two modifications thereof: (1) the replacement of identity couplings of connections with multimaximal couplings, and (2) the replacement of systems of random variables with their split representations. Both these modifications have been justified in previous CbD publications [10]. We will recapitulate these justifications briefly.

The only alternative to multimaximal coupling proposed in the literature as a generalization of identity couplings is the notion of a globally maximal coupling. Such a coupling maximizes the probability of

$$S_q^{c_1} = S_q^{c_2} = \ldots = S_q^{c_k}, \tag{21}$$

where $\{c_1, \ldots, c_n\}$ are all contexts in which q is measured (i.e., $\left\{S_q^{c_1}, \ldots, S_q^{c_k}\right\}$ is a coupling of an entire connection). This generalization was adopted in an earlier version of CbD, but abandoned later as it fails to satisfy the following principle.

Noncontextual Nestedness Any subsystem of a noncontextual system is noncontextual.

A subsystem of a system is created by removing certain random variables from the system. Consider, e.g., the system

R_1^1	R_2^1	$c=1$
R_1^2	R_2^2	$c=2$
R_1^3	R_2^3	$c=3$
R_1^4	R_2^4	$c=4$
$q=1$	$q=2$	$\mathcal{R}$

(22)

with binary random variables whose bunches are distributed as

	$R_2^1=1$	$R_2^1=0$
$R_1^1=1$	1/2	0
$R_1^1=0$	0	1/2

	$R_2^2=1$	$R_2^2=0$
$R_1^2=1$	0	1/2
$R_1^2=0$	1/2	0

	$R_2^3=1$	$R_2^3=0$
$R_1^3=1$	1	0
$R_1^3=0$	0	0

	$R_2^4=1$	$R_2^4=0$
$R_1^4=1$	0	0
$R_1^4=0$	0	1

. (23)

It is easy to show that the maximal probability of $S_q^1 = S_q^2 = S_q^3 = S_q^4$ is zero for both $q=1$ and $q=2$. Consequently, any coupling of $\mathcal{R}$ has globally maximal connections. If we adopt the definition of contextuality based on globally maximal couplings, then this system is noncontextual. At the same time, the subsystem

R_1^1	R_2^1	$c=1$
R_1^2	R_2^2	$c=2$
$q=1$	$q=2$	$\mathcal{R}' \subset \mathcal{R}$

(24)

is, by the same definition (which in this case coincides with our Definition 8), contextual. In fact, it has the maximal degree of contextuality among all cyclic systems of rank 2 [15]. This example also shows that the use of globally maximal connections violates another reasonable principle, formulated next.

Deterministic Redundancy Any deterministic random variable can be deleted from a system without affecting its contextuality status; and for any $(q,c)\notin\prec$ (q is not measured in c), one can add a deterministic R_q^c without affecting the system's contextuality status.

As we use multimaximality to define contextuality, the Noncontextual Nestedness principle is satisfied trivially, and it is shown in [11] that the Deterministic Redundancy principle is satisfied as well. However, these and other constraints stipulated in this paper do not determine multimaximality uniquely. A generalization of multimaximally connected couplings, dubbed C-*couplings*, has in fact been considered [11, 44], in which maximality of $\Pr\left[S_q^c, S_q^{c'}\right]$ is replaced by an arbitrary property C that every pair $\left(S_q^c, S_q^{c'}\right)$ in a C-coupling S has to satisfy. Any C-coupling satisfies the principles of Noncontextual Nestedness and Deterministic Redundancy. At present, however, we do not know reasonable alternatives to the maximality of $\Pr\left[S_q^c, S_q^{c'}\right]$ as a realization of property C.

The main reason why CbD requires dichotomization is that outside the class of binary random variables the notion of contextuality does not satisfy the following principle.

Coarse-graining A noncontextual system remains noncontextual following coarse-graining of its random variables.

A coarse-graining is a measurable function $f_q : E_q \to E'_q$, for $q\in Q$. Thus, a coarse-graining maps R_q^c into another random variable $f_q\left(R_q^c\right)$ by lumping together certain elements of E_q (the set of values of R_q^c), doing this in the same way for all variables in a connection. Dichotomization is a special case of coarse-graining, with $E'_q=\{0,1\}$.

Consider the following system:

R_1^1	R_2^1	$c=1$
R_1^2	R_2^2	$c=2$
$q=1$	$q=2$	$\mathcal{R}$

, (25)

with the bunches distributed as

	$R_2^1=1$	$R_2^1=2$	$R_2^1=3$	$R_2^1=4$
$R_1^1=1$	$1/2$	0	0	0
$R_1^1=2$	0	0	0	0
$R_1^1=3$	0	0	$1/2$	0
$R_1^1=4$	0	0	0	0

	$R_2^2=1$	$R_2^2=2$	$R_2^2=3$	$R_2^2=4$
$R_1^2=1$	0	0	0	0
$R_1^2=2$	0	0	0	$1/2$
$R_1^2=3$	0	0	0	0
$R_1^2=4$	0	$1/2$	0	0

. (26)

This system is noncontextual, because any coupling thereof has (multi)maximal connections. However, if we coarse-grain (here, dichotomize) them by

$$f_1 : \downarrow \begin{matrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 0 & 0 \end{matrix}, f_2 : \downarrow \begin{matrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 0 & 0 \end{matrix}, \tag{27}$$

the bunches of the new system will be distributed as

	$R_2^1=1$	$R_2^1=0$
$R_1^1=1$	$1/2$	0
$R_1^1=0$	0	$1/2$

	$R_2^2=1$	$R_2^2=0$
$R_1^2=1$	0	$1/2$
$R_1^2=0$	$1/2$	0

, (28)

and this system, as already stated in the previous example, is contextual.

By contrast, if a system consists of binary random variables, the Coarse-graining principle is satisfied trivially. A coarse-graining of a binary variable maps it into itself (modulo renaming its values) or into a deterministic variable, attaining a single value with probability 1. The latter cannot violate the Coarse-graining principle because the CbD definition of contextuality satisfies the Deterministic Redundancy principle.

If the system to be analyzed is consistently connected, multimaximal couplings reduce to identity couplings, and (non)contextuality in CbD properly specializes to the traditional understanding of (non)contextuality (provided the latter is rigorously stated in terms of couplings). The choice of a split representation for a consistently connected system is inconsequential, and may even be omitted as a matter of convenience.

5 How to choose dichotomizations?

The intuition behind how one has to do coarse-graining in general and dichotomization in particular is simple: allowable "lumping" should only lump together "contiguous" sets of values. This intuition is captured by the notion of (pre-topologically, or V-) *linked sets*.[6]

We define a *symmetrical Fréchet V-space* (see [43] for a general theory of Fréchet V-spaces) as a non-empty set E endowed with a collection $\mathbb{V}$ of nonempty subsets of E, called *vicinities*. A vicinity V can be called a vicinity of any element of V, and every element of E has to have a vicinity. The term "symmetrical" reflects the fact that if y is in a vicinity of x, then x is in the vicinity of y. A topological space is a symmetrical V-space with additional properties that we do not need to use.

Let us illustrate this and related concepts on a simple example. In a psychophysical experiment described in [34], a small visual object (a "dot") could be in one of five positions, as shown,

$$\boxed{\begin{array}{ccc} & * & \\ * & * & * \\ & * & \end{array}}\,, \tag{29}$$

and an observer had to identify the position as *center*, *left*, *right*, *up*, or *down*. Thus the response of the observer was a 5-valued random variable, and we take this set of 5 values as E. Let us associate to each point x of E as its vicinities all sets consisting of x and its one-step-away neighbor. For instance, the point $left$ has the vicinities $V_1 = \{left, up\}$, $V_2 = \{left, center\}$, and $V_3 = \{left, down\}$.

To define V-linked sets, we need to remind the concept of *limit points* (generalized to V-spaces). Given a V-space E, a point $x \in E$ is a limit point of a set $F \subseteq E$ if every vicinity of x contains a point of F other than x.

Definition 12. A subset F of a V-space E is V-linked if
(i) F is a singleton or a vicinity of some point in E;
(ii) F is a union of a V-linked set and a subset of its limit points;
(iii) F is a union of V-linked sets with a nonempty intersection.

When dealing with a random variable R_q^c whose set of values E_q is endowed with a sigma algebra Σ_q, the latter does not generally determine the choice of a V-space $\mathbb{V}_q$ for E_q uniquely (and vice versa). However, in "ordinary" cases, we have a natural choice of $\mathbb{V}_q$ and a natural choice of Σ_q for E_q that satisfy the following definition.

Definition 13. Let a random variable R have a set of possible values E endowed with a V-space $\mathbb{V}$ and a sigma-algebra Σ. The variable R is said to be *ordinary* if Σ is the smallest sigma-algebra containing all the vicinities in $\mathbb{V}$.

In our example (29), one can check that the smallest sigma-algebra containing all the vicinities is the power set of E (because every singleton can be obtained by appropriate intersections of the vicinities).

We are ready now to stipulate the definition that guides our choice of dichotomizations.

[6] This is essentially a weak form of pre-topological connectedness, but we avoid using the latter word to prevent confusing it with its use in CbD, in such terms as "multimaximally connected" or "consistently connected," derived from the term "connection" for the set of random variables sharing a content.

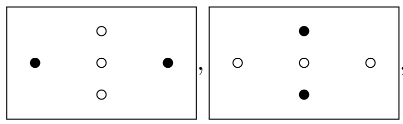
Definition 14. An allowable coarse-graining of V-space E is a surjection $f : E \to E'$ such that E' is a V-space, and

(i) for any V-linked subset X of E, $f(X)$ is a V-linked subset of E', and

(ii) for any V-linked subset Y of E', $f^{-1}(Y)$ is a V-linked subset of E.

A dichotomization is a mapping $f : E \to E'$ where the vicinities in $E' = \{0, 1\}$ are taken to be $\{0\}$, $\{1\}$, and $\{0, 1\}$. So for any E, the dichotomization is allowable if and only if $D_0 = f^{-1}(0)$, $D_1 = F^{-1}(1)$, and $E = F^{-1}(\{0, 1\})$ are V-linked.

Thus, in our example (29), there are 15 distinct partitions of the set into two subsets, and all of them are allowable except for



(30)

where the filled circles form non-linked sets D_0.

The proof of the following statement is obvious.

Theorem 15. *Allowable coarse-grainings are closed under compositions, that is, if* $f : E \to E'$ *and* $g : E' \to E''$ *are allowable coarse-grainings, then* $g \circ f : E \to E''$ *is an allowable coarse-graining. In particular, every allowable dichotomization* $d : E' \to \{0, 1\}$ *of an allowably coarse-grained V-space* $E' = f(E)$ *yields an allowable dichotomization* $d \circ f : E \to \{0, 1\}$ *of the original space* E*.*

It follows that if one forms the split representation of a given system $\mathcal{R}$ by all allowable dichotomizations of each connection, then the Coarse-graining principle is satisfied. Indeed, a split representation of a coarse-grained system is merely a subsystem of the split representation of the original system, because of which if the latter is noncontextual, then so is the former.

In the case of random variables with linearly ordered sets of values $E \subseteq \mathbb{R}$, the natural vicinities of $x \in E$ can be chosen as all intervals $\{z : a < z < b\}$ containing x, and the natural sigma-algebra is the Borel sigma-algebra. The only linked subsets of E are intervals. Thence the allowable dichotomizations are cuts:

$$\{D_0(a) = \{x : x \leq a\}, D_1(a) = \{x : x > a\}\} \tag{31}$$

and

$$\{D'_0(a) = \{x : x < a\}, D'_1(a) = \{x : x \geq a\}\}, \tag{32}$$

for all $a \in E$. One of these two types can be dropped if the other is used, as shown in the next section.

In the case E is a region of $\mathbb{R}^n$, the situation is more complex, as one may associate with it many "natural" but "uninteresting" V-spaces, making too many types of dichotomizations allowable. This leads to all inconsistently connected systems being contextual, in violation of the Analyticity principle. However, a variable with values in $\mathbb{R}^n$ can always be treated as n jointly distributed real-valued variables, in which case the choice reduces to the one previously considered. At present, we do not know whether there are other approaches to $\mathbb{R}^n$-valued variables that comply with Analyticity.

In the case of a categorical random variable, $E = \{1, \ldots, r\}$, its V-space involves all possible subsets, the same as its sigma-algebra. Definition 14 then allows for all possible dichotomizations.

There seems to be no need to multiply examples, as they are easily construable.

6 Cut dichotomizations of real-valued random variables

Consider a *single connection* $\{R_q^1, \ldots, R_q^n\}$ of a system, with all random variables being defined on the set of reals endowed with the usual (Borel) V-space and the Borel sigma-algebra. This includes variables with continuous distribution function, but also a variety of discrete linearly ordered random variables, such as spin measurements in quantum physics, which have ordering $(-\frac{1}{2}, \frac{1}{2})$ for spin-1/2 particles, $(-1,0,1)$ for spin-1 particles, $(-\frac{3}{2},-\frac{1}{2},\frac{1}{2},\frac{3}{2})$ for spin $3/2$-particles, etc. Thus, virtually all measurements in physics can be modeled using real-valued random variables.

Since the content q is fixed, we can drop this subscript and present our connection as $\left\{R^1,\dots,R^n\right\}$. In accordance with Section 5, we replace $\left\{R^1,\dots,R^n\right\}$ with the set of its cuts, that is, we form the system of binary random variables

$$R^k_{(-\infty,x]} := [R^k \le x] := \begin{cases} 1 & \text{if } R^k \le x, \\ 0 & \text{otherwise,} \end{cases} \tag{33}$$

for $k = 1,\dots,n$ (rows) and $x \in \mathbb{R}$ (columns). We could also have chosen them as

$$R^k_{(-\infty,x)} := [R^k < x] := \begin{cases} 1 & \text{if } R^k < x, \\ 0 & \text{otherwise,} \end{cases} \tag{34}$$

but it would not make any difference. Indeed, the set of points

$$\mathbb{R}_0 = \left\{x \in \mathbb{R} : \Pr[R^k \le x] > \Pr[R^k < x] \text{ for some } k \in \{1,\dots,n\}\right\} \tag{35}$$

is at most countable. Let us indicate the elements of $\mathbb{R}-\mathbb{R}_0$ by $\bar{x}$. Consider any coupling $\left\{S^1,\dots,S^n\right\}$ of $\left\{R^1,\dots,R^n\right\}$. Let $a \in \mathbb{R}$ and $i,i' \in \{1,\dots,n\}$ be arbitrary. Using the right-continuity of the distribution functions,

$$\Pr\left[S^i \le a, S^{i'} \le a\right] = \lim_{\bar{x}\to a^+} \Pr\left[S^i \le \bar{x}, S^{i'} \le \bar{x}\right], \tag{36}$$

for any $a \in \mathbb{R}$.

But then

$$\Pr\left[S^i \le \bar{x}, S^{i'} \le \bar{x}\right] = \min\left\{\Pr\left[S^i \le \bar{x}\right], \Pr\left[S^{i'} \le \bar{x}\right]\right\} \tag{37}$$

for all $\bar{x} \in \mathbb{R}-\mathbb{R}_0$ implies

$$\Pr\left[S^i \le a, S^{i'} \le a\right] = \min\left\{\Pr\left[S^i \le a\right], \Pr\left[S^i \le a\right]\right\}, \tag{38}$$

for all $a \in \mathbb{R}$. By Theorem 3 (statement 4a), this means that if the cuts are defined by (33), multimaximality of the split representation of $\left\{S^1,\dots,S^n\right\}$ is implied by the multimaximality of the same split representation from which all connections corresponding to $x \in \mathbb{R}_0$ are removed. Since the reverse implication is trivial, we can replace the implication with equivalence. We can analogously prove the same for the split representations defined by (34), using the left-continuity instead of the right-continuity.

Let us therefore choose (33) for subsequent analysis, and let us write $R^k_{(-\infty,x]}$ more conveniently as R^k_x.

Theorem 16. *The split representation of a single connection formed by cuts as given by (33) is noncontextual.*

Proof. This system has the coupling $\{S^k_x : k = 1,\dots,n,\ x \in \mathbb{R}\}$ where

$$S^k_x = [F^{-1}_k(U) \le x] = [U \le F_k(x)],$$

U is a $[0,1]$ uniform random variable, and F_k and F^{-1}_k are, respectively, the cumulative distribution function and the quantile function of R^k. As $F^{-1}_k(U) \overset{d}{=} R^k$, the first equality implies that this is indeed a coupling of the system. The second equality implies, by Theorem 5, that this coupling is multimaximal. As the whole coupling is multimaximal, its connections are also multimaximal, and the system is noncontextual. □

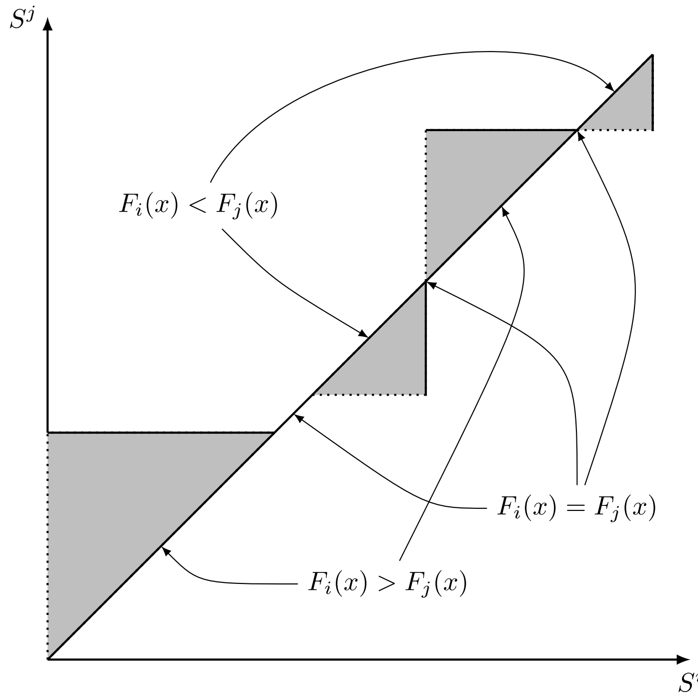


Figure 1: Illustration for Theorem 17. The shaded areas and solid lines contain the support of the joint distribution of $\left(S^i, S^j\right)$ whose split representation has maximal connections.

This is quite a difference from considering all possible dichotomizations, which leads to all inconsistently connected single connections to be contextual [10]. Since a single connection can be viewed as a system of random variables, generally inconsistently connected, the theorem shows that split representations formed by cuts satisfy the Analyticity principle. Of course, a system consisting of more than one connection may very well be contextual.

Consider an arbitrary coupling $\left\{S^1, \dots, S^n\right\}$ of a single connection $\left\{R^1, \dots, R^n\right\}$ such that the split representation of this coupling is multimaximally connected. Theorem 16 says such couplings exist, but the specific coupling constructed in this theorem is not the only possible multimaximally connected coupling. We will now analyze the constraints a coupling $\{S_x^k : k = 1, \dots, n,\ x \in \mathbb{R}\}$ with multimaximal connections imposes on the joint distributions of the coupling $\left\{S^1, \dots, S^n\right\}$. Let us choose two arbitrary elements of the coupling and denote them S^i and S^j, and let F_i and F_j be their respective distribution functions (i.e., the distribution functions of R^i and R^j).

Suppose we have a cut point $x \in \mathbb{R}$ such that $F_i(x) > F_j(x)$. For the dichotomized variables $S_x^i = [S^i \le x]$ and $S_x^j = [S^j \le x]$ this means $\Pr[S_x^i = 1] > \Pr[S_x^j = 1]$, and by Theorem 3 (statement 4c) the pair $\{S_x^i, S_x^j\}$ is maximal if and only if $\Pr[S_x^i = 0, S_x^j = 1] = 0$. This is equivalent to $\Pr\left[S^i > x,\, S^j \le x\right] = 0$, i.e., the joint distribution of (S^i, S^j) vanishes on the set $(x, \infty) \times (-\infty, x]$.

By symmetry, for a cut point $x \in \mathbb{R}$ such that $F_i(x) < F_j(x)$, we have $\{S^i, S^j\}$ maximal if and only if the joint distribution of (S^i, S^j) vanishes on the set $(-\infty, x] \times (x, \infty)$.

If $F_i(x) = F_j(x)$, we have $\Pr[S_x^i = 1] = \Pr[S_x^j = 1]$, and $\{S^i, S^j\}$ is maximal if and only if $\Pr[S_x^i = 0, S_x^j = 1] = \Pr[S_x^i = 1, S_x^j = 0] = 0$, implying that the joint distribution of (S^i, S^j) vanishes in the set $(-\infty, x] \times (x, \infty) \cup (x, \infty) \times (-\infty, x]$.

Let us denote

$$K=\bigcup_{x\in\mathbb{R}}\underbrace{\begin{cases}(-\infty,x]\times(x,\infty)\cup(x,\infty)\times(-\infty,x] & \text{if } F_i(x)=F_j(x),\\ (-\infty,x]\times(x,\infty) & \text{if } F_i(x)<F_j(x),\\ (x,\infty)\times(-\infty,x] & \text{if } F_i(x)>F_j(x).\end{cases}}_{=K_x} \tag{39}$$

The union does not change if instead of all $x \in \mathbb{R}$, we take it only over the union of a countable dense subset of $\mathbb{R}$ and the (at most countable set of) boundary points of the sets $\{x \in \mathbb{R} : F_i(x) < F_j(x)\}$ and $\{x \in \mathbb{R} : F_i(x) > F_j(x)\}$. Since the distribution of (S^i, S^j) vanishes on each K_x, it also vanishes for the countable union of such sets. Thus, we have the following result.

Theorem 17. *The split representation of a coupling $\{S^1, \ldots, S^n\}$ of the single connection $\{R^1, \ldots, R^n\}$ of a system is multimaximally connected if and only if, for any $i, j \in \{1, \ldots, n\}$, the joint distribution of (S^i, S^j) vanishes on the set K given by (39).*

The region left after removing the set indicated in the theorem is shown in Figure 1 for a situation when each of the sets $\{x \in \mathbb{R} : F_i(x) = F_j(x)\}$, $\{x \in \mathbb{R} : F_i(x) < F_j(x)\}$, and $\{x \in \mathbb{R} : F_i(x) > F_j(x)\}$ is a union of disjoint intervals (including isolated single-point ones). The statement of Theorem 17 holds, however, in complete generality.

7 Split representation for categorical random variables

For categorical random variables, Definition 14 leads us to use all possible dichotomizations for split representations of systems. The following notion was introduced in [10].

Definition 18. Given two probability mass functions p and q on the set $\{1, \ldots, k\}$ we say that p *nominally dominates* q if and only if $p(i) < q(i)$ for at most one index $i \in \{1, \ldots, k\}$. If A and B are random variables such that the distribution of A nominally dominates the distribution of B we write $A \succcurlyeq B$.

The significance of this notion is due to the following result obtained in [10]. (As we consider single connections of systems, or single-connection systems, in the remainder of this section, we continue to drop fixed subscripts indicating contents in their notation, writing $\{R^1, \ldots, R^n\}$ instead of $\{R^1_q, \ldots, R^n_q\}$.)

Theorem 19. *The split representation of a single connection $\{R^1, R^2\}$ of two categorical random variables with values in $\{1, \ldots, k\}$ is noncontextual if and only if $R^1 \succcurlyeq R^2$ or $R^1 \preccurlyeq R^2$.*

Since $k \leq 3$ implies that $R^1 \succcurlyeq R^2$ or $R^1 \preccurlyeq R^2$ always holds, the split representation of a connection of two categorical random variables is always noncontextual for $k = 3$. For more than two random variables, this is no longer the case. For instance, we have the following observation.

Example 20. The split representation of all possible dichotomizations of a system consisting of a single connection $\{R^1, R^2, R^3\}$ with values distributed as

	1	2	3
R^1	$1/2$	$1/2$	0
R^2	0	$1/2$	$1/2$
R^3	$1/2$	0	$1/2$

is contextual. This is the same system that was used as an example of a set of multi-valued random variables that does not have a multimaximal coupling in [44]— noncontextuality of the split representation of all possible dichotomizations of a connection implies the existence of a multimaximal coupling of the original connection.

Let us consider next a connection $\mathcal{R} = \{R^1, \dots, R^n\}$ with the value set $\{1, \dots, k\}$, $k \geq 3$. The split representation is contextual only if $R^c \not\succcurlyeq R^{c'}$ or $R^c \not\preccurlyeq R^{c'}$ for all pairs $c, c' \in \{1, \dots, n\}$. This is a necessary condition only. We obtain a rather weak sufficient condition if we impose the following stringent constraints on the ordering of the probability distributions. Let us call $\mathcal{R}$ *dominance-aligned* if, for some permutation $\{k_1, \dots, k_n\}$ of $\{1, \dots, n\}$, the ordering

$$\Pr[R^{k_1} = i] \leq \Pr[R^{k_2} = i] \leq \cdots \leq \Pr[R^{k_n} = i] \tag{40}$$

holds for all but one value of $i \in \{1, \dots, n\}$. It is clear that for the exceptional value of i (which, with no loss of generality, can be taken as $i = 1$), the ordering is opposite,

$$\Pr[R^{k_1} = i] \geq \Pr[R^{k_2} = i] \geq \cdots \geq \Pr[R^{k_n} = i]. \tag{41}$$

Without loss of generality then $\{k_1, \dots, k_n\}$ can be taken to be $\{1, \dots, n\}$, which we will assume in the following proposition.

Theorem 21. *If a single connection $\mathcal{R} = \{R^1, \dots, R^n\}$ is dominance-aligned, the split representation of $\mathcal{R}$ is noncontextual.*

Proof. Let us consider the split representation consisting of the splits

$$R^c_W := [R^c \in W]$$

for all nonempty $W \subset \{2, \dots, n\}$. All other splits are complements of these so this is a complete set of splits. Choose a coupling S of $\mathcal{R}$ such that the events

$$S^1 = i, S^2 = i, \dots, S^n = i \tag{*}$$

form a nested sequence of sets in the domain space of S for each $i = 2, \dots, k$. It follows that the sequence of events

$$S^1 \in W, S^2 \in W, \dots, S^n \in W \tag{**}$$

forms a nested sequence for each nonempty $W \subset \{2, \dots, k\}$, since these sequences are termwise unions of the nested sequences (*). Then, in the split representation of the coupling S, each column S_W, $W \subset \{2,\dots, n\}$, is multimaximal, as the sequence of events

$$S^1_W = 1, S^2_W = 1, \dots, S^n_W = 1$$

corresponds to the nested sequence (**). □

The well-alignedness condition is far from being necessary, as shown in the example below.

Example 22. The split representation of the single-connection system

	a	b	c	d
R^1	.7	.1	.1	.1
R^2	.1	.5	.2	.2
R^3	.2	.2	.3	.3

is noncontextual as it has the coupling

	.1	.1	.1	.1	.1	.1	.1	.1	.1	.1
S^1	a	a	a	a	a	a	a	b	c	d
S^2	b	b	b	b	c	d	a	b	c	d
S^3	b	a	c	d	c	d	a	b	c	d

whose split representation can be verified to be multimaximally connected (see the theorem below for a general condition for this). However, the exceptional index is a for $R^2 \succcurlyeq R^1$ (and for $R^3 \succcurlyeq R^1$) and b for $R^3 \succcurlyeq R^2$.

Theorem 23. *Let $S = \{S^1, \dots, S^n\}$ be a coupling of the single connection $\{R^1, \dots, R^n\}$ with the value set $\{1, \dots, k\}$, $k \geq 3$, and let the index l enumerate all value combinations of S. Let $S^i(l)$ denote the value of S^i in the combination of values indexed by l. With reference to the matrix*

	$\cdots$	l	$\cdots$	l'	$\cdots$
$\vdots$		$\vdots$		$\vdots$	
S^i	$\cdots$	$S^i(l) = x$	$\cdots$	$S^i(l') = y$	$\cdots$
$\vdots$		$\vdots$		$\vdots$	
$S^{i'}$	$\cdots$	$S^{i'}(l) = z$	$\cdots$	$S^{i'}(l') = w$	$\cdots$
$\vdots$		$\vdots$		$\vdots$	
probability	$\cdots$	p_l	$\cdots$	$p_{l'}$	$\cdots$

,

the split representation of S is multimaximally connected if and only if there are no indices i, i', l, l' such that $p_l, p_{l'} > 0$ and

$$x \neq y \neq w \neq z \neq x$$

(which does not preclude $x = w$ and $z = y$).

Proof. If such i, i', l, l' exist, then $\Pr[S^i_{\{x,w\}} = 1, S^{i'}_{\{x,w\}} = 0] \geq p_l > 0$ and $\Pr[S^i_{\{x,w\}} = 0, S^{i'}_{\{x,w\}} = 1] \geq p_{l'} > 0$, which implies by Theorem 3 (statement 4c) that $(S^1_{\{x,w\}}, \dots, S^n_{\{x,w\}})$ is not a multimaximal coupling. Conversely, assume that for some $W \subset \{1, \dots, k\}$ the coupling $(S^1_W, \dots, S^n_W)$ is not multimaximal. This means, by Theorem 3 (statement 4c), that for some i, i' we have $\Pr[S^i_W = 1, S^{i'}_W = 0] > 0$ and $\Pr[S^i_W = 0, S^{i'}_W = 1] > 0$, which further implies that there exist indices l, l' with $p_l, p_{l'} > 0$ such that

$$\underbrace{S^i(l)}_{=:x}, \underbrace{S^{i'}(l')}_{=:w} \in W \not\ni \underbrace{S^{i'}(l)}_{=:z}, \underbrace{S^i(l')}_{=:y}.$$

This implies $x \neq y \neq w \neq z \neq x$. □

Note that in this proof, if the indices i, i', l, l' satisfying the stipulated conditions exist, we always can choose $W = \{x, w\}$ such that $\left(S^i_W, S^{i'}_W\right)$ is not maximal. This W is a two-element or one-element set. In [10], a subsystem of a split representation of system consisting of the one-element or two-element dichotomizations is called a *1-2 subsystem.* We have therefore the following consequence of the theorem.

Corollary 24. *The split representation of a single connection based on all possible dichotomizations is noncontextual if and only if its 1-2-subsystem is noncontextual.*

This generalizes the analogous result obtained in [10] for connections consisting of two random variables.

8 Conclusion

To summarize, the CbD notion of (non)contextuality, based on multimaximality and dichotomizations, satisfies the principles of Analyticity, Noncontextual Nestedness, Deterministic Redundancy, and Coarse-graining. It properly specializes to the traditional notion when applied to consistently connected systems, and contains as a special case the well-developed theory of cyclic systems of binary random variables. We have formulated a general principle by which we choose allowable dichotomizations. It requires that both parts of a dichotomization of the set of possible values E_q be linked subsets of E_q. For the broad class of random variables we called ordinary, this uniquely determines the set

λ_q of all allowable dichotomizations of E_q, and we have presented applications of this principle to real-valued and categorical random variables.

Clearly, we have not provided an exhaustive list of principles or desiderata for a well-constructed theory of contextuality. We may need to explicate additional principles to be able to deduce the CbD theory (or perhaps a generalization thereof) axiomatically, as the only solution.

Conflict of Interests Author declares no conflicts of interests

Data Availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

[1] D. Bohm, Y. Aharonov. Discussion of experimental proof for the paradox of Einstein, Rosen and Podolski. Phys. Rev. 108, 1070-1076 (1957).

[2] J. Bell. On the Einstein-Podolsky-Rosen paradox. Phys. 1, 195 (1964).

[3] J. Bell. On the problem of hidden variables in quantum mechanics. Rev. Modern Phys. 38, 447 (1966).

[4] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. Phys. Rev. Lett. 23, 880 (1969).

[5] E. N. Dzhafarov, J. V. Kujala, and J.-Å. Larsson. Contextuality in three types of quantum-mechanical systems. Foundations of Physics 7, 762-782 (2015).

[6] S. Abramsky, A. Brandenburger The sheaf-theoretic structure of non-locality and contextuality. New J. Phys. 13: 113036-113075 (2011).

[7] S. Abramsky, R. S. Barbosa, S. Mansfield. The contextual fraction as a measure of contextuality. Phys. Rev. Lett. 119: 050504 (2017).

[8] J. V. Kujala, E. N. Dzhafarov, and J.-Å. Larsson. Necessary and sufficient conditions for extended noncontextuality in a broad class of quantum mechanical systems. Phys. Rev. Lett. 115, 150401 (2015).

[9] J. V. Kujala and E. N. Dzhafarov. Proof of a conjecture on contextuality in cyclic systems with binary variables. Found. Phys. 46, 282 (2016).

[10] E. N. Dzhafarov, V. H. Cervantes, and J. V. Kujala. Contextuality in canonical systems of random variables. Phil. Trans. Roy. Soc. A 375, 20160389 (2017).

[11] E. N. Dzhafarov. Replacing nothing with something special: Contextuality-by-Default and dummy measurements. In *Quantum Foundations, Probability and Information* (pp 39-44). Springer, Berlin (2017).

[12] E. N. Dzhafarov and J. V. Kujala. Probabilistic foundations of contextuality. Fortsch. Phys. - Prog. Phys. 65, 1600040 (1-11) (2017).

[13] E. N. Dzhafarov. On joint distributions, counterfactual values, and hidden variables in understanding contextuality. Phil. Trans. Roy. Soc. A 377, 20190144 (2019).

[14] J. V. Kujala and E. N. Dzhafarov. Measures of contextuality and noncontextuality, Phil. Trans. Roy. Soc. A 377, 20190149 (2019).

[15] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes, Contextuality and noncontextuality measures and generalized Bell inequalities for cyclic systems. Phys. Rev. A 101, 042119 (2020). [Erratum Notes in Physical Review A 101, 069902, 2020, and Physical Review A 103, 059901, 2021.]

[16] E. N. Dzhafarov, J. V. Kujala Contextuality analysis of the double slit experiment (with a glimpse into three slits). Entropy 20, 278, doi:10.3390/e20040278 (2018).

[17] E. N. Dzhafarov, R. Zhang, and J. V. Kujala. Is there contextuality in behavioral and social systems? Phil. Trans. Roy. Soc. A 374, 20150099 (2016).

[18] Z. Wang, T. Solloway, R. M. Shiffrin, and J. R. Busemeyer. Context effects produced by question orders reveal quantum nature of human judgments. Proc. Natl. Acad. Sci. 111, 9431 (2014).

[19] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger. Experimental non-classicality of an indivisible quantum system. Nature 474, 490 (2011).

[20] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky. A simple test for hidden variables in spin-1 system. Phys. Rev. Lett. 101, 020403 (2008).

[21] M. Arias, G. Canas, E. S. Gomez, J. B. Barra, G. B. Xavier, G. Lima, V. D'Ambrosio, F. Baccari, F. Sciarrino, A. Cabello. Testing noncontextuality inequalities that are building blocks of quantum correlations. Phys. Rev. A 92:032126 (2015).

[22] A. Crespi, M. Bentivegna, I. Pitsios, D. Rusc, D. Poderini, G. Carvacho, V. D'Ambrosio, A. Cabello, F. Sciarrino, R. Osellame. Single-photon quantum contextuality on a chip. ACS Phot. 4:2807-2812 (2017).

[23] X. Zhan, P. Kurzyński, D. Kaszlikowski, K. Wang, Zh. Bian, Y. Zhang, P. Xue. Experimental detection of information deficit in a photonic contextuality scenario. Phys. Rev. Lett. 119:220403 (2017).

[24] C. Flühmann, V. Negnevitsky, M. Marinelli, J. P. Home Sequential modular position and momentum measurements of a trapped ion mechanical oscillator. Phys. Rev. X 8:021001 (2018).

[25] F. M. Leupold, M. Malinowski, C. Zhang, V. Negnevitsky, J. Alonso, A. Cabello, J. P. Home. Sustained state-independent quantum contextual correlations from a single ion. Phys. Rev. Lett. 120:180401 (2018).

[26] M. Malinowski, C. Zhang, F. M. Leupold, J. Alonso, J. P. Home, A. Cabello. Probing the limits of correlations in an indivisible quantum system. arXiv:1712.06494v2 (2018).

[27] G. Bacciagaluppi. Einsten, Bohm, and Leggett-Garg. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). *Contextuality from Quantum Physics to Psychology* (pp. 63-76). New Jersey: World Scientific (2015).

[28] G. Bacciagaluppi. Leggett-Garg inequalities, pilot waves and contextuality. Int. J. Quant. Found. 1, 1 (2015).

[29] A. J. Leggett and A. Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? Phys. Rev. Lett. 54, 857 (1985).

[30] J. Kofler and C. Brukner. Condition for macroscopic realism beyond the Leggett-Garg inequalities. Phys. Rev. A 87, 052115 (2013).

[31] C. Budroni. *Temporal Quantum Correlations and Hidden Variable Models*. Springer: Heidelberg (2016).

[32] V. H. Cervantes, E. N. Dzhafarov (2018). Snow Queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices. Decision 5:193-204.

[33] I. Basieva, V. H. Cervantes, E. N. Dzhafarov, A. Khrennikov. True contextuality beats direct influences in human decision making. J. Exp. Psych.: Gen. 148, 1925-1937 (2019).

[34] E. N. Dzhafarov and V. H. Cervantes. True contextuality in a psychophysical experiment. J. Math. Psych. 91, 119-127 (2019).

[35] H. Thorisson. *Coupling, Stationarity, and Regeneration*. New York: Springer (2000).

[36] D. Williams. *Probability with martingales*. Cambridge University Press (1991).

[37] P. Suppes and M. Zanotti. When are probabilistic explanations possible? Synth. 48, 191 (1981).

[38] A. Fine. Joint distributions, quantum correlations, and commuting observables. J. Math. Phys. 23, 1306 (1982).

[39] M. Araújo, M. T. Quintino, C. Budroni, M.T. Cunha, A Cabello. All noncontextuality inequalities for the n-cycle scenario, Phys. Rev. A 88: 022118 (2013).

[40] A. Cabello. Simple explanation of the quantum violation of a fundamental inequality. Phys. Rev. Lett., 110: 060402 (2013).

[41] P. Kurzynski, A. Cabello, D. Kaszlikowski. Fundamental monogamy relation between contextuality and nonlocality. Phys. Rev. Lett. 112: 100401 (2014).

[42] P. Kurzynski, R. Ramanathan, D. Kaszlikowski. Entropic test of quantum contextuality. Phys. Rev. Lett. 109: 020404 (2012).

[43] W. Sierpinski. *General Topology*. University of Toronto Press: Toronto (1956).

[44] E. N. Dzhafarov, J. V. Kujala. Contextuality-by-Default 2.0: Systems with binary random variables. In J.A. de Barros, B. Coecke, E. Pothos (Eds.) *Lecture Notes in Computer Science* 10106, 16-32 (2017).