Contextuality in Canonical Systems of Random Variables

Ehtibar N. Dzhafarov,[1] Víctor H. Cervantes,[2] and Janne V. Kujala[3]

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

Abstract

Random variables representing measurements, broadly understood to include any responses to any inputs, form a system in which each of them is uniquely identified by its content (that which it measures) and its context (the conditions under which it is recorded). Two random variables are jointly distributed if and only if they share a context. In a canonical representation of a system, all random variables are binary, and every content-sharing pair of random variables has a unique maximal coupling (the joint distribution imposed on them so that they coincide with maximal possible probability). The system is contextual if these maximal couplings are incompatible with the joint distributions of the context-sharing random variables. We propose to represent any system of measurements in a canonical form and to consider the system contextual if and only if its canonical representation is contextual. As an illustration, we establish a criterion for contextuality of the canonical system consisting of all dichotomizations of a single pair of content-sharing categorical random variables.

KEYWORDS: canonical systems, contextuality, dichotomization, direct influences, measurements.

1 Introduction

We begin by recapitulating the basics of our theory of "quantum-like" contextuality, and then explain how this theory is developed in this paper. The name of the theory is *Contextuality-by-Default* (CbD), and its recent accounts can be found in Ref. [10–12].

Remark 1.1. We use the following two notation conventions throughout the paper: (1) owing to its frequent occurrence we abbreviate the term *random variable* as *rv* (*rvs* in plural); and (2) we unconventionally capitalize the words *conteNt* and *conteXt* to prevent their confusion in reading.

The matrix below represents the smallest possible version of what we call a *cyclic system* [15, 23–25]:

$$\begin{array}{|c|c|c}\cline{1-2} R_1^1 & R_2^1 & c=1\\ \cline{1-2} R_1^2 & R_2^2 & c=2\\ \cline{1-2} \multicolumn{1}{c}{q=1} & \multicolumn{1}{c}{q=2} & \boxed{\mathcal{R}}\end{array} .$$

Each of the rvs R_q^c represents measurements of one of two properties, $q = 1$ or $q = 2$, under one of two conditions, $c = 1$ or $c = 2$. The "properties" q can also be called "objects," "inputs," "stimuli," etc., depending on the application, and we refer to q generically as the *conteNt* of the measurement R_q^c. The superscript c in R_q^c describes how and under what circumstances q is measured, including what other conteNts are measured together with q. We refer to c generically (and traditionally) as the *conteXt* of the measurement R_q^c. The conteNt-conteXt pair (q, c) provides a *unique identification* of R_q^c within the system of measurements $\mathcal{R}$. In addition, being an rv, R_q^c is characterized by its *distribution.* In this paper, consideration is confined to *categorical rvs*, those with finite numbers of values. The term "measurement" is understood very broadly, to include any response to any input or stimulus.

Let us begin with the simplest case of the system $\mathcal{R}$, when all four rvs R_q^c are *binary.* In quantum physics, R_q^c may describe a measurement of spin along one of two fixed axes, $q = 1$ or $q = 2$, in a spin-$^1\!/_2$ particle. In psychology, R_q^c may describe a response to one of two Yes-No questions, $q = 1$ or $q = 2$. In both applications, in conteXt $c = 1$ one measures first $q = 1$ and then $q = 2$; in conteXt $c = 2$ the measurements are made in the opposite order. The rvs sharing a conteXt c are recorded in pairs, (R_1^c, R_2^c), which means that they are *jointly distributed* and can be viewed as a single (here, four-valued) rv. No such joint distribution is defined for rvs in different conteXts, such as R_2^1 and R_1^2. They are *stochastically unrelated* (to each other): one cannot ask about the probability of an "event" $\left[R_2^1 = x, R_1^2 = y\right]$, as no such "event" is defined. In particular, two conteNt-sharing rvs, R_q^1 and R_q^2, are always stochastically unrelated, hence they can never be considered one and the same rv, even if they are identically distributed (see Ref. [11] for a detailed probabilistic analysis).

In both applications mentioned, the distributions of R_q^1 and R_q^2 are de facto different. In the quantum-mechanical example, the first spin measurement generally changes the state of the particle [4]. Assuming identical preparations in both conteXts c, therefore, the state of the particle when a q-spin is measured first will be different from that when it is measured second. In the behavioral example, one's response to a question asked second will generally be influenced by the question asked first [27, 29]. This creates obvious *conteXt-dependence* of the measurements, but this is not what we call *contextuality* in our theory. The original meaning of the term in quantum mechanics, when translated into the language of probability theory (as in Refs. [10, 11, 13] and, with caveats, [6, 17, 19, 20, 25, 26, 28]), is that measurements of one and the same physical property q have to be represented by different rvs depending on what other properties are being measured together with q — even when the laws of physics exclude all *direct interactions* (energy/information transfer) between the measurements. By extension, when such direct interactions are present, as they are in our two applications of the system $\mathcal{R}$, we speak of contextuality only if the dependence of R_q^c on c is greater, in a well-defined sense, than just the changes in its distribution. Contextuality is a non-causal aspect of conteXt-dependence, revealed in the probabilistic relations between different measurements rather than in their individual distributions.

This is how this understanding is implemented in CbD. We characterize the conteXt-induced changes in the individual distributions, i.e., the difference between those of R_q^1 and R_q^2, by *maximally coupling* them. This means that we replace R_q^1 and R_q^2 with jointly distributed T_q^1 and T_q^2 that have the same respective individual distributions, and among all such couplings we find one with the maximal value of $\Pr\left[T_q^1 = T_q^2\right]$. This maximal coupling $\left(T_q^1, T_q^2\right)$ always exists and is unique. The next step is to see if there exists an *overall coupling* S of $\mathcal{R}$, a jointly distributed quadruple with elements corresponding to those of $\mathcal{R}$,

$$\begin{array}{|c|c|l} \cline{1-2} S_1^1 & S_2^1 & c=1 \\ \cline{1-2} S_1^2 & S_2^2 & c=2 \\ \cline{1-2} \multicolumn{1}{c}{q=1} & \multicolumn{1}{c}{q=2} & \boxed{S} \end{array} \quad ,$$

such that its rows (S_1^c, S_2^c) are distributed as the rows of $\mathcal{R}$ and its columns $\left(S_q^1, S_q^2\right)$ are distributed as the maximal couplings $\left(T_q^1, T_q^2\right)$ of the columns of $\mathcal{R}$. If such a *maximally-connected* coupling S does not exist, one can say that the within-conteXt (row-wise) relations prevent different measurements of the same conteNt (column-wise) to be as close to each other as this is allowed by the direct influences alone. Put differently, the relations of R_q^1 and R_q^2 with their same-conteXt counterparts force them, if imposed a joint distribution upon, to coincide less frequently than if these relations are ignored. The system then is deemed *contextual.* Conversely, if the coupling S above exists, the within-conteXt relations do not make the measurements of R_q^1 and R_q^2 any more dissimilar than required by the direct influences: the system is *noncontextual.*

The (non)existence of S is determined by a simple linear programing procedure [10, 15]: in our example, S has 2^4 possible values, and we find out if they can be assigned nonnegative numbers (probability masses) that sum to the given row-wise probabilities $\Pr\left[R_1^c = x, R_2^c = y\right]$ and the computed column-wise probabilities $\Pr\left[T_q^1 = x, T_q^2 = y\right]$. There is also a simple criterion (inequality) for the existence of a solution for this system of equations [15, 24, 25]. Using it one can show, e.g., that in our quantum-mechanical application the system $\mathcal{R}$ is always noncontextual, and this is also true for the behavioral application if one adopts the model proposed in Ref. [29] (see Ref. [16] for details). Mathematically, however, the system $\mathcal{R}$ can be contextual, and if it is, CbD provides a simple way of computing the *degree* of its contextuality [10]: one replaces the probability masses in the above linear programing task with *quasi-probabilities*, allowed to be negative, and finds among the solutions the minimum sum of their absolute values (see Section 2.3).

Although most of these principles and procedures of CbD have been formulated for arbitrary systems of measurements [10, 13], they only work without complications with systems that satisfy the following two constraints: (A) they contain only binary rvs, and (B) there are no more than two rvs sharing a conteNt (i.e., occupying the same column). What we propose in this paper is to always present a system of measurements in a *canonical form*, which is in essence one with the properties A and B. The cyclic systems form a subclass of canonical systems, rich enough to cover most experimental paradigms of traditional interest in quantum-mechanical and behavioral contextuality studies [10, 13–16, 25], but far from satisfactory generality.

What are the complications one faces if a system does not satisfy the properties A and B? Consider the system below, with all its rvs binary but with three rather than two of them in each column:

$$\begin{array}{|c|c|l} \cline{1-2} R_1^1 & R_2^1 & c=1 \\ \cline{1-2} R_1^2 & R_2^2 & c=2 \\ \cline{1-2} R_1^3 & R_2^3 & c=3 \\ \cline{1-2} \multicolumn{1}{c}{q=1} & \multicolumn{1}{c}{q=2} & \boxed{\mathcal{R}'} \end{array}$$

How does CbD apply here? In the earlier version of theory (summarized in Refs. [10, 13]) we computed the couplings $\left(T_q^1, T_q^2, T_q^3\right)$ of each column that maximize $\Pr\left[T_q^1 = T_q^2 = T_q^3\right]$. One problem with this approach is that the maximal coupling $\left(T_q^1, T_q^2, T_q^3\right)$, while it always exists, is not defined uniquely. What should be the contextuality analysis of $\mathcal{R}'$ if the within-conteXt (row-wise) distributions are compatible with some but not all combinations of the maximal couplings for the two columns? Shall one then speak of a partial (non)contextuality? Originally we proposed to consider a system noncontextual if it is compatible with at least one of these pairs of maximal couplings, but in addition to being arbitrary, this leads to another complication: it may then very well happen that the system $\mathcal{R}'$ is noncontextual but one of its subsystems, e.g. $\mathcal{R}$, is contextual. This is contrary to one's intuition of noncontextuality.

In the most recent publications therefore [11, 12] we modified our approach into "CbD 2.0," by positing that a coupling for conteNt-sharing measurements should be computed so that it maximizes the probability of coincidence for every *pair* (equivalently, every subset) of them. In our case, this means maximization of $\Pr\left[T_q^1 = T_q^2\right]$, $\Pr\left[T_q^2 = T_q^3\right]$, and $\Pr\left[T_q^1 = T_q^3\right]$ (it is in fact sufficient to maximize only certain pairs rather than all of them, but this is not critical here). Such a coupling $\left(T_q^1, T_q^2, T_q^3\right)$ is called *multimaximal.* With only binary rvs involved, a multimaximal coupling always exists and is unique; and a subsystem of a noncontextual system then is always noncontextual.

Returning to system $\mathcal{R}$, consider now the situation when the measurements involved are not dichotomous. For example, let the two successive spin measurements along axes $q = 1$ and $q = 2$ be made on a hypothetical spin-2 particle, with the measurement outcomes denoted $\{-2, -1, 0, 1, 2\}$. In the behavioral application, let the questions asked allow 5 answers each, labeled in the same way. A maximal coupling in this situation exists for each column of $\mathcal{R}$, but not uniquely. This takes us back to the problem of what one should do if the row-wise distributions are compatible with some but not all pairs of these maximal couplings. Another problem is even harder. If the system is deemed noncontextual, one may consider it desirable that it remain noncontextual after some of the measurement outcomes are "lumped together." Thus, one may wish to consider $\{-2, -1, 0, 1, 2\}$ in terms of "negative-zero-positive," lumping together -2 with -1 and 2 with 1. Or one may wish to look at the outcomes in terms of "zero-nonzero." As it turns out, a noncontextual system may become contextual after such coarsening of some of its measurements.

Both these problems can be resolved if we agree that *every measurement included in the system, empirically recorded or computed from those empirically recorded, should be represented by a set of binary rvs.* Let us denote by D_{qW}^c the Bernoulli rv that equals 1 if the value of R_q^c is within the subset W of its possible values. We call D_{qW}^c a *split* (of the original rv). We posit that a measurement with k distinct values should always be represented by k "detectors" of these values, i.e. the splits with one-element subsets W. Thus, in our system $\mathcal{R}$, each measurement R_q^c should be replaced with the jointly distributed splits

$$\left(D_{q\{-2\}}^c, D_{q\{-1\}}^c, D_{q\{0\}}^c, D_{q\{1\}}^c, D_{q\{2\}}^c\right).$$

If one is also interested in the coarsening of $R_{q'}^c$ into values "negative-zero-positive," then the list should be expanded into

$$\left(D_{q\{-2\}}^c, D_{q\{-1\}}^c, D_{q\{0\}}^c, D_{q\{1\}}^c, D_{q\{2\}}^c, D_{q\{-2,-1\}}^c, D_{q\{1,2\}}^c\right).$$

If one wishes to include *all* possible coarsenings of the original rvs in $\mathcal{R}$, then the set of binary rvs should consist of *all* possible splits. Since every dichotomization creating a split should be applied to all rvs sharing a conteNt, one ends up replacing the system $\mathcal{R}$ with

$$\begin{array}{|c|c|c|c|c|c|c|c|l}
\cline{1-8}
D_{1\{-2\}}^1 & \cdots & D_{1\{2\}}^1 & D_{1\{-2,-1\}}^1 & \cdots & D_{1\{1,2\}}^1 & \cdots & D_{2\{1,2\}}^1 & c=1 \\
\cline{1-8}
D_{1\{-2\}}^2 & \cdots & D_{1\{2\}}^2 & D_{1\{-2,-1\}}^2 & \cdots & D_{1\{1,2\}}^2 & \cdots & D_{2\{1,2\}}^2 & c=2 \\
\cline{1-8}
\multicolumn{1}{c}{q=1\,\{-2\}} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{q=1\,\{2\}} & \multicolumn{1}{c}{q=1\,\{-2,-1\}} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{q=1\,\{1,2\}} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{q=2\,\{1,2\}} & \boxed{\mathcal{D}}
\end{array}$$

There are $(2^5 - 2)/2 = 15$ distinct dichotomizations of the set $\{-2, -1, 0, 1, 2\}$, and the 15 subsets W in D_{qW}^c should be chosen to avoid duplication, such as in $D_{q\{0,1\}}^c$ and $D_{q\{-2,-1,2\}}^c$. Once duplication is prevented, however, all splits of all rvs one is interested in should be included. It is irrelevant that some of them can be presented as functions of the others. In fact, any split of our R_q^c can be presented as a function of just three splits, chosen, e.g., as

$$D_{q'}^c = D_{q\{-1,1\}}^c, D_{q''}^c = D_{q\{0,1\}}^c, D_{q'''}^c = D_{q\{2\}}^c.$$

It is easy to show, however, that in the subsystem

$$\begin{array}{|c|c|c||c|l}
\cline{1-4}
D_{1'}^1 & D_{1''}^1 & D_{1'''}^1 & f\left(D_{1'}^1, D_{1''}^1, D_{1'''}^1\right) & c=1 \\
\cline{1-4}
D_{1'}^2 & D_{1''}^2 & D_{1'''}^2 & f\left(D_{1'}^2, D_{1''}^2, D_{1'''}^2\right) & c=2 \\
\cline{1-4}
\multicolumn{1}{c}{q=1'} & \multicolumn{1}{c}{q=1''} & \multicolumn{1}{c}{q=1'''} & \multicolumn{1}{c}{q^*} & \boxed{\mathcal{D}'}
\end{array}$$

of the system $\mathcal{D}$, the f-transformation of the maximal couplings of the first three columns, since these couplings are not jointly distributed, would not determine the coupling of the fourth column, let alone ensure that this coupling is maximal.

There is no general prescription as to which rvs should or should not be included in the system representing an empirical set of measurements: what one includes (e.g., what coarsenings of the rvs already in play one considers) reflects what aspects of the empirical situation one is interested in. Once a set of rvs is chosen, however, we uniquely form their splits and place them in a canonical system.

The remainder of the paper is organized as follows. In Section 2, we present the abstract version of CbD applicable to all possible systems of categorical (and not only categorical) rvs. In Section 3, we formalize the idea of representing any system of rvs by their splits and applying contextuality analysis to these representations only. In Section 4, we investigate the representation of all coarsenings of a single pair of conteNt-sharing rvs by all possible splits. In the concluding section we explain why one might wish to consider only some rather than all possible splits.

Remark 1.2. The proofs of the formal propositions in the paper, unless obvious or referenced as presented elsewhere, are given in the supplementary file S, together with additional theorems and examples.

2 Formal Theory of Contextuality

2.1 Basic notions

The definition of a system of rvs requires two nonempty finite sets, a set of *conteNts* Q and a set of *conteXts* C. There is a relation

$$\prec\ \subseteq Q \times C, \tag{1}$$

such that the projections of $\prec$ into Q and C equal Q and C, respectively (this means that for every $q \in Q$ there is a $c \in C$, and vice versa, such that $q \prec c$). We read both $q \prec c$ and $c \succ q$ as "*q is measured in c.*"

A categorical rv is one with a finite set of values and its power set as the codomain sigma-algebra. A system of (categorical) rvs is a double-indexed set (we use calligraphic letters for sets of random variables)

$$\mathcal{R} = \left\{ R_q^c : q \in Q, c \in C, q \prec c \right\}, \tag{2}$$

such that (i) any R_q^c and $R_q^{c'}$ have the same set of possible values; (ii) R_q^c and $R_{q'}^{c'}$ are jointly distributed if $c = c'$; and (iii) if $c \neq c'$, R_q^c and $R_{q'}^{c'}$ are *stochastically unrelated* (possess no joint distribution). For any $c \in C$ the subset

$$\mathcal{R}^c = \left\{ R_q^c : q \in Q, q \prec c \right\} = R^c \tag{3}$$

of $\mathcal{R}$ is called a *bunch* (of rvs) corresponding to c. Since the elements of a bunch are jointly distributed, the bunch is a (categorical) rv in its own right, so it can be also written as R^c. Note that we do not distinguish the representations of $\mathcal{R}$ as (2) and as

$$\mathcal{R} = \left\{ R^c : c \in C \right\}. \tag{4}$$

(See Refs. [10, 11] for a detailed probabilisitic analysis.)

For any $q \in Q$, the subset

$$\mathcal{R}_q = \left\{ R_q^c : c \in C, q \prec c \right\} \tag{5}$$

of $\mathcal{R}$ is called a *connection* (between the bunches of rvs) corresponding to q. Any two elements of a connection are stochastically unrelated, so it is not an rv.

2.2 General definition of (non)contextuality

A (probabilistic) *coupling* Y of a set of rvs $\{X_1, \ldots, X_n\}$ is a set of jointly distributed $\{Y_1, \ldots, Y_n\}$ such that $Y_i \sim X_i$ for $i = 1, \ldots, n$. The tilde $\sim$ stands for "has the same distribution as."

An (overall) coupling S of a system $\mathcal{R}$ in (2) is a coupling of its bunches. That is, it is an rv

$$S = \left\{ S^c : c \in C \right\} \tag{6}$$

(with jointly distributed components) such that $S^c \sim R^c$, for any $c \in C$. This implies that

$$S^c = \left\{ S_q^c : q \in Q, q \prec c \right\} \tag{7}$$

is a set of jointly distributed rvs in a one-to-one correspondence with the identically labeled elements of $\mathcal{R}$.

For a given $q \in Q$, a coupling T_q of a connection $\mathcal{R}_q$ is an rv

$$T_q = \left\{ T_q^c : c \in C, q \prec c \right\} \tag{8}$$

such that $T_q^c \sim R_q^c$. In particular, if S is a coupling of $\mathcal{R}$, then

$$S_q = \left\{ S_q^c : c \in C, q \prec c \right\} \tag{9}$$

is a coupling of $\mathcal{R}_q$, for any $q \in Q$.

Definition 2.1. Given a set $\mathcal{T} = \{T^c : c \in C\}$ of couplings for all connections in a system $\mathcal{R}$, the system is said to be *noncontextual with respect to* $\mathcal{T}$ if $\mathcal{R}$ has a coupling S with $S_q \sim T_q$ for any $q \in Q$. Otherwise $\mathcal{R}$ is said to be *contextual with respect to* $\mathcal{T}$.

Put differently, $\mathcal{R}$ is noncontextual with respect to $\mathcal{T}$ if and only if there is a jointly distributed set

$$S = \left\{ S_q^c : q \in Q, c \in C, q \prec c \right\}, \tag{10}$$

such that, for every $c \in C$, $S^c \sim R^c$, and for every $q \in Q$, $S_q \sim T_q$. A coupling S with this property is called $\mathcal{T}$*-connected.*

If the couplings T_q are characterized by some property C such that one and only one coupling T_q satisfies this property for any given connection $\mathcal{R}_q$, then the definition can be rephrased as follows:

Definition 2.2. $\mathcal{R}$ is said to be *noncontextual with respect to property* C if it has a C*-connected* coupling S, defined as one with S_q satisfying C for any $q \in Q$. Otherwise $\mathcal{R}$ is said to be *contextual with respect to* C.

Remark 2.3. In Section 3.3 we will use the property of *(multi)maximality* to play the role of C, and the couplings in question then are referred to as *(multi)maximally-connected.*

2.3 Degree of contextuality

A *quasi-distribution* on a finite set V is a function $V \to \mathbb{R}$ (real numbers) such that the numbers assigned to the elements of V sum to 1. We will refer to these numbers as *quasi-probability masses.* A *quasi-rv* X is defined analogously to an rv but with a quasi-distribution instead of a distribution.

A *quasicoupling* X of $\mathcal{R}$ is defined as a quasi-rv

$$X = \left\{ X_q^c : q \in Q, c \in C, q \prec c \right\}, \tag{11}$$

such that $X^c \sim R^c$ for every $c \in C$. We have the following results.

Theorem 2.4 ([10] Theorem 6.1)**.** *For any system $\mathcal{R}$ and any set $\mathcal{T}$ of couplings for the connections of $\mathcal{R}$, there is a quasi-coupling X of $\mathcal{R}$ such that $X_q = \left\{ X_q^c : c \in C, q \prec c \right\} \sim T_q$ for any $q \in Q$.*

The *total variation* of X is denoted by $\|X\|$ and defined as the sum of the absolute values of the quasi-probability masses assigned to all values of X.

Theorem 2.5 ([10] Section 6.3)**.** *The total variation $\|X\|$ reaches its minimum in the class of all quasi-couplings X satisfying the conditions of Theorem 2.4.*

If $\min \|X\|$ is 1, then all quasi-probability masses are nonnegative, and the system $\mathcal{R}$ is noncontextual with respect to $\mathcal{T}$. If $\min \|X\| > 1$, then the system is contextual with respect to $\mathcal{T}$, and $\min \|X\| - 1$ can be taken as a (universally applicable) measure of the *degree of contextuality.*

3 Splits and Canonical Representations

3.1 Expansions of the original system

One is often interested not only in a system of empirically measured rvs $\mathcal{R}$ but also in some transformations thereof. Each such a transformation $F_{q_1,\ldots,q_k}$ is labeled by a set of conteNts, $q_1, \ldots, q_k$, and it takes as its arguments the rvs $R_{q_1}^c, \ldots, R_{q_k}^c$ in each conteXt c such that $c \succ q_1, \ldots, q_k$. The outcome,

$$R_{q^*}^c = F_{q_1,\ldots,q_k}\left(R_{q_1}^c, \ldots, R_{q_k}^c\right), \tag{12}$$

is an rv interpreted as measuring a new conteNt q^* in the conteXt c. One is free to choose any such transformations and form the corresponding new conteNts, as there can be no rules mandating what one should be interested in measuring.

Using various transformations to add new conteNts and new rvs to the original system *expands* it into a larger system. Two types of expansions that are of particular interest are *expansion-through-joining* and *expansion-through-coarsening.* Joining is defined as

$$R_{q_1}^c, \ldots, R_{q_k}^c \longmapsto \left(R_{q_1}^c, \ldots, R_{q_k}^c\right) = R_{q'}^c, \tag{13}$$

whereas coarsening is transformation

$$R_q^c \longmapsto F_q\left(R_q^c\right) = R_{q''}^c. \tag{14}$$

In fact any other transformation $F_{q_1,\ldots,q_k}\left(R_{q_1}^c, \ldots, R_{q_k}^c\right)$ can be presented as joining followed by coarsening.

Example 3.1 (Joining)**.** Consider the system

$$\begin{array}{|c|c|c|l}
\hline
R_1^1 & R_2^1 & \cdot & c=1\\ \hline
R_1^2 & R_2^2 & \cdot & c=2\\ \hline
R_1^3 & \cdot & R_3^3 & c=3\\ \hline
\cdot & R_2^4 & R_3^4 & c=4\\ \hline
\multicolumn{1}{c}{q=1} & \multicolumn{1}{c}{q=2} & \multicolumn{1}{c}{q=3} & \boxed{\mathcal{R}}
\end{array} .$$

It contains the jointly distributed R_1^1, R_2^1 and also the jointly distributed R_1^2, R_2^2, but in determining the maximal couplings of R_1^1, R_1^2 and of R_2^1, R_2^2 in the first and second columns these row-wise joints are not utilized. In some applications this would be unacceptable (e.g., in the theory of selective influences [8, 9] and in the approach advocated by Abramsky and colleagues [1, 2] this is never acceptable), and then the following expansion has to be used:

$$\begin{array}{|c|c|c|c|r}
\hline
R_1^1 & R_2^1 & \cdot & \left(R_1^1, R_2^1\right) & c=1\\ \hline
R_1^2 & R_2^2 & \cdot & \left(R_1^2, R_2^2\right) & 2\\ \hline
R_1^3 & \cdot & R_3^3 & \cdot & 3\\ \hline
\cdot & R_2^4 & R_3^4 & \cdot & 4\\ \hline
\multicolumn{1}{c}{q=1} & \multicolumn{1}{c}{2} & \multicolumn{1}{c}{3} & \multicolumn{1}{c}{12} & \boxed{\mathcal{R}^*}
\end{array} .\square$$

Example 3.2 (Coarsening)**.** If V is a set of possible values of R_q^c, then $U = F_q(V)$ is the set of possible values of the rv $R_{q^*}^c = F_q\left(R_q^c\right)$. This rv is a coarsening of R_q^c. Note that any rv is its own coarsening. As the way one labels the values of U is usually irrelevant, each such function F_q can be presented as a partition of V. Consider, e.g., the "mini"-system

$$\begin{array}{|c|l}
\hline
R_q^1 & c=1\\ \hline
R_q^2 & c=2\\ \hline
\multicolumn{1}{c}{q} & \boxed{\mathcal{R}}
\end{array} ,$$

and let the two rvs take values on $\{1,2,3,4,5\}$. If these values are considered ordered, $1 < \ldots < 5$, one may be interested in all possible partitions of $\{1,2,3,4,5\}$ into subsets of consecutive numbers, such as $\{12\,|\,34\,|\,5\}$, $\{1\,|\,2345\}$, etc. There are 15 such partitions (counting $\{1\,|\,2\,|\,3\,|\,4\,|\,5\}$ that defines the original rvs R_q^c, but excluding the trivial partition $\{12345\}$). If the values $1,2,3,4,5$ are treated as unordered labels, one might consider all possible nontrivial partitions, such as $\{\{14\},\{25\},\{3\}\}$, $\{\{145\},\{23\}\}$, etc. There are 51 such partitions. In either of these two coarsening schemes the partitions can be ordered in some way, and the respective expanded systems then become

$$\begin{array}{|c|c|c|c|l}
\hline
R_q^1 & R_{q1'}^1 & \cdots & R_{q14'}^1 & c=1\\ \hline
R_q^2 & R_{q1'}^2 & \cdots & R_{q14'}^2 & c=2\\ \hline
\multicolumn{1}{c}{q} & \multicolumn{1}{c}{q1'} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{q14'} & \boxed{\mathcal{R}'}
\end{array}
\quad \text{and} \quad
\begin{array}{|c|c|c|c|l}
\hline
R_q^1 & R_{q1''}^1 & \cdots & R_{q50''}^1 & c=1\\ \hline
R_q^2 & R_{q1''}^2 & \cdots & R_{q50''}^2 & c=2\\ \hline
\multicolumn{1}{c}{q} & \multicolumn{1}{c}{q1''} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{q50''} & \boxed{\mathcal{R}''}
\end{array}$$

Remark 3.3. Although the number of the states (combinations of the values of the elements) of the bunch R^c in $\mathcal{R}'$ and especially in $\mathcal{R}^{''}$ is very large, the support of each bunch (the set of the states with nonzero probabilities) has the same size as that of the initial random variable R_q^c in $\mathcal{R}$ (i.e., in our example, it cannot exceed 5). This follows from the facts that each event $R_q^c = x$ uniquely defines the state of R^c in $\mathcal{R}'$ and in $\mathcal{R}^{''}$, and that $\sum_x \Pr\left[R_q^c = x\right] = 1$. $\square$

3.2 Dichotomizations and canonical/split representations

Definition 3.4. A *dichotomization* of a set V is a function $f : V \to \{0,1\}$. Applying such an f to an rv R with the set of possible values V, we get a binary rv $f(R)$. We call this $f(R)$ a *split* of the original R.

If R_q^c is an element of a system $\mathcal{R}$, let us agree to identify $f\left(R_q^c\right)$ as D_{qW}^c, where $W = f^{-1}(1)$, with the understanding that D_{qW}^c and $D_{q(V-W)}^c$ are indistinguishable. To make the choice definitive, we always choose W as the smaller of W and $V - W$; in the case they have the same number of elements, we order the elements of V, say $1 < 2 < \ldots < k$, and then choose W as lexicographically preceding $V - W$.

With $V = \{1, 2, \ldots, k\}$, the jointly distributed set of splits

$$\left\{D_{q\{1\}}^c, D_{q\{2\}}^c, \ldots, D_{q\{k\}}^c\right\} \tag{15}$$

is called the *split representation* of R_q^c. If $k = 2$, then R_q^c is its own split representation, because $D_{q\{1\}}^c$ and $D_{q\{2\}}^c$ are indistinguishable.

Definition 3.5. The system $\mathcal{D}$ obtained from a system $\mathcal{R}$ by replacing each of its elements by its split representations is called the *canonical* (or *split*) *representation* of $\mathcal{R}$.

Example 3.6 (continuing Example 3.1)**.** Let all rvs in $\mathcal{R}$ be binary, 0/1, whence $\left(R_1^1, R_2^1\right)$ and $\left(R_1^2, R_2^2\right)$ in $\mathcal{R}^*$ have 4 values each: 00, 01, 10, 11. Replacing them with the split representations and observing that the first three columns do not change, we get the following canonical representation of $\mathcal{R}^*$:

$D_1^1 = R_1^1$	$D_2^1 = R_2^1$	$\cdot$	$D_{12\{00\}}^1$	$D_{12\{01\}}^1$	$D_{12\{10\}}^1$	$D_{12\{11\}}^1$	$c = 1$
$D_1^2 = R_1^2$	$D_2^2 = R_2^2$	$\cdot$	$D_{12\{00\}}^2$	$D_{12\{01\}}^2$	$D_{12\{10\}}^2$	$D_{12\{11\}}^2$	2
$D_1^3 = R_1^3$	$\cdot$	$D_3^3 = R_3^3$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	3 .□
	$D_2^4 = R_2^4$	$D_3^4 = R_3^4$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	4
$q = 1$	2	3	$12\{00\}$	$12\{01\}$	$12\{10\}$	$12\{11\}$	$\boxed{\mathcal{D}^*}$

Example 3.7 (continuing Example 3.2)**.** For the system $\mathcal{R}'$, it is clear that the split representations of the 15 coarsenings of R_q^c variously overlap: e.g., $D_{q\{3\}}^1$ belongs to the split representations of R_q^1 and of the coarsenings defined by the partitions $\{12 \mid 3 \mid 45\}$, $\{1 \mid 2 \mid 3 \mid 45\}$, and $\{12 \mid 3 \mid 4 \mid 5\}$. Following our rules, W in the splits D_{qW}^c comprising the split representation of $\mathcal{R}'$ are (when written as strings) 1, 2, 3, 4, 5, 12, 23, 34, 45, and 15 (note that, e.g., the split of the coarsening $\{1|23|4|5\}$ with $W = \{1, 23\}$ should be denoted $D_{q\{1,23\}}^1$ according to our definitions, but this is the same random variable as $D_{q\{45\}}^1$ which we have included in the list). For the system $\mathcal{R}''$ the canonical representation, obviously, consists of all possible splits of R_q^c. It will be the target of the analysis presented in Section 4. □

3.3 Multimaximality for canonical representations

If each connection in a canonical representation $\mathcal{D}$ contains just two rvs, one can compute unique maximal couplings for all of these connections. The determination of whether $\mathcal{D}^*$ is (non)contextual then can proceed in compliance with the general theory presented in Section 2.2, and amounts to determining if $\mathcal{D}^*$ has a *maximally-connected* coupling S (see Remark 2.3). If no such coupling exists, the computation of the degree of contextuality in $\mathcal{D}^*$ can be done in compliance with Section 2.3.

In a more general case, however, with an arbitrary number of rvs in each connection, maximal couplings should be replaced with computing what we call *multimaximal couplings* [11,12].

Definition 3.8. A coupling T_q of a connection $\mathcal{D}_q$ of a split representation $\mathcal{D}$ is called *multimaximal* if, for any $c, c' \in C$ such that $c, c' \succ q$, $\Pr\left[T_q^c = T_q^{c'}\right]$ is maximal over all possible couplings of $\mathcal{D}_q$. (If the connection contains two rvs, its multimaximal coupling is simply maximal.)

A multimaximal coupling is known to have the following properties.

Multimax1: The multimaximal coupling exists and is unique for any connection $\mathcal{D}_q$ ([12] Corollary 1).

Multmax2: T_q is a multimaximal coupling of $\mathcal{D}_q$ if and only if any subset of T_q is a maximal coupling for the corresponding subset of $\mathcal{D}_q$ ([12] Theorem 5; [11] Theorem 2.3).

Multimax3: In a connection $\mathcal{D}_q$, if $\{c_1, \ldots, c_n\}$ is the set of all $c \succ q$ enumerated so that

$$\Pr\left[D_q^{c_1} = 1\right] \leq \ldots \leq \Pr\left[D_q^{c_n} = 1\right],$$

then T_q is a multimaximal coupling of $\mathcal{D}_q$ if and only if $\Pr\left[T_q^{c_i} = T_q^{c_{i+1}}\right]$ is maximal for $i = 1, \ldots, n-1$, over all possible couplings of $\mathcal{D}_q$ ([11] Theorem 2.3).

4 The Largest Canonical Representation of a Two-Element Connection

We consider here the case when one is interested in all possible coarsenings of the rvs in a system. The canonical/split representation of the system then contains all splits of all rvs. We will investigate in detail a fragment of the original (expanded) system involving just two k-valued rvs within a single connection:

$$\begin{array}{|c|c}
\hline
R_1^1 & c=1 \\
\hline
R_1^2 & c=2 \\
\hline
\multicolumn{1}{c}{q=1} & \boxed{\mathcal{R}}
\end{array}$$

The canonical system with all splits of these k-valued rvs is

$$\begin{array}{r|c|c|c|c|c}
\cline{2-5}
D^1: & D^1_{W1} & D^1_{W2} & & D^1_{W(2^{k-1}-1)} & c=1 \\
\cline{2-5}
D^2: & D^2_{W1} & D^2_{W2} & \cdots & D^2_{W(2^{k-1}-1)} & c=2 \\
\cline{2-5}
\multicolumn{1}{c}{} & \multicolumn{1}{c}{q=W1} & \multicolumn{1}{c}{W2} & \multicolumn{1}{c}{\cdots} & \multicolumn{1}{c}{W\left(2^{k-1}-1\right)} & \boxed{\mathcal{D}}
\end{array}$$

where $W1$, $W2$, etc. are the subsets $f^{-1}(1)$ chosen as explained in Section 3.2 from the $2^{k-1}-1$ distinct dichotomizations f of $\{1,\ldots,k\}$. The number $2^{k-1}-1$ is arrived at by taking the number of all subsets, subtracting 2 improper subsets, and dividing by 2 because one chooses only one of W and $\{1,2,\ldots,k\}-W$. The goal is to determine whether $\mathcal{D}$ is contextual. If it is, then any canonical system that includes $\mathcal{D}$ as its subsystem (i.e., represents an original system with $\mathcal{R}$ as part of one of its connections) is contextual.

The two original rvs have distributions

$$\Pr\left[R_1^1=i\right]=p_i,\ \Pr\left[R_1^2=i\right]=q_i\ , i=1,2,\ldots,k. \tag{16}$$

A state (or value) of a bunch in the system $\mathcal{D}$ is a vector of $2^{k-1}-1$ zeroes and ones. However, the support of each of the bunches in system $\mathcal{D}$ consists of at most k corresponding states, and we can enumerate them by any k symbols, say, $1,2,\ldots,k$, as in the original variable:

$$\Pr\left[D^1=i\right]=p_i,\ \Pr\left[D^2=i\right]=q_i\ , i=1,2,\ldots,k, \tag{17}$$

As a result, $\mathcal{D}=\left\{D^1,D^2\right\}$ has k^2 possible states that we can denote ij, with $i,j\in\{1,2,\ldots,k\}$. A coupling $S=\left(S_q^1,S_q^2\right)$ of $\mathcal{D}$ assigns probabilities

$$r_{ij}=\Pr\left[S_q^1=i,S_q^2=j\right],\ i,j\in\{1,\ldots,k\}, \tag{18}$$

to these k^2 states so that they satisfy $2k$ linear constraints imposed by (16),

$$\sum_{j=1}^{k}r_{ij}=p_i,\ \sum_{i=1}^{k}r_{ij}=q_j,\ i,j\in\{1,\ldots,k\}. \tag{19}$$

If S is maximally-connected, then it should also satisfy $2^{k-1}-1$ linear constraints imposed by the maximal couplings of the corresponding connections. Specifically, if $W=\{i_1,\ldots,i_m\}\subset\{1,\ldots,k\}$, then the maximal coupling $\left(S_W^1,S_W^2\right)$ of $\left(D_W^1,D_W^2\right)$ is distributed as

$$\left.\begin{array}{r}
\Pr\left[S_W^1=1\right]=\Pr\left[D_W^1=1\right]=p_{i_1}+p_{i_2}+\ldots+p_{i_m}\\
\Pr\left[S_W^2=1\right]=\Pr\left[D_W^2=1\right]=q_{i_1}+q_{i_2}+\ldots+q_{i_m}\\
\Pr\left[S_W^1=S_W^2=1\right]=\min\left(p_{i_1}+p_{i_2}+\ldots+p_{i_m},q_{i_1}+q_{i_2}+\ldots+q_{i_m}\right)
\end{array}\right]. \tag{20}$$

Let us use the term m-split to designate any split D_W with an m-element set W ($m\leq k/2$). Thus, D_W with $W=\{i\}$ is a 1-split, with $W=\{i,j\}$ it is a 2-split, and the higher-order splits appear beginning with $k>5$. Theorem 4.3 and its corollaries below show that in determining whether the system $\mathcal{D}$ is contextual one needs to consider only the 1-splits and 2-splits. Let us use the term *1-2 system* for this subsystem of $\mathcal{D}$. An overall coupling S of $\mathcal{D}$ contains as its part a maximally-connected coupling of the 1-2 system if and only if the probabilities r_{ij} in (18) satisfy (20) for $m=1$ and $m=2$:

$$r_{ii}=\min\left(p_i,q_i\right),\ i\in\{1,\ldots,k\} \tag{21}$$

and

$$r_{ii}+r_{ij}+r_{ji}+r_{jj}=\min\left(p_i+p_j,q_i+q_j\right),\ i,j\in\{1,\ldots,k\},i<j. \tag{22}$$

That is, a maximally-connected coupling of the 1-2 system is described by the $3k+\binom{k}{2}$ linear equations (19), (21), and (22). We have therefore the following necessary condition for noncontextuality of $\mathcal{D}$.

Theorem 4.1. *If the system $\mathcal{D}$ is noncontextual, then the $3k + \binom{k}{2}$ linear equations (19), (21), and (22) are satisfied.*

Remark 4.2. Note that $3k + \binom{k}{2} < k^2$ for $k > 5$. (For completeness only, Theorem S.1 in the supplementary file S shows that the rank of this system of equations is $2k - 1 + \binom{k}{2}$.)

Theorem 4.3. *In a maximally-connected coupling S of $\mathcal{D}$ with $k > 5$, the distributions of the 1-splits and 2-splits uniquely determine the probabilities of all higher-order splits. Specifically, for any $2 < m \leq k/2$, and any $W = \{i_1, \ldots, i_m\} \subset \{1, \ldots, k\}$, the probability that the corresponding m-split equals 1 is*

$$\begin{array}{l}\min\left(p_{i_1} + p_{i_2} + \ldots + p_{i_m}, q_{i_1} + q_{i_2} + \ldots + q_{i_m}\right) = \sum_{j=1}^{m} \min\left(p_{i_j}, q_{i_j}\right)\\ + \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m} \left[\min\left(p_{i_j} + p_{i_{j'}}, q_{i_j} + q_{i_{j'}}\right) - \min\left(p_{i_j}, q_{i_j}\right) - \min\left(p_{i_{j'}}, q_{i_{j'}}\right)\right].\end{array} \tag{23}$$

It is easy to find numerical examples of the distributions of R_1^1 and R_1^2 for which (23) is violated (see Example S.2 in the supplementary file S). As shown below, however, (23) cannot be violated if a maximally-connected coupling for the 1-2 system exists. It follows from the fact that the statement of Theorem 4.1 can be reversed: (19), (21), and (22) imply that $\mathcal{D}$ is noncontextual. We establish this fact by first characterizing the distributions of R_1^1 and R_1^2 for a noncontextual 1-2 system (Theorem 4.4 with Corollary 4.5), and then showing that (23) always holds for such distributions (Theorem 4.6).

Theorem 4.4. *A maximally-connected coupling for a 1-2 system is unique if it exists. In this coupling, the only pairs of ij in (18) that may have nonzero probabilities assigned to them are the diagonal states $\{11, 22, \ldots, kk\}$ and either the states $\{i1, i2, \ldots, ik\}$ for a single fixed i or the states $\{1j, 2j, \ldots, kj\}$ for a single fixed j $(i, j = 1, \ldots, k)$.*

Assuming, with no loss of generality, that the single fixed i or the single fixed j in the formulation above is 2, the theorem says that the nonzero probabilities assigned to the states of the maximally-connected coupling (shown below for $k = 4$) could only occupy the cells marked with asterisks:

$$\begin{array}{c|c|c|c|c|} & 1 & 2 & 3 & 4 \\ \hline 1 & * & * & 0 & 0 \\ \hline 2 & 0 & * & 0 & 0 \\ \hline 3 & 0 & * & * & 0 \\ \hline 4 & 0 & * & 0 & * \\ \hline \end{array} \quad \text{or} \quad \begin{array}{c|c|c|c|c|} & 1 & 2 & 3 & 4 \\ \hline 1 & * & 0 & 0 & 0 \\ \hline 2 & * & * & * & * \\ \hline 3 & 0 & 0 & * & 0 \\ \hline 4 & 0 & 0 & 0 & * \\ \hline \end{array}.$$

Corollary 4.5. *The 1-2 system for the original rvs R_1^1, R_1^2 has a maximally-connected coupling if and only if either $p_i > q_i$ for no more than one i (this single possible i being the single fixed i in the formulation of the theorem), or $p_j < q_j$ for no more than one j (this single possible j being the single fixed j in the formulation of the theorem), $i, j \in \{1, \ldots, k\}$.*

The relationship between $(p_1, \ldots, p_k)$ and $(q_1, \ldots, q_k)$ described in this corollary is some form of stochastic dominance for categorical rvs, but it does not seem to have been previously identified. We propose to say that R_1^1 *nominally dominates* R_1^2 if $p_i < q_i$ for no more than one value of $i = 1, \ldots, k$ (i.e., $p_i \geq q_i$ for at least $k - 1$ of them). Two categorical rvs nominally dominate each other if and only if either they are identically distributed or $k = 2$. Using this notion, and combining Corollary 4.5 with Theorems 4.1 and 4.4, we get the main result of this section.

Theorem 4.6. *The system $\mathcal{D}$ is noncontextual if and only if its 1-2 subsystem is noncontextual, i.e., if and only if one of the* R_1^1 and R_1^2 *nominally dominates the other.*

5 Concluding remarks

Contextuality analysis of an empirical situation involves the following sequence of steps:

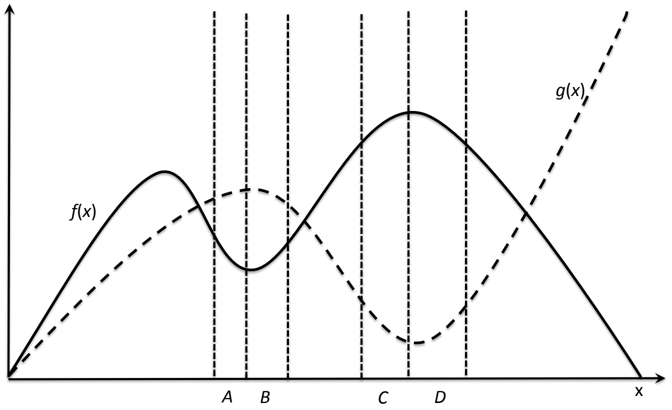
$$\begin{array}{c}\text{empirical}\\ \text{measurements}\end{array} \longrightarrow \begin{array}{c}\text{initial}\\ \text{system of rvs}\end{array} \longrightarrow \begin{array}{c}\text{expanded}\\ \text{system of rvs}\end{array} \longrightarrow \begin{array}{c}\text{canonical/split}\\ \text{representation}\end{array}$$

In the initial system, measurements are represented by rvs each of which generally has multiple values. Expansion means adding to the system new conteNts with corresponding connections (conteNt-sharing rvs) computed as functions of the existing connections. In a canonical representation of the system all rvs are binary, and the connections are coupled multimaximally, meaning essentially that one deals with their elements pairwise. The issue of contextuality is reduced to that of compatibility of the unique couplings for pairs of conteNt-sharing rvs with the known distributions of the conteXt-sharing bunches of rvs. Coupling the connections multimaximally ensures that a noncontextual system has all its subsystems noncontextual too.

The canonical system of rvs is uniquely determined by the expanded system, but the latter is inherently non-unique, it depends on what aspects of the empirical situation one wishes to include in the system. Thus, it is one's choice rather than a general rule whether one considers a multi-valued measurement as representable by *all* or only *some* of its possible coarsenings. If one chooses all coarsenings, the split/canonical representation involves all dichotomizations, and then Theorem 4.6 says that the canonical system is noncontextual only if, for any pair of rvs $R_q^c, R_q^{c'}$ in the expanded system, one of them, say R_q^c, "nominally dominates" the other. This domination means that $\Pr\left[R_q^c = x\right] < \Pr\left[R_q^{c'} = x\right]$ holds for no more than one value x of these rvs: a stringent necessary condition for noncontextuality, likely to be violated in many empirical systems.

This is of special interest for contextuality studies outside quantum physics. Historically, the search for non-quantum contextual systems was motivated by the possibility of applying quantum-theoretic formalisms in such fields as biology [3], psychology [5, 21, 29], economics [18, 21], and political science [22]. In CbD, the notion of contextuality is not tied to quantum formalisms in any special way. The possibility of non-quantum contextual systems here is motivated by treating contextuality as an abstract probabilistic issue: there are no a priori reasons why a system of rvs describing, say, human behavior could not be contextual if it is qualitatively (i.e., up to specific probability values) the same as a contextual one describing particle spins. Nevertheless, all known to us systems with dichotomous responses investigated for potential contextuality (with the exception of one, very recent experiment) have been found to be noncontextual [7, 14, 16]. The use of canonical representations with dichotomizations of multiple-choice responses offers new possibilities.

In some cases, however, the use of all possible dichotomizations is not justifiable. Notably, if the values of an rv are linearly ordered, $x_1 < x_2 < \ldots, x_N$, it may be natural to only allow dichotomizations f with $f^{-1}(1)$ containing several successive values, $\{x_l, x_{l+1}, \ldots, x_L\}$, for some $l, L \in \{1, \ldots, N\}$. An even stronger restriction would be to only allow "cuts," with $f^{-1}(1) = \{x_l, x_{l+1}, \ldots, x_N\}$ or $\{x_1, x_2, \ldots, x_{l-1}\}$.



Stronger restrictions on possible dichotomizations translate into stronger restrictions on the pairs $R_q^c, R_{q'}^c$ whose canonical representation is contextual. This fact is especially important if one considers expanding CbD beyond categorical rvs. Thus, it is easy to see that if one considers all possible dichotomizations of two conteNt-sharing rvs with continuous densities on the set of real numbers, then the system will be contextual whenever the two distributions are not identical. Let the densities of these rvs be $f(x)$ and $g(x)$ shown in the graphic above. If the set of all splits of these rvs forms a noncontextual system, then any discretization of these rvs should satisfy Corollary 4.5 to Theorem 4.4. That is, for any $k > 2$ and any partition $H_1, \ldots, H_k$ of the set of reals into intervals, we should have either

$$\begin{array}{c} \int_{H_i} f(x)\,dx < \int_{H_i} g(x)\,dx \text{ for no more than one of } i = 1, \ldots, k, \\ \text{or} \\ \int_{H_i} f(x)\,dx > \int_{H_i} g(x)\,dx \text{ for no more than one of } i = 1, \ldots, k. \end{array} \quad (24)$$

This is, however, impossible unless $f(x) = g(x)$. If they are different, then f exceeds g on some interval, and g exceeds f on some other interval. If we take any two subintervals within each of these intervals (in the graphic they are denoted by A, B and C, D), any partition $H_1, \ldots, H_k$ that includes A, B, C, D will violate (24). The development of the theory of canonical representations with variously restricted sets of splits is a task for future work.

Data accessibility. See Remark 1.2.

Competing interests. We have no financial or non-financial competing interests.

Authors' contributions. All authors significantly contributed to the development of the theory and drafting of the paper.

Acknowledgments. We have greatly benefited from discussions with Matt Jones, Samson Abramsky, Rui Soares Barbosa, and Pawel Kurzynski.

Funding statement. This research has been supported by AFOSR grant FA9550-14-1-0318.

References

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

[2] Abramsky S, Barbosa RS, Kishida K, Lal R, Mansfield, S 2015 Contextuality, cohomology and paradox. Comp. Sci. Log. 2015: 211-228.

[3] Asano M, Basieva I, Khrennikov A, Ohya M, Tanaka Y, Yamato I 2015 Quantum information biology: from information interpretation of quantum mechanics to applications in molecular biology and cognitive psychology. Found. Phys, 45, 1362-1378.

[4] Bacciagaluppi G 2015 Leggett-Garg inequalities, pilot waves and contextuality. Int. J. Quant. Found. 1, 1-17.

[5] Busemeyer JR, Bruza PD. 2012 Quantum Cognition and Decision. Cambridge, UK: Cambridge University Press.

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

[7] Cervantes VH, Dzhafarov EN Advanced analysis of quantum contextuality in a psychophysical double-detection experiment. To be published in Journal of Mathematical Psychology.

[8] Dzhafarov EN, Kujala JV. 2012 Selectivity in probabilistic causality: Where psychology runs into quantum physics. J. Math. Psychol. 56, 54-63.

[9] Dzhafarov EN, Kujala JV 2016 Probability, random variables, and selectivity. In W. Batchelder, H. Colonius, E.N. Dzhafarov, J. Myung (Eds) pp. 85-150. The New Handbook of Mathematical Psychology. Cambridge University Press.

[10] Dzhafarov EN, Kujala JV 2016 Context-content systems of random variables: The contextuality-by-default theory. J. Math. Psych. 74, 11-33.

[11] Dzhafarov EN, Kujala JV 2017 Probabilistic foundations of contextuality. Fortschr. Phys. - Prog. Phys 65, 1600040.

[12] Dzhafarov EN, Kujala JV 2017 Contextuality-by-Default 2.0: Systems with binary random variables. In J.A. de Barros, B. Coecke, E. Pothos (Eds.) Lect. Not. Comp. Sci. 10106, 16-32.

[13] Dzhafarov EN, Kujala, JV, Cervantes VH 2016 Contextuality-by-Default: A brief overview of ideas, concepts, and terminology. In H. Atmanspacher, T. Filk, E. Pothos (Eds.) Lect. Not. Comp. Sci. 9535, 12-23.

[14] Dzhafarov EN, Kujala JV, Cervantes VH, Zhang R, Jones M 2016 On contextuality in behavioral data. Phil. Trans. Roy. Soc. A 374: 20150234.

[15] Dzhafarov EN, Kujala JV, Larsson J-Å 2015 Contextuality in three types of quantum-mechanical systems. Found. Phys. 7, 762--782.

[16] Dzhafarov EN, Zhang R, Kujala JV 2016 Is there contextuality in behavioral and social systems? Phil. Trans. Roy. Soc. A 374, 20150099.

[17] Fine A. 1982 Joint distributions, quantum correlations, and commuting observables. J. Math. Phys. 23, 1306–1310. (doi:10.1063/1.525514)

[18] Haven E, Khrennikov A. 2012 Quantum Social Science. Cambridge, UK: Cambridge University Press.

[19] Khrennikov A. 2005 The principle of supplementarity: A contextual probabilistic viewpoint to complementarity, the interference of probabilities, and the incompatibility of variables in quantum mechanics. Found. Phys., 35, 1655 - 1693.

[20] Khrennikov A 2009 Contextual approach to quantum formalism. Springer: Berlin.

[21] Khrennikov A 2010 Ubiquitous Quantum Structure: From Psychology to Finance. Springer: Berlin.

[22] Khrennikova P, Haven E 2016 Instability of political preferences and the role of mass media: a representation in quantum framework . Phil. Trans. Roy. Soc. A 374, 20150106.

[23] Kujala JV, Dzhafarov EN 2015 Probabilistic Contextuality in EPR/Bohm-type systems with signaling allowed. In E.N. Dzhafarov, S. Jordan, R. Zhang, V.H. Cervantes (Eds.) Contextuality from Quantum Physics to Psychology, pp. 287-308. New Jersey: World Scientific.

[24] Kujala JV, Dzhafarov EN 2016 Proof of a conjecture on contextuality in cyclic systems with binary variables. Found. Phys. 46, 282-299.

[25] Kujala JV, Dzhafarov EN, Larsson J-Å 2015 Necessary and sufficient conditions for maximal noncontextuality in a broad class of quantum mechanical systems. Phys. Rev. Lett. 115, 150401.

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

[27] Moore DW. 2002 Measuring new types of question-order effects. Public Opin. Quart. 66, 80-91.

[28] Suppes P, Zanotti M. 1981 When are probabilistic explanations possible? Synthese 48, 191.

[29] Wang Z, Busemeyer JR. 2013 A quantum question order model supported by empirical tests of an a priori and precise prediction.Top. Cogn. Sci. 5, 689--710.

S Supplementary Text to “Contextuality in Canonical Systems of Random Variables” by Ehtibar N. Dzhafarov, Víctor H. Cervantes, and Janne V. Kujala (Phil. Trans. Roy. Soc. A xxx, 10.1098/rsta.2016.0389)

Theorem S.1 (Section 4, Remark 4.2)**.** *The rank of the system of linear equations (19)-(21)-(22) is* $2k-1+\binom{k}{2}$.

Proof of Theorem S.1. This system of linear equations can be written as

$$\mathbf{M}\times\mathbf{X}=\mathbf{P},$$

where

$$\mathbf{P}^{T}=\left(\begin{array}{c}\overbrace{p_{1},\ldots,p_{k}}^{k},\overbrace{q_{1},\ldots,q_{k}}^{k},\overbrace{\min\left(p_{1},q_{1}\right),\ldots,\min\left(p_{k},q_{k}\right)}^{k},\\ \\ \overbrace{\min\left(p_{1}+p_{2},q_{1}+q_{2}\right),\ldots,\min\left(p_{k-1}+p_{k},q_{k-1}+q_{k}\right)}^{\binom{k}{2}}\end{array}\right),$$

$$\mathbf{X}^{T}=\left\{x_{ij}:i,j\in\{1,\ldots,k\}\right\},$$

and $\mathbf{M}$ is a Boolean matrix. The $\left(k+k+k+\binom{k}{2}\right)$ rows of matrix $\mathbf{M}$ correspond to the elements of $\mathbf{P}$ and can be labeled as

$$\left(\overbrace{\mathbf{r}_{1\cdot},\ldots,\mathbf{r}_{k\cdot}}^{k},\overbrace{\mathbf{r}_{\cdot 1},\ldots,\mathbf{r}_{\cdot k}}^{k},\overbrace{\mathbf{r}_{11},\ldots,\mathbf{r}_{kk}}^{k},\overbrace{\mathbf{r}_{12},\ldots,\mathbf{r}_{k-1,k}}^{\binom{k}{2}}\right),$$

whereas the k^{2} columns of $\mathbf{M}$ correspond to the elements of $\mathbf{X}$ and can be labeled as

$$\left\{\mathbf{c}_{ij}:i,j\in\{1,\ldots,k\}\right\}.$$

Thus, if $k=4$, the matrix $\mathbf{M}$ is

$$\begin{array}{ccccccccccccccccc}
{}^{\mathbf{c}}_{\mathbf{r}} & 11 & 12 & 13 & 14 & 21 & 22 & 23 & 24 & 31 & 32 & 33 & 34 & 41 & 42 & 43 & 44\\
1\cdot & 1 & 1 & 1 & 1 & & & & & & & & & & & & \\
2\cdot & & & & & 1 & 1 & 1 & 1 & & & & & & & & \\
3\cdot & & & & & & & & & 1 & 1 & 1 & 1 & & & & \\
4\cdot & & & & & & & & & & & & & 1 & 1 & 1 & 1\\
\cdot 1 & 1 & & & & 1 & & & & 1 & & & & 1 & & & \\
\cdot 2 & & 1 & & & & 1 & & & & 1 & & & & 1 & & \\
\cdot 3 & & & 1 & & & & 1 & & & & 1 & & & & 1 & \\
\cdot 4 & & & & 1 & & & & 1 & & & & 1 & & & & 1\\
11 & 1 & & & & & & & & & & & & & & & \\
22 & & & & & & 1 & & & & & & & & & & \\
33 & & & & & & & & & & & 1 & & & & & \\
44 & & & & & & & & & & & & & & & & 1\\
12 & 1 & 1 & & & 1 & 1 & & & & & & & & & & \\
13 & 1 & & 1 & & & & & & 1 & & 1 & & & & & \\
14 & 1 & & & 1 & & & & & & & & & 1 & & & 1\\
23 & & & & & & 1 & 1 & & & 1 & 1 & & & & & \\
24 & & & & & & 1 & & 1 & & & & & & 1 & & 1\\
34 & & & & & & & & & & & 1 & 1 & & & 1 & 1
\end{array}$$

We will continue to illustrate the steps of the proof using this matrix. We begin by adding to $\mathbf{M}$ the row $\mathbf{r}_{all}$ with all cells equal to 1, and denote the new matrix $\mathbf{M}'$.

$$\begin{array}{c|cccccccccccccccc}
{}_{\mathbf{r}}{}^{\mathbf{c}} & 11 & 12 & 13 & 14 & 21 & 22 & 23 & 24 & 31 & 32 & 33 & 34 & 41 & 42 & 43 & 44 \\
1\cdot & 1 & 1 & 1 & 1 & & & & & & & & & & & & \\
2\cdot & & & & & 1 & 1 & 1 & 1 & & & & & & & & \\
3\cdot & & & & & & & & & 1 & 1 & 1 & 1 & & & & \\
4\cdot & & & & & & & & & & & & & 1 & 1 & 1 & 1 \\
\cdot 1 & 1 & & & & 1 & & & & 1 & & & & 1 & & & \\
\cdot 2 & & 1 & & & & 1 & & & & 1 & & & & 1 & & \\
\cdot 3 & & & 1 & & & & 1 & & & & 1 & & & & 1 & \\
\cdot 4 & & & & 1 & & & & 1 & & & & 1 & & & & 1 \\
11 & 1 & & & & & & & & & & & & & & & \\
22 & & & & & & 1 & & & & & & & & & & \\
33 & & & & & & & & & & & 1 & & & & & \\
44 & & & & & & & & & & & & & & & & 1 \\
12 & 1 & 1 & & & 1 & 1 & & & & & & & & & & \\
13 & 1 & & 1 & & & & & & 1 & & 1 & & & & & \\
14 & 1 & & & 1 & & & & & & & & & 1 & & & 1 \\
23 & & & & & & 1 & 1 & & & 1 & 1 & & & & & \\
24 & & & & & & 1 & & 1 & & & & & & 1 & & 1 \\
34 & & & & & & & & & & & 1 & 1 & & & 1 & 1 \\
\mathit{all} & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}$$

This does not change the rank of the matrix since $\mathbf{r}_{all}$ is the sum of all $\mathbf{r}_{\cdot i}$. Then we observe that the rows $\mathbf{r}_{k\cdot}$, $\mathbf{r}_{\cdot k}$, and all $\mathbf{r}_{ik}$ with $i < k$ can be deleted as they are linear combinations of the remaining rows of $\mathbf{M}'$. Indeed, it can be checked directly that

$$\mathbf{r}_{k\cdot} = \mathbf{r}_{all} - \sum_{i=1}^{k-1} \mathbf{r}_{i\cdot},$$

$$\mathbf{r}_{\cdot k} = \mathbf{r}_{all} - \sum_{i=1}^{k-1} \mathbf{r}_{\cdot i},$$

$$(\mathbf{r}_{ik} - \mathbf{r}_{ii} - \mathbf{r}_{kk}) = (\mathbf{r}_{i\cdot} - \mathbf{r}_{ii}) + (\mathbf{r}_{\cdot i} - \mathbf{r}_{ii}) - \sum_{l<i} (\mathbf{r}_{li} - \mathbf{r}_{ll} - \mathbf{r}_{ii}) - \sum_{l>i}^{l<k} (\mathbf{r}_{il} - \mathbf{r}_{ii} - \mathbf{r}_{ll}),$$

for all $i < k$. Moreover, one can also delete $\mathbf{r}_{kk}$, because

$$\sum_{i<j<k} (\mathbf{r}_{ij} - \mathbf{r}_{ii} - \mathbf{r}_{jj}) + \sum_{i<k} (\mathbf{r}_{ik} - \mathbf{r}_{ii} - \mathbf{r}_{kk}) + \sum_{i<k} \mathbf{r}_{ii} + \mathbf{r}_{kk} = \mathbf{r}_{all}.$$

Let the resulting matrix be $\mathbf{M}''$:

$$\begin{array}{c|cccccccccccccccc}
{}_{\mathbf{r}}{}^{\mathbf{c}} & 11 & 12 & 13 & 14 & 21 & 22 & 23 & 24 & 31 & 32 & 33 & 34 & 41 & 42 & 43 & 44 \\
1\cdot & 1 & 1 & 1 & 1 & & & & & & & & & & & & \\
2\cdot & & & & & 1 & 1 & 1 & 1 & & & & & & & & \\
3\cdot & & & & & & & & & 1 & 1 & 1 & 1 & & & & \\
\cdot 1 & 1 & & & & 1 & & & & 1 & & & & 1 & & & \\
\cdot 2 & & 1 & & & & 1 & & & & 1 & & & & 1 & & \\
\cdot 3 & & & 1 & & & & 1 & & & & 1 & & & & 1 & \\
11 & 1 & & & & & & & & & & & & & & & \\
22 & & & & & & 1 & & & & & & & & & & \\
33 & & & & & & & & & & & 1 & & & & & \\
12 & 1 & 1 & & & 1 & 1 & & & & & & & & & & \\
13 & 1 & & 1 & & & & & & 1 & & 1 & & & & & \\
23 & & & & & & 1 & 1 & & & 1 & 1 & & & & & \\
\mathit{all} & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}$$

This matrix contains

$$\overset{initial}{3k+\binom{k}{2}}-\underbrace{3}_{\mathbf{r}_{k\cdot},\mathbf{r}_{\cdot k},\mathbf{r}_{kk}}-\overbrace{(k-1)}^{all\ \mathbf{r}_{ik},i<k}+\underbrace{1}_{\mathbf{r}_{all}}=2k-1+\binom{k}{2}$$

rows. We prove that this matrix is of full row rank. Consider equation

$$\sum_{all\ \mathbf{r}\text{ in }\mathbf{M}''}\alpha_{\mathbf{r}}\mathbf{r}=0.$$

We use the following principle: if a row $\mathbf{r}$ intersects a columns whose only nonzero entry is in the row $\mathbf{r}$, then $\alpha_{\mathbf{r}}=0$, and we can delete the row $\mathbf{r}$ from the matrix, decreasing the row rank of the matrix by 1. The following statements can be directly verified.

$\mathbf{r}_{all}$ can be deleted because column $\mathbf{c}_{kk}$ has its only 1 in $\mathbf{r}_{all}$.

$\mathbf{c}$ $\mathbf{r}$	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
1·	1	1	1	1												
2·					1	1	1	1								
3·									1	1	1	1				
·1	1				1				1				1			
·2		1				1				1				1		
·3			1				1				1				1	
11	1															
22						1										
33											1					
12	1	1			1	1										
13	1		1						1		1					
23						1	1			1	1					

Then each of $\mathbf{r}_{\cdot i}$can be deleted because the column $\mathbf{c}_{ki}$ has its only 1 in $\mathbf{r}_{\cdot i}$ $(i=1,\ldots,k-1)$.

$\mathbf{c}$ $\mathbf{r}$	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
1·	1	1	1	1												
2·					1	1	1	1								
3·									1	1	1	1				
11	1															
22						1										
33											1					
12	1	1			1	1										
13	1		1						1		1					
23						1	1			1	1					

Then each of $\mathbf{r}_{i\cdot}$ can be deleted because the column $\mathbf{c}_{ik}$ has its only 1 in $\mathbf{r}_{i\cdot}$ $(i=1,\ldots,k-1)$.

$\mathbf{c}$ $\mathbf{r}$	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
11	1															
22						1										
33											1					
12	1	1			1	1										
13	1		1						1		1					
23						1	1			1	1					

Then each of $\mathbf{r}_{ij}$ can be deleted because the column $\mathbf{c}_{ji}$ has its only 1 in $\mathbf{r}_{ij}$ $(i,j\in\{1,\ldots,k-1\},i<j)$.

$\mathbf{c}$ $\mathbf{r}$	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
11	1															
22						1										
33											1					

This leaves only $\mathbf{r}_{11}, \dots, \mathbf{r}_{(k-1)(k-1)}$ that are obviously linearly independent. □

Theorem (Section 4, Theorem 4.3)**.** *In a maximally-connected coupling S of $\mathcal{D}$ with $k > 5$, the distributions of the 1-splits and 2-splits uniquely determine the probabilities of all higher-order splits. Specifically, for any $2 < m \leq k/2$, and any $W = \{i_1, \dots, i_m\} \subset \{1, \dots, k\}$, the probability that the corresponding m-split equals 1 is*

$$\begin{array}{l} \min\left(p_{i_1} + p_{i_2} + \ldots + p_{i_m}, q_{i_1} + q_{i_2} + \ldots + q_{i_m}\right) = \sum_{j=1}^{m} \min\left(p_{i_j}, q_{i_j}\right) \\ + \sum_{j=1}^{m-1} \sum_{j'=j+1}^{m} \left[\min\left(p_{i_j} + p_{i_{j'}}, q_{i_j} + q_{i_{j'}}\right) - \min\left(p_{i_j}, q_{i_j}\right) - \min\left(p_{i_{j'}}, q_{i_{j'}}\right)\right]. \end{array} \tag{S.1}$$

Proof of Theorem 4.3. From (21) and (22),

$$\begin{array}{rcl} r_{12} + r_{21} & = & \min\left(p_1 + p_2, q_1 + q_2\right) - \min\left(p_1, q_1\right) - \min\left(p_2, q_2\right) \\ \vdots & \vdots & \vdots \\ r_{ij} + r_{ji} & = & \min\left(p_i + p_j, q_i + q_j\right) - \min\left(p_i, q_i\right) - \min\left(p_j, q_j\right) \qquad (i < j). \\ \vdots & \vdots & \vdots \\ r_{(k-1)k} + r_{k(k-1)} & = & \min\left(p_{k-1} + p_k, q_{k-1} + q_k\right) - \min\left(p_{k-1}, q_{k-1}\right) - \min\left(p_k, q_k\right) \end{array}$$

Consider an m-split with $2 < m \leq k/2$, and assume without loss of generality that $W = (1, \dots, m)$. We have

$$\sum_{i=1}^{m} \sum_{j=1}^{m} r_{ij} = \min\left(p_1 + \ldots + p_m, q_1 + \ldots + q_m\right). \tag{S.2}$$

The left-hand-side sum can be presented as

$$\sum_{i=1}^{m} r_{ii} + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \left(r_{ij} + r_{ji}\right)$$

$$= \sum_{i=1}^{m} \min\left(p_i, q_i\right) + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \left[\min\left(p_i + p_j, q_i + q_j\right) - \min\left(p_i, q_i\right) - \min\left(p_j, q_j\right)\right],$$

whence we get (23). □

Example S.2 (showing that the relation (23) may be violated, see Section 4.)**.** If

$$\begin{array}{r} R_1^1 = \\ \text{prob. mass } p = \end{array} \begin{array}{|c|c|c|c|c|c|} \hline 1 & 2 & 3 & 4 & 0 & 0 \\ \hline .6 & .1 & .1 & .2 & 0 & 0 \\ \hline \end{array}, \qquad \begin{array}{r} R_1^2 = \\ \text{prob. mass } q = \end{array} \begin{array}{|c|c|c|c|c|c|} \hline 1 & 2 & 3 & 4 & 0 & 0 \\ \hline .2 & .3 & .4 & .1 & 0 & 0 \\ \hline \end{array},$$

then

$$\overbrace{\min\left(p_1 + p_2 + p_3, q_1 + q_2 + q_3\right)}^{.8}$$

$$\neq \left.\begin{array}{rl} \min\left(p_1, q_1\right) & .2 \\ +\min\left(p_2, q_2\right) & .1 \\ +\min\left(p_3, q_3\right) & .1 \\ +\min\left(p_1 + p_2, q_1 + q_2\right) - \min\left(p_1, q_1\right) - \min\left(p_2, q_2\right) & .5 - .2 - .1 \\ +\min\left(p_1 + p_3, q_1 + q_3\right) - \min\left(p_1, q_1\right) - \min\left(p_3, q_3\right) & .6 - .2 - .1 \\ +\min\left(p_2 + p_3, q_2 + q_3\right) - \min\left(p_2, q_2\right) - \min\left(p_3, q_3\right) & .2 - .1 - .1 \end{array}\right\} = .5 \quad \square$$

Theorem (Section 4, Theorem 4.4)**.** *A maximally-connected coupling for a 1-2 system is unique if it exists. In this coupling, the only pairs of ij in (18) that may have nonzero probabilities assigned to them are the diagonal states $\{11, 22, \dots, kk\}$ and either the states $\{i1, i2, \dots, ik\}$ for a single fixed i or the states $\{1j, 2j, \dots, kj\}$ for a single fixed j $(i, j = 1, \dots, k)$.*

Proof of Theorem 4.4. (The matrices illustrating the proof are shown for $k > 6$ but the theorem is valid for all $k > 1$.) If the only nonzero entries in the matrix are in the main diagonal, the theorem is trivially true. Assume therefore that $r_{ij} > 0$ for some $i \neq j$. Without loss of generality, we can assume that $r_{12} > 0$ and $p_1 + p_2 \leq q_1 + q_2$. Indeed, if some $r_{ij} > 0$, we can always rename the values so that $i = 1$ and $j = 2$; and if $p_1 + p_2 > q_1 + q_2$, then we can simply rename all ps into qs and vice versa. In the following we will use the expression "r_{ij} is p-minimized" if $p_i + p_j \leq q_i + q_j$, and "r_{ij} is q-minimized" if $p_i + p_j \geq q_i + q_j$ (in both cases, $i \neq j$).

We have (the empty cells are those whose value is to be determined later)

$$\begin{array}{c|c|c|c|c|c|c|c||c}
 & 1 & 2 & 3 & 4 & 5 & 6 & \dots & \\ \hline
1 & r_{11} & r_{12} > 0 & & & & & & p_1 \\ \hline
2 & r_{21} & r_{22} & & & & & & p_2 \\ \hline
3 & & & r_{33} & & & & & \\ \hline
4 & & & & r_{44} & & & & \\ \hline
5 & & & & & r_{55} & & & \\ \hline
6 & & & & & & r_{66} & & \\ \hline
\vdots & & & & & & & \ddots & \vdots \\ \hline
 & q_1 & q_2 & & & & & \dots &
\end{array}.$$

From (21)-(22), $r_{11} + r_{12} + r_{21} + r_{22} = \min\{p_1 + p_2 q_1 + q_2\}$, and since r_{12} is p-minimized, $r_{11} + r_{12} + r_{21} + r_{22} = p_1 + p_2$. This means

$$\begin{array}{c|c|c|c|c|c|c|c||c}
 & 1 & 2 & 3 & 4 & 5 & 6 & \dots & \\ \hline
1 & r_{11} & r_{12} > 0 & 0 & 0 & 0 & 0 & \mathbf{0} & p_1 = r_{11} + r_{12} \\ \hline
2 & r_{21} & r_{22} & 0 & 0 & 0 & 0 & \mathbf{0} & p_2 = r_{21} + r_{22} \\ \hline
3 & & & r_{33} & & & & & \\ \hline
4 & & & & r_{44} & & & & \\ \hline
5 & & & & & r_{55} & & & \\ \hline
6 & & & & & & r_{66} & & \\ \hline
\vdots & & & & & & & \ddots & \vdots \\ \hline
 & q_1 \geq r_{11} + r_{21} & q_2 \geq r_{12} + r_{22} & & & & & \dots &
\end{array}.$$

We also should have

$$\begin{array}{c|c|c|c|c|c|c|c||c}
 & 1 & 2 & 3 & 4 & 5 & 6 & \dots & \\ \hline
1 & r_{11} & r_{12} > 0 & 0 & 0 & 0 & 0 & \mathbf{0} & p_1 = r_{11} + r_{12} \\ \hline
2 & 0 & r_{22} & 0 & 0 & 0 & 0 & \mathbf{0} & p_2 = r_{22} \\ \hline
3 & 0 & & r_{33} & & & & & \\ \hline
4 & 0 & & & r_{44} & & & & \\ \hline
5 & 0 & & & & r_{55} & & & \\ \hline
6 & 0 & & & & & r_{66} & & \\ \hline
\vdots & \mathbf{0} & & & & & & \ddots & \vdots \\ \hline
 & q_1 = r_{11} & q_2 \geq r_{12} + r_{22} & & & & & \dots &
\end{array}$$

because $r_{11} = \min\{p_1, q_1\}$ and $r_{11} < p_1$.

Generalizing, we have established the following rules:

(R1) If $r_{ij} > 0$ and it is p-minimized, then all non-diagonal elements in the rows i and j are zero except for r_{ij}, and all non-diagonal elements in the column i are zero.

(R2) (By symmetry, on exchanging ps and qs) If $r_{ij} > 0$ and it is q-minimized, then all non-diagonal elements in the columns i and j are zero except for r_{ij}, and all non-diagonal elements in the row j are zero.

Returning to our special arrangement of the rows and columns, let us prove now that all r_{1j} with $j > 2$ are q-minimized. Assume the contrary, and with no loss of generality, let $r_{15} = 0$ be p-minimized. This would mean that

$$r_{15} + r_{51} = p_1 + p_5 - r_{11} - r_{55} = r_{12} + p_5 - r_{55} = 0,$$

which could only be true if $r_{12} = 0$, which it is not.

	1	2	3	4	5	6	$\dots$	
1	r_{11}	$r_{12} > 0$	0 $q{-}min$	0 $q{-}min$	0 $q{-}min$	0 $q{-}min$	**0** $q{-}min$	$p_1 = r_{11} + r_{12}$
2	0	r_{22}	0	0	0	0	**0**	$p_2 = r_{22}$
3	0		r_{33}					
4	0			r_{44}				
5	0				r_{55}			p_5
6	0					r_{66}		
$\vdots$	**0**						$\ddots$	$\vdots$
	$q_1 = r_{11}$	$q_2 \geq r_{12} + r_{22}$					$\dots$	

Generalizing, we have established two additional rules:

(R3) If r_{ij} and $r_{ij'}$ are both p-minimized (for pairwise distinct i, j, j'), then they are both zero (because if one of them is not, say $r_{ij} > 0$, then $r_{ij'} = 0$ and it must be q-minimized).

(R4) (By symmetry, on exchanging ps and qs) If r_{ij} and $r_{i'j}$ are both q-minimized (for pairwise distinct i, i', j), then they are both zero.

Returning to our special arrangement of the rows and columns, it follows that nowhere in the matrix can we have $r_{ij} > 0$ ($i > 2$) which is q-minimized. Indeed, if $j > 2$, then this would have contradicted R4 (because the zeros in the first row are all q-minimized), and if $j = 2$, it would have contradicted R2 (because $r_{12} > 0$).

Let us prove now that if $j > 2$ and $i > 2$ and $i \neq j$, then there is no $r_{ij} > 0$ that is p-minimized. Assume the contrary: $r_{ij} > 0$ and q-minimized, and consider r_{2i}, r_{i2}. With no loss of generality, let (i, j)=$(4, 6)$. In accordance with R1, we fill in the 4th and the 6th rows with zeros, and we fill in the 4th column with zeros too:

	1	2	3	4	5	6	$\dots$	
1	r_{11}	$r_{12} > 0$	0	0	0	0	**0**	$p_1 = r_{11} + r_{12}$
2	0	r_{22}	0	0	0	0	**0**	$p_2 = r_{22}$
3	0		r_{33}	0				
4	0	0	0	r_{44}	0	$r_{46} > 0$	**0**	$p_4 = r_{44} + r_{46}$
5	0			0	r_{55}			
6	0	0	0	$r_{64} = 0$	0	r_{66}	**0**	$p_6 = r_{66}$
$\vdots$	**0**			**0**			$\ddots$	$\vdots$
	$q_1 = r_{11}$	$q_2 \geq r_{12} + r_{22}$		$q_4 = r_{44}$		$q_6 \geq r_{46} + r_{66}$	$\dots$	

Then r_{24}, r_{42} are both zero, whence $\min(p_2 + p_4, q_2 + q_4)$ must equal $r_{22} + r_{44}$ to be a maximal coupling. But

$$\min(p_2 + p_4, q_2 + q_4) = \min(r_{22} + r_{44} + r_{46}, r_{12} + r_{22} + r_{44} + x) > r_{22} + r_{44},$$

since both r_{12} and r_{46} are positive, a contradiction.

We come to the conclusion that the only positive non-diagonal elements in the matrix can be in the column 2 (and they are all p-minimized).

	1	2	3	4	5	6	$\dots$	
1	r_{11}	$r_{12} > 0$	0	0	0	0	**0**	$p_1 = r_{11} + r_{12}$
2	0	r_{22}	0	0	0	0	**0**	$p_2 = r_{22}$
3	0	$r_{32} \geq 0$	r_{33}	0	0	0	**0**	$p_3 = r_{32} + r_{33}$
4	0	$r_{42} \geq 0$	0	r_{44}	0	0	**0**	$p_4 = r_{42} + r_{44}$
5	0	$r_{52} \geq 0$	0	0	r_{55}	0	**0**	$p_5 = r_{52} + r_{55}$
6	0	$r_{62} \geq 0$	0	0	0	r_{66}	**0**	$p_6 = r_{62} + r_{66}$
$\vdots$	**0**	$\vdots$	**0**	**0**	**0**	**0**	$\ddots$	$\vdots$
	$q_1 = r_{11}$	$q_2 \geq r_{12} + r_{22}$	$q_3 = r_{33}$	$q_4 = r_{44}$	$q_5 = r_{55}$	$q_6 = r_{66}$	$\dots$	

Generalizing, let $r_{ij} > 0$ and $i \neq j$. Then, if r_{ij} is p-minimized, all non-diagonal elements of the matrix outside column j are zero (and the non-diagonal elements in the jth column are p-minimized); if r_{ij} is q-minimized, then all non-diagonal elements of the matrix outside row i are zero (and the non-diagonal elements in the ith row are q-minimized).

It is easy to check that such a construction is always internally consistent. □

Corollary (Section 4, Corollary 4.5)**.** *The 1-2 system for the original rvs R_1^1, R_1^2 has a maximally-connected coupling if and only if either $p_i > q_i$ for no more than one i (this single possible i being the single fixed i in the formulation of the theorem), or $p_j < q_j$ for no more than one j (this single possible j being the single fixed j in the formulation of the theorem), $i, j \in \{1, \dots, k\}$.*

Proof of Corollary 4.5. The "only if" part is obvious. To demonstrate the "if" part, consider (without loss of generality) the arrangement

	1	2	3	4	5	6	…	
1							…	$p_1 \geq q_1$
2							…	p_2
3							…	$p_3 \geq q_3$
4							…	$p_4 \geq q_4$
5							…	$p_5 \geq q_5$
6							…	$p_6 \geq q_6$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮⋮	⋮
	q_1	$q_2 \geq p_2$	q_3	q_4	q_5	q_6	…	

and fill it in as

	1	2	3	4	5	6	…	
1	q_1	$p_1 - q_1$	0	0	0	0	**0**	$p_1 \geq q_1$
2	0	p_2	0	0	0	0	**0**	p_2
3	0	$p_3 - q_3$	q_3	0	0	0	**0**	$p_3 \geq q_3$
4	0	$p_4 - q_4$	0	q_4	0	0	**0**	$p_4 \geq q_4$
5	0	$p_5 - q_5$	0	0	q_5	0	**0**	$p_5 \geq q_5$
6	0	$p_6 - q_6$	0	0	0	q_6	**0**	$p_6 \geq q_6$
⋮	**0**	⋮	**0**	**0**	**0**	**0**	⋮⋮	⋮
	q_1	$q_2 \geq p_2$	q_3	q_4	q_5	q_6	…	

with the empty cells filled in with zeros. Check that (a) all rows sum to the marginals; (b) the second column sums to

$$\sum_{i=1}^{k} p_i - \left(\sum_{i=1}^{k} q_i - q_2 \right) = q_2;$$

(c) the rest of the columns sum to the marginals; (d) all r_{ii} are $\min(p_i, q_i)$; and (e) for all pairs r_{ij} $(i \neq j)$ the sums $r_{ii} + r_{ij} + r_{ji} + r_{jj}$ equal $\min(p_i + p_j, q_i + q_j)$. The latter is proved by considering first all $j \neq 2$, where it is obvious, and then $j = 2$ where the computation is, for $i \neq 2$,

$$r_{ii} + r_{i2} + r_{2i} + r_{22} = q_i + (p_i - q_i) + 0 + p_2 = p_i + p_2,$$

as it should be because the values in the second column are to be p-minimized. □

Theorem (Section 4, Theorem 4.6)**.** *The system $\mathcal{D}$ is noncontextual if and only if its 1-2 subsystem is noncontextual, i.e., if and only if one of the R_1^1* and *R_1^2 nominally dominates the other.*

Proof of Theorem 4.6. The "only if" part is Theorem 4.1. All we need to proof the "if " part is to check that the relation (23) holds. Assume the arrangement is as in the previous corollary. Consider first any set $i_1, \dots, i_m$ that does not include 2:

$$\min(p_{i_1} + p_{i_2} + \dots + p_{i_m}, q_{i_1} + q_{i_2} + \dots + q_{i_m}) = q_{i_1} + q_{i_2} + \dots + q_{i_m},$$

$$\sum_{j=1}^{m} \min\left(p_{i_j}, q_{i_j}\right) = q_{i_1} + q_{i_2} + \dots + q_{i_m},$$

$$\min\left(p_{i_j} + p_{i_{j'}}, q_{i_j} + q_{i_{j'}}\right) - \min\left(p_{i_j}, q_{i_j}\right) - \min\left(p_{i_{j'}}, q_{i_{j'}}\right) = 0.$$

So, (23) holds. If one of the indices (let it be i_1) is 2, then

$$q_2 + q_{i_2} + \ldots + q_{i_m} = \left(p_2 + \sum_{x \neq 2} (p_x - q_x) \right) + q_{i_2} + \ldots + q_{i_m} > p_2 + p_{i_2} + \ldots + p_{i_m},$$

so

$$\min \left(p_2 + p_{i_2} + \ldots + p_{i_m}, q_2 + q_{i_2} + \ldots + q_{i_m}\right) = p_2 + p_{i_2} + \ldots + p_{i_m}.$$

We also have

$$\sum_{j=1}^{m} \min \left(p_{i_j}, q_{i_j}\right) = p_2 + q_{i_2} + \ldots + q_{i_m},$$

and for any $j \neq 2, j' \neq 2$,

$$\min \left(p_{i_j} + p_{i_{j'}}, q_{i_j} + q_{i_{j'}}\right) - \min \left(p_{i_j}, q_{i_j}\right) - \min \left(p_{i_{j'}}, q_{i_{j'}}\right) = 0,$$

$$\min \left(p_2 + p_{i_j}, q_2 + q_{i_j}\right) - \min \left(p_2, q_2\right) - \min \left(p_{i_j}, q_{i_j}\right) = p_{i_j} - q_{i_j}.$$

Since index $i_1 = 2$ is paired with each of $i_2, \ldots, i_m$ only once, the right-hand side in (23) is

$$p_2 + q_{i_2} + (p_{i_2} - q_{i_2}) + \ldots + q_{i_m} + (p_{i_m} - q_{i_m}) = p_2 + p_{i_2} + \ldots + p_{i_m}.$$

$\square$