

Research



Cite this article: Dzhafarov EN, Cervantes VH, Kujala JV. 2017 Contextuality in canonical systems of random variables. *Phil. Trans. R. Soc. A* **375**: 20160389. <http://dx.doi.org/10.1098/rsta.2016.0389>

Accepted: 19 July 2017

One contribution of 15 to a theme issue ‘Second quantum revolution: foundational questions’.

Subject Areas:

probability theory, quantum physics, psychology

Keywords:

canonical systems, contextuality, dichotomization, direct influences, measurements

Author for correspondence:

Ehtibar N. Dzhafarov
e-mail: ehtibar@purdue.edu

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.3865489>.

Contextuality in canonical systems of random variables

Ehtibar N. Dzhafarov¹, Víctor H. Cervantes¹ and Janne V. Kujala²

¹Psychological Sciences, Purdue University, West Lafayette, IN, USA

²Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland

END, 0000-0003-1909-7706

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.

This article is part of the themed issue ‘Second quantum revolution: foundational questions’.

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 [1–3].

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* [4–7]:

$$\begin{array}{cc|c}
 R_1^1 & R_2^1 & c = 1 \\
 \hline
 R_1^2 & R_2^2 & c = 2 \\
 \hline
 q = 1 & q = 2 & \mathcal{R}
 \end{array}$$

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- $\frac{1}{2}$ 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’ $[R_2^1 = x, R_1^2 = y]$, 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 [1] 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 [8]. 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 behavioural example, one’s response to a question asked second will generally be influenced by the question asked first [9,10]. 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 [1,3,11] and, with caveats, [6,12–17]), 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 $\Pr [T_q^1 = T_q^2]$. This maximal coupling (T_q^1, T_q^2) 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} ,

$$\begin{array}{cc|c}
 S_1^1 & S_2^1 & c = 1 \\
 \hline
 S_1^2 & S_2^2 & c = 2 \\
 \hline
 q = 1 & q = 2 & \boxed{S} ,
 \end{array}$$

such that its rows (S_1^c, S_2^c) are distributed as the rows of \mathcal{R} and its columns (S_q^1, S_q^2) are distributed as the maximal couplings (T_q^1, T_q^2) of the columns of \mathcal{R} . If such a *maximally connected* coupling S does not exist, one can say that the within-conteNt (row-wise) relations prevent different measurements of the same conteNt (column-wise) from being 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-conteNt counterparts force them, if imposed a joint distribution on, 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-conteNt 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 *non-contextual*.

The (non)existence of S is determined by a simple linear programming procedure [3,4]: in our example, S has 2^4 possible values, and we find out if they can be assigned non-negative numbers (probability masses) that sum to the given row-wise probabilities $\Pr[R_1^c = x, R_2^c = y]$ and the computed column-wise probabilities $\Pr[T_q^1 = x, T_q^2 = y]$. There is also a simple criterion (inequality) for the existence of a solution for this system of equations [4–6]. Using it one can show, e.g. that in our quantum-mechanical application the system \mathcal{R} is always non-contextual, and this is also true for the behavioural application if one adopts the model proposed in [9] (see [18] 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 [3]: 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 §2c).

Although most of these principles and procedures of CbD have been formulated for arbitrary systems of measurements [3,11], 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 behavioural contextuality studies [3,4,6,11,18,19], 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:

$$\begin{array}{cc|c}
 R_1^1 & R_2^1 & c = 1 \\
 \hline
 R_1^2 & R_2^2 & c = 2 \\
 \hline
 R_1^3 & R_2^3 & c = 3 \\
 \hline
 q = 1 & q = 2 & \boxed{\mathcal{R}'}.
 \end{array}$$

How does CbD apply here? In the earlier version of the theory (summarized in [3,11]), we computed the couplings (T_q^1, T_q^2, T_q^3) of each column that maximize $\Pr[T_q^1 = T_q^2 = T_q^3]$. One problem with this approach is that the maximal coupling (T_q^1, T_q^2, T_q^3) , while it always exists, is not defined uniquely. What should be the contextuality analysis of \mathcal{R}' if the within-conteNt (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 non-contextual 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 non-contextual but one of its subsystems, e.g. \mathcal{R} , is contextual. This is contrary to one's intuition of non-contextuality.

In the most recent publications therefore [1,2], 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 $\Pr [T_q^1 = T_q^2]$, $\Pr [T_q^2 = T_q^3]$ and $\Pr [T_q^1 = T_q^3]$ (it is in fact sufficient to maximize only certain pairs rather than all of them, but this is not critical here). Such a coupling (T_q^1, T_q^2, T_q^3) is called *multimaximal*. With only binary rvs involved, a multimaximal coupling always exists and is unique; and a subsystem of a non-contextual system then is always non-contextual.

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 by $\{-2, -1, 0, 1, 2\}$. In the behavioural application, let the questions asked allow five answers each, labelled 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 non-contextual, one may consider it desirable that it remain non-contextual 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-non-zero.' As it turns out, a non-contextual 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

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

If one is also interested in the coarsening of R_q^c into values 'negative-zero-positive', then the list should be expanded into

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

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. As every dichotomization creating a split should be applied to all rvs sharing a conteNt, one ends up replacing the system \mathcal{R} with

$D_{1\{-2\}}^1$	\cdots	$D_{1\{2\}}^1$	$D_{1\{-2,-1\}}^1$	\cdots	$D_{1\{1,2\}}^1$	\cdots	$D_{2\{1,2\}}^1$	$c = 1$
$D_{1\{-2\}}^2$	\cdots	$D_{1\{2\}}^2$	$D_{1\{-2,-1\}}^2$	\cdots	$D_{1\{1,2\}}^2$	\cdots	$D_{2\{1,2\}}^2$	$c = 2$
$q = 1\{-2\}$	\cdots	$q = 1\{2\}$	$q = 1\{-2, -1\}$	\cdots	$q = 1\{1, 2\}$	\cdots	$q = 2\{1, 2\}$	\mathcal{D}

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

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

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

$D_{1'}^1$	$D_{1''}^1$	$D_{1'''}^1$	$f(D_{1'}^1, D_{1''}^1, D_{1'''}^1)$	$c = 1$
$D_{1'}^2$	$D_{1''}^2$	$D_{1'''}^2$	$f(D_{1'}^2, D_{1''}^2, D_{1'''}^2)$	$c = 2$
$q = 1'$	$q = 1''$	$q = 1'''$	q^*	\mathcal{D}'

of the system \mathcal{D} , the f -transformation of the maximal couplings of the first three columns, because 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 §2, we present the abstract version of CbD applicable to all possible systems of categorical (and not only categorical) rvs. In §3, we formalize the idea of representing any system of rvs by their splits and applying contextuality analysis to these representations only. In §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 electronic supplementary material, file S, together with additional theorems and examples.

2. Formal theory of contextuality

(a) Basic notions

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

$$\prec \subseteq Q \times C, \quad (2.1)$$

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)

$$\mathcal{R} = \{R_q^c : q \in Q, c \in C, q \prec c\}, \quad (2.2)$$

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

$$\mathcal{R}^c = \{R_q^c : q \in Q, q \prec c\} = R^c \quad (2.3)$$

of \mathcal{R} is called a *bunch* (of rvs) corresponding to c . As 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 (2.2) and as

$$\mathcal{R} = \{R^c : c \in C\}. \quad (2.4)$$

(See [1,3] for a detailed probabilistic analysis.)

For any $q \in Q$, the subset

$$\mathcal{R}_q = \{R_q^c : c \in C, q \prec c\} \quad (2.5)$$

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.

(b) General definition of (non)contextuality

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

An (overall) coupling S of a system \mathcal{R} in (2.2) is a coupling of its bunches. That is, it is an rv

$$S = \{S^c : c \in C\}, \quad (2.6)$$

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

$$S^c = \{S_q^c : q \in Q, q \prec c\} \quad (2.7)$$

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

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

$$T_q = \{T_q^c : c \in C, q \prec c\}, \quad (2.8)$$

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

$$S_q = \{S_q^c : c \in C, q \prec c\} \quad (2.9)$$

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 *non-contextual 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 non-contextual with respect to \mathcal{T} if and only if there is a jointly distributed set

$$S = \{S_q^c : q \in Q, c \in C, q \prec c\}, \quad (2.10)$$

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

Remark 2.3. In §3c, we will use the property of (multi)maximality to play the role of \mathcal{C} , and the couplings in question then are referred to as (multi)maximally connected.

(c) Degree of contextuality

A *quasi-distribution* on a finite set V is a function $V \rightarrow \mathbb{R}$ (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 *quasi-coupling* X of \mathcal{R} is defined as a quasi-rv

$$X = \{X_q^c : q \in Q, c \in C, q \prec c\}, \quad (2.11)$$

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

Theorem 2.4 ([3, 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 $X_q = \{X_q^c : c \in C, q \prec c\} \sim T_q$ 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 ([3, 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 non-negative, and the system \mathcal{R} is non-contextual 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

(a) 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, \dots, q_k} is labelled by a set of conteNts, q_1, \dots, q_k , and it takes as its arguments the rvs $R_{q_1}^c, \dots, R_{q_k}^c$ in each conteXt c such that $c \succ q_1, \dots, q_k$. The outcome,

$$R_{q^*}^c = F_{q_1, \dots, q_k}(R_{q_1}^c, \dots, R_{q_k}^c) \quad (3.1)$$

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

$$R_{q_1}^c, \dots, R_{q_k}^c \mapsto (R_{q_1}^c, \dots, R_{q_k}^c) = R_{q'}^c, \quad (3.2)$$

whereas coarsening is transformation

$$R_q^c \mapsto F_q(R_q^c) = R_{q''}^c. \quad (3.3)$$

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

Example 3.1 (joining). Consider the system

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

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_2^1 and of R_2^1, R_2^2 in the first and second columns, these row-wise joints are not used. In some applications, this would be unacceptable (e.g. in the theory of selective influences [20,21] and in the approach advocated by Abramsky and colleagues [22,23]

this is never acceptable), and then the following expansion has to be used:

R_1^1	R_2^1	·	(R_1^1, R_2^1)	$c = 1$
R_1^2	R_2^2	·	(R_1^2, R_2^2)	2
R_1^3	·	R_3^3	·	3
·	R_2^4	R_3^4	·	4
$q = 1$	2	3	12	\mathcal{R}^*

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 $R_q^c = F_q(R_q^c)$. 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

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

and let the two rvs take values on $\{1, 2, 3, 4, 5\}$. If these values are considered ordered, $1 < \dots < 5$, one may be interested in all possible partitions of $\{1, 2, 3, 4, 5\}$ into subsets of consecutive numbers, such as $\{12 \mid 34 \mid 5\}$, $\{1 \mid 2345\}$, etc. There are 15 such partitions (counting $\{1 \mid 2 \mid 3 \mid 4 \mid 5\}$ 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 non-trivial 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

R_q^1	$R_{q1'}^1$	·	$R_{q14'}^1$	$c = 1$	and	R_q^1	$R_{q1''}^1$	·	$R_{q50''}^1$	$c = 1$
R_q^2	$R_{q1'}^2$	·	$R_{q14'}^2$	$c = 2$		\mathcal{R}'	R_q^2	$R_{q1''}^2$	·	$R_{q50''}^2$
q	$q1'$	·	$q14'$			q	$q1''$	·	$q50''$	\mathcal{R}''

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 non-zero 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 $\sum_x \Pr [R_q^c = x] = 1$.

(b) Dichotomizations and canonical/split representations

Definition 3.4. A *dichotomization* of a set V is a function $f: V \rightarrow \{0, 1\}$. 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 $f(R_q^c)$ 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 case they have the same number of elements, we order the elements of V , say $1 < 2 < \dots < k$, and then choose W as lexicographically preceding $V - W$.

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

$$\{D_{q(1)}^c, D_{q(2)}^c, \dots, D_{q(k)}^c\} \quad (3.4)$$

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 (R_1^1, R_2^1) and (R_1^2, R_2^2) in \mathcal{R}^* have four values each: 00, 01, 10 and 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}^* :

$D_1^1 = R_1^1$	$D_2^1 = R_2^1$	·	$D_{12\{00\}}^1$	$D_{12\{01\}}^1$	$D_{12\{10\}}^1$	$D_{12\{11\}}^1$	$c = 1$
$D_1^2 = R_1^2$	$D_2^2 = R_2^2$	·	$D_{12\{00\}}^2$	$D_{12\{01\}}^2$	$D_{12\{10\}}^2$	$D_{12\{11\}}^2$	2
$D_1^3 = R_1^3$	·	$D_3^3 = R_3^3$	·	·	·	·	3
	$D_2^4 = R_2^4$	$D_3^4 = R_3^4$	·	·	·	·	4
$q = 1$	2	3	12{00}	12{01}	12{10}	12{11}	\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 $\{12 | 3 | 45\}$, $\{1 | 2 | 3 | 45\}$ and $\{12 | 3 | 4 | 5\}$. 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 as $D_{q\{1,23\}}^1$ 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 §4.

(c) 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 §2b, 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 §2c.

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* [1,2].

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 $c, c' \succ q$, $\Pr [T_q^c = T_q^{c'}]$ 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 ([2] Corollary 1).

Multimax2: 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 ([2, Theorem 5]; [1, Theorem 2.3]).

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

$$\Pr [D_q^{c_1} = 1] \leq \dots \leq \Pr [D_q^{c_n} = 1],$$

then T_q is a multimaximal coupling of \mathcal{D}_q if and only if $\Pr [T_q^{c_i} = T_q^{c_{i+1}}]$ is maximal for $i = 1, \dots, n - 1$, over all possible couplings of \mathcal{D}_q ([1, 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:

$$\begin{array}{|c|} \hline R_1^1 \\ \hline R_1^2 \\ \hline \end{array} \quad \begin{array}{l} c = 1 \\ c = 2 \end{array} \\
 q = 1 \quad \boxed{\mathcal{R}}.$$

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

$$\begin{array}{|c|c|c|c|} \hline D^1: & D_{W1}^1 & D_{W2}^1 & D_{W(2^{k-1}-1)}^1 \\ \hline D^2: & D_{W1}^2 & D_{W2}^2 & D_{W(2^{k-1}-1)}^2 \\ \hline & q = W1 & W2 & \dots \quad W(2^{k-1} - 1) \\ \hline & & & \boxed{\mathcal{D}} \end{array},$$

where $W1, W2$, etc. are the subsets $f^{-1}(1)$ chosen as explained in §3b from the $2^{k-1} - 1$ distinct dichotomizations f of $\{1, \dots, k\}$. 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 $\{1, 2, \dots, k\} - W$. 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

$$\Pr [R_1^1 = i] = p_i \quad \text{and} \quad \Pr [R_1^2 = i] = q_i, \quad i = 1, 2, \dots, k. \quad (4.1)$$

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, $1, 2, \dots, k$, as in the original variable:

$$\Pr [D^1 = i] = p_i \quad \text{and} \quad \Pr [D^2 = i] = q_i, \quad i = 1, 2, \dots, k. \quad (4.2)$$

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

$$r_{ij} = \Pr [S_q^1 = i, S_q^2 = j], \quad i, j \in \{1, \dots, k\}, \quad (4.3)$$

to these k^2 states so that they satisfy $2k$ linear constraints imposed by (4.1),

$$\sum_{j=1}^k r_{ij} = p_i \quad \text{and} \quad \sum_{i=1}^k r_{ij} = q_j, \quad i, j \in \{1, \dots, k\}. \quad (4.4)$$

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 $W = \{i_1, \dots, i_m\} \subset \{1, \dots, k\}$, then the maximal coupling (S_W^1, S_W^2) of (D_W^1, D_W^2) is distributed as

$$\left. \begin{array}{l} \Pr [S_W^1 = 1] = \Pr [D_W^1 = 1] = p_{i_1} + p_{i_2} + \dots + p_{i_m} \\ \Pr [S_W^2 = 1] = \Pr [D_W^2 = 1] = q_{i_1} + q_{i_2} + \dots + q_{i_m} \end{array} \right\}. \quad (4.5)$$

and $\Pr [S_W^1 = S_W^2 = 1] = \min(p_{i_1} + p_{i_2} + \dots + p_{i_m}, q_{i_1} + q_{i_2} + \dots + q_{i_m})$

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 (4.3) satisfy (4.5) for $m = 1$ and $m = 2$:

$$r_{ii} = \min(p_i, q_i), \quad i \in \{1, \dots, k\} \quad (4.6)$$

and

$$r_{ii} + r_{ij} + r_{ji} + r_{jj} = \min(p_i + p_j, q_i + q_j), \quad i, j \in \{1, \dots, k\}, \quad i < j. \quad (4.7)$$

That is, a maximally connected coupling of the 1–2 system is described by the $3k + \binom{k}{2}$ linear equations (4.4), (4.6) and (4.7). We have therefore the following necessary condition for non-contextuality of \mathcal{D} .

Theorem 4.1. *If the system \mathcal{D} is non-contextual, then the $3k + \binom{k}{2}$ linear equations (4.4), (4.6) and (4.7) are satisfied.*

Remark 4.2. Note that $3k + \binom{k}{2} < k^2$ for $k > 5$. (For completeness only, theorem S.1 in electronic supplementary material, 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 $W = \{i_1, \dots, i_m\} \subset \{1, \dots, k\}$, the probability that the corresponding m -split equals 1 is*

$$\begin{aligned} \min(p_{i_1} + p_{i_2} + \dots + p_{i_m}, q_{i_1} + q_{i_2} + \dots + q_{i_m}) &= \sum_{j=1}^m \min(p_{i_j}, q_{i_j}) \\ &+ \sum_{j=1}^{m-1} \sum_{j'=j+1}^m [\min(p_{i_j} + p_{i_{j'}}, q_{i_j} + q_{i_{j'}}) - \min(p_{i_j}, q_{i_j}) - \min(p_{i_{j'}}, q_{i_{j'}})]. \end{aligned} \quad (4.8)$$

It is easy to find numerical examples of the distributions of R_1^1 and R_2^1 for which (4.8) is violated (see example S.2 in electronic supplementary material, file S). As shown below, however, (4.8) 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: (4.4), (4.6) and (4.7) imply that \mathcal{D} is non-contextual. We establish this fact by first characterizing the distributions of R_1^1 and R_2^1 for a non-contextual 1–2 system (theorem 4.4 with corollary 4.5), and then showing that (4.8) 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 (4.3) that may have non-zero 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 ($i, j = 1, \dots, k$).*

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 non-zero probabilities assigned to the states of the maximally connected coupling (shown below for $k = 4$) could only occupy the cells marked with asterisks:

	1	2	3	4		1	2	3	4
1	*	*	0	0	or	1	*	0	0
2	0	*	0	0		2	*	*	*
3	0	*	*	0		3	0	0	*
4	0	*	0	*		4	0	0	*

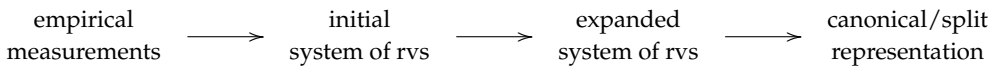
Corollary 4.5. *The 1–2 system for the original rvs R_1^1, R_2^1 has a maximally connected coupling if and only if either $p_i > q_i$ for no more than one i (this single possible i being the single fixed i in the formulation of the theorem), or $p_j < q_j$ for no more than one j (this single possible j being the single fixed j in the formulation of the theorem), $i, j \in \{1, \dots, k\}$.*

The relationship between (p_1, \dots, p_k) and (q_1, \dots, 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 $i = 1, \dots, k$ (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 non-contextual if and only if its 1–2 subsystem is non-contextual, 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:

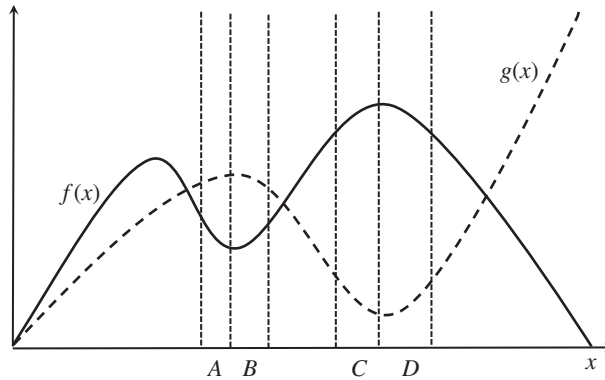


In the initial system, measurements are represented by rvs each of which generally has multiple values. Expansion means adding to the system new conteNts with corresponding connections (conteNt-sharing rvs) computed as functions of the existing connections. In a canonical representation of the system all rvs are binary, and the connections are coupled multimaximally, meaning essentially that one deals with their elements pair-wise. 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 non-contextual system has all its subsystems non-contextual 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 non-contextual 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 $\Pr [R_q^c = x] < \Pr [R_q^{c'} = x]$ holds for no more than one value x of these rvs: a stringent necessary condition for non-contextuality, 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 [24], psychology [9,25,26], economics [26,27] and political science [28]. 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 behaviour 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 non-contextual [18,19,29]. 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, $x_1 < x_2 < \dots, x_N$, it may be natural to only allow dichotomizations f with $f^{-1}(1)$ containing several successive values, $\{x_l, x_{l+1}, \dots, x_L\}$, for some $l, L \in \{1, \dots, N\}$. An even stronger restriction would be to only allow 'cuts', with $f^{-1}(1) = \{x_l, x_{l+1}, \dots, x_N\}$ or $\{x_1, x_2, \dots, x_{l-1}\}$.



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 non-contextual 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, \dots, H_k of the set of reals into intervals, we should have either

$$\int_{H_i} f(x) dx < \int_{H_i} g(x) dx \text{ for no more than one of } i = 1, \dots, k$$

or

$$\int_{H_i} f(x) dx > \int_{H_i} g(x) dx \text{ for no more than one of } i = 1, \dots, k. \quad (5.1)$$

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, \dots, H_k that includes A, B, C, D will violate (5.1). 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.

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

Competing interests. We declare we have no competing interests.

Funding. This research has been supported by AFOSR grant no. FA9550-14-1-0318 and Purdue Graduate School Summer Research Grant.

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

References

1. Dzhafarov EN, Kujala JV. 2017 Probabilistic foundations of contextuality. *Fortschr. Phys. Prog. Phys.* **65**, 1600040.
2. Dzhafarov EN, Kujala JV. 2017 Contextuality-by-Default 2.0: systems with binary random variables. In (eds JA de Barros, B Coecke, E Pothos) *Lect. Not. Comp. Sci.*, no. 10106, pp. 16–32. Berlin, Germany: Springer.
3. Dzhafarov EN, Kujala JV. 2016 Context-content systems of random variables: the contextuality-by-Default theory. *J. Math. Psych.* **74**, 11–33. (doi:10.1016/j.jmp.2016.04.010)
4. Dzhafarov EN, Kujala JV, Larsson J-Å. 2015 Contextuality in three types of quantum-mechanical systems. *Found. Phys.* **7**, 762–782. (doi:10.1007/s10701-015-9882-9)
5. Kujala JV, Dzhafarov EN. 2016 Proof of a conjecture on contextuality in cyclic systems with binary variables. *Found. Phys.* **46**, 282–299. (doi:10.1007/s10701-015-9964-8)

6. 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. (doi:10.1103/PhysRevLett.115.150401)
7. Kujala JV, Dzhafarov EN. 2015 Probabilistic contextuality in EPR/Bohm-type systems with signaling allowed. In *Contextuality from quantum physics to psychology* (eds EN Dzhafarov, S Jordan, R Zhang, VH Cervantes), pp. 287–308. Hackensack, NJ: World Scientific.
8. Bacciagaluppi G. 2015 Leggett-Garg inequalities, pilot waves and contextuality. *Int. J. Quant. Found.* **1**, 1–17.
9. 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. (doi:10.1111/tops.12040)
10. Moore DW. 2002 Measuring new types of question-order effects. *Public Opin. Quart.* **66**, 80–91. (doi:10.1086/338631)
11. Dzhafarov EN, Kujala JV, Cervantes VH 2016 Contextuality-by-Default: a brief overview of ideas, concepts, and terminology. In (eds H Atmanspacher, T Filk, E Pothos) *Lect. Not. Comp. Sci.* no. 9535, pp. 12–23. Berlin, Germany: Springer.
12. Cabello A. 2013 Simple explanation of the quantum violation of a fundamental inequality. *Phys. Rev. Lett.* **110**, 060402. (doi:10.1103/PhysRevLett.110.060402)
13. Kurzynski P, Ramanathan R, Kaszlikowski D. 2012 Entropic test of quantum contextuality. *Phys. Rev. Lett.* **109**, 020404. (doi:10.1103/PhysRevLett.109.020404)
14. Khrennikov A 2009 *Contextual approach to quantum formalism*. Berlin, Germany: Springer.
15. 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. (doi:10.1007/s10701-005-6511-z)
16. Fine A. 1982 Joint distributions, quantum correlations, and commuting observables. *J. Math. Phys.* **23**, 1306–1310. (doi:10.1063/1.525514)
17. Suppes P, Zanotti M. 1981 When are probabilistic explanations possible? *Synthese* **48**, 191–199. (doi:10.1007/BF01063886)
18. Dzhafarov EN, Zhang R, Kujala JV. 2016 Is there contextuality in behavioral and social systems?. *Phil. Trans. R. Soc. A* **374**, 20150099. (doi:10.1098/rsta.2015.0099)
19. Dzhafarov EN, Kujala JV, Cervantes VH, Zhang R, Jones M. 2016 On contextuality in behavioral data. *Phil. Trans. R. Soc. A* **374**, 20150234. (doi:10.1098/rsta.2015.0234)
20. Dzhafarov EN, Kujala JV. 2016 Probability, random variables, and selectivity. In *The new handbook of mathematical psychology* (eds W Batchelder, H Colonius, EN Dzhafarov, J Myung) pp. 85–150. Cambridge, UK: Cambridge University Press.
21. Dzhafarov EN, Kujala JV. 2012 Selectivity in probabilistic causality: where psychology runs into quantum physics. *J. Math. Psychol.* **56**, 54–63. (doi:10.1016/j.jmp.2011.12.003)
22. Abramsky S, Barbosa RS, Kishida K, Lal R, Mansfield S. 2015 Contextuality, cohomology and paradox. *Comp. Sci. Log.* **2015**, 211–228. (doi:10.4230/LIPIcs.CSL.2015.211)
23. Abramsky S, Brandenburger A. 2011 The sheaf-theoretic structure of non-locality and contextuality. *New J. Phys.* **13**, 113 036–113 075. (doi:10.1088/1367-2630/13/11/113036)
24. 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. (doi:10.1007/s10701-015-9929-y)
25. Busemeyer JR, Bruza PD 2012 *Quantum cognition and decision*. Cambridge, UK: Cambridge University Press.
26. Khrennikov A 2010 *Ubiquitous quantum structure: from psychology to finance*. Berlin, Germany: Springer.
27. Haven E, Khrennikov A. 2012 *Quantum social science*. Cambridge, UK: Cambridge University Press.
28. Khrennikova P, Haven E. 2016 Instability of political preferences and the role of mass media: a representation in quantum framework. *Phil. Trans. R. Soc. A* **374**, 20150106. (doi:10.1098/rsta.2015.0106)
29. Cervantes VH, Dzhafarov EN. 2017 Advanced analysis of quantum contextuality in a psychophysical double-detection experiment. *J. Math. Psychol.* **79**, 77–84. (doi:10.1016/j.jmp.2017.03.003)