Author(s): Dzhafarov, Ehtibar N.; Kujala, Janne

Title: Probabilistic foundations of contextuality

Year: 2017

Version:

Please cite the original version:

Dzhafarov, E. N., & Kujala, J. (2017). Probabilistic foundations of contextuality. Fortschritte der Physik - Progress of Physics, 65(6-8), Article 1600040. https://doi.org/10.1002/prop.201600040

Probabilistic Foundations of Contextuality

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

[1] *Purdue University, USA*

[2] *University of Jyväskylä, Finland*

Contextuality is usually defined as absence of a joint distribution for a set of measurements (random variables) with known joint distributions of some of its subsets. However, if these subsets of measurements are not disjoint, contextuality is mathematically impossible even if one generally allows (as one must) for random variables not to be jointly distributed. To avoid contradictions one has to adopt the Contextuality-by-Default approach: measurements made in different contexts are always distinct and stochastically unrelated to each other. Contextuality is reformulated then in terms of the (im)possibility of imposing on all the measurements in a system a joint distribution of a particular kind: such that any measurements of one and the same property made in different contexts satisfy a specified property, $\mathcal{C}$. In the traditional analysis of contextuality $\mathcal{C}$ means "are equal to each other with probability 1". However, if the system of measurements violates the "no-disturbance principle", due to signaling or experimental biases, then the meaning of $\mathcal{C}$ has to be generalized, and the proposed generalization is "are equal to each other with maximal possible probability" (applied to any set of measurements of one and the same property). This approach is illustrated on arbitrary systems of binary measurements, including most of quantum systems of traditional interest in contextuality studies (irrespective of whether the "no-disturbance" principle holds in them).

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

I. WHAT IS CONTEXTUALITY?

I.1. Measurements as random variables

It appears to be an established view in quantum mechanics that the notion of contextuality equally applies to systems of measurements of very different physical nature, ranging from the Kochen-Specker (KS) [6, 28, 38] and Klyachko-Can-Binicoglu-Shumovsky (KCBS) systems [27] with their generalizations [37] to the Einsten-Podolsky-Rosen-Bohm-Bell (EPR-BB) systems [4, 7, 22] and Suppes-Zanotti-Leggett-Garg ones (SZLG) [23, 41]. Even though these systems are variously discussed in terms of nonlocality, complementarity, or macroscopic realism and noninvasiveness (in addition to "contextuality proper"), most contemporary researchers agree that, from a mathematical point of view, their contextuality analysis is always about one and the same determination:

> **(S1)** existence or nonexistence of a joint distribution for the random variables representing all the measurements in play [22, 24–26, 33, 34, 41].

A measurement is a random variable characterized (distinguished from other measurements in the same system) by what it measures (a certain property q, a quantum observable) and by the context c in which it measures q. The context c is usually (but not necessarily, as we will see below) defined by the set of all observables being measured simultaneously with q. As an example, the diagram below shows the formal structure of the KCBS measurement system:

$$\begin{bmatrix} \text{(KCBS)} & c_1 & c_2 & c_3 & c_4 & c_5 \\ & & & & & \\ q_1 & \star & & & & \star \\ q_2 & \star & \star & & & \\ q_3 & & \star & \star & & \\ q_4 & & & \star & \star & \\ q_5 & & & & \star & \star \\ & & & & & \end{bmatrix} \tag{1}$$

The measured properties $q_1, \ldots, q_5$ are represented by projection operators, and each context is defined by two projection operators being measured together. A star symbol in the cell (q_i, c_j) indicates the binary measurement R of property q_i in context c_j. The two measurements within each context are recorded as a pair, so they possess an empirically well-defined joint distribution. For instance, the two measurements in context c_1, in addition to having well-defined expected values when taken separately, also have a well-defined correlation.

In this paper the consideration is confined to systems with finite numbers of properties q and contexts c, and in all examples and applications the measurements are assumed to be binary.

It is instructive to look at some more complex paradigms in the same format as (1). The matrices in Fig. 1 represents the systems of measurements devised in Refs. [6] and [38] to prove two versions of the Kochen-Specker theorem [28]. The KS-3D matrix demonstrates the fact that contexts need not be defined by sets of observables alone but may involve additional conditions.

* E-mail: ehtibar@purdue.edu

† E-mail: jvk@iki.fi

$$
\left[\begin{array}{cccccccccc}
\text{KS-4D} & c_1 & c_2 & c_3 & c_4 & c_5 & c_6 & c_7 & c_8 & c_9 \\
 & & & & & & & & & \\
q_{0001} & \star & \star & & & & & & & \\
q_{0010} & \star & & & & \star & & & & \\
q_{1100} & \star & & \star & & & & & & \\
q_{1\bar{1}00} & \star & & & & & & \star & & \\
q_{0100} & & \star & & & \star & & & & \\
q_{1010} & & \star & & & & & & \star & \\
q_{10\bar{1}0} & & \star & & \star & & & & & \\
q_{1\bar{1}1\bar{1}} & & & \star & \star & & & & & \\
q_{1\bar{1}\bar{1}1} & & & \star & & & \star & & & \\
q_{0011} & & & \star & & & & \star & & \\
q_{1111} & & & & \star & & \star & & & \\
q_{010\bar{1}} & & & & \star & & & & \star & \\
q_{1001} & & & & & \star & & & & \star \\
q_{100\bar{1}} & & & & & \star & \star & & & \\
q_{01\bar{1}0} & & & & & & \star & & & \star \\
q_{11\bar{1}1} & & & & & & & \star & \star & \\
q_{111\bar{1}} & & & & & & & \star & & \star \\
q_{\bar{1}111} & & & & & & & & \star & \star
\end{array}\right]
\qquad
\left[\begin{array}{ccccccccccc}
\text{KS-3D} & c_{001} & c_{101} & c_{011} & c_{1\bar{1}2} & c_{102} & c_{211} & c_{201} & c_{112} & c_{012} & c_{121} \\
 & & & & & & & & & & \\
q_{100} & \boxed{\star} & & \star & & & & & & \star & \\
q_{010} & \star & \star & & & \star & & \star & & & \\
q_{110} & \star & & & \star & & & & & & \\
q_{1\bar{1}0} & \star & & & & & & & \star & & \\
q_{\bar{1}01} & & \star & & & & & & & & \star \\
q_{0\bar{1}1} & & & \star & & & \star & & & & \\
q_{\bar{1}12} & & & & \star & & & & & & \\
q_{\bar{2}01} & & & & \star & \star & & & & & \\
q_{021} & & & & \boxed{\star} & & & & & & \\
q_{\bar{2}11} & & & & & \star & \star & & & & \\
q_{\bar{1}02} & & & & & & \star & \star & & & \\
q_{\bar{1}\bar{1}2} & & & & & & & \star & \star & & \\
q_{0\bar{2}1} & & & & & & & & \star & \star & \\
q_{1\bar{2}1} & & & & & & & & & \star & \star \\
q_{0\bar{1}2} & & & & & & & & & & \boxed{\star}
\end{array}\right]
$$

Figure 1. A measurement system devised in Ref. [6] (left) and a modification of one devised in Ref. [38] (right) to prove the Kochen-Specker theorem in, respectively, 4D and 3D real Hilbert space. The q's are projection operators corresponding to radius-vectors with coordinates shown by subscripts (with $\bar{1}$ denoting -1, 2 denoting $\sqrt{2}$, and $\bar{2}$ denoting $-\sqrt{2}$). In the KS-4D matrix a context is defined by four observables measured together, chosen so that one and only one of them can yield value 1 when measured. In the KS-3D matrix a context c_{ijk} is defined by several observables measured together with the observable q_{ijk}, under the additional constraint that the measurement of q_{ijk} yields value 1, from which it follows that the measurements of all other observables within the context yield value 0. (The boxed measurements form an orthogonal system that leads to a contradiction proving the theorem.)

I.2. How is contextuality understood traditionally?

A traditional contextuality analysis of a system like KCBS, KS-4D, or KS-3D is based on the following assumption (almost always adopted implicitly):

Noncontextual Identification:: The random variable R representing a measurement is uniquely identified by the property q it measures, i.e., it is one and the same random variable in each context in which it appears.

In relation to the KCBS system, Noncontextual Identification means that it can be presented as (writing R_i in place of R_{q_i})

$$
\left[\begin{array}{cccccc}
\text{(KCBS)} & c_1 & c_2 & c_3 & c_4 & c_5 \\
 & & & & & \\
q_1 & R_1 & & & & R_1 \\
q_2 & R_2 & R_2 & & & \\
q_3 & & R_3 & R_3 & & \\
q_4 & & & R_4 & R_4 & \\
q_5 & & & & R_5 & R_5
\end{array}\right] \tag{2}
$$

The system is considered noncontextual (contextual) if one can (respectively, cannot) find a joint distribution for all the random variables in the system that agrees with the observed joint distributions (those within individual contexts) as its marginals. Thus, in (2), we seek a joint distribution of $R_1, \ldots, R_5$ that agrees with the observed joint distributions of (R_1, R_2) in context c_1, (R_2, R_3) in context c_2, etc.

Now, if such an overall distribution for (2) exists, then the observed joint distributions of (R_1, R_2), (R_2, R_3), ..., (R_5, R_1) must be related to each other in a certain way [27]. Denoting the possible values of the random variables by $+1$ and -1, one such relationship (a consequence of a more general formulation given in Section II.4) can be presented as

$$
-\langle R_1R_2\rangle - \langle R_2R_3\rangle - \langle R_3R_4\rangle - \langle R_4R_5\rangle - \langle R_5R_1\rangle \leq 3. \tag{3}
$$

We know that with an appropriate choice of the projection operators $q_1, \ldots, q_5$, both quantum theory [27] and experimental data [32, 35, 36] show that this inequality is violated. This is considered as a reductio ad absurdum proof that in such cases an overall joint distribution in question does not exist (while within each individual context, of course, the two random are jointly distributed).

In relation to the matrices shown in Fig. 1, the analysis appears to be even simpler. In the KS-4D matrix, replacing both star symbols in each row by one and the same random variable, we conclude that in any overall joint distribution of all random variables the number of the cells assigned value 1 is even. But each context, due to the choice of the projection operators, has one and only one measurement resulting in 1, and the number of contexts is odd — a contradiction. In the KS-3D matrix, replacing all star symbols in each row by one and the same random variable, we get a system in which all random variables, if jointly distributed, attain value 0 with probability 1. This is, however, impossible, because the observables marked by boxes form an orthogonal basis, whence one (and only one) of them must have value 1. Again, in both systems of Fig. 1, the usual interpretation is that the contradiction proves that an overall joint distribution does not exist, while the measurements within each context are jointly distributed.

The situation, however, is not that simple.

I.3. How can an overall joint distribution fail to exist?

A naive understanding of classical, Kolmogorovian probability theory (KPT) is that it requires that any two random variables have a joint distribution. With this understanding, the set of five random variables in (2) cannot violate a condition like (3), and since we know this is possible, classical probability theory must be inapplicable to quantum systems. A universal-domain-space model, however, is untenable for purely mathematical reasons (think, e.g., of the possible cardinality of the universal sample set); it should be dismissed irrespective of any quantum-mechanical considerations [11, 12, 14]. In fact, stochastically unrelated random variables (those possessing no joint distribution) arise very naturally in KPT, by the procedure of conditionalization [13]: the (conditional) random variables Y and Z in the tree diagram below,

$$\begin{tikzcd} & X \arrow[dl, "if\ X\in A"'] \arrow[dr, "if\ X\notin A"] & \\ Y & & Z \end{tikzcd}$$

(with A indicating any event in the distribution of X) have no joint distribution. One cannot, e.g., meaningfully speak of the correlation between Y and Z, or probability with which they equal each other.

There was considerable conceptual work done to explain that KPT allows for stochastically unrelated random variables, and that the variables in different contexts are defined on different probability spaces [24–26]. This view is one that is supposed to justify the interpretations of the contextuality of the KCBS and KS systems mentioned in the previous section. However, this justification is invalid: KPT must indeed be endowed with multiple domain probability spaces, but still, the five random variables in (2) or the 18 and 15 random variables in the respective matrices of Fig. 1 simply cannot fail to be jointly distributed. The reason for this is that the relation "are jointly distributed" in KPT is transitive (in fact, it is an equivalence relation).

Let us understand this clearly by recapitulating certain basics about random variables. A random variable X is identified by a triple consisting of

1. a domain probability space $(S_{Dom}, \Sigma_{Dom}, \mu_{Dom})$,
2. a codomain measurable space (S_{Cod}, Σ_{Cod}), and
3. a measurable function $f : S_{Dom} \to S_{Cod}$.

Here, S_{Dom}, S_{Cod} are sets, $\Sigma_{Dom}, \Sigma_{Cod}$ are respective sigma algebras of subsets of these sets, μ_{Dom} is a probability measure on Σ_{Dom}, and f is called a measurable function because for any $A \in \Sigma_{Cod}$, its pre-image $f^{-1}(A)$ belongs to Σ_{Dom} and therefore has a probability, $\mu_{Dom}\left(f^{-1}(A)\right)$. This probability is interpreted as the probability with which random variable X falls in set A:

$$\mu_{Dom}\left(f^{-1}(A)\right) = \Pr\left[X \in A\right]. \tag{4}$$

Two random variables

$$X \overset{def}{\equiv} \left(\left(S_{Dom}, \Sigma_{Dom}, \mu_{Dom}\right), \left(S_{Cod}, \Sigma_{Cod}\right), f\right) \tag{5}$$

and

$$X' \overset{def}{\equiv} \left(\left(S'_{Dom}, \Sigma'_{Dom}, \mu'_{Dom}\right), \left(S'_{Cod}, \Sigma'_{Cod}\right), f'\right), \tag{6}$$

are jointly distributed if and only if their domain spaces coincide:

$$\begin{aligned}\left(S_{Dom}, \Sigma_{Dom}, \mu_{Dom}\right) &= \left(S'_{Dom}, \Sigma'_{Dom}, \mu'_{Dom}\right) \\ &= \left(S, \Sigma, \mu\right).\end{aligned} \tag{7}$$

Then, for any $A \in \Sigma_{Cod}$ and any $A' \in \Sigma'_{Cod}$, the intersection of $f^{-1}(A)$ and $f'^{-1}(A')$ belongs to Σ, and the measure μ of this intersection is interpreted as the joint probability of $X \in A$ and $X' \in A'$:

$$\mu\left(f^{-1}(A) \cap f'^{-1}(A')\right) = \Pr\left[X \in A, X' \in A'\right]. \tag{8}$$

This criterion of the existence of a joint distribution is generalized to any set of random variables. In particular, if we add to X and X' a random variable

$$X'' \overset{def}{\equiv} \left(\left(S''_{Dom}, \Sigma''_{Dom}, \mu''_{Dom}\right), \left(S''_{Cod}, \Sigma''_{Cod}\right), f''\right), \tag{9}$$

the three random variables are jointly distributed if and only if

$$\begin{aligned}\left(S_{Dom}, \Sigma_{Dom}, \mu_{Dom}\right) &= \left(S'_{Dom}, \Sigma'_{Dom}, \mu'_{Dom}\right) \\ = \left(S''_{Dom}, \Sigma''_{Dom}, \mu''_{Dom}\right) &= \left(S, \Sigma, \mu\right),\end{aligned} \tag{10}$$

and then, for any $A \in \Sigma_{Cod}$, $A' \in \Sigma'_{Cod}$, and $A'' \in \Sigma''_{Cod}$,

$$\begin{aligned}&\mu\left(f^{-1}(A) \cap f'^{-1}(A') \cap f''^{-1}(A'')\right) \\ &= \Pr\left[X \in A, X' \in A', X'' \in A''\right].\end{aligned} \tag{11}$$

Now, it is easy to see that the existence of a joint distribution for X, X', X'', determined by (10), is implied by the conjunction of the existence of a joint distribution for X, X' with the existence of a joint distribution for X', X''. Indeed, the latter means

$$\begin{aligned}(S'_{Dom}, \Sigma'_{Dom}, \mu'_{Dom}) &= (S''_{Dom}, \Sigma''_{Dom}, \mu''_{Dom}) \\ &= (S', \Sigma', \mu'),\end{aligned} \tag{12}$$

and then (S', Σ', μ') must coincide with (S, Σ, μ) because the equalities in (7) and in (12) share $(S'_{Dom}, \Sigma'_{Dom}, \mu'_{Dom})$.

Applying this simple but fundamental fact to, e.g., (2), we see that R_1 and R_2 in context c_1 are jointly distributed, and so are R_2 and R_3 in context c_2, whence R_1, R_2, and R_3 are jointly distributed. Continuing in this fashion, all five random variables in (2) are jointly distributed.

The same conclusion can be arrived at by using another way of thinking about joint distributions. Two or more random variables are jointly distributed if and only if they can be presented as functions of one and the same random variable. In special cases this has been shown in Refs. [22] and [41], for a general version see Ref. [10]. The "if" part of the statement is true trivially, and the "only if" part follows form the fact that a set of jointly distributed random variables is a random variable whose components are its measurable functions (projections). Applying this view to R_1 and R_2 in context c_1, there is a random variable R_{12} such that $R_1 = f_1(R_{12})$ and $R_2 = f_2(R_{12})$. Analogously, for some R_{23}, $R_2 = g_2(R_{23})$ and $R_3 = g_3(R_{23})$. If R_{12} and R_{23} are stochastically unrelated, then so are $g_2(R_{23})$ and $f_2(R_{12})$, which is impossible as they are both equal to R_2. Hence R_{12} and R_{23} are jointly distributed, i.e., they are functions of some R_{123}. But then so are R_1, R_2, and R_3, and they are jointly distributed. Continuing in this fashion we see that all five random variables in (2) are jointly distributed.

Generalizing, given a system of measurements, consider a graph with contexts as its nodes, such that two contexts are connected by an edge if and only if the two sets of (jointly distributed) random variables corresponding to these contexts are not disjoint. Then, if such a graph contains a path through all the nodes, the system of random variables must have an overall joint distribution. Thus, for the KCBS system in (1), such a path is

$$c_1 \,\rule[0.5ex]{3em}{0.4pt}\, c_2 - c_3 - c_4 \,\rule[0.5ex]{3em}{0.4pt}\, c_5\ . \tag{13}$$

For the KS-4D and KS-3D matrices in Fig. 1, examples of such paths are, respectively,

$$c_1 \,\rule[0.5ex]{3em}{0.4pt}\, c_2 - c_4 - c_3 - c_6 - c_5 - c_9 - c_8 \,\rule[0.5ex]{3em}{0.4pt}\, c_7 \tag{14}$$

and

$$c_2 \,\rule[0.5ex]{3em}{0.4pt}\, c_5 - c_4 - c_1 - c_3 - c_6 - c_7 - c_8 - c_9 \,\rule[0.5ex]{3em}{0.4pt}\, c_{10} \tag{15}$$

Although this is immaterial for the argument, all these paths happen to be Hamiltonian paths (passing through each node only once), and even Hamiltonian cycles.

We arrive at the conclusion that within the confines of KPT the existence of overall joint distributions is guaranteed for all quantum-mechanical systems that are of traditional interest in contextuality studies. At the same time, in all such systems, with an appropriate choice of parameters, the existence of such a joint distribution can be shown to be impossible as it leads to a contradiction. Does this mean that the naive conclusion that KPT is inapplicable to quantum phenomena is correct after all?

I.4. Foundational paradox or reductio ad absurdum with respect to a hidden assumption?

The notion of contextuality is formulated and derivations of tests like (3) are made (or can always be made) entirely in the language of KPT. Abandoning KPT would amount to abandoning the issue of contextuality altogether. The quantum probability theory (or its generalizations, such as c^*-algebra) allows one to describe behavior of quantum systems, but it does not make a qualitative distinction between systems that satisfy conditions like (3) and those that violate them. The computations follow precisely the same rules (based on the angles between the directions into which operators $q_1, \ldots, q_5$ project) whether the left-hand side of (3) turns out to be, say, 2.99 or 3.01. These computations do not answer (and do not address) the question of whether the random variables representing the measurements of $q_1, \ldots, q_5$ in various contexts are jointly distributed.

Most importantly, from a logical point of view, there is no reason for abandoning KPT because of the contradiction arrived at in the previous section. If one introduces three numbers with $x = -y$, $y = -z$, $z = -x$ and derives a contradiction, one does not abandon algebra, one simply considers this a proof that such three numbers do not exist. The denial of an overall joint distribution in KPT is an attempt to follow the same (correct) logic, it is simply mistaken in the choice of the culprit. The true culprit, the assumption that creates a contradiction and has to be dropped by reductio ad absurdum, is Noncontextual Identification. One assumes that measurements of the same property in different contexts are represented by one and the same random variable; then the random variables in different contexts overlap; then, by the transitivity of the relation "are defined on the same probability space", all random variables in the system are jointly distributed; but no overall joint distribution is compatible with certain properties of the random variables within individual contexts; ergo, the assumption that measurements of the same property in different contexts are represented by one and the same random variable is wrong.

We should therefore abandon the assumption of Noncontextual Identification and replace it with the principle of

Contextual Identification:: A random variable R is uniquely identified by the property q it measures and the context c in which it is measured, i.e., random variables in different contexts are different random variables.

This is the starting point of a variant (or implementation) of KPT that we dubbed Contextuality-by-Default (CbD) [12, 14, 15, 17, 19, 20, 30–32]. Applying it to the KCBS example, the matrix (1) should be filled in as

$$\left[\begin{array}{cccccc} \text{(KCBS)} & c_1 & c_2 & c_3 & c_4 & c_5 \\ & & & & & \\ q_1 & R_1^1 & & & & R_1^5 \\ q_2 & R_2^1 & R_2^2 & & & \\ q_3 & & R_3^2 & R_3^3 & & \\ q_4 & & & R_4^3 & R_4^4 & \\ q_5 & & & & R_5^4 & R_5^5 \\ & & & & & \end{array}\right] \tag{16}$$

Here, $R_{q_i}^{c_j}$ (written as R_i^j for simplicity) is a unique (within the given system) identification of the random random variable as measuring q_i in context c_j. We need not ask why R_1^1 is not the same as random variable R_1^5: they are different by definition ("by default").

I.5. How does one redefine the traditional notion of (non)contextuality in CbD?

Representation (16) dissolves the contradiction arrived at in Section I.3: the sets of measurements made in different contexts are necessarily disjoint, and (since we allow for multiple domain probability spaces) it is very well possible that, e.g., $\left(R_1^1, R_2^1\right)$ and $\left(R_2^2, R^2\right)$ are stochastically unrelated to each other. This leads to no contradictions. CbD goes a step further by stipulating that $\left(R_1^1, R_2^1\right)$ and $\left(R_2^2, R^2\right)$ are *necessarily* stochastically unrelated to each other. This is a companion principle to Contextual Identification in CbD:

Stochastic (Un)Relatedness:: Random variables R_q^c and $R_{q'}^{c'}$ are jointly distributed if and only if $c = c'$ (otherwise they are stochastically unrelated).

The reason for this is very simple: while there is an empirical meaning for R_q^c and $R_{q'}^c$ (in the same context) co-occurring, there is no such meaning for R_q^c and $R_{q'}^{c'}$, whether $q = q'$ or not [12, 14, 15, 19, 32]. We can still speak of what joint distributions could be compatible with (imposable on) R_q^c and $R_{q'}^{c'}$, but not of what this joint distribution "is".

Following this observation, contextuality (or lack thereof) should be formulated in terms of joint distributions imposable on the stochastically unrelated random variables. In KPT (and CbD as its variant), "imposition" of a joint distribution on a set of random variables $X, Y, \ldots, Z$ means constructing a probabilistic coupling for them [42]: it is defined as any jointly distributed $(X', Y', \ldots, Z')$ (which can be viewed as a "single" random variable with the components as its functions) such that X' is distributed as X, Y' as Y, ..., Z' as Z. A coupling for the system of measurements (16) is a random variable S defined as

$$S = \left(S_i^j : j = 1, \ldots, 5, i = 1, \ldots, 5, q_i \text{ measured in context } c_j\right), \tag{17}$$

with $\left(S_i^j, S_{i'}^j\right)$ distributed as $\left(R_i^j, R_{i'}^j\right)$ for any q_i and $q_{i'}$ measured in the same context c_j. See Refs. [14, 15, 19, 32] for a more detailed discussion of couplings.

A seeming difficulty arising here is that couplings for systems like (16), with disjoint sets of measurements in different contexts, always exist. This means that formulation S1 is no longer appropriate as a basis for definition of (non)contextual systems. In CbD therefore this formulation is modified: the determination to be made in contextuality analysis is

> **(S2)** existence or nonexistence, for the set of all measurements in the system, of a coupling in which measurements of one and the same property in different contexts are stochastically related to each other in a specified way, $\mathcal{C}$.

(Strictly speaking, we have to speak of "sub-couplings corresponding to sets of measurements of the same property," but we adopt a simpler language here, hoping not to cause confusion.) Different meanings of $\mathcal{C}$ correspond to different meanings of contextuality. The traditional meaning is obtained by choosing

$$\mathcal{C} = \text{"are equal with probability 1"}. \tag{18}$$

The definition of contextuality that corresponds to this meaning is as follows.

Definition I.1 (Traditional meaning of contextuality, in CbD language)**.** A system of measurements is noncontextual if it has a coupling in which any measurements of one and the same property in different contexts are equal to each other with probability 1. If such a coupling does not exist, the system is contextual.

In other words, the system of measurements

$$R = \left\{R_j^i : \text{all } (i,j) \text{ such that } q_i \text{ is measured in context } c_j\right\} \tag{19}$$

(which is not a random variable because its components are not jointly distributed) is noncontextual if and only if one can find for it a coupling

$$S = \left(S_j^i : \text{all } (i,j) \text{ such that } q_i \text{ is measured in context } c_j\right) \tag{20}$$

(which is a random variable) in which, for any q_i and any set $c_{j_1}, \ldots, c_{j_k}$ of contexts in which q_i is measured,

$$\Pr\left[S_i^{j_1} = S_i^{j_2} = \ldots = S_i^{j_k}\right] = 1. \tag{21}$$

Clearly, this property is equivalent to its more restrictive form: for any q_i and any sequence of contexts in which q_i is measured,

$$\Pr\left[S_i^j = S_i^{j'}\right] = 1 \tag{22}$$

for any two neighboring contexts in the sequence.

II. CONTEXTUALITY IN ARBITRARY SYSTEMS OF BINARY MEASUREMENTS

II.1. Consistently and inconsistently connected systems

Definition I.1 is sufficient in the idealized quantum scenarios where any two random variables measuring the same property are identically distributed. We call systems of measurements with this property consistently connected, and the systems in which consistent connectedness is violated we call inconsistently connected. For instance, the EPR-BB paradigm [4, 5, 7], on denoting Alice's axis choices q_1 and q_3 and Bob's choices q_2 and q_4, is formally represented by the matrix

$$\begin{array}{c|cccc} \text{(EPR-BB)} & c_1 & c_2 & c_3 & c_4 \\ \hline q_1 & R_1^1 & & & R_1^4 \\ q_2 & R_2^1 & R_2^2 & & \\ q_3 & & R_3^2 & R_3^3 & \\ q_4 & & & R_4^3 & R_4^4 \end{array} \tag{23}$$

The identical distribution of, say, R_1^1 (Alice's spin measurement along axis q_1 when Bob's choice of axis is q_2) and R_1^4 (Alice's measurement along axis q_1 when Bob's choice of axis is q_4) can be ensured by space-like separation of the two particles. Even then, however, context-dependent biases in experimental set-up and in recording of the simultaneous measurements are possible, so one should still speak of an idealization, even if a safe one.

The Suppes-Zanotti-Leggett-Garg paradigm [23, 41] is formally represented by the matrix

$$\begin{array}{c|ccc} \text{(SZLG)} & c_1 & c_2 & c_3 \\ \hline q_1 & R_1^1 & & R_1^3 \\ q_2 & R_2^1 & R_2^2 & \\ q_3 & & R_3^2 & R_3^3 \end{array} \tag{24}$$

The q_1, q_2, q_3 here are measurements of an observable at three points of time, assumed to be performable in pairs only. The assumption of consistent connectedness here is justified by Leggett and Garg [23] by invoking "noninvasiveness" argument, but the existence of "signaling in time" [2, 3, 29] cannot be eliminated except when the system is in a perfectly balanced mixed state.

Returning to the KCBS paradigm represented by (16), consistent connectedness here, as well as in other KS-type paradigms, can be justified by invoking the "no-disturbance" principle [34, 39]. A factual experiment, however [35], shows that non-negligible violations of consistent connectedness do occur [32, 36].

If one adds to this discussion the desirability of using the notion of contextuality outside physics [1, 18, 20], where inconsistent connectedness is essentially a universal rule, it becomes clear that Definition I.1 is too limited in scope. Indeed, any amount of inconsistent connectedness, however small, makes a system contextual by this definition. This would mean, among other things, that contextuality is ubiquitous in classical physics, as it is easy to construct classical systems whose inconsistent connectedness is caused by direct influence upon a measurement of a property by other properties measured in the same context. See Refs. [20, 21] for examples.

II.2. Multimaximality constraint for binary measurements

As mentioned in Section I.5, different choices of the constrain $\mathcal{C}$ in formulation S2 correspond to different meanings of contextuality. To generalize Definition I.1 to include inconsistently connected systems, we have to generalize the meaning of $\mathcal{C}$ in (18). The generalization proposed in CbD is

$$\mathcal{C} = \text{"are equal with maximal possible probability".} \tag{25}$$

Definition II.1 (Contextuality in arbitrary systems with binary measurements)**.** A system of binary measurements is noncontextual if it has a coupling in which, given any set of measurements of one and the same property in different contexts, they are equal to each other with the maximal possible probability. If such a coupling does not exist, the system is contextual.

In other words, the system (19) is noncontextual if and only if one can find for it a coupling (20) in which, for any q_i and any set $c_{j_1}, \ldots, c_{j_k}$ of contexts in which q_i is measured, the value of

$$\Pr\left[S_i^{j_1} = S_i^{j_2} = \ldots = S_i^{j_k}\right] \tag{26}$$

is maximal among all possible couplings of the system.

Remark. This use of the constraint $\mathcal{C}$ was recently proposed in Ref. [16]. In the earlier version of CbD this constraint was only applied to the set of *all* measurements of one and the same property rather than to *any* set of such measurements.

Following the standard terminology of CbD, let us use the term "connection" to refer to the set of all measurements of a given property,

$$R_q = \left\{R_q^c : q \text{ is measured in context } c\right\}. \tag{27}$$

Note that R_q is not a random variable as its components are not jointly distributed. These components, however, always have the same set of possible values identically related to the empirical measurement procedure. For instance, if $R_q^c = +1$ and $R_q^c = -1$ mean, respectively "spin-up" and "spin-down" for a particular axis in a spin-1/2 particle, then all random variables $R_q^{c'}$ in the same connection (with subscript q) have to have the same values with the same interpretation. For another q (say, another axis) the values of the measurements in a connection may very well be denoted differently, e.g., $+1$ and -1 may have their meanings exchanged.

Each connection R_q can be taken separately and viewed as a system of measurement in its own right, with a single measured property. A coupling

$$S_q = \left(S_q^c : q \text{ is measured in context } c\right) \tag{28}$$

for R_q satisfying Definition II.1 is called a multimaximal coupling (as it is a maximal coupling [42] for any subset of the connection).

II.3. Existence and uniqueness of multimaximal couplings for binary measurements

Contextuality analysis is greatly simplified by the following result, proved in Ref. [16].

Theorem II.2 (Ref. [16])**.** *For a connection R_q consisting of $k > 1$ binary measurements a multimaximal coupling exists and is unique. If one denotes the measurement outcomes $+1/-1$ and arranges the contexts in the connection so that*

$$p_1 = \Pr\left[R_q^1 = 1\right] \leq \ldots \leq \Pr\left[R_q^k = 1\right] = p_k, \tag{29}$$

then the multimaximal coupling S_q is defined by

$$\Pr\left[S_q^1 = \ldots = S_q^k = 1\right] = p_1,$$

$$\Pr\left[S_q^1 = \ldots = S_q^l = -1\ ;\ S_q^{l+1} = \ldots = S_q^k = 1\right] = p_{l+1} - p_l, \quad (\text{for } l = 1, \ldots, k-1)$$

$$\Pr\left[S_q^1 = \ldots = S_q^k = 1\right] = 1 - p_k, \tag{30}$$

with all other combinations of values having probability zero.

In this theorem, for any $1 \leq l < m \leq k$, $\left(S_q^l, S_q^m\right)$ is a maximal (hence also multimaximal) coupling for $\left(R_q^l, R_q^m\right)$, i.e.,

$$\Pr\left[S_q^l = 1, S_q^m = 1\right] = p_l. \tag{31}$$

Note that we continue to assume ordering (29). Let $\left(S_q^1, \ldots, S_q^k\right)$ be a coupling of $R_q^1, \ldots, R_q^k$ such that $\left(S_q^l, S_q^{l+1}\right)$ is the (multi)maximal coupling for R_q^l, R_q^{l+1}, for any $l = 1, \ldots, k-1$. Such a coupling exists, because the multimaximal coupling of $R_q^1, \ldots, R_q^k$ is one such coupling.

In fact, as it turns out, it is the only such coupling. Indeed, it follows from (31) that, for any $l < m$,

$$\Pr\left[S_q^l = 1, S_q^m = -1\right] = 0, \tag{32}$$

This, it turn, implies, for any $l < l'$,

$$\Pr\left[S_q^l = S_q^{l'} = 1 \text{ and, for some } m > l, S^m = -1\right] = 0. \tag{33}$$

Then, for any $l = 1, \ldots, k-1$,

$$\Pr\left[S_q^l = S^{l+1} = \ldots = S_q^k = 1\right] = \Pr\left[S_q^l = 1, S_q^{l+1} = 1\right] = p_l, \tag{34}$$

and for $l = k$,

$$\Pr\left[S_q^k = 1\right] = p_k,$$

whence (30) follows by straightforward algebra.

This establishes

Theorem II.3. *S_q is the multimaximal coupling for a connection R_q consisting of $k > 1$ binary measurements arranged as in (29) if and only if the subcouplings $\left(S_q^l, S_q^{l+1}\right)$ are (multi)maximal couplings for pairs R_q^l, R_q^{l+1}.*

In other words, in Definition II.1, the requirement that (26) be maximal for any set of contexts, can be replaced with the requirement that, for any q_i the value of

$$\Pr\left[S_i^j = S_i^{j'}\right] \tag{35}$$

be maximal for all consecutive pairs of contexts c_j and $c_{j'}$ in the ordering (29) for all contexts in which q_i is measured. This parallels (except there we can choose the ordering arbitrarily) the possibility of replacing (21) with (22) in the traditional definition of contextuality.

II.4. Properties of systems of binary measurements

The following propositions follow trivially from Definition II.1 and Theorem II.2.

(1) A noncontextual system remains noncontextual if one deletes some of its components (measurements).

(2) For a consistently connected system, Definition II.1 specializes to Definition I.1.

(3) For the important class of cyclic systems of binary measurements the theory specializes to one published in Refs. [9, 15, 17, 19, 31, 32]. A cyclic system of rank $n > 1$ is representable by a graph

$$\begin{tikzcd} q_1 \arrow[r, no head] & q_2 - \cdots - q_{n-1} \arrow[r, no head] & q_n \arrow[ll, bend left, no head] \end{tikzcd}, \tag{36}$$

in which two properties are connected if and only if they are measured in the same context. Of the systems mentioned earlier, KCBS (16), EPR-BB (23), and SZLG (24) are cyclic (of ranks 5, 4, and 3, respectively), while the two KS systems in Fig. 1 are not cyclic.

Theorem II.4 (Ref. [31]). *A cyclic system of rank $n > 1$ with binary $(+1/-1)$ measurements is noncontextual if and only if*

$$\max_{(\iota_1,\ldots,\iota_k)\in\{-1,1\}^n:\prod_{i=1}^n \iota_i=-1} \sum_{i=1}^n \iota_i \left\langle R_i^i R_{i\oplus 1}^i \right\rangle \leq n-2+\sum_{i=1}^n \left|\left\langle R_i^i\right\rangle - \left\langle R_i^{i\ominus 1}\right\rangle\right|, \tag{37}$$

where

$$i\oplus 1 = \begin{cases} i+1 & \text{if } 1\leq i<n \\ 1 & \text{if } i=n \end{cases}, i\ominus 1 = \begin{cases} i-1 & \text{if } 1< i\leq n \\ n & \text{if } i=1 \end{cases}. \tag{38}$$

In the left-hand side of the inequality, each expected product is taken with plus or minus, keeping the number of the minuses odd, and the largest of all such linear combinations is compared to the right-hand side. The latter reduces to $n-2$ in the case of consistent connectedness. Inequality (3) used above as an example is obtained by picking one of the linear combinations for $n = 5$ and assuming consistent connectedness.

III. SUMMARIZING WITH MAGIC BOXES

Contextuality was originally introduced by Specker in Ref. [40] in the form of a parable whose gist is as follows: there are three boxes, q_a, q_b, q_c, each of which may or may not contain a gem; they can only be opened two at a time; and some magic ensures that when opening any two of them, one and only one of them contains a gem.

A usual way of conceptualizing this situation would be to denote by A the random variable "contents of box q_a", by B the random variable "contents of box q_b", and define C analogously. Let presence of a gem be encoded by $+1$ and absence by -1. Then the magic of the boxes translates into statements

$$\begin{aligned} &\Pr[A=-B]=1,\\ &\Pr[-B=C]=1,\\ &\Pr[C=-A]=1. \end{aligned} \tag{39}$$

The traditional reasoning then proceeds as follows: in any joint distribution of A, B, C, these three statements should be satisfied jointly; since this is readily seen leading to a contradiction, $\Pr[A=-A]=1$, the "assumption" that there is a joint distribution of A, B, C should be rejected (and the system declared contextual).

However, as we argued in this paper, this reasoning is incorrect: if A and B are jointly distributed (and they are, because otherwise $\Pr[A=-B]$ is undefined), and if so are B and C (or A and C), then A, B, C must be jointly distributed, by the definition of jointly distributed random variables in probability theory. The correct conclusion therefore is not that the "assumption" of the overall joint distribution is wrong, but rather that A, B, C satisfying (39) do not exist — in precisely the same meaning in which one would say that there are no three numbers x, y, z such that $x=-y=z=-x$. Assuming that the three magic boxes is our empirical reality, A, B, C as defined above do not provide a possible description thereof. One has to conclude that the random variable describing "contents of box q_a" in the context of opening q_a and q_b is not the same as the random variable describing "contents of box q_a" in the context of opening q_a and q_c. They are two distinct random variables, "by default" (i.e., even before we know anything of their correlations with other random variables), because they are recorded under mutually incompatible conditions. And for the same reason, they possess no empirically defined joint distribution.

A noncontradictory description of the three magic boxes we arrive at involves six random variables,

$$\begin{bmatrix} \text{(magic)} & c_{ab} & c_{bc} & c_{ca} \\ q_a & R_a^{ab} & & R_a^{ca} \\ q_b & R_b^{ab} & R_b^{bc} & \\ q_c & & R_c^{bc} & R_c^{ca} \end{bmatrix}, \tag{40}$$

forming a special case of the cyclic SZLG system in (24). If the magic system is consistently connected, i.e., if

$$\begin{aligned} &\Pr\left[R_a^{ab}=1\right]=\Pr\left[R_a^{ca}=1\right],\\ &\Pr\left[R_b^{ab}=1\right]=\Pr\left[R_b^{bc}=1\right],\\ &\Pr\left[R_c^{bc}=1\right]=\Pr\left[R_c^{ca}=1\right], \end{aligned} \tag{41}$$

then the system is indeed contextual by the criterion in Theorem II.4.

However, and this is one of the advantages offered by CbD, once we have correctly identified the six random variables in (40), there is no special reason why one should assume consistent connectedness. One can consider all possible distributions satisfying (39) and find out that the system is noncontextual if and only if

$$\left|\left\langle R_a^{ab}\right\rangle - \left\langle R_a^{ca}\right\rangle\right| + \left|\left\langle R_b^{ab}\right\rangle - \left\langle R_b^{bc}\right\rangle\right| + \left|\left\langle R_c^{bc}\right\rangle - \left\langle R_c^{ca}\right\rangle\right| \geq 2. \tag{42}$$

In particular, the system is noncontextual if for at least one pair of boxes the gem always appears in a particular one of them. If this is true for all three pairs of boxes, the system is deterministic, and any deterministic system is noncontextual [20]. And so on, one can proceed investigating this magic system from a variety of angles.

Contextual analysis based on CbD is not only mathematically more rigorous than the common understanding, it also appears more interesting.

Acknowledgments.

This research has been supported by NSF grant SES-1155956 and AFOSR grant FA9550-14-1-0318.

[1] Asano, M., Hashimoto, T., Khrennikov, A.Yu., Ohya, M., Tanaka, T. (2014). Violation of contextual generalization of the Leggett-Garg inequality for recognition of ambiguous figures. Physica Scripta T 163:014006.
[2] Bacciagaluppi, G. (2015). Leggett-Garg inequalities, pilot waves and contextuality. International Journal of Quantum Foundations 1, 1-17.
[3] Bacciagaluppi, G. (2016). Einsten, Bohm, and Leggett-Garg. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology, pp. 63-76. New Jersey: World Scientific.
[4] Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. Physics 1: 195-200.
[5] Bell, J. (1966). On the problem of hidden variables in quantum mechanics. Review of Modern Physics 38, 447-453.
[6] Cabello, A., Estebaranz, J. M., & Alcaine, G. G. (1996). Bell-Kochen-Specker theorem: A proof with 18 vectors. Physics Letters A 212:183.
[7] Clauser, J.F., Horne, M.A., Shimony, A., & Holt, R.A. (1969). Proposed experiment to test local hidden-variable theories. Physical Review Letters 23:880–884.
[8] de Barros, J.A., Oas, G. (2014). Negative probabilities and counter-factual reasoning in quantum cognition. Physica Scripta T163:014008.
[9] de Barros, J.A., Dzhafarov, E.N., Kujala, J.V., Oas, G. (2015). Measuring Observable Quantum Contextuality. Lecture Notes in Computer Science 9535, 36-47.
[10] Dzhafarov, E.N., & Kujala, J.V. (2010). The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality Frontiers in Quantitative Psychology and Measurement 1:151 doi:10.3389/fpsyg.2010.00151.
[11] Dzhafarov, E.N., & Kujala, J.V. (2014). A qualified Kolmogorovian account of probabilistic contextuality. Lecture Notes in Computer Science 8369, 201-212.
[12] Dzhafarov, E.N., & Kujala, J.V. (2014). Contextuality is about identity of random variables. Physica Scripta T163, 014009.
[13] Dzhafarov, E.N., & Kujala, J.V. (2014). Embedding quantum into classical: contextualization vs conditionalization PLoS ONE 9(3):e92818
[14] Dzhafarov, E.N., & Kujala, J.V. (2016). Conversations on contextuality. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology, pp. 1-22. New Jersey: World Scientific.
[15] Dzhafarov, E.N., & Kujala, J.V. (2016). Context-content systems of random variables: The Contextuality-by-Default theory. To appear in Journal of Mathematical Psychology [arXiv:1511.03516].
[16] Dzhafarov, E.N., & Kujala, J.V. (2016). Contextuality by Default 2.0: Systems with binary random variables. To appear in Lecture Notes in Computer Science [arXiv:1604.04799].
[17] Dzhafarov, E.N., Kujala, J.V., Cervantes, V.H. (2016). Contextuality-by-Default: A brief overview of ideas, concepts, and terminology. Lecture Notes in Computer Science 9535, 12-23.
[18] Dzhafarov, E.N., Zhang, & R., Kujala, J.V. (2015). Is there contextuality in behavioral and social systems? Philosophical Transactions of the Royal Society A 374: 20150099.
[19] Dzhafarov, E.N., Kujala, J.V., & Larsson, J.-A. (2105). Contextuality in three types of quantum-mechanical systems. Foundations of Physics 7, 762-782.
[20] Dzhafarov, E.N., Kujala, J.V., Cervantes, V.H., Zhang, R., & Jones, M. (2016). On contextuality in behavioral data. Philosophical Transactions of the Royal Society A 374: 20150234.
[21] Filk, T. (2016). It is the theory which decides what we can observe. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology, pp. 77-92. New Jersey: World Scientific.
[22] Fine, A. (1982). Hidden variables, joint probability, and the Bell inequalities. Physical Review Letters 48: 291–295.
[23] Leggett, A.J., & Garg A. (1985). Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? Physical Review Letters 54: 857–860.
[24] Khrennikov, A. Yu. (2008). Bell–Bool einequality: Nonlocality or probabilistic incompatibility of random variables? Entropy 10: 19–32.
[25] Khrennikov, A. Yu. (2008). EPR–Bohm experiment and Bell's inequality: Quantum physics meets probability theory. Theoretical and Mathematical Physics 157: 1448–1460.
[26] Khrennikov, A. Yu. (2009). Contextual approach to quantum formalism. In:Fundamental Theories of Physics, vol. 160. Springer, Dordrecht.
[27] Klyachko, A.A., Can, M.A., Binicioglu, S., & Shumovsky,A.S. (2008). A simple test for hidden variables in spin-1 system. Physical Review Letters 101:020403.
[28] Kochen, S., & Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. Journal of Mathematics and Mechanics, 17:59–87.
[29] Kofler, J., & Brukner, C. (2013). Condition for macroscopic realism beyond the Leggett-Garg inequalities, Physical Review A 87:052115.
[30] Kujala, J.V., & Dzhafarov, E.N. (2016). Probabilistic Contextuality in EPR/Bohm-type systems with signaling allowed. In E.N. Dzhafarov, S. Jordan, R. Zhang, V. Cervantes (Eds). Contextuality from Quantum Physics to Psychology, pp. 287-308. New Jersey: World Scientific.
[31] Kujala, J.V., Dzhafarov, E.N. (2016). Proof of a conjecture on contextuality in cyclic systems with binary variables. Foundations of Physics, 46: 282-299.
[32] Kujala, J.V. , & Dzhafarov, E.N., & Larsson, J.-A. (2015). Necessary and sufficient conditions for maximal contextuality in a broad class of quantum mechanical systems. Physical Review Letters 115:150401.
[33] Kurzynski, P., Ramanathan, R., & Kaszlikowski, D. (2012). Entropic test of quantum contextuality. Physical Review Letters 109:020404.
[34] Kurzynski, P., Cabello, A., & Kaszlikowski, D. (2014). Fundamental monogamy relation between contextuality and nonlocality. Physical Review Letters 112:100401.
[35] Lapkiewicz, R., Li, P., Schaeff, C., Langford, N.K., Ramelow, S., Wiesniak, M., & Zeilinger, A. (2011). Experimental non-classicality of an indivisible quantum system. Nature 474: 490.

[36] (2013). Lapkiewicz, R., Li, P., Schaeff, C., Langford, N.K., Ramelow, S., Wiesniak, M., & Zeilinger, A.Comment on "Two fundamental experimental tests of nonclassicality with qutrits". arXiv:1305.5529.

[37] Liang, Y.-C., Spekkens, R. W., Wiseman, H. M. (2011). Specker's parable of the overprotective seer: A road to contextuality, nonlocality and complementarity. Physics Reports 506, 1-39.

[38] Peres, A. (1995). Quantum Theory: Concepts and Methods. Dordrecht: Kluwer.

[39] Ramanathan, R., Soeda, A., Kurzynski, P., & Kaszlikowski, D. (2012). Generalized monogamy of contextual inequalities from the no-disturbance principle. Physical Review Letters 109:050404.

[40] Specker, E. (1960). Die Logik Nicht Gleichzeitig Entscheidbarer Aussagen. Dialectica 14: 239– 246 (English translation by M.P. Seevinck available as arXiv:1103.4537.)

[41] Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? Synthese 48:191–199.

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