The Contextuality-by-Default View of the Sheaf-Theoretic Approach to Contextuality

Ehtibar N. Dzhafarov

Purdue University, ehtibar@purdue.edu

Abstract

The Sheaf-Theoretic Contextuality (STC) theory developed by Abramsky and colleagues is a very general account of whether multiply overlapping subsets of a set, each of which is endowed with certain "local" structure, can be viewed as inheriting this structure from a global structure imposed on the entire set. A fundamental requirement of STC is that any intersection of subsets inherit one and the same structure from all intersecting subsets. I show that when STC is applied to systems of random variables, it can be recast in the language of the Contextuality-by-Default (CbD) theory, and this allows one to extend STC to arbitrary systems, in which the requirement in question (called "consistent connectedness" in CbD) is not necessarily satisfied. When applied to possibilistic systems, such as systems of logical statements with unknown truth values, the problem arises of distinguishing lack of consistent connectedness from contextuality. I show that it can be resolved by considering systems with multiple possible deterministic realizations as quasi-probabilistic systems with epistemic (or Bayesian) probabilities assigned to the realizations. Although STC and CbD have distinct native languages and distinct aims and means, the conceptual modifications presented in this paper seem to make them essentially coextensive.

KEYWORDS: contextual fraction, contextuality, consistent connectedness, dichotomization, inconsistent connectedness, measures of contextuality.

1 Introduction

1.1. Contextuality-by-Default (CbD) and Sheaf-Theoretic Contextuality (STC) are two general approaches to establishing and measuring (non)contextuality of systems of measurements. The word "measurement" is understood very broadly, including various relations between inputs and outputs in a physical entity, between databases and records, between logical statements and their truth values, etc. The two theories employ distinct mathematical languages, and were designed with different aims in mind. With respect to their applicability areas, when dealing with probabilisitic scenarios, STC is confined to special systems, called "strongly consistently connected" in CbD. By contrast, CbD was designed with the primary purpose to apply to arbitrary probabilistic systems. On the other hand, CbD is confined to probabilistic systems, with all deterministic systems being trivially noncontextual. By contrast, STC offers contextuality analysis of systems that are inherently deterministic, finding there interesting cases of contextual systems. These comparative characterizations are incomplete, and I will elaborate them as we proceed.

1.2. The two theories are represented by numerous publications, and their representative exposition can be found in Refs. [1–5] for STC and Refs. [10, 11, 14, 20, 22] for CbD. Familiarity with the two theories would be helpful in reading this paper, but all relevant results and concepts, including those mentioned in this introduction, will be explained below. The reader is to be warned, however, that I will not endeavor here what might be a worthy project for future work: to systematically present the two theories using their native languages and establish correspondences and differences between them. Rather, as the title of the paper suggests, STC will only be presented here from the point of view of CbD, in the CbD language.

1.3. I will show that by doing this one can easily extend STC, including the important notion of (non)contextual fraction, to apply to inconsistently connected systems. More precisely, the CbD language allows one to redefine inconsistently connected systems into consistently connected ones, so that STC can apply to them.

1.4. I will argue that there is a conceptual problem in applying original STC, with its commitment to strong consistent connectedness, to inherently deterministic systems, such as systems of statements with definitive truth values. A way of dealing with this issue I propose is to distinguish completely specified deterministic systems (that are always noncontextual) and systems with multiple possible deterministic realizations. In the latter case, by assigning Bayesian priors to these realizations one renders such systems quasi-probabilistic, disentangling thereby contextuality and inconsistent connectedness. This construction allows one to extend CbD to the contextuality analysis of deterministic systems in the spirit of STC.

2 CbD: Conceptual set-up

2.1. In CbD, the object of contextuality analysis is a *system of random variables* representing what generically can be called measurements. Depending on application, the random variables may describe outputs of physical measurements, responses to inputs, answers to questions, etc. The random variables in a system are assumed to be *dichotomous* (say, +1/-1), for reasons discussed in Sections 5.5 and 5.6. This means that any measurement is presented as simultaneous answers to a set of Yes/No questions, such as "Is the measured value less than 5?".

2.2. The question a random variable answers is referred to as the *content* of the random variable. A set of random variables forms a system if they are labeled both by their contents and by their *contexts*. A context of a variable includes all conditions recorded *together* with this variable, where "together" can mean any empirical procedure by which the observed values of the random variables and conditions within the context are paired. For instance, if a random variable R is recorded together with two other random variables, this fact is part of the context of R. If the order of recording these random variables is itself systematically recorded, this is part of the context of R too.

2.3. The terminology is: if content q is measured in context c, which is written

$$q \prec c,$$

the outcome of the measurement is the (dichotomous) random variable R_q^c.

2.4. If the sets of the contents and contexts are finite (*as we are going to assume throughout this paper*), the system can be represented by a matrix like this:

$$\begin{array}{|c|c|c|c||c|}
\hline
R_1^1 & R_2^1 & & & c^1 \\
\hline
 & R_2^2 & R_3^2 & R_4^2 & c^2 \\
\hline
R_1^3 & & R_3^3 & & c^3 \\
\hline
R_1^4 & & & R_4^4 & c^4 \\
\hline
R_1^5 & R_2^5 & R_3^5 & & c^5 \\
\hline
\hline
q_1 & q_2 & q_3 & q_4 & \mathcal{R} \\
\hline
\end{array} \,. \qquad (1)$$

This system has four contents variously measured in five contexts, and R_i^j is the abbreviation for $R_{q=q_i}^{c=c^j}$. I will use this system as an example throughout the paper.

2.5. The following are the two basic properties of any system.

CbD1: All random variables sharing a context are jointly distributed (i.e., they can be presented as measurable functions on one and the same probability space).

CbD2: Any two random variables in different contexts are stochastically unrelated, i.e., they are defined on distinct probability spaces.

2.6. If any two random variables in the system that have the same content are identically distributed, writing this as

$$R_q^c \sim R_q^{c'},$$

the system is called (*simply*) *consistently connected*. CbD does not assume this property: generally, a system can be *inconsistently connected* (and this term is often used in the meaning of "not necessarily consistently connected").

2.7. The following is the main definition in CbD.

Definition 1. *A system is noncontextual if it has a multimaximally connected coupling. Otherwise it is contextual.*

The terminology in the definition is deciphered in the next three sections (2.8-2.10)

2.8. Let

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

be a system (with Q and C being sets of contents and contexts, respectively). A *coupling* of this system is a correspondingly labeled set of random variables

$$\mathcal{S} = \left\{ S_q^c : c \in C, q \in Q, q \prec c \right\}, \tag{3}$$

such that

(a) it is jointly distributed;

(b) its context-wise marginals are distributed in the same way as the corresponding subsets of the original system.

That is, for any $c \in C$,

$$\left\{ R_q^c : q \in Q, q \prec c \right\} \sim \left\{ S_q^c : q \in Q, q \prec c \right\}. \tag{4}$$

2.9. For instance, the jointly distributed random variables

$$
\begin{array}{|c|c|c|c||c|}
\hline
S_1^1 & S_2^1 & & & c^1 \\
\hline
 & S_2^2 & S_3^2 & S_4^2 & c^2 \\
\hline
S_1^3 & & S_3^3 & & c^3 \\
\hline
S_1^4 & & & S_4^4 & c^4 \\
\hline
S_1^5 & S_2^5 & S_3^5 & & c^5 \\
\hline
\hline
q_1 & q_2 & q_3 & q_4 & \mathcal{S} \\
\hline
\end{array} \tag{5}
$$

form a coupling of system $\mathcal{R}$ in (1) if the corresponding row-wise distributions in $\mathcal{S}$ and $\mathcal{R}$ coincide.

2.10. Now, the coupling $\mathcal{S}$ is *multimaximally connected* if, for any two content-sharing random variables $S_q^c, S_q^{c'}$ it contains, the probability of $S_q^c = S_q^{c'}$ is maximal among all possible couplings of $\mathcal{R}$. Equivalently, the probability of $S_q^c = S_q^{c'}$ in a multimaximal coupling is maximal given the marginal distributions of $S_q^c \sim R_q^c$ and $S_q^{c'} \sim R_q^{c'}$. For system (1) with coupling (5), this means the simultaneous maximization of the probabilities of

$$S_1^1 = S_1^3, S_1^1 = S_1^4, S_1^3 = S_1^4, S_3^2 = S_3^3, \ldots$$

2.11. Two couplings of the same system that have the same distribution are considered equivalent and are not distinguished. In other words, the domain probability space of a coupling with a given distribution can be chosen arbitrarily. The most economic choice of the domain space is the distribution itself, with all random variables being defined as the componentwise projections of the identity function on this space. That is, (3) can be viewed as the identity function on the probability space

$$(X_S, \Sigma_S, \mu_S) = \left\{ \{-1, 1\}^{|\prec|}, 2^{\{-1,1\}^{|\prec|}}, \mu_S : 2^{\{-1,1\}^{|\prec|}} \to [0, 1] \right\}, \tag{6}$$

where $|\prec|$ is the cardinality of the relation $\prec$, and μ_S is the probability measure defined by the (joint) probability mass function

$$\Pr\left[S_q^c = s_q^c : c \in C, q \in Q, q \prec c\right],$$

with $s_q^c = -1, 1$. For uncountably infinite Q and/or C the definition should be modified in well-known ways, but we have agreed to consider finite Q, C only.

2.12. Note that any jointly distributed set of random variables is a random variable. So the set of all variables within context c can be written as

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

and a coupling $\mathcal{S}$ in (3) can be written as a random variable S.

3 STC in the CbD Language

3.1. The language of STC is very different, and it manages to avoid even mentioning random variables. Thus, our example system (1) could be represented in STC as

<table>
<tr><td colspan="2">$p^1_{\bar{1}\bar{1}}$</td><td colspan="2">$p^1_{\bar{1}1}$</td><td colspan="2">$p^1_{1\bar{1}}$</td><td colspan="2">p^1_{11}</td><td>(q_1, q_2)</td></tr>
<tr><td>$p^2_{\bar{1}\bar{1}\bar{1}}$</td><td>$p^2_{\bar{1}\bar{1}1}$</td><td>$p^2_{\bar{1}1\bar{1}}$</td><td>$p^2_{\bar{1}11}$</td><td>$p^2_{1\bar{1}\bar{1}}$</td><td>$p^2_{1\bar{1}1}$</td><td>$p^2_{11\bar{1}}$</td><td>$p^2_{111}$</td><td>$(q_2, q_3, q_4)$</td></tr>
<tr><td colspan="2">$p^3_{\bar{1}\bar{1}}$</td><td colspan="2">$p^3_{\bar{1}1}$</td><td colspan="2">$p^3_{1\bar{1}}$</td><td colspan="2">p^3_{11}</td><td>(q_1, q_3)</td></tr>
<tr><td colspan="2">$p^4_{\bar{1}\bar{1}}$</td><td colspan="2">$p^4_{\bar{1}1}$</td><td colspan="2">$p^4_{1\bar{1}}$</td><td colspan="2">p^4_{11}</td><td>(q_1, q_4)</td></tr>
<tr><td>$p^5_{\bar{1}\bar{1}\bar{1}}$</td><td>$p^5_{\bar{1}\bar{1}1}$</td><td>$p^5_{\bar{1}1\bar{1}}$</td><td>$p^5_{\bar{1}11}$</td><td>$p^5_{1\bar{1}\bar{1}}$</td><td>$p^5_{1\bar{1}1}$</td><td>$p^5_{11\bar{1}}$</td><td>$p^5_{111}$</td><td>$(q_1, q_2, q_3)$</td></tr>
<tr><td colspan="2">q_1</td><td colspan="2">q_2</td><td colspan="2">q_3</td><td colspan="2">q_4</td><td>$\mathcal{S}$</td></tr>
</table>

, (7)

where p^i is the probability distribution in the ith context, and the subscripts represent combinations of values 1 and $-1 \equiv \bar{1}$. However, logically, the values $(-1, -1)$ in $p^1_{\bar{1}\bar{1}}$ are not values of the contents (q_1, q_2) (called in STC "measurements", "observables", or simply "variables"). They are values of the random variables not being mentioned.

3.2. The reason this does not lead to complications is that STC is committed to requiring that a system of random variable amenable to contextuality analysis be *strongly consistently connected.* This means that for any pair of contexts c, c', we have

$$\left\{R_q^c : q \prec c, q \prec c'\right\} \sim \left\{R_q^{c'} : q \prec c, q \prec c'\right\}, \tag{8}$$

i.e., the joint distributions for identically subscripted random variables are identical. Thus, in our example (1),

$$\left\{R_1^5, R_2^5\right\} \sim \left\{R_1^1, R_2^1\right\}$$

and

$$\left\{R_2^5, R_3^5\right\} \sim \left\{R_2^2, R_3^2\right\}.$$

Abramsky and colleagues consider this property fundamental (see, e.g., Ref. [5]), and it is indeed indispensable if one is to use the language of sheafs.

3.3. It should be noted, however, that in quantum-physical experiments even simple consistent connectedness (which is obviously implied by strong one) is routinely violated (see, e.g., Ref. [21], and for more references, Ref. [9]). In some non-physical applications, notably in human behavior, inconsistent connectedness is a universal rule [8, 15].

3.4. Note that random variables may, in particular, be deterministic, i.e. they may attain a single value with probability 1. If the requirement of consistent connectedness is applied to such variables, then it translates into any two content-sharing variables being equal to one and the same value,

$$R_q^c \equiv r \Longleftrightarrow R_{q'}^c \equiv r. \tag{9}$$

We will see later, in Section 6, that this constraint creates a difficulty when STC deals with deterministic systems.

3.5. The property of strong consistent connectedness allows Abramsky and colleagues to define a context simply by the set of contents measured together. Thus, in the example (1), c^1 would be defined as the context in which we measure $\{q_1, q_2\}$, c^2 as the context in which we measure $\{q_2, q_3, q_4\}$, etc. In STC, there cannot be distinct contexts with the same set of random variables in them, because their joint distributions would have to be the same.

3.6. To illustrate the effect of the restriction imposed by STC on the CbD framework, consider the system

$$\begin{array}{|c|c||c|}\hline R_1^1 & R_2^1 & c^1 \\ \hline R_1^2 & R_2^2 & c^2 \\ \hline\hline q_1 & q_2 & \mathcal{C}_2 \\ \hline \end{array}\,. \tag{10}$$

In STC, it can only represent the same context repeated twice, which makes the system trivially noncontextual. By contrast, in CbD, this so-called cyclic system of rank 2 is the smallest nontrivial system, one that can be contextual or noncontextual depending on the distributions involved [20]. In particular, the difference between the two contexts may be the order in which the two contents are measured ($q_1 \to q_2$ and $q_2 \to q_1$) [15].

3.7. The assumption of consistent connectedness (not necessarily strong one) simplifies the definition of (non)contextuality.

Definition 2 (Equivalent of STC definition)**.** A consistently connected system is noncontextual if it has an *identically connected coupling*. Otherwise it is *contextual*.

A coupling $\mathcal{S}$ is identically connected if, for any q, c, c' such that $q \prec c, c'$,

$$\Pr\left[S_q^c = S_q^{c'}\right] = 1. \tag{11}$$

3.8. This is, clearly, a special case of a multimaximal coupling: the maximal probability of $S_q^c = S_q^{c'}$ in such a coupling is 1 if and only if $R_q^c \sim R_q^{c'}$. For all practical purposes, it allows one to think of the random variables R_q^c as "context-independent," and many authors would even denote them as R_q. This is, however, a bad practice that leads to a logical contradiction [9, 10]. STC avoids this difficulty by not mentioning random variables at all, and systematically labeling probability distributions by their contexts [9].

3.9. Definition 2 does not explicitly require that consistent connectedness of the system be strong. This, however, makes little difference if one is only interested in determining whether a system is contextual. It is easy to show the following.

Theorem 3. *A simply consistently connected system that is not strongly consistently connected is contextual.*

Proof. In an identically connected coupling of the system, $\Pr\left[S_q^c = S_q^{c'}\right] = 1$ and $\Pr\left[S_{q'}^c = S_{q'}^{c'}\right] = 1$ imply

$$\Pr\left[\left(S_q^c, S_{q'}^c\right) = \left(S_q^{c'}, S_{q'}^{c'}\right)\right] = 1,$$

which is only possible if $\left(R_q^c, R_{q'}^c\right) \sim \left(R_q^{c'}, R_{q'}^{c'}\right)$. □

One can think of the strong consistent connectedness requirement in STC as a provision excluding this "guaranteed" contextuality from consideration.

4 Contextual Fraction

4.1. There are several possible ways of measuring the degree of contextuality in CbD [22], but in STC the measure of choice is contextual fraction. I present it, as everything else in this paper, in the language of CbD. We need a few general probabilistic notions first.

4.2. An *incomplete probability space*, or *α-probability space* (where $0 \le \alpha \le 1$) is defined as a measure space (X, Σ, μ) with $\mu(X) = \alpha$. The meaning of the components of the space (set X, sigma-algebra Σ, and sigma-additive measure μ) is standard. Any measurable function Z defined on this space is called an *incomplete (random) variable*, or a *(random) α-variable*. It is essentially an ordinary random variable: for any measurable set D in the codomain of Z, $\Pr[Z \in D]$ is defined as $\mu\left(Z^{-1}(D)\right)$ and referred to as the probability of Z

falling in D. The only difference is that if $D = Z(X)$ then $\Pr[Z \in D] = \alpha$. The rest of the concepts related to α-variables (e.g., their joint distribution) are the same as for true random variables (with $\alpha = 1$). In Feller's classical monograph [16] incomplete random variables are called "defective". Other terms, such as "improper", are used too.

4.3. Let $0 \le \alpha \le \beta \le 1$, and let Z_α and Z_β be an α-variable and a β-variable, respectively, with the same codomain. We say that Z_α is *majorized* by Z_β, if for every measurable set D in their common codomain,

$$\Pr[Z_\alpha \in D] \le \Pr[Z_\beta \in D]. \tag{12}$$

We write then

$$Z_\alpha \lesssim Z_\beta. \tag{13}$$

4.4. An *incomplete (or α-) coupling* of a system of random variables $\mathcal{R}$ is a correspondingly indexed set

$${}^{\alpha}\mathcal{S} = \left\{ {}^{\alpha}S_q^c : c \in C, q \in Q, q \prec c \right\} \tag{14}$$

of jointly distributed α-variables such that, for any context $c \in C$,

$$\left\{ {}^{\alpha}S_q^c : q \in Q, q \prec c \right\} \lesssim \left\{ R_q^c : q \in Q, q \prec c \right\}. \tag{15}$$

An α-coupling is identically connected if

$$\Pr\left[{}^{\alpha}S_q^c \ne {}^{\alpha}S_q^{c'} \right] = 0, \tag{16}$$

for any $q \prec c, c'$. Clearly, an identically connected α-coupling may exist only for a consistently connected system, and the latter is noncontextual if and only if it has an identically connected 1-coupling.

4.5. The following theorem allows one to introduce a measure of contextuality.

Theorem 4. *Any consistently connected system has an identically connected $\alpha_{\max}$-coupling ($0 \le \alpha_{\max} \le 1$), such that the system has no identically connected α-couplings with $\alpha > \alpha_{\max}$.*

This property can be proved by employing the linear programming representation of the relation between an α-coupling and the context-wise distributions in the system. This representation is essentially the same as one routinely used in both STC [3,4] and CbD [11,22].

Proof. We represent the system $\mathcal{R}$ and an α-coupling by probability vectors $\mathbf{r}$ and $\mathbf{s}$, respectively, such that

$$\mathbf{B}\mathbf{s} \le \mathbf{r} \text{ (componentwise)}, \tag{17}$$

with the following meaning of the terms. The entries of $\mathbf{r}$ are context-wise joint probabilities

$$\Pr\left[R_q^c = r_q : q \in Q, q \prec c\right], \tag{18}$$

across all $c \in C$ and all combinations of $r_q = +1/-1$. The entries of $\mathbf{z}$ are joint probabilities

$$\Pr\left[{}^{\alpha}S_q^c = s_q : c \in C, q \in Q, q \prec c\right], \tag{19}$$

across all combinations of $s_q = +1/-1$. $\mathbf{B}$ is a Boolean matrix ("incidence matrix" in Ref. [4]) with rows indexed by the values of c, and, for each c, by the combinations of r_q-values in (18). Its columns are indexed by the combinations of s_q-values in (19). A cell of $\mathbf{B}$ is filled with 1 if its s_q-combination contains its r_q-combination for the corresponding q-values, otherwise the cell is filled with 0. The set of all α-couplings of $\mathcal{R}$ is represented by the (obviously nonempty) polytope

$$\mathbb{Z} = \{\mathbf{s} : \mathbf{B}\mathbf{s} \leq \mathbf{r}, \mathbf{s} \geq 0, \mathbf{1} \cdot \mathbf{s} \leq 1\}. \tag{20}$$

Every linear functional, including $\mathbf{1} \cdot \mathbf{s}$, attains its extrema within this polytope, and the maximum value of $\mathbf{1} \cdot \mathbf{s}$ is taken as $\alpha_{\max}$. □

4.6. The quantity $\alpha_{\max}$ is called *noncontextual fraction*, and $1 - \alpha_{\max}$ is called *contextual fraction*. It is easy to see that if $\alpha_{\max} = 1$, then $\mathbf{B}\mathbf{s} = \mathbf{r}$, and the system is noncontextual. Otherwise it is contextual, and the contextual fraction is a natural measure of the degree of contextuality.

4.7. Abramsky and colleagues single out the case $\alpha_{\max} = 0$, calling such systems *strongly contextual*. In strongly contextual systems, every possible combination of s_q-values has the probability of zero. This is the situation one encounters with the Kochen-Specker systems [19] and with the Popescu-Rohrlich boxes [23].

5 Consistified Systems

5.1. Can STC, and contextual fraction in particular, be generalized to arbitrary, generally inconsistently connected, systems? It turns out this can be done by a simple procedure that converts an inconsistently connected system $\mathcal{R}$ into a *contextually equivalent* consistently connected one, $\mathcal{R}^{\ddagger}$. Contextual equivalence means that $\mathcal{R}$ is contextual if and only if so is $\mathcal{R}^{\ddagger}$, and that if $\mathcal{R}$ is consistently connected, then $\mathcal{R}^{\ddagger}$ has the same value of contextual fraction.

5.2. The discussion of the procedure of *consistification* is helped by two additional CbD terms. Let us call all random variables sharing a context a *bunch*,

$$\mathcal{R}^c = \left\{R_q^c : q \in Q, q \prec c\right\} = R^c, \tag{21}$$

and all random variables sharing a content a *connection*,

$$\mathcal{R}_q = \left\{ R_q^c : c \in C, q \prec c \right\}. \tag{22}$$

The terminology is intuitive: a bunch is jointly distributed, and different bunches are disjoint, but the fact that some random variables in different contexts have the same content creates connections between the bunches.

5.3. Within a connection (22) the random variables are stochastically unrelated, but they can be coupled by

$$\mathcal{T}_q = \left\{ T_q^c : c \in C, q \prec c \right\} = T_q, \tag{23}$$

and among all such couplings one can seek a *multimaximal coupling*, one that maximizes the probabilities for all equalities

$$T_q^c = T_q^{c'}, q \prec c, c'. \tag{24}$$

5.4. If, as we have agreed, all random variables in the system are dichotomous, then we have the following result, proved in [13].

Theorem 5. *Any connection has one and only one multimaximal coupling.* $\mathcal{T}_q$ *is a multimaximal coupling of the connection* $\mathcal{R}_q$ *if and only if any subset of* $\mathcal{T}_q$ *is a maximal coupling of the corresponding subset of* $\mathcal{R}_q$.

The second statement means that, for any part

$$\mathcal{R}'_q = \left\{ R_q^c : c \in \{c_1, \ldots, c_k\} \subseteq C, q \prec c \right\},$$

of the connection (22), the event

$$T_q^{c_1} = \ldots = T_q^{c_k}$$

has the maximal possible probability among all possible couplings of $\mathcal{R}'_q$ (or, equivalently, given the marginal distributions of $T_q^{c_i} \sim R_q^{c_i}$, $i = 1, \ldots, k$).

5.5. The theorem above is one of the two reasons why CbD subjects any set of measurements to dichotomization before making it a system of random variables amenable to contextuality analysis. The consistification procedure to be described is based on the possibility to find unique multimaximal couplings for all connections.

5.6. For completeness, I should mention the second, and main, reason for the dichotomization of all random variables: it prevents the otherwise possible situation when coarse-graining of the random variables in a noncontextual system makes it contextual [11]. Clearly, a good theory of contextuality should not have this property. Note that dichotomization does not lose any information extractable from random variables before they are dichotomized. It does not even increase the size of a system, if size is measured by the cardinality of the supports of the system's bunches.

5.7. The idea of consistification is to treat the multimaximal couplings of connections as if they were additional bunches. This is implicit in any CbD-based algorithm for establishing or measuring contextuality [20–22]. Explicitly, however, it was first described by Amaral and coauthors [6]. However, Amaral and coauthors use maximal couplings instead of the multimaximal ones (as we did in the older version of CbD, e.g., in [12]), and they allow for multivalued variables. The difference between the two types of couplings of a set $\{X_1, \ldots, X_n\}$ is that in the multimaximal coupling $\{Y_1, \ldots, Y_n\}$ we maximize probabilities of all equalities $Y_i = Y_j$ (whence it follows, by Theorem 5, that we also maximize the probability of $Y_{i_1} = Y_{i_2} = \ldots = Y_{i_k}$ for any subset of $\{Y_1, \ldots, Y_n\}$), whereas a maximal coupling $\{Z_1, \ldots, Z_n\}$ only maximizes the probability of the single chain equality $Z_1 = Z_2 = \ldots = Z_n$. Maximal couplings generally are not unique, even for dichotomous variables (if there are more than two of them). Since measures of (non)contextuality generally depend on what couplings are being used, the approach advocated in Ref. [6] faces the problem of choice. In addition, a system declared noncontextual using maximal rather than multimaximal couplings may have contextual subsystems, obtained by dropping some of the variables. I consider this possibility highly undesirable for a theory of contextuality.

5.8. The particular consistification scheme presented below is an elaboration of one described to me by Janne Kujala (personal communication, November 2018).

5.9. Given an arbitrary system $\mathcal{R}$, the new system $\mathcal{R}^{\ddagger}$ has a set $Q^{\ddagger}$ of "new" contents, a set $C^{\ddagger}$ of "new contexts", and a "new" is-measured-in relation $\prec^{\ddagger}$. The corresponding constructs in the original system, Q, C, and $\prec$, will be called "old".

5.10. For each random variable R_j^i in $\mathcal{R}$ we form a new content, denoted q_j^i. The set of all new contents is

$$Q^{\ddagger} = \left\{ q_j^i : c^i \in C, q_j \in Q, q_j \prec c^i \right\}. \tag{25}$$

The number of the new contents is the cardinality of $\prec$, which cannot exceed $|C \times Q|$.

5.11. New contexts are formed as the set

$$C^{\ddagger} = C \sqcup Q,$$

and their number is $|C| + |Q|$.

5.12. The new is-measured-in relation is

$$\prec^{\ddagger} = \left\{ \left(q_j^i, c^i\right) : c^i \in C, q_j \in Q, q_j \prec c^i \right\} \sqcup \left\{ \left(q_j^i, q_j\right) : c^i \in C, q_j \in Q, q_j \prec c^i \right\}. \tag{26}$$

That is, a new content q_j^i is measured in the new contexts c^i and q_j only.

5.13. Each $\left(q_j^i, c^i\right)$-cell contains the old random variables R_j^i. The new bunch

$$R^i = \left\{R_j^i : q_j^i \in Q^\ddagger, q_j^i \prec^\ddagger c^i\right\} \tag{27}$$

coincides with the old bunch

$$R^i = \left\{R_j^i : q_j \in Q, q_j \prec c^i\right\}. \tag{28}$$

5.14. Each $\left(q_j^i, q_j\right)$-cell contains a new random variable V_j^i whose distribution is the same as that of R_j^i. The bunch

$$V^j = \left\{V_j^i : q_j^i \in Q^\ddagger, q_j^i \prec^\ddagger q_j\right\} \tag{29}$$

is the multimaximal coupling of the old connection

$$\mathcal{R}_j = \left\{R_j^i : c^i \in C, q_j \prec c^i\right\}. \tag{30}$$

5.15. Using our examples (1) and (10), the corresponding consistified systems are

R_1^1	R_2^1											c^1
		R_2^2	R_3^2	R_4^2								c^2
					R_1^3	R_3^3						c^3
							R_1^4	R_4^4				c^4
									R_1^5	R_2^5	R_3^5	c^5
V_1^1					V_1^3		V_1^4		V_4^4			q_1
	V_2^1	V_2^2								V_2^5		q_2
			V_3^2			V_3^3					V_3^5	q_3
				V_4^2				V_4^4				q_4
q_1^1	q_2^1	q_2^2	q_3^2	q_4^2	q_1^3	q_3^3	q_1^4	q_4^4	q_1^5	q_2^5	q_3^5	$\mathcal{A}^\ddagger$

(31)

and

$$
\begin{array}{|c|c|c|c||c|}
\hline
R_1^1 & R_2^1 & & & c^1 \\
\hline
 & & R_1^2 & R_2^2 & c^2 \\
\hline
V_1^1 & & V_1^2 & & q_1 \\
\hline
 & V_2^1 & & V_2^2 & q_2 \\
\hline
\hline
q_1^1 & q_2^1 & q_1^2 & q_2^2 & \mathcal{C}_2^{\ddagger} \\
\hline
\end{array}\ . \tag{32}
$$

5.16. Note the following properties of all consistified systems.

1. Bunches corresponding to different old contexts, $c^i, c^{i'}$, are disjoint.
2. Bunches corresponding to different old contents, $q_j, q_{j'}$, are disjoint.
3. A bunch corresponding to an old content, q_j, and a bunch corresponding to an old context, c^i, have at most one connection between them, $\{R_j^i, V_j^i\}$.
4. The connection corresponding to any new content q_j^i contains precisely two random variables, R_j^i and V_j^i, with the same distribution.

5.17. Property 3 above means that in a consistified system the notions of simple consistent connectedness and strong consistent connectedness coincide. As mentioned earlier, in Section 3.6, system $\mathcal{C}_2$ in (10) will only be considered in STC if the two bunches $\{R_1^1, R_2^1\}$ and $\{R_1^2, R_2^2\}$ are identically distributed, which would make this system trivial. By contrast, if one adopts the CbD-based consistification of $\mathcal{C}_2$, in (32), its STC analysis will be the same as CbD's.

5.18. The following fact ensures that the generalization of the contextual fraction coincides with the original one in the case of consistently connected systems.

Theorem 6. *If a system $\mathcal{R}$ is consistently connected, then its contextual fraction is the same as that of $\mathcal{R}^{\ddagger}$. (In particular, $\mathcal{R}$ is contextual if and only if so is $\mathcal{R}^{\ddagger}$.)*

Proof. Immediately follows from the observation that any state of an α-coupling in which the values corresponding to R_q^c and V_q^c are different has the probability zero. □

5.19. For completeness, I formulate the following as a formal statement. Recall the definition of contextual equivalence in Section 5.1.

Theorem 7. *Any system $\mathcal{R}$ is contextually equivalent to its consistification $\mathcal{R}^{\ddagger}$.*

Proof. Follows from the previous theorem, and the obvious fact that $\mathcal{R}$ and $\mathcal{R}^{\ddagger}$ have the same linear programming representation (see Ref. [22] for a detailed description of the latter). $\square$

6 Deterministic Systems

6.1. Deterministic systems can be viewed as systems of random variables whose distributions attain specific values with probability 1. In CbD, therefore, they are treated as a special case of systems of random variables, with the following general result.

Theorem 8. *Any deterministic system is noncontextual.*

Proof. A deterministic system

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

where $\equiv$ means equality with probability 1, has a single overall coupling,

$$\mathcal{S} = \left\{ S_q^c \equiv r_q^c : c \in C, q \in Q, q \prec c \right\},$$

with all S_q^c defined on an arbitrary probability space. Since $\left\{ S_q^c \equiv r_q^c, S_q^{c'} \equiv r_q^{c'} \right\}$ is the only coupling of $\left\{ R_q^c \equiv r_q^c, R_q^{c'} \equiv r_q^{c'} \right\}$, the probability of $S_q^c = S_q^{c'}$ (0 or 1) is maximal possible, whence $\mathcal{S}$ is multimaximally connected. $\square$

6.2. This simple observation seems to put CbD at odds with STC, where the theoretical ideas formulated in algebraic and topological terms are not restricted to random variables. Consider, e.g., two deterministic systems that have the same $\prec$-format as the system $\mathcal{C}_2$ in example (10):

$$\begin{array}{|c|c||c|}\hline R_1^1 \equiv 1 & R_2^1 \equiv -1 & c^1 \\ \hline R_1^2 \equiv 1 & R_2^2 \equiv -1 & c^2 \\ \hline q_1 & q_2 & \mathcal{C}_{2.1} \\ \hline \end{array} \text{ and } \begin{array}{|c|c||c|}\hline R_1^1 \equiv 1 & R_2^1 \equiv -1 & c^1 \\ \hline R_1^2 \equiv 1 & R_2^2 \equiv 1 & c^2 \\ \hline q_1 & q_2 & \mathcal{C}_{2.2} \\ \hline \end{array}. \quad (33)$$

System $\mathcal{C}_{2.1}$ is consistently connected, which in a deterministic system means it is strongly consistently connected. It is therefore trivially noncontextual. System $\mathcal{C}_{2.2}$ is inconsistently connected. Strictly speaking, therefore, the original STC analysis should not be applicable to this system, as it violates the fundamental assumption underlying STC.

6.3. If we use the extended version of STC, with the help of consistification, we get

$$\begin{array}{|c|c|c|c||c|}
\hline
R_1^1 \equiv 1 & R_2^1 \equiv -1 & & & c^1 \\
\hline
 & & R_1^2 \equiv 1 & R_2^2 \equiv 1 & c^2 \\
\hline
V_1^1 \equiv 1 & & V_1^2 \equiv 1 & & q_1 \\
\hline
 & V_2^1 \equiv -1 & & V_2^2 \equiv 1 & q_2 \\
\hline
\hline
q_1^1 & q_2^1 & q_1^2 & q_2^2 & \mathcal{C}^{\ddagger}_{2.2} \\
\hline
\end{array}\;. \tag{34}$$

This system is trivially noncontextual by the STC/CbD definition.

6.4. This reasoning would apply to any deterministic system: if it is consistently connected (or consistified), it is trivially noncontextual, and if it is inconsistently connected, STC should place it outside its sphere of applicability (or consistify it). In other words, STC with consistification and CbD treat deterministic systems identically (finding them noncontextual). There is, of course, a simple way out: to complement STC with the additional stipulation that all inconsistently connected systems are contextual. STC would then have to allow for contextual systems whose degree of contextuality cannot be measured by contextual fraction. I do not think this simple way out is intellectually satisfactory.

6.5. Here is a good place to mention that CbD treats inconsistent connectedness and contextuality as fundamentally different concepts. Inconsistent connectedness, i.e. the difference in the distributions of R_q^c and $R_q^{c'}$, is interpreted as the result of direct influences of the contexts upon the measurements. In the case of physical systems, one can say that some elements of the contexts c and c' differently affect (in the causal sense) the measurement of the content q. In quantum physics this is reflected by such notions as "signaling" or (a better term) "disturbance". Contextuality, by contrast, is non-causal, and reflects the differences between random variables R_q^c and $R_q^{c'}$ that are above and beyond the differences in their distributions.

6.6. This interpretation is philosophically based on the *no-conspiracy principle* [8], according to which in "not-precariously-unstable" and "not-deliberately-contrived" systems, no differences in the direct influences exerted by the elements of context are hidden. Being hidden means that these differences are present but are not reflected in the differences of the distributions. For instance, if R_q^c and $R_q^{c'}$ attain values 1 and -1 with probability $\frac{1}{2}$ each, and if c' by some causal mechanism reverses (multiplies by -1) each value of $R_q^{c'}$, then this influence will be hidden, as it will not affect the distribution of $R_q^{c'}$. The no-conspiracy principle says this should not be expected to happen, and if it does, should be expected to disappear by slight modifications of the experimental set-up. The principle is closely related to the "no-fine-tuning" principle advocated by Cavalcanti [7] (see a detailed analysis of these principles by Jones in Ref. [18]).

6.7. With this in mind, let us consider an especially elegant application of STC to an inherently deterministic system, described in Ref. [2]. This is a system whose contents are statements referencing each other's truth value and forming a version of the Liar antinomy. I will consider the version with three statements, although any larger number will be analyzed similarly:

R_1^1	R_2^1		c^1
	R_2^2	R_3^2	c^2
R_1^3		R_3^3	c^3
$q_1 =$ "q_2 is true"	$q_2 =$ "q_3 is true"	$q_3 =$ "q_1 is false"	$\mathcal{L}_3$

. (35)

The contexts combine the statements one of which references the other, and the R_q^c is the truth value (1 or -1) of statement q in context c.

6.8. We could have considered the smaller system

R_1^1	R_2^1	c^1
R_1^2	R_2^2	c^2
$q_1 =$ "q_2 is true"	$q_2 =$ "q_1 is false"	$\mathcal{L}_2$

, (36)

representing a more familiar classical form of the antinomy, but the contexts in $\mathcal{L}_3$ are easier to interpret, as the direction of inference there need not be specified. The interpretation is even more complicated with the classical form $q =$ "q is false", although the reasoning below is still applicable.

6.9. We can posit that each statement in a given context should have one definitive truth value, and this makes $\mathcal{L}_3$ a deterministic system. In Ref. [2] this system is characterized as strongly contextual, based on the impossibility to assign the truth values in a context-independent way. However, we know that the original version of STC is predicated on the assumption of strong consistent connectedness, whereas any deterministic realization of $\mathcal{L}_3$ (precisely because no context-independent assignment of truth values exists) is inconsistently connected. Consider one of the eight such deterministic versions, corresponding to the usual conceptualization of the Liar Antinomy:

$R_1^1 \equiv 1$	$R_2^1 \equiv 1$		c^1
	$R_2^2 \equiv 1$	$R_3^2 \equiv 1$	c^2
$R_1^3 \equiv -1$		$R_3^3 \equiv 1$	c^3
$q_1 =$ "q_2 is true"	$q_2 =$ "q_3 is true"	$q_3 =$ "q_1 is false"	$\mathcal{L}_{3.1}$

. (37)

The arguments related to system $\mathcal{C}_{2.2}$ in (33) apply here fully. I see no reasonable way a system like $\mathcal{L}_{3.1}$ can be treated as contextual, either in CbD or in STC.

6.10. There is, however, another way of looking at system $\mathcal{L}_3$. It seems to be very much in the spirit of how it is treated in Ref. [2]. Moreover, it corresponds to the traditional presentation of the Liar antinomy: suppose $R_1^1 \equiv 1$, then it follows that $R_1^1 \equiv -1$; now suppose $R_1^1 \equiv -1$, then it follows that $R_1^1 \equiv 1$. With the stipulation that each statement in a given context should have one definitive truth value, $\mathcal{L}_3$ can indeed be just one of the eight deterministic (and inconsistently connected) systems of which $\mathcal{L}_{3.1}$ is one. However, we do not know which of these eight systems to choose, and we can consider all eight of them as variants of $\mathcal{L}_3$:

1	1		c^1
	1	1	c^2
-1		1	c^3
q_1	q_2	q_3	$\mathcal{L}_{3.1}$

-1	-1		c^1
	-1	-1	c^2
1		-1	c^3
q_1	q_2	q_3	$\mathcal{L}_{3.2}$

-1	-1		c^1
	1	1	c^2
-1		1	c^3
q_1	q_2	q_3	$\mathcal{L}_{3.3}$

etc.

6.11. One can assign Bayesian (or epistemic) probabilities to these possibilities, a natural choice here being to assign them uniformly. This renders the system probabilistic in the epistemic sense, with

$$\left\langle R_q^c \right\rangle_B = 0, \tag{38}$$

for all the "epistemically-random" variables (indicated by the subscript B), and

$$\left\langle R_1^1 R_2^1 \right\rangle_B = \left\langle R_2^2 R_3^2 \right\rangle_B = -\left\langle R_3^3 R_1^3 \right\rangle_B = 1. \tag{39}$$

This is a Bayesian analogue of a rank 3 cyclic system that is consistently connected and forms a Popescu-Rohrlich box. Its contextuality, both in CbD and STC, is maximal. In particular, when measured by contextual fraction, it is strong ($\alpha_{\max} = 0$), in accordance with how Abramsky and colleagues view it.

6.12. This Bayesian procedure can be applied to any deterministic system with more than one possible deterministic realization. The procedure will render the system quasi-probabilistic and, at least in all the simple cases I can think of, consistently connected and contextual. More work is needed to elaborate this approach.

7 Conclusion

7.1. We have seen that STC can be extended to apply to inconsistently connected systems, using CbD-based multimaximal couplings to consistify these systems. We have also seen that the Bayesian rendering of the deterministic systems with multiple possible realizations allows STC to circumvent the difficulty associated with inconsistent connectedness of each of these realizations. It simultaneously extends CbD to such systems and allows CbD to treat them in the spirit of STC, forming thereby another bridge between the two theories.

7.2. Together, the consistification and the Bayesian treatment make STC and CbD essentially coextensive, with a major proviso: one has to agree to represent all measurement outcomes in a system as sets of jointly distributed dichotomous random variables. Dichotomization of a system is always possible, so it is more of a language choice than a restriction of applicability. Dealing only with dichotomous variables allows one to avoid a variety of difficulties [11], but no proof exists that they could not be avoided by other means.

7.3. Finally, nothing in this paper implies that CbD can be replaced with STC, or vice versa. Each of the two theories has its own aims and means. Thus, logical aspects of contextuality, especially in the possibilistic proofs of contextuality, are significantly more salient in STC than CbD, adding to the former's aesthetic elegance. Perhaps the use of the Bayesian/epistemic random variables, as discussed above, might offer CbD a way to "catch up" in this respect. STC in turn might benefit from using the language of random variables for proof purposes. For instance, the fact that the existence of a hidden variable model for a system of random variables is equivalent to the existence of their joint distribution (from which it follows, in particular, that nonlocality is a special case of contextuality) is true almost by definition if the language of random variables is used explicitly. It is a non-trivial, perhaps even surprising fact, however, if one considers the systems of random variables in terms of their distributions only [4, 17]. It would be good if the equivalences established in this paper helped the two theories to more freely borrow from each other's native languages, follow each other's directions of research, and use each other's proof techniques.

References

[1] Abramsky S. (2014). Contextual semantics: From quantum mechanics to logic, databases, constraints, and complexity. Bull. Eur. Ass. Theor. Comp. Sci. 113, 137--163.

[2] Abramsky S., Barbosa R.S., Kishida K., Lal R., Mansfield S. (2017). Contextuality, cohomology and paradox. arXiv:1502.03097.

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

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

[5] Abramsky S., Brandenburger A., Savochkin, A. (2014). No-signalling is equivalent to free choice of measurements. In M.J. Hoban, B. Coecke (Eds.): 10th International Workshop on Quantum Physics and Logic (QPL 2013) EPTCS 171 (pp. 1–9).

[6] Amaral B., Duarte C., Oliveira R.I. (2018). Necessary conditions for extended noncontextuality in general sets of random variables. J. Math. Phys. 59, 072202.

[7] Cavalcanti E.G. (2018). Classical causal models for Bell and Kochen-Specker inequality violations require fine-tuning. Phys. Rev. X 8, 021018.

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

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

[10] Dzhafarov E.N., Kujala J.V. (2017). Probabilistic foundations of contextuality. Fort. Phys. 65, 1-11.

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

[12] Dzhafarov E.N., Kujala J.V. (2016). Context-content systems of random variables: The Contextuality-by- Default theory. J. Math. Psych. 74, 11-33.

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

[14] Dzhafarov, E.N.; Kujala, J.V.; Cervantes, V.H. Contextuality and Noncontextuality Measures and Generalized Bell Inequalities for Cyclic Systems. *Phys. Rev. A* **2020**, 101, 042118 (Erratum note in *Phys. Rev. A* 101:069902).

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

[16] Feller W. (1968). An introduction to probability theory and its applications, Vol. 1. NewYork: Wiley.

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

[18] Jones M. (2019). Causal and probabilistic approaches to contextuality. Phil. Trans. Roy.Soc. A 377, 20190133.

[19] Kochen S., Specker E.P. (1967). The problem of hidden variables in quantum mechanics. J. Math. Mech. 17: 59–87.

[20] Kujala J.V., Dzhafarov E.N. (2016). Proof of a conjecture on contextuality in cyclic systems with binary variables. Found. Phys. 46, 282-299.

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

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

[23] Popescu S., Rohrlich D. (1994). Quantum nonlocality as an axiom. Found. Phys. 24, 379–385.