Necessary and Sufficient Conditions for Extended Noncontextuality in a Broad Class of Quantum Mechanical Systems

Janne V. Kujala,[1,*] Ehtibar N. Dzhafarov,[2,†] and Jan-Åke Larsson[3,‡]

[1]*Department of Mathematical Information Technology, University of Jyväskylä, Jyväskylä, Finland*
[2]*Department of Psychological Sciences, Purdue University, West Lafayette, Indiana, USA*
[3]*Department of Electrical Engineering, Linköping University, 58183 Linköping, Sweden*

The notion of (non)contextuality pertains to sets of properties measured one subset (context) at a time. We extend this notion to include so-called inconsistently connected systems, in which the measurements of a given property in different contexts may have different distributions, due to contextual biases in experimental design or physical interactions (signaling): a system of measurements has a maximally noncontextual description if they can be imposed a joint distribution on in which the measurements of any one property in different contexts are equal to each other with the maximal probability allowed by their different distributions. We derive necessary and sufficient conditions for the existence of such a description in a broad class of systems including Klyachko-Can-Binicioğlu-Shumvosky-type (KCBS), EPR-Bell-type, and Leggett-Garg-type systems. Because these conditions allow for inconsistent connectedness, they are applicable to real experiments. We illustrate this by analyzing an experiment by Lapkiewicz and colleagues aimed at testing contextuality in a KCBS-type system.

Keywords: CHSH inequalities; contextuality; criterion of contextuality; Klyachko-Can-Binicioğlu-Shumvosky inequalities; Leggett-Garg inequalities; measurement bias; measurement errors; probabilistic couplings; signaling.

The notion of (non)contextuality in Quantum Mechanics (QM) relates the outcome of a measurement of a physical property q to the choice of properties $q', q'', \ldots$ co-measured with q [1]. The set of co-measured properties $q, q', q'', \ldots$ forms a *measurement context* for each of its members. The traditional understanding of a contextual QM system is that if the measurement of each property q in it is represented by a random variable R_q, then the random variables representing all properties in the system do not have a joint distribution.

We use here a different formulation, which, although formally equivalent, lends itself to more productive development [5–10]. We label all measurements contextually: this means that a property q is represented by different random variables R_q^c depending on the context $c = \{q, q', q'', \ldots\}$. We say that the system has *a noncontextual* description if there exists a joint distribution of these random variables in which any two of them, $R_q^{c_1}$ and $R_q^{c_2}$, representing the same property q in different contexts, are equal with probability 1. If no such description exists we say that the system is *contextual.* Note that the existence of a joint distribution of several random variables is equivalent to the possibility of presenting them as functions of a single, "hidden" variable λ [2, 3, 5, 11].

This formulation applies to systems in which the random variables $R_q^{c_1}, R_q^{c_2}, \ldots$ representing a given property in different contexts always have the same distribution. We call such systems *consistently connected*, because we call the set of all such variables $R_q^{c_1}, R_q^{c_2}, \ldots$ for a given q a *connection.* If the properties forming any given context are space-time separated, consistent connectedness coincides with the no-signaling condition [12]. The central aim of this paper is to extend the notion of contextuality to the cases of *inconsistent connectedness,* where the measurements of a given property may have different distributions in different contexts. This may happen due to a contextually biased measurement design or due to physical influences exerted on R_q^c by elements of context c other than q.

The criterion of (necessary and sufficient conditions for) contextuality we derive below is formulated for inconsistently connected systems, treating consistent connectedness as a special case. This makes it applicable to real experimental data. For example, the experiment in Ref. [20] testing the Klyachko-Can-Binicioğlu-Shumvosky (KCBS) inequality [21] exhibits inconsistent connectedness, necessitating a sophisticated work-around to establish contextuality (see Refs. [22, 23]). Below, we apply our extended notion to the same data to establish contextuality directly, with no work-arounds. Another example is Leggett-Garg (LG) systems [17], where our approach allows for the possibility that later measurements may be affected by previous settings ("signaling in time," [18, 19]). Finally, in EPR-Bell-type systems [13, 14] our approach allows for the possibility that Alice's measurements are affected by Bob's settings [15] when they are time-like separated; and even with space-like separation, the same effect can be caused by systematic errors [16].

Earlier treatments.— In the Kochen-Specker theorem [1] or its variants [24, 25], contexts are chosen so that each property enters in more than one context, and in each context, according to QM, one and only one of the measurements has a nonzero value. The proof of contextuality, using our language, consists in showing that the variables R_q^c cannot be jointly assigned values consistent with this constraint so that all the variables representing the same property q are assigned the same value. An experimental

* To whom correspondence should be addressed. E-mail: jvk@iki.fi
† E-mail: ehtibar@purdue.edu
‡ E-mail: jan-ake.larsson@liu.se

test of contextuality here consists in simply showing that the observables it specifies can be measured in the contexts it specifies, and that the QM constraint in question is satisfied.

There has been recent work translating the value assignment proofs into probabilistic inequalities (sometimes called *Kochen-Specker inequalities*) giving necessary conditions for noncontextuality [5, 26]. Inequalities that do not use value-assignment restrictions but only the assumption of noncontextuality are known as *noncontextuality inequalities* [21, 27, 28]. Bell inequalities [3, 13, 14, 29, 30] and LG inequalities [2, 17] are also established through noncontextuality [31], motivated by specific physical considerations (locality and noninvasive measurement, resp.).

An extension of the notion of (non)contextuality that allows for inconsistent connectedness was suggested in Refs. [5, 32]. However, the error probability proposed in those papers as a measure of context-dependent change in a random variable cannot be measured experimentally. The suggestion in both Refs. [5, 32] is to estimate the accuracy of the measurement and from that argue for a particular value of the error probability. For example, Ref. [32] uses the quantum description of the system for the estimate (quantum tomography), but there is no clear reason why or how the quantum error model would be related to that of the proposed noncontextual description. A noncontextuality test should not mix the two descriptions, as it attempts to show their fundamental differences.

In this paper we generalize the definition of contextuality in a different manner, to allow for inconsistent connectedness while only using directly measurable quantities. We derive a criterion of (non)contextuality for a broad class of systems that includes as special cases the systems intensively studied in the recent literature on contextuality: KCBS, EPR-Bell, and LG systems [21, 33, 34], with their inconsistently connected versions [35, 36].

Basic Concepts and Definitions.— We begin by formalizing the notation and terminology. Consider a finite set of distinct *physical properties* $Q = \{q_1, \ldots, q_n\}$. These properties are measured in subsets of Q called *contexts*, $c_1, \ldots, c_m$. Let C denote the set of all contexts, and C_q the set of all contexts containing a given property q.

The result of measuring property q in context c is a random variable R_q^c . The result of jointly measuring all properties within a given context $c \in C$ is a set of jointly distributed random variables $R^c = \{R_q^c : q \in c\}$.

No two random variables in different contexts, $R_q^c, R_{q'}^{c'}$, $c \neq c'$, are jointly distributed, they are *stochastically unrelated* [9, 10]. The set of random variables representing the same property q in different contexts is called a *connection* (for q). So the elements of a connection $\{R_q^c : c \in C_q\}$ are pairwise stochastically unrelated. If all random variables within each connection are identically distributed, the system is called *consistently connected*; if it is not necessarily so, it is *inconsistently connected.* Consistent connectedness is also known in QM as the *Gleason property* [37], outside physics as *marginal selectivity* [9], and Ref. [38] lists some dozen names for the same notion; a recent addition to the list is *no-disturbance principle* [39, 40].

The set Q of all properties together with the set C of all contexts and the set $\{R^c : c \in C\}$ of all sets of random variables representing contexts is referred to as a *system.* In the systems we consider here the set of properties q is finite (whence the set of contexts c is finite too), and each random variable has a finite number of possible values (e.g., spin measurement outcomes).

We introduce next the notion of a (probabilistic) *coupling* of all the random variables R_q^c in our system [41]. Intuitively, this is simply a joint distribution imposed, or "forced" on all of them (recall that they include stochastically unrelated variables from different contexts). Formally, a coupling of $\{R_q^c : q \in c \in C\}$ is any jointly distributed set of random variables $S = \{S_q^c : q \in c \in C\}$ such that, for every $c \in C$, $\{S_q^c : q \in c\} \sim \{R_q^c : q \in c\}$, where $\sim$ stands for "has the same (joint) distribution as." One can also speak of a coupling for any subset of the random variables R_q^c. Thus, fixing a property q, a coupling of a connection $\{R_q^c : c \in C_q\}$ is any jointly distributed $\{X_q^c : c \in C_q\}$ such that $X_q^c \sim R_q^c$ for all contexts $c \in C_q$. Note that if S is a coupling of all R_q^c, then every marginal (jointly distributed subset) $\{S_q^c : c \in C_q\}$ of S is a coupling of the corresponding connection $\{R_q^c : c \in C_q\}$.

Expressed in this language, the traditional approach is to consider a system *noncontextual* if there is a coupling S of the random variables R_q^c, such that for every property q the random variables in $\{S_q^c : c \in C_q\}$ are equal to each other with probability 1. That is, for every possible coupling S of the random variables R_q^c and every property q we consider the marginal $\{S_q^c : c \in C_q\}$ corresponding to a connection $\{R_q^c : c \in C_q\}$, and we compute

$$\Pr\left[S_q^{c_{q1}} = \cdots = S_q^{c_{qn_q}}\right],\ \{c_{q1}, \ldots, c_{qn_q}\} = C_q. \quad (1)$$

If there exists a coupling S for which this probability equals 1 for all q, this S provides a noncontextual description for our system. Otherwise, if in every possible coupling S the probability in question is less than 1 for some properties q, the system is considered *contextual.*

This understanding, however, only involves consistently connected systems. As mentioned in the introduction, a system may be inconsistently connected due to systematic biases or interactions (such as "signaling in time" in LG systems). If for some q and some contexts $c, c' \in C_q$, the distribution of R_q^c and $R_q^{c'}$ are not the same, then $\Pr\left[S_q^c = S_q^{c'}\right]$ cannot equal 1 in any coupling S. There would be nothing wrong if one chose to say that any such inconsistently connected system is therefore contextual, but contextuality due to systematic measurement errors or signaling is clearly a special, trivial kind of contextuality. One should be interested in whether the system exhibits any contextuality that is not reducible to (or explainable by) the factors that make distributions of random variables within a connection different. For systems in general therefore we propose a different definition.

Definition 1. A system has a *maximally noncontextual description* if there is a coupling S of the random variables

R_q^c, such that for any q the random variables $\{S_q^c : c \in C_q\}$ in S are equal to each other with the maximum probability allowed by the individual distributions of R_q^c.

To explain, consider a connection $\{R_q^c : c \in C_q\}$ in isolation, and let $\{X_q^c : c \in C_q\}$ be its coupling. Among all such couplings there must be *maximal* ones, those in which the probability that all variables in $\{X_q^c : c \in C_q\}$ are equal to each other is maximal possible, given the distributions of $X_q^c \sim R_q^c$. If a connection consists of two dichotomic (± 1) variables R_q^1 and R_q^2, and $\{X_q^1, X_q^2\}$ is its coupling (i.e., X_q^1, X_q^2 are jointly distributed with $\langle X_q^1 \rangle = \langle R_q^1 \rangle$, $\langle X_q^2 \rangle = \langle R_q^2 \rangle$), then by Lemma A3 in Supplementary Material, the maximal possible expectation $\langle X_q^1 X_q^2 \rangle$ is $1 - |\langle R_q^1 \rangle - \langle R_q^2 \rangle|$; a coupling $\{X_q^1, X_q^2\}$ with this expectation is maximal. Now take every possible coupling S of all our random variables R_q^c, consider the marginals $\{S_q^c : c \in C_q\}$ corresponding to connections $\{R_q^c : c \in C_q\}$, and for each of these marginals compute the probability (1). If there is a coupling S in which this probability equals its maximal possible value for every q, this S provides a maximally noncontextual description for our system. For consistently connected systems Definition 1 reduces to the traditional understanding: the maximal probability with which all variables in $\{X_q^c : c \in C_q\}$ can be equal to each other is 1 if all these variables are identically distributed.

Cyclic systems of dichotomic random variables.— We focus now on systems in which: (S1) each context consists of precisely two distinct properties; (S2) each property belongs to precisely two distinct contexts; and (S3) each random variable representing a property is dichotomic (± 1). As shown in Lemma A1 (Supplementary Material), a set of properties satisfying S1–S2 can be arranged into one or more distinct cycles $q_1 \to q_2 \to \ldots \to q_k \to q_1$, in which any two successive properties form a context. Without loss of generality we will assume that we deal with a *single-cycle* arrangement $q_1 \to q_2 \to \ldots \to q_n \to q_1$ of all the properties $\{q_1, \ldots, q_n\}$. The number n is referred to as the *rank* of the system.

A schematic representation of a cyclic system is shown in Figure 1. The LG paradigm exemplifies a cyclic system of rank $n = 3$, on labeling the observables q_1, q_2, q_3 measured chronologically. The contexts $\{q_1, q_2\}, \{q_2, q_3\}, \{q_3, q_1\}$ here are represented by, respectively, pairs $(R_1^1, R_2^1), (R_2^2, R_3^2), (R_3^3, R_1^3)$ with observed joint distributions, whereas $(R_1^1, R_1^3), (R_2^2, R_2^1), (R_3^3, R_3^2)$ are connections for q_1, q_2, q_3, respectively. The EPR-Bell paradigm exemplifies a cyclic system of rank $n = 4$, on labeling the observables q_1, q_3 for Alice and q_2, q_4 for Bob. Cyclic systems of rank $n = 5$ are exemplified by the KCBS paradigm, on labeling the vertices of the KCBS pentagram by $q_1 \to q_2 \to q_3 \to q_4 \to q_5$.

(Non)Contextuality Criterion.— For any n, and any $x_1, \ldots, x_n \in \mathbb{R}$, we define the function

$$\mathsf{s}_1(x_1, \ldots, x_n) = \max_{\iota_1, \ldots, \iota_n \in \{-1,1\}, \prod_k \iota_k = -1} \sum_k \iota_k x_k. \qquad (2)$$

The maximum is taken over all combinations of ± 1 coefficients $\iota_1, \ldots, \iota_n$ containing odd numbers of -1's. The following is our main theorem.

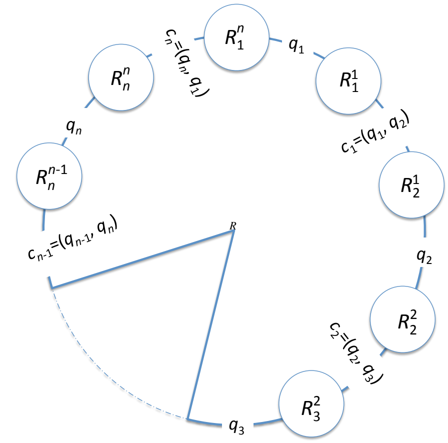

Figure 1. A schematic representation of a cyclic (single-cycle) system of rank $n > 1$. The properties $q_1, \ldots, q_n, q_1$ form a circle, any two successive properties $(q_i, q_{i\oplus 1})$ form a context, denoted c_i ($\oplus$ is clockwise shift $1 \mapsto 2 \mapsto \ldots \mapsto n \mapsto 1$). In a given context c_i the random variable representing q_i is denoted R_i^i, and the one representing $q_{i\oplus 1}$ is denoted $R_{i\oplus 1}^i$. Each property q_i therefore is represented by two random variables: R_i^i (when q_i is measured in context c_i) and $R_i^{i\ominus 1}$ (when q_i is measured in context $c_{i\ominus 1}$). The pair $(R_i^{i\ominus 1}, R_i^i)$ is a connection for q_i, and the pair $(R_i^i, R_{i\oplus 1}^i)$ represents the context c_i.

Theorem 2. *A cyclic system of rank $n > 1$ with dichotomic random variables (see Figure 1) has a maximally noncontextual description if and only if*

$$\mathsf{s}_1\left(\langle R_i^i R_{i\oplus 1}^i \rangle, 1 - |\langle R_i^i \rangle - \langle R_i^{i\ominus 1} \rangle| : i = 1, \ldots, n\right) \le 2n - 2 \qquad (3)$$

(s_1 here having $2n$ arguments, each entry being taken with $i = 1, \ldots, n$).

See Supplementary Material for the proof. In (3), $\langle R_i^i R_{i\oplus 1}^i \rangle$ are the quantum correlations observed within contexts, whereas $1 - |\langle R_i^i \rangle - \langle R_i^{i\ominus 1} \rangle|$ are the maximal values for the unobservable correlations within the couplings of connections. If the system is consistently connected, i.e., $\langle R_i^i \rangle = \langle R_i^{i\ominus 1} \rangle$, then these maximal values equal 1. By Corollary A10, the criterion (3) then reduces to the formula

$$\mathsf{s}_1\left(\langle R_i^i R_{i\oplus 1}^i \rangle : i = 1, \ldots, n\right) \le n - 2, \qquad (4)$$

well-known for $n = 3$ (the LG inequality in the form derived in Ref. [2]) and for $n = 4$ (CHSH inequalities [29]). For $n = 5$, (4) contains the KCBS inequality (which by Corollary A.11 is not only necessary but also sufficient for the existence of a maximally noncontextual description). Finally, for any even $n \ge 4$, inequality (4) contains the *chained Bell inequalities* studied in Refs. [43, 44]. It is known that for $n > 4$ the chained Bell inequalities are not criteria, the latter requiring many more inequalities [45–48].

Generally, some of the terms $\langle R_i^i \rangle - \langle R_i^{i\ominus 1} \rangle$ in (3) may be nonzero. Thus, in an LG system ($n = 3$), if inconsistency is due to "signaling in time" [18, 19], these may include $\langle R_2^2 \rangle - \langle R_2^1 \rangle$ and $\langle R_3^3 \rangle - \langle R_3^2 \rangle$ but not $\langle R_1^1 \rangle - \langle R_1^3 \rangle$, because q_1 cannot be influenced by later events. However, $\langle R_1^1 \rangle - \langle R_1^3 \rangle$ may be nonzero due to contextual biases in design, if something in the procedure of measuring q_1 is different depending on whether the next measurement is going to be of q_2 or q_3.

An application to experimental data.— To illustrate the applicability of our theory to real experiments, consider the data from the KCBS experiment of Ref. [20]. The experiment uses a single photon in a quantum overlap of three optical modes (paths) as an indivisible quantum system. Readout is performed through single-photon detectors that terminate the three paths. Context is chosen through "activation" of transformations, by rotating a waveplate that precedes each beamsplitter to change the behavior of two out of three paths. Each transformation leaves one path untouched, which serves as justification for consistent connectedness of the corresponding measurements, $\langle R_i^i \rangle = \langle R_i^{i\ominus 1} \rangle$, so that the target inequality is (4) for $n = 5$.

R_1^1 and R_1^5 are recorded in different experimental setups with zero or four polarizing beamsplitters "activated". These outputs have significantly different distributions: from Ref. [20] Table 1, $\langle R_1^1 \rangle = .136(6)$, $\langle R_1^5 \rangle = .172(4)$, and taking them as means and standard errors of 20 replications, the standard t-test with df $= 19$ is significant at 0.1%. Lapkiewicz et al., deal with this by introducing in (4) a correction term involving $\langle R_1^1 R_1^5 \rangle$. They estimate $\langle R_1^1 R_1^5 \rangle$ by identifying R_1^1 with R_1', an output measured in a separate context and in a special manner: instead of photon detections it is measured by blocking two paths early in the setup. While this results in a well-motivated experimental test, the identification of R_1' with R_1^1 involves additional assumptions [22, 23]. Furthermore, Lapkiewicz et al. have to discount the fact that the assumption $\langle R_i^i \rangle = \langle R_i^{i\ominus 1} \rangle$ can also be challenged for $i = 4$: the same t-test as above for $\langle R_4^4 \rangle = .122(4)$ and $\langle R_4^3 \rangle = .142(4)$ is significant at 1%. We see that the traditional approach adopted in Ref. [20] encounters considerable experimental and analytic difficulties due to the necessity of avoiding inconsistent connectedness.

Our theory allows one to analyze the data directly as found in the measurement record. It is convenient to do this by using the inequality

$$\mathsf{s}_1\left(\langle R_i^i R_{i\oplus 1}^i \rangle : i = 1, \dots, n \right) - \sum_{i=1}^{n} \left| \langle R_i^i \rangle - \langle R_i^{i\ominus 1} \rangle \right| \le n - 2, \tag{5}$$

which, by Corollary A9, follows from the criterion (3) [42]. One way of using it is to construct a conservative $100\,(1-\alpha)\,\%$ confidence interval with, say, $\alpha = 10^{-10}$ for the left-hand side of (5) with $n = 5$ and show that its lower endpoint exceeds $n - 2 = 3$. One can, e.g., construct 10 Bonferroni $100\,(1-\alpha/10)\,\%$ confidence intervals for each of the approximately normally distributed terms $\langle R_i^i R_{i\oplus 1}^i \rangle$ and $\langle R_i^i \rangle - \langle R_i^{i\ominus 1} \rangle$ ($i = 1, \dots, 5$), with respective error terms read or computed from Table 1 of Ref. [20], and then determine the range of (5). Treating each estimated term as the mean of 20 observations, we have $t_{1-\alpha/10}\,(19) < 14$ and so a conservative confidence interval for each term is given by $\pm 14 \times$ standard error. Using these intervals, we can calculate the conservative $100\left(1 - 10^{-10}\right)\%$ confidence interval for (5) as

$$\begin{aligned} &\mathsf{s}_1\Big(\overbrace{\langle R_1^1 R_2^1 \rangle}^{-.805\pm.028}, \overbrace{\langle R_2^2 R_3^2 \rangle}^{-.804\pm.042}, \overbrace{\langle R_3^3 R_4^3 \rangle}^{-.709\pm.042}, \overbrace{\langle R_4^4 R_5^4 \rangle}^{-.810\pm.028}, \overbrace{\langle R_5^5 R_1^5 \rangle}^{-.766\pm.028} \Big) \\ &- \Big| \underbrace{\langle R_1^1 \rangle - \langle R_1^5 \rangle}_{-.036\pm.101} \Big| - \Big| \underbrace{\langle R_2^2 \rangle - \langle R_2^1 \rangle}_{-.004\pm.140} \Big| - \Big| \underbrace{\langle R_3^3 \rangle - \langle R_3^2 \rangle}_{.006\pm.126} \Big| - \Big| \underbrace{\langle R_4^4 \rangle - \langle R_4^3 \rangle}_{-.020\pm.080} \Big| - \Big| \underbrace{\langle R_5^5 \rangle - \langle R_5^4 \rangle}_{-.006\pm.080} \Big| = [3.127, 4.062]. \end{aligned} \tag{6}$$

The system is contextual. The conclusion is the same as in Ref. [20], but we arrive at it by a shorter and more robust route.

Conclusion.— We have derived a criterion of (non)contextuality applicable to cyclic systems of arbitrary ranks. Even for consistently connected systems this criterion has not been previously known for ranks $n \ge 5$ (KCBS and higher-rank systems). However, it is the inclusion of inconsistently connected systems that is of special interest, because it makes the theory applicable to real experiments. A "system" is not just a system of properties being measured, but also a system of measurement procedures being used, with possible contextual biases and unaccounted-for interactions. Our analysis opens the possibility of studying contextuality without attempting to eliminate these first, whether by statistical analysis or by improved experimental procedure.

ACKNOWLEDGMENTS

This work is supported by NSF grant SES-1155956, AFOSR grant FA9550-14-1-0318, A. von Humboldt Foundation, and FQXi through Silicon Valley Community Foundation. We thank J. Acacio de Barros, Gary Oas, Samson Abramsky, Guido Bacciagaluppi, Adán Cabello, Andrei Khrennikov, and Lasse Leskelä for numerous discussions.

[1] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. Journal of Mathematics and Mechanics, 17:59–87, 1967.
[2] P. Suppes and M. Zanotti. When are probabilistic explanations possible? Synthese 48:191–199, 1981.
[3] A. Fine. Hidden variables, joint probability, and the Bell inequalities. Physical Review Letters 48:291–295, 1982.
[4] P. Kurzynski, R. Ramanathan, and D. Kaszlikowski. Entropic test of quantum contextuality. Physical Review Letters 109:020404, 2012.
[5] J.-Å. Larsson. A Kochen-Specker inequality. Europhysics Letters, 58(6):799–805, 2002.
[6] E.N. Dzhafarov and J.V. Kujala. All-possible-couplings approach to measuring probabilistic context. PLoS ONE 8(5):e61712. doi:10.1371/journal.pone.0061712, 2013.
[7] E.N. Dzhafarov and J.V. Kujala. No-Forcing and No-Matching theorems for classical probability applied to quantum mechanics. Foundations of Physics 44:248–265, 2014.
[8] E.N. Dzhafarov and J.V. Kujala. Embedding quantum into classical: contextualization vs conditionalization. PLoS One 9(3):e92818. doi:10.1371/journal.pone.0092818, 2014.
[9] E.N. Dzhafarov and J.V. Kujala. A qualified Kolmogorovian account of probabilistic contextuality. Lecture Notes in Computer Science 8369:201–212, 2014.
[10] E.N. Dzhafarov and J.V. Kujala. Contextuality is about identity of random variables. Physica Scripta T163, 014009, 2014 (available as arXiv:1405.2116).
[11] E.N. Dzhafarov and J.V. Kujala. The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. Frontiers in Psychology 1:151 doi:10.3389/fpsyg.2010.00151, 2010.
[12] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. Foundations of Physics 24:379–385, 1994.
[13] J. Bell. On the Einstein-Podolsky-Rosen paradox. Physics 1:195-200, 1964.
[14] J. Bell. On the problem of hidden variables in quantum mechanics. Review of Modern Physic 38:447-453, 1966.
[15] D. Bacon and B. F. Toner. Bell Inequalities with auxiliary communication. Physical Review Letters 90:157904, 2003.
[16] G. Adenier and A. Yu. Khrennikov. Is the fair sampling assumption supported by EPR experiments? Journal of Physics B: Atomic, Molecular and Optical Physics 40:131, 2007.
[17] A.J. Leggett and A. Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? Physical Review Letters, 54:857–860, 1985.
[18] J. Kofler and Č. Brukner. Condition for macroscopic realism beyond the Leggett-Garg inequalities. Physical Review A 87:052115, 2013.
[19] G. Bacciagaluppi. Leggett-Garg inequalities, pilot waves and contextuality. International Journal of Quantum Foundations 1, 1-17, 2015.
[20] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger. Experimental nonclassicality of an indivisible quantum system. Nature 474: 490–93, 2011.
[21] A.A. Klyachko, M.A. Can, S. Binicioğlu, and A.S. Shumovsky. Simple test for hidden variables in spin-1 systems. Physical Review Letters, 101(2):020403, 2008.
[22] J. Ahrens, E. Amselem, A. Cabello, and M. Bourennane. Two fundamental experimental tests of nonclassicality with qutrits. Scientific Reports 3, 2013.
[23] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger. Comment on "Two Fundamental Experimental Tests of Nonclassicality with Qutrits". arXiv:1305.5529, 2013.
[24] A. Peres. Quantum Theory: Concepts and Methods, Dordrecht: Kluwer, 1995.
[25] A. Cabello, J. Estebaranz, and G. Garcìa-Alcaine. Bell-Kochen-Specker Theorem: A Proof with 18 vectors", Physics Letters A 212:183–87, 1996.
[26] C. Simon, Č. Brukner, and A. Zeilinger. Hidden-variable theorems for real experiments. Physical Review Letters, 86(20):4427–4430, 2001.
[27] A. Cabello. Experimentally testable state-independent quantum contextuality. Physical Review Letters, 101(21):210401, 2008.
[28] S. Yu and C.H. Oh. State-Independent proof of Kochen-Specker theorem with 13 rays. Physical Review Letters, 108(3):030402, 2012.
[29] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt. Proposed experiment to test local hidden-variable theories. Physical Review Letters 23:880–884, 1969.
[30] J.F. Clauser and M.A. Horne. Experimental consequences of objective local theories. Physical Review D 10:526–535, 1974.
[31] Mermin, N. D. Hidden variables and the two theorems of John Bell. Rev. Mod. Phys. 65, 803–815 (1993).
[32] A. Winter. What does an experimental test of quantum contextuality prove or disprove? Journal of Physics A: Mathematical and Theoretical, 47(42):424031, 2014.
[33] A. Cabello. Simple explanation of the quantum violation of a fundamental inequality. Physical Review Letters, 110:060402, 2013.
[34] A. Cabello, S. Severini, and A. Winter. Graph-theoretic approach to quantum correlations. Physical Review Letters, 112:040401, 2014.
[35] E.N. Dzhafarov and J.V. Kujala. Generalizing Bell-type and Leggett-Garg-type inequalities to systems with signaling. arXiv:1407.2886, 2014.
[36] E.N. Dzhafarov, E.N., J.V. Kujala, and J.-Å. Larsson. Contextuality in three types of quantum-mechanical systems. Foundations of Physics 2015, DOI 10.1007/s10701-015-9882-9.
[37] A. Cabello, S. Severini, and A. Winter. (Non-)Contextuality of physical theories as an axiom. Physical Review Letters 112:040401, 2014.
[38] J. Cereceda. Quantum mechanical probabilities and general probabilistic constraints for Einstein–Podolsky–Rosen–Bohm experiments. Foundations of Physics Letters 13: 427–442, 2000.
[39] R. Ramanathan, A. Soeda, P. Kurzynski, and D. Kasznlikowski. Physical Review Letters 109:050404, 2012.
[40] P. Kurzynski, A. Cabello, and D. Kaszlikowski. Fundamental monogamy relation between contextuality and nonlocality. Physical Review Letters 112:100401, 2014.
[41] H. Thorisson. Coupling, Stationarity, and Regeneration. New York: Springer, 20002.
[42] This formula is in fact equivalent to (3), as conjectured in Ref. [36] and proved in Ref. [49].
[43] P. Pearle. Hidden-variable example based upon data rejection. Physical Review D2, 1418–1425, 1970.
[44] S. L. Braunstein and C. M. Caves. Wringing out better Bell inequalities. Annals of Physics 202, 22–56, 1990.

[45] R. F. Werner and M. M. Wolf. All-multipartite Bell-correlation inequalities for two dichotomic observables per site. Physical Review A 64, 032112, 2001.

[46] R. F. Werner and M. M. Wolf. Bell inequalities and entanglement. Quantum Information and Computation 1, 1-25, 2001.

[47] R. M. Basoalto and I. C. Percival. BellTest and CHSH experiments with more than two settings. Journal of Physics A: Mathematical & General 36, 7411–7423, 2003.

[48] Dzhafarov, E.N., Kujala, J.V.: Selectivity in probabilistic causality: Where psychology runs into quantum physics. Journal of Mathematical Psychology 56, 54-63, 2012.

[49] J.V. Kujala and E.N. Dzhafarov. Proof of a conjecture on contextuality in cyclic systems with binary variables. arXiv:1503.02181.

SUPPLEMENTARY MATERIAL TO "NECESSARY AND SUFFICIENT CONDITIONS FOR MAXIMAL NONCONTEXTUALITY IN A BROAD CLASS OF QUANTUM MECHANICAL SYSTEMS." PROOF OF THE MAIN CRITERION AND ITS CONSEQUENCES

The (non)contextuality criterion derived in this main text is a corollary to Theorem A.8 proved below. We first need the following simple result (see properties S1 and S2 formulated in section *Cyclic systems of dichotomic random variables*):

Lemma A.1. *In a system satisfying S1-S2, the physical properties $\{q_1,\ldots,q\}$ can be (re)indexed and arranged in one or more non-overlapping cycles*

$$\left(q_{11},\ldots,q_{1n_1},q_{11}\right),\left(q_{21},\ldots,q_{2n_2},q_{21}\right),\ldots,\left(q_{k1},\ldots,q_{kn_k},q_{k1}\right), \tag{A.1}$$

with $n_1+\ldots+n_k=n$ and $n_i>2$ $(i=1,\ldots,k)$, such that any two successive properties in each cycle form a context.

Proof. Apparent from Figure A.2. □

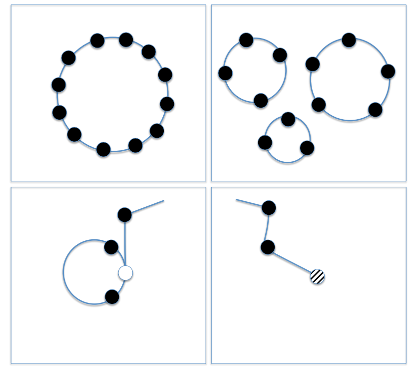

Figure A.2. In a system satisfying S1-S2 the properties being measured (represented by small circles) can be arranged in one (top left) or more (top right) cycles in which any two successive elements form a context. The bottom panels show that no other arrangements are possible: the patterned circle participates in less than two contexts, the open circle belongs to more than two contexts.

Our proof of Theorem A.8 uses the fact that the connections and context representations enter a circular system symmetrically, so that it is possible to view circular systems as a circular arrangement of random variables $A_1,\ldots,A_n,A_1$ in which any two successive variables have a joint distribution (see Figure A.3).

We need some auxiliary results. In addition to s_1 defined in the main text, we use function

$$\mathsf{s}_0\left(x_1,\ldots,x_n\right)=\max_{\iota_1,\ldots,\iota_n\in\{-1,1\},\prod_k\iota_k=1}\sum_k\iota_kx_k, \tag{A.2}$$

in which the maximum is taken over all combinations of ±1 coefficients $\iota_1,\ldots,\iota_n$ containing even numbers of -1's.

Lemma A.2. *For any $a_1,\ldots,a_n,b_1,\ldots,b_m\in\mathbb{R}$,*

$$\begin{aligned}&\mathsf{s}_1(a_1,\ldots,a_n,b_1,\ldots,b_m)\\&=\max\left\{\begin{array}{c}\mathsf{s}_0(a_1,\ldots,a_n)+\mathsf{s}_1(b_1,\ldots,b_m),\\ \mathsf{s}_1(a_1,\ldots,a_n)+\mathsf{s}_0(b_1,\ldots,b_m)\end{array}\right\},\end{aligned} \tag{A.3}$$

and

$$\begin{aligned}&\mathsf{s}_0(a_1,\ldots,a_n,b_1,\ldots,b_m)\\&=\max\left\{\begin{array}{c}\mathsf{s}_0(a_1,\ldots,a_n)+\mathsf{s}_0(b_1,\ldots,b_m),\\ \mathsf{s}_1(a_1,\ldots,a_n)+\mathsf{s}_1(b_1,\ldots,b_m)\end{array}\right\}.\end{aligned} \tag{A.4}$$

The proof is obvious.

Lemma A.3. *Jointly distributed ±1-valued random variables A and B with given expectations $\langle A\rangle,\langle B\rangle,\langle AB\rangle$ exist if and only if*

$$\begin{array}{c}-1\le\langle A\rangle\le1,\\-1\le\langle B\rangle\le1,\\|\langle A\rangle+\langle B\rangle|-1\le\langle AB\rangle\le1-|\langle A\rangle-\langle B\rangle|.\end{array} \tag{A.5}$$

Proof. For jointly distributed (A,B), from the table of probabilities

	$B=+1$	$B=-1$	
$A=+1$	r	$p-r$	p
$A=-1$	$q-r$	$1-p-q+r$	$1-p$
	q	$1-q$	

it is clear that

$$\max\left(p+q-1,0\right)\le r\le\min\left(p,q\right). \tag{A.6}$$

Since

$$\begin{aligned}\langle AB\rangle&=1-2p-2q+4r,\\\langle A\rangle&=2p-1,\\\langle B\rangle&=2q-1,\end{aligned}$$

straightforward algebra leads to (A.5). Conversely, expressing p,q,r through $\langle A\rangle,\langle B\rangle,\langle AB\rangle$, (A.5) implies (A.6), and then all probabilities in the table above are well defined. □

Lemma A.4. *Jointly distributed ±1-valued random variables A, B, and C with given expectations $\langle A\rangle$, $\langle B\rangle$, $\langle C\rangle$ $\langle AB\rangle$, $\langle AC\rangle$, $\langle BC\rangle$ exist if and only if these expectations satisfy Lemma A.3 and*

$$\mathsf{s}_1\left(\langle AB\rangle,\langle BC\rangle,\langle CA\rangle\right)\le1. \tag{A.7}$$

Proof. $\langle AB\rangle$, $\langle A\rangle$, and $\langle B\rangle$ satisfying Lemma A.3 uniquely determine $\Pr\left[A=1,B=1\right]$; and analogously for $\Pr\left[B=1,C=1\right]$ and $\Pr\left[C=1,A=1\right]$. A joint distribution of (A,B,C) is determined by 8 probabilities $p_{abc}=\Pr\left[A=a,B=b,C=c\right]$, $a,b,c\in\{-1,1\}$. It has the given expectations if and only if the 8 probabilities p_{abc} satisfy 7 equations

$$\textstyle\sum_{b,c}p_{1bc}=\Pr\left[A'=1\right],\quad\sum_{c}p_{11c}=\Pr\left[A'=1,B'=1\right],$$

$$\textstyle\sum_{a,c}p_{a1c}=\Pr\left[B'=1\right],\quad\sum_{a}p_{a11}=\Pr\left[B'=1,C'=1\right],$$

$$\textstyle\sum_{a,b}p_{ab1}=\Pr\left[C'=1\right],\quad\sum_{b}p_{1b1}=\Pr\left[C'=1,A'=1\right],$$

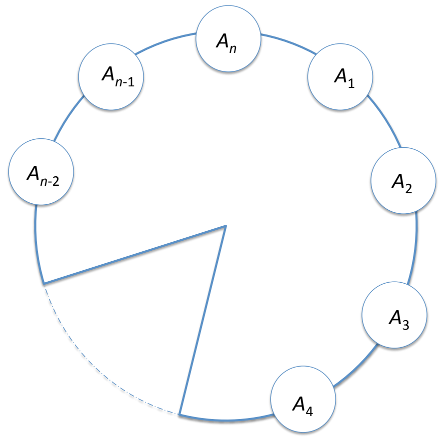

Figure A.3. The arrangement of random variables with no distinction made between context representations and connections: n random variables any successive two of which have a joint distribution. This structure is characterized in Theorem A.8.

$$\sum_{a,b,c} p_{abc} = 1.$$

The statement of the lemma obtains by any algorithm (facet enumeration and reduction) analogous to that described in Text S3 of Ref. [1]. □

Remark A.5. One can also obtain the proof by using Fine's theorem [2], presenting it as (using Fine's notation for the random variables)

$$\mathsf{s}_1 \left(\langle A_1B_1 \rangle , \langle A_1B_2 \rangle , \langle A_2B_1 \rangle , \langle A_2B_2 \rangle \right) \leq 2,$$

and then putting $A_1 = B_1 = A$, $B_2 = B$, and $A_2 = C$.

Lemma A.6. *Jointly distributed arbitrary random variables A, B, C with given 2-marginal distributions of (A, B) and (B, C) exist if and only if these 2-marginals agree for the distribution of B.*

Proof. The necessity is obvious. The sufficiency obtains by the Markov rule

$$\begin{aligned} &\Pr[A = a, B = b, C = c] \\ &= \Pr[C = c \mid B = b] \Pr[B = b \mid A = a] \Pr[A = a], \end{aligned}$$

for any possible values a, b, c of, respectively, A, B, C. □

Corollary A.7 (to Lemma A.6)**.** *Jointly distributed ± 1-valued random variables $A_1, \ldots, A_n$ with given expectations $\langle A_1 \rangle , \ldots , \langle A_n \rangle$, $\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle$ exist if and only if these expectations satisfy Lemma A.3.*

Theorem A.8. *Jointly distributed ± 1-valued random variables $A_1, \ldots, A_n$ $(n \geq 3)$ with given expectations*

$$\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle , \langle A_nA_1 \rangle \tag{A.8}$$

exist if and only if these expectations satisfy Lemma A.3 and

$$\mathsf{s}_1 \left(\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle , \langle A_nA_1 \rangle \right) \leq n - 2. \tag{A.9}$$

Proof. For $n = 3$ the statement follows from Lemma A.4.

Assume that the statement holds up to and including some $n \geq 3$. We will prove that

(i) jointly distributed ± 1-valued random variables $A_1, \ldots, A_{n+1}$ with given expectations

$$\langle A_1 \rangle , \ldots , \langle A_{n+1} \rangle , \langle A_1A_2 \rangle , \ldots , \langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle$$

exist if and only if

(ii) these expectations satisfy Lemma A.3 and

(iii) they satisfy

$$\mathsf{s}_1 \left(\langle A_1A_2 \rangle , \ldots , \langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle \right) \leq n - 1.$$

Since Statement (ii) is an obvious consequence of Statement (i), we only need to prove that if Statement (ii) is satisfied, then Statements (i) and (iii) are equivalent. So we assume Statement (ii).

By Lemma A.6, jointly distributed $(A_2, \ldots, A_{n-1}), (A_1, A_n), A_{n+1}$ exist if and only if there are jointly distributed $(A_2, \ldots, A_{n-1}), (A_1, A_n)$ and $(A_1, A_n), A_{n+1}$, with one and the same jointly distributed (A_1, A_n). Hence Statement (i) holds if and only if, for some $\langle A_nA_1 \rangle$ satisfying Lemma A.3, $(A_1, \ldots, A_n)$ exists with expectations $\langle A_1 \rangle , \ldots , \langle A_n \rangle , \langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle , \langle A_nA_1 \rangle$, and (A_n, A_{n+1}, A_1) exists with expectations $\langle A_1 \rangle , \langle A_n \rangle , \langle A_{n+1} \rangle , \langle A_1A_n \rangle , \langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle$. Therefore, by the induction hypothesis, Statement (i) holds if and only if

$$\begin{aligned} \mathsf{s}_1 \left(\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle , \langle A_nA_1 \rangle \right) &\leq n - 2, \\ \mathsf{s}_1 \left(\langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle , \langle A_nA_1 \rangle \right) &\leq 1. \end{aligned}$$

Applying now Lemma A.2 to these inequalities and adding the condition of Lemma A.3 for the consistency of $\langle A_nA_1 \rangle$ with $\langle A_n \rangle$ and $\langle A_1 \rangle$, we obtain the following system

$$\begin{aligned} &\mathsf{s}_1 \left(\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle \right) + \langle A_nA_1 \rangle \leq n - 2, \\ &\mathsf{s}_0 \left(\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle \right) - \langle A_nA_1 \rangle \leq n - 2, \\ &\mathsf{s}_1 \left(\langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle \right) + \langle A_nA_1 \rangle \leq 1, \\ &\mathsf{s}_0 \left(\langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle \right) - \langle A_nA_1 \rangle \leq 1, \\ &\left| \langle A_n \rangle + \langle A_1 \rangle \right| - 1 \leq \langle A_nA_1 \rangle \leq 1 - \left| \langle A_n \rangle - \langle A_1 \rangle \right| . \end{aligned}$$

Statement (i) holds if and only if this system is satisfied, for some real value of $\langle A_nA_1 \rangle$. And it is satisfied if and only if

$$\left\{ \begin{array}{c} \mathsf{s}_0 \left(\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle \right) - n + 2 \\ \mathsf{s}_0 \left(\langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle \right) - 1 \\ \left| \langle A_n \rangle + \langle A_1 \rangle \right| - 1 \end{array} \right.$$

$$\leq \left\{ \begin{array}{c} n - 2 - \mathsf{s}_1 \left(\langle A_1A_2 \rangle , \ldots , \langle A_{n-1}A_n \rangle \right) , \\ 1 - \mathsf{s}_1 \left(\langle A_nA_{n+1} \rangle , \langle A_{n+1}A_1 \rangle \right) , \\ 1 - \left| \langle A_n \rangle - \langle A_1 \rangle \right| , \end{array} \right.$$

with the inequality holding for any left-hand expression combined with any right-hand expression. The inequalities with matching rows are satisfied always: the first two because

$$\mathsf{s}_0(a_1\dots,a_n)+\mathsf{s}_1(a_1,\dots,a_n)=2\sum_{k=1}^{n}|a_k|-2\min_k|a_k|\le 2n-2$$

for $a_1,\dots,a_n\in[-1,1]$; the third one due to the fact that

$$|a+b|+|a-b|=\max(2|a|,2|b|)\le 2$$

for $a,b\in[-1,1]$. This leaves the following six inequalities

$$\mathsf{s}_0(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle)-n+2 \le 1-\mathsf{s}_1(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle),$$

$$\mathsf{s}_0(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle)-n+2\le 1-|\langle A_n\rangle-\langle A_1\rangle|,$$

$$\mathsf{s}_0(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle)-1 \le n-2-\mathsf{s}_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle),$$

$$|\langle A_n\rangle+\langle A_1\rangle|-1\le n-2-\mathsf{s}_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle),$$

$$|\langle A_n\rangle+\langle A_1\rangle|-1\le n-2-\mathsf{s}_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle),$$

$$|\langle A_n\rangle+\langle A_1\rangle|-1\le 1-\mathsf{s}_1(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle).$$

They simplify to

$$s_0(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle) +s_1(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle)\le n-1,$$

$$s_0(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle)+|\langle A_n\rangle-\langle A_1\rangle|\le n-1,$$

$$s_0(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle) +s_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle)\le n-1,$$

$$s_0(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle)+|\langle A_n\rangle-\langle A_1\rangle|\le 2,$$

$$s_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle)+|\langle A_n\rangle+\langle A_1\rangle|\le n-1,$$

$$s_1(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle)+|\langle A_n\rangle+\langle A_1\rangle|\le 2,$$

and we combine pairs of inequalities using Lemma A.2 to obtain

$$s_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle,\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle)\le n-1, \tag{A.10}$$

$$s_1(\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle,\langle A_n\rangle,\langle A_1\rangle)\le n-1, \tag{A.11}$$

$$s_1(\langle A_nA_{n+1}\rangle,\langle A_{n+1}A_1\rangle,\langle A_n\rangle,\langle A_1\rangle)\le 2. \tag{A.12}$$

These three inequalities are satisfied if and only if Statement (i) holds. In particular, Statement (i) implies (A.10), and this completes the proof by induction of the necessity part of the theorem: for any $n>1$, if $A_1,\dots,A_n$ are jointly distributed with expectations (A.8) then these expectations satisfy (A.9) (and Lemma A.3).

Now, Corollary A.7 implies that a joint distribution of $A_1,\dots,A_n$ with expectations $\langle A_1\rangle,\dots,\langle A_n\rangle$, $\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle$ (satisfying Lemma A.3) always exists. If we close this chain into a cycle by introducing a constant variable $A_{n+1}\equiv 1$, we get $n+1$ jointly distributed variables with expectations $\langle A_1\rangle,\dots,\langle A_n\rangle,\langle A_{n+1}\rangle=1$, $\langle A_1A_2\rangle,\dots,\langle A_{n-1}A_n\rangle,\langle A_nA_{n+1}\rangle=\langle A_n\rangle,\langle A_{n+1}A_1\rangle=\langle A_1\rangle$. Applying to it the just established necessary part of the theorem, we conclude that (A.11) always holds. Similarly, considering the chain A_n,A_{n+1},A_1 (whose joint distribution always exists) and adding the constant variable $A'\equiv 1$ to close the chain into a cycle, the necessary condition implies (A.12) with $A'\equiv 1$. Thus, (A.12) also holds always, leaving just (A.10) as the equivalent condition for Statement (i). □

Proof of the main criterion (Theorem 4). From Theorem A.8, contexts $\{(R_i^i,R_{i\oplus 1}^i):i=1,\dots,n\}$ and connections $\{(R_i^i,R_i^{i\ominus 1}):i=1,\dots,n\}$ with specified expectations $\{\langle R_i^iR_{i\oplus 1}^i\rangle:i=1,\dots,n\}$ and $\{\langle R_i^{i\ominus 1}R_i^i\rangle:i=1,\dots,n\}$ (subject to Lemma A.3) can be imposed a joint distribution upon if and only if

$$\mathsf{s}_1\left(\langle R_i^iR_{i\oplus 1}^i\rangle,\langle R_i^{i\ominus 1}R_i^i\rangle:i=1,\dots,n\right)\le 2n-2. \tag{A.13}$$

As the variables of the connection $(R_i^i,R_i^{i\ominus 1})$ are dichotomic, the probability of them being equal can be written as $\Pr\left[R_i^i=R_i^{i\ominus 1}\right]=\left(1+\langle R_i^{i\ominus 1}R_i^i\rangle\right)/2$ and so this probability is maximized if and only if the expectation $\langle R_i^{i\ominus 1}R_i^i\rangle$ is maximized. By Lemma A.3, $1-\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|$ is the maximum possible value of $\langle R_i^{i\ominus 1}R_i^i\rangle$ given the distributions of R_i^i and $R_i^{i\ominus 1}$ (determined by $\langle R_i^i\rangle$ and $\langle R_i^{i\ominus 1}\rangle$). The statement of the theorem now follows from Definition 2. □

Corollary A.9 (to Theorem 4)**.** *A cyclic system of rank $n>1$ with dichotomic random variables has a maximally noncontextual description only if*

$$\mathsf{s}_1\left(\langle R_i^iR_{i\oplus 1}^i\rangle:i=1,\dots,n\right)-\sum_{i=1}^{n}\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|\le n-2. \tag{A.14}$$

Proof. Using Lemma A.2,

$$\begin{aligned}2n-2&\ge \mathsf{s}_1\left(\langle R_i^iR_{i\oplus 1}^i\rangle,1-\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|:i=1,\dots,n\right)\\ &\ge \mathsf{s}_1\left(\langle R_i^iR_{i\oplus 1}^i\rangle:i=1,\dots,n\right)\\ &\quad+\mathsf{s}_0\left(1-\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|:i=1,\dots,n\right),\end{aligned}$$

and

$$\begin{aligned}&\mathsf{s}_0\left(1-\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|:i=1,\dots,n\right)\\ &\ge\sum_{i=1}^{n}\left(1-\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|\right)\\ &=n-\sum_{i=1}^{n}\left|\langle R_i^i\rangle-\langle R_i^{i\ominus 1}\rangle\right|.\end{aligned}$$

□

Corollary A.10 (to Theorem 4.)**.** *A cyclic consistently-connected system of rank $n > 1$ with dichotomic random variables has a maximally noncontextual description if and only if*

$$\mathsf{s}_1\left(\left\langle R_i^i R_{i\oplus 1}^i\right\rangle : i = 1, \ldots, n\right) \leq n - 2. \tag{A.15}$$

Proof. For consistently-connected systems, the main criterion has the form

$$2n - 2 \geq \mathsf{s}_1\Big(\left\langle R_1^1 R_2^1\right\rangle, \ldots, \left\langle R_n^n R_1^n\right\rangle, \underbrace{1, \ldots, 1}_{n \text{ times}}\Big). \tag{A.16}$$

The form A.15 follows from the easily verifiable general formula

$$\mathsf{s}_1\left(x_1, \ldots, x_k\right) = \sum_{i=1}^{k} |x_i| - 2[x_1 \cdots x_k > 0] \min\left(|x_1|, \ldots, |x_k|\right),$$

where [...] is the Iverson bracket, equal to 1 if the predicate within it is true, and zero otherwise, □

Corollary A.11 (to Theorem 4)**.** *A cyclic consistently-connected system of rank $n = 5$ with dichotomic random variables and with*

$$\Pr\left[R_i^i = 1, R_{i\oplus 1}^i = 1\right] = 0, \quad i = 1, \ldots, 5,$$

has a maximally noncontextual description if and only if the original KCBS inequality holds,

$$K = \sum_{i=1}^{5} p_i \leq 2, \tag{A.17}$$

where $p_i = \Pr\left[R_i^i = 1\right] = \Pr\left[R_i^{i\ominus 1} = 1\right]$, $i = 1, \ldots, 5$.

The expression (A.17) for K follows from (A.15) and $\left\langle R_i^i R_{i\oplus 1}^i\right\rangle = 1 - 2\left(p_i + p_{i\oplus 1}\right)$. The proof that (A.17) and (A.15) are equivalent is obtained by considering linear combinations $L = \sum_{k=1}^{5} \iota_k \left[1 - 2\left(p_k + p_{k\oplus 1}\right)\right]$ with 1, 3, and 5 negative ι_k's, in accordance with the definition of s_1, and showing that paired inequalities $L > 3, K \leq 2$ and $L \leq 3, K > 2$ have no solutions satisfying $p_i \geq 0$ and $p_i + p_{i\oplus 1} \leq 1$, $i = 1, \ldots, 5$.

---

[1] E.N. Dzhafarov and J.V. Kujala. All-possible-couplings approach to measuring probabilistic context. PLoS ONE 8(5):e61712. doi:10.1371/journal.pone.0061712, 2013.

[2] A. Fine. Hidden variables, joint probability, and the Bell inequalities. Physical Review Letters 48:291–295, 1982.