

# Measuring Observable Quantum Contextuality

J. Acacio de Barros<sup>1</sup>, Ehtibar N. Dzhafarov<sup>2</sup>, Janne V. Kujala<sup>3</sup>, Gary Oas<sup>4</sup>

<sup>1</sup>San Francisco State University  
barros@sfsu.edu

<sup>2</sup>Purdue University  
ehtibar@purdue.edu

<sup>3</sup>University of Jyväskylä  
jvk@iki.fi

<sup>3</sup>Stanford University  
oas@stanford.edu

**Abstract.** Contextuality is a central property in comparative analysis of classical, quantum, and supercorrelated systems. We examine and compare two well-motivated approaches to contextuality. One approach (“contextuality-by-default”) is based on the idea that one and the same physical property measured under different conditions (contexts) is represented by different random variables. The other approach is based on the idea that while a physical property is represented by a single random variable irrespective of its context, the joint distributions of the random variables describing the system can involve negative (quasi-)probabilities. We show that in the Leggett-Garg and EPR-Bell systems, the two measures essentially coincide.

## 1 Introduction

Contextuality is a key feature of quantum systems, as no noncontextual hidden-variable theory exists that is consistent with quantum theory. This feature has been at the core of recent research in quantum information, such as attempts to identify the underlying principles for the quantum boundary. Despite its importance, there seem to be no universally accepted measure of contextuality (see, e.g., the different approaches in Refs. [6, 12, 15–20, 23, 25, 29, 30, 32, 40]). Here, we consider and compare two measures inspired by the idea that contextuality means the impossibility of finding a joint probability distribution (jpd) for different sets of random variables with some elements in common. One measure (denoted below by  $\Delta_{\min}$ ) is based on extended sets of context-indexed random variables; its precursors can be found in Refs. [12, 14, 17, 32, 36, 40, 41] and in its current form it is presented in Refs. [15, 18–21, 30, 31]. The other measure (denoted below by  $\Gamma_{\min}$ ) is based on negative (quasi-)probabilities dating back to Dirac, and recently explored in connection to contextuality in Refs. [1, 7, 10, 11, 34, 38].

As an example of contextuality, let there be three properties of a system,  $P$ ,  $Q$ , and  $R$ , whose measurement outcomes are represented by the random variables  $\mathbf{P}$ ,  $\mathbf{Q}$ , and  $\mathbf{R}$ <sup>1</sup>. Assume we can never observe  $P$ ,  $Q$ , and  $R$  simultaneously, but only in pairwise combinations,  $(\mathbf{P}, \mathbf{Q})$ ,  $(\mathbf{P}, \mathbf{R})$ , or  $(\mathbf{Q}, \mathbf{R})$ . We may think of each pair as recorded under a different experimental condition providing a context. The system exhibits contextuality if one cannot find a jpd of  $(\mathbf{P}, \mathbf{Q}, \mathbf{R})$  that agrees with the observed distributions of  $(\mathbf{P}, \mathbf{Q})$ ,  $(\mathbf{P}, \mathbf{R})$ , and  $(\mathbf{Q}, \mathbf{R})$  as its marginals. The two approaches to be considered in this paper deal with this situation differently. The negative probabilities (NP) approach relaxes the notion of a jpd by allowing some (unobservable) joint probabilities for  $(\mathbf{P}, \mathbf{Q}, \mathbf{R})$  to be negative. The “contextuality-by-default” (CbD) approach treats random variables recorded under different conditions as different “by default”, so that, e.g., property  $P$  in the context of experiment  $(\mathbf{P}, \mathbf{Q})$  is represented by some random variable  $\mathbf{P}_A$ , and in the context  $(\mathbf{P}, \mathbf{R})$  by another random variable,  $\mathbf{P}_B$ . Denoting the three contexts by  $A, B, C$ , this yields three pairs of contextually labeled random variables,  $(\mathbf{P}_A, \mathbf{Q}_A)$ ,  $(\mathbf{P}_B, \mathbf{R}_B)$ , and  $(\mathbf{Q}_C, \mathbf{R}_C)$ , and in the CbD approach the joint distribution imposed on them allows, say,  $\mathbf{P}_A$  and  $\mathbf{P}_B$  to be unequal with some probability.

Here we compare the NP and CbD approaches applied to the simplest contextual case possible, with three pairwise correlated random variables, and to the standard EPR-Bell experiment. We show that for such examples the two measures of contextuality coincide (although they differ for more complex systems). The details of proofs and computations used in the main text are presented in Appendix.

## 2 Negative Probabilities (NP)

Using our above example, with  $\mathbf{P}, \mathbf{Q}, \mathbf{R}$  observed in pairs, in the NP approach one ascribes to the vector  $(\mathbf{P}, \mathbf{Q}, \mathbf{R})$  a joint quasi-distribution by means of assigning to each possible combination  $w = (p, q, r)$  a real number  $\mu(w)$  (possibly negative), such that

$$\begin{aligned} \sum_r \mu(w) &= \Pr[\mathbf{P} = p, \mathbf{Q} = q], \\ \sum_q \mu(w) &= \Pr[\mathbf{P} = p, \mathbf{R} = r], \\ \sum_p \mu(w) &= \Pr[\mathbf{Q} = q, \mathbf{R} = r]. \end{aligned} \tag{1}$$

Such  $\mu$  exists if and only if the no-signaling condition (built into EPR paradigms with spacelike separation) is satisfied [2, 3, 34], i.e., the distribution of, say,  $\mathbf{P}$  is the same in  $(\mathbf{P}, \mathbf{Q})$  and in  $(\mathbf{P}, \mathbf{R})$ .<sup>2</sup> The numbers  $\mu(w)$  can then be interpreted as quasi-probabilities of events  $\{w\}$ , with the quasi-probability of any other event (subset of  $w$  values) being computed by additivity, inducing thereby a signed

<sup>1</sup> This example has the same structure as the Leggett-Garg system [33]. A special version of it was examined by Suppes and Zanotti [39], and also by Specker [37].

<sup>2</sup> No-signaling condition is a fundamental limitation of any approach with noncontextually labeled random variables, including NP.

measure [24] on the set of all events. The quasi-probability of the entire set of  $w$  will then be necessarily equal to unity, because, e.g.,

$$1 = \sum_{p,q} \Pr[\mathbf{P} = p, \mathbf{Q} = q] = \sum_w \mu(w). \quad (2)$$

The function  $\mu$  is generally not unique. In our approach [7, 10] we restrict the class of possible  $\mu$  to those as close as possible to a proper jpd by requiring that the L1 norm of the probability distribution, defined by  $M = \sum_w |\mu(w)|$ , be minimized. This ensures that if the class of all possible  $\mu$  satisfying (1) contains proper probability distributions, the chosen  $\mu$  will have to be one of them. Since in this case  $|\mu(w)| = \mu(w)$  for all  $w$ , the minimum of  $M$  is 1. If (and only if) no proper probability distribution exists, then the minimum of  $M$  exceeds 1. As a result, the smallest possible value  $\Gamma_{\min}$  of  $M - 1$  can be taken as a measure of contextuality.

### 3 Contextuality-by-Default (CbD)

A more direct approach to contextuality [12, 14–21, 30, 31] is to posit that the identity of a random variable is determined by all systematically recorded conditions under which it is observed. Thus, in  $(\mathbf{P}_A, \mathbf{Q}_A)$ ,  $(\mathbf{P}_B, \mathbf{R}_B)$ , and  $(\mathbf{Q}_C, \mathbf{R}_C)$  of our example, any random variable in any of the pairs is *a priori* different from and stochastically unrelated to any random variable in any other pair [15, 17], but a jpd can always be imposed on the six random variables. In other words, one can always assign probability masses  $\lambda$  to  $v = (p_A, p_B, q_A, q_C, r_B, r_C)$  in such a way that

$$\begin{aligned} \sum_{p_B, q_C, r_B, r_C} \lambda(v) &= \Pr[\mathbf{P}_A = p_A, \mathbf{Q}_A = q_A], \\ \sum_{p_A, q_A, q_C, r_C} \lambda(v) &= \Pr[\mathbf{P}_B = p_B, \mathbf{R}_B = r_B], \\ \sum_{p_A, p_B, q_A, r_B} \lambda(v) &= \Pr[\mathbf{Q}_C = q_C, \mathbf{R}_C = r_C]. \end{aligned} \quad (3)$$

The noncontextuality hypothesis for  $\mathbf{P}_A, \mathbf{Q}_A, \mathbf{R}_B$  and  $\mathbf{P}_B, \mathbf{Q}_C, \mathbf{R}_C$  is that among these jpds  $\lambda$  we can find at least one for which  $\Pr[\mathbf{P}_A \neq \mathbf{P}_B] = \Pr[\mathbf{Q}_A \neq \mathbf{Q}_C] = \Pr[\mathbf{R}_B \neq \mathbf{R}_C] = 0$ , which is equivalent to  $\Delta = \Pr[\mathbf{P}_A \neq \mathbf{P}_B] + \Pr[\mathbf{Q}_A \neq \mathbf{Q}_C] + \Pr[\mathbf{R}_B \neq \mathbf{R}_C] = 0$ . Such a jpd need not exist, and then the smallest possible value  $\Delta_{\min}$  of  $\Delta$  for which a jpd of  $(\mathbf{P}_A, \mathbf{P}_B, \mathbf{Q}_A, \mathbf{Q}_C, \mathbf{R}_B, \mathbf{R}_C)$  exists can be taken as a measure of contextuality.<sup>3</sup>

The CdB approach has its precursors in the literature: various aspects of the contextual indexation of random variables and probabilities of the kind shown are considered in Refs. [12, 17, 26–28, 32, 36, 40, 41]. The principal difference, however, is in the use of minimization of  $\Delta$  under the assumption that a jpd exists. This is a well-defined mathematical problem, solvable in principle for any set of distributions observed empirically. We will now compare and interrelate the two approaches, NP and CdB, by applying them to the Leggett-Garg and the EPR-Bell setups.

<sup>3</sup> This formulation is predicated on no-signaling, which we assume throughout this paper. CdB has been generalized to situations when this condition is violated [4, 18–21, 30, 31].

## 4 Leggett-Garg

Let us consider Leggett and Garg's  $\pm 1$ -valued random variables,  $\mathbf{Q}_1$ ,  $\mathbf{Q}_2$ , and  $\mathbf{Q}_3$  [33]. Applying the NP approach, we seek signed probabilities  $\mu$  for  $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$  that are consistent with the observed correlations  $\langle \mathbf{Q}_i \mathbf{Q}_j \rangle$  and individual expectations  $\langle \mathbf{Q}_i \rangle$ , with the smallest possible value of the L1 norm  $M \equiv \sum_w |\mu(w)|$ , where  $w$  denotes all possible combinations of values  $(q_1, q_2, q_3)$  for  $(\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3)$ . Here, we use the standard notation  $\langle \cdot \rangle$  for the expectation operator. This problem can be easily solved, as we only have  $2^3$  atomic elements  $w$ :  $(1, 1, 1)$ ,  $(1, 1, -1)$ ,  $\dots$ ,  $(-1, -1, -1)$ . Thus, for  $\mathbf{Q}_1$ ,  $\mathbf{Q}_2$ , and  $\mathbf{Q}_3$ , the minimal L1 norm  $1 + \Gamma_{\min}$  satisfies

$$\Gamma_{\min} = \max \left\{ 0, -\frac{1}{2} + \frac{1}{2} S_{LG} \right\}, \quad (4)$$

where  $S_{LG}$  is defined as

$$S_{LG} \equiv \max_{\#^- = 1, 3} \{ \pm \langle \mathbf{Q}_1 \mathbf{Q}_2 \rangle \pm \langle \mathbf{Q}_1 \mathbf{Q}_3 \rangle \pm \langle \mathbf{Q}_2 \mathbf{Q}_3 \rangle \}, \quad (5)$$

where each  $\pm$  in the expression should be replaced with  $+$  or  $-$ , and  $\#^-$  indicates the possible numbers of minuses. Notice that  $S_{LG} \leq 1$ , which is equivalent to  $\Gamma_{\min} = 0$ , is a necessary and sufficient condition for the existence of a proper jpd.

Turning now to the Cbd approach, we create a set of six random variables

$$\mathbf{Q}_{1,2}, \mathbf{Q}_{1,3}, \mathbf{Q}_{2,1}, \mathbf{Q}_{2,3}, \mathbf{Q}_{3,1}, \mathbf{Q}_{3,2}, \quad (6)$$

each indexed by the measurement conditions under which it is recorded: for any two random variables recorded at moments  $t_i$  and  $t_j$ , with  $i < j$ , the  $\mathbf{Q}_{i,j}$  designates the earlier variable and  $\mathbf{Q}_{j,i}$  the later one. We have thus three pairs of variables with known jpd:

$$(\mathbf{Q}_{1,2}, \mathbf{Q}_{2,1}), (\mathbf{Q}_{1,3}, \mathbf{Q}_{3,1}), (\mathbf{Q}_{2,3}, \mathbf{Q}_{3,2}). \quad (7)$$

A jpd can always be constructed for these pairs (e.g., they can always be connected as stochastically independent pairs), but we seek a jpd with the smallest value  $\Delta_{\min}$  of

$$\Delta = \Pr[\mathbf{Q}_{1,2} \neq \mathbf{Q}_{1,3}] + \Pr[\mathbf{Q}_{2,1} \neq \mathbf{Q}_{2,3}] + \Pr[\mathbf{Q}_{3,1} \neq \mathbf{Q}_{3,2}]. \quad (8)$$

A classical joint exists for  $\mathbf{Q}_1$ ,  $\mathbf{Q}_2$ , and  $\mathbf{Q}_3$  (no contextuality) if and only if a joint exists for (7) with  $\Delta = 0$ . The more we depart from the classical joint, the larger the minimum value  $\Delta_{\min}$ . Thus,  $\Delta_{\min}$  can serve as a measure of contextuality.

Requiring a jpd consistent with (7) means to assign a probability to each of the  $2^6$  possible values of these random variables,

$$\mathbf{Q}_{1,2} = \pm 1, \mathbf{Q}_{1,3} = \pm 1, \dots, \mathbf{Q}_{3,2} = \pm 1, \quad (9)$$

constrained by being nonnegative and summing to the observed probabilities. For instance, the probabilities assigned to all combinations with  $\mathbf{Q}_{1,2} = 1$  and  $\mathbf{Q}_{2,1} = -1$  should sum to the observed  $\Pr[\mathbf{Q}_{1,2} = 1, \mathbf{Q}_{2,1} = -1]$ . A computer-assisted Fourier-Motzkin elimination algorithm gives the following analytic expression for the minimum value of  $\Delta$  consistent with the observable pairs (7):

$$\Delta_{\min} = \max \left\{ 0, -\frac{1}{2} + \frac{1}{2} S_{LG} \right\}. \quad (10)$$

This is a special case of the result in Ref. [19, 20, 30].

Comparing the general expressions (4) for  $\Gamma_{\min}$  and (10) for  $\Delta_{\min}$  we see that the two simply coincide:

$$\Delta_{\min} = \Gamma_{\min}. \quad (11)$$

## 5 EPR-Bell

We now turn to the EPR-Bell case where Alice and Bob have each two distinct settings, 1 and 2, corresponding to four observable random variables  $\mathbf{A}_1$ ,  $\mathbf{A}_2$ ,  $\mathbf{B}_1$ , and  $\mathbf{B}_2$ . This notation implicitly contains the assumption that the identity of Alice's measurements as random variables does not depend on Bob's settings, and vice versa. It is well known [22] that under the no-signaling conditions the existence of the jpd is equivalent to the CHSH inequalities being satisfied. Applying the NP approach, the minimal L1 norm of the probability distribution is given by [34]

$$\Gamma_{\min} = \max \left\{ 0, \frac{1}{2} S_{CHSH} - 1 \right\}, \quad (12)$$

where

$$S_{CHSH} = \max_{\# = -1, 3} \{ \pm \langle \mathbf{A}_{1,1} \mathbf{B}_{1,1} \rangle \pm \langle \mathbf{A}_{1,2} \mathbf{B}_{1,2} \rangle \pm \langle \mathbf{A}_{2,1} \mathbf{B}_{2,1} \rangle \pm \langle \mathbf{A}_{2,2} \mathbf{B}_{2,2} \rangle \}. \quad (13)$$

Here  $\Gamma_{\min} = 0$  corresponds to the CHSH inequalities, and  $\Gamma_{\min} > 0$  to contextuality.

Turning now to the CbD approach, we have four pairs of random variables,

$$(\mathbf{A}_{1,1}, \mathbf{B}_{1,1}), (\mathbf{A}_{1,2}, \mathbf{B}_{1,2}), (\mathbf{A}_{2,1}, \mathbf{B}_{2,1}), (\mathbf{A}_{2,2}, \mathbf{B}_{2,2}). \quad (14)$$

Here,  $\mathbf{A}_{i,j}$  denotes Alice's measurement under her setting  $i = 1, 2$  when Bob's setting is  $j = 1, 2$ , and analogously for  $\mathbf{B}_{i,j}$ . We seek a jpd with the smallest value  $\Delta_{\min}$  of

$$\Pr[\mathbf{A}_{1,1} \neq \mathbf{A}_{1,2}] + \Pr[\mathbf{A}_{2,1} \neq \mathbf{A}_{2,2}] + \Pr[\mathbf{B}_{1,1} \neq \mathbf{B}_{2,1}] + \Pr[\mathbf{B}_{1,2} \neq \mathbf{B}_{2,2}]. \quad (15)$$

No contextuality means  $\Delta_{\min} = 0$ . A computer assisted Fourier-Motzkin elimination algorithm yields (this is a special case of the result in Ref. [19, 20, 30])

$$\Delta_{\min} = \max \left\{ 0, \frac{1}{2} S_{CHSH} - 1 \right\}. \quad (16)$$

We have the same simple coincidence the two measures as in the case of the Leggett-Garg systems,

$$\Delta_{\min} = \Gamma_{\min}. \quad (17)$$

## 6 Final remarks

We have discussed two ways to measure contextuality. The direct approach, named Contextuality-by-Default (CbD), assigns to each random variable an index related to their context. If a system is noncontextual, a jpd can be imposed on the random variables so that any two of them representing the same property in different contexts always have the same values. If the system is contextual, the minimum value of  $\Delta$  in (10)-(16) across all possible jpd's has the interpretation of how close a variable can be in two different contexts: the larger the value the greater contextuality, zero representing a necessary and sufficient condition for no contextuality.

The other approach maintains the original set of random variables, but requires negative (quasi-)probabilities. This leads to nonmonotonicity (i.e., a set of outcomes can have a smaller probability than some of its proper subsets), which is a characteristic of quantum interference. The departure from a proper probability distribution is measured by  $\Gamma_{\min}$  in the minimum L1 norm  $1 + \Gamma_{\min}$ . Similar to the CbD approach, we use here a minimization principle that gives the closest probability distribution to an ideal (but impossible) jpd. The value of  $\Gamma_{\min}$  has the interpretation of how contextual the system is: a necessary and sufficient condition for no contextuality is  $\Gamma_{\min} = 0$ , and the larger the value of  $\Gamma_{\min}$ , the more contextual the system is.

As we have seen, in the case of EPR-Bell and Leggett-Garg systems the two approaches lead to simple coincidence,  $\Delta_{\min} = \Gamma_{\min}$ . The two measures,  $\Gamma_{\min}$  and  $\Delta_{\min}$ , can be computed for any given system, and they do not coincide for more complex systems. Thus, for a bi-partite system with three settings for Alice and for Bob our computations show that  $\Gamma_{\min}$  has a value of 1 for all Popescu-Rohrlich (PR) boxes [5,35], whereas for some PR boxes  $\Delta_{\min} = 2$  and for others  $\Delta_{\min} = 1$ . Still more complex systems exhibit still richer patterns of values for  $\Gamma_{\min}$  and  $\Delta_{\min}$ .

Of the two measures of contextuality,  $\Gamma_{\min}$  is computationally much simpler, as it involves fewer random variables and a simpler set of conditions (no non-negativity constraints). However, CbD has the advantage of being more general than NP, as it can include cases where no NP distributions exist due to violations of the no-signaling condition [18, 20, 30].

*Acknowledgments.* This work was supported by NSF grant SES-1155956, AFOSR grant FA9550-14-1-0318, and A. von Humboldt Foundation. The authors are grateful to Samson Abramsky, Guido Bacciagaluppi, Andrei Khrennikov, Jan-Åke Larsson, and Patrick Suppes for helpful discussions.

## A Proofs of statements

In this appendix, we describe how the analytic results of the main text were obtained for each of the expressions (16), (12), (10), and (4) of the main text.

### A.1 EPR-Bell: Contextuality-by-Default

Following the computations of Dzhafarov and Kujala [12, Text S3], or the more general formulation in Ref. [19], it can be shown that the observable distributions with probabilities given by the matrices

$$\begin{array}{c}
 \begin{array}{|c|} \hline \mathbf{A}_{1,1} = +1 \\ \hline \mathbf{A}_{1,1} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{B}_{1,1} = +1 & \mathbf{B}_{1,1} = -1 \\ \hline p_{1,1} & a_1 - p_{1,1} \\ \hline b_1 - p_{1,1} & 1 - a_1 - b_1 + p_{1,1} \\ \hline b_1 & 1 - b_1 \\ \hline \end{array}
 \begin{array}{|c|} \hline a_1 \\ \hline 1 - a_1 \\ \hline \end{array}
 \qquad
 \begin{array}{|c|} \hline \mathbf{A}_{1,2} = +1 \\ \hline \mathbf{A}_{1,2} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{B}_{1,2} = +1 & \mathbf{B}_{1,2} = -1 \\ \hline p_{1,2} & a_1 - p_{1,2} \\ \hline b_2 - p_{1,2} & 1 - a_1 - b_2 + p_{1,2} \\ \hline b_2 & 1 - b_2 \\ \hline \end{array}
 \begin{array}{|c|} \hline a_1 \\ \hline 1 - a_1 \\ \hline \end{array}
 \\
 \\
 \begin{array}{|c|} \hline \mathbf{A}_{2,1} = +1 \\ \hline \mathbf{A}_{2,1} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{B}_{2,1} = +1 & \mathbf{B}_{2,1} = -1 \\ \hline p_{2,1} & a_2 - p_{2,1} \\ \hline b_1 - p_{2,1} & 1 - a_2 - b_1 + p_{2,1} \\ \hline b_1 & 1 - b_1 \\ \hline \end{array}
 \begin{array}{|c|} \hline a_2 \\ \hline 1 - a_2 \\ \hline \end{array}
 \qquad
 \begin{array}{|c|} \hline \mathbf{A}_{2,2} = +1 \\ \hline \mathbf{A}_{2,2} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{B}_{2,2} = +1 & \mathbf{B}_{2,2} = -1 \\ \hline p_{2,2} & a_2 - p_{2,2} \\ \hline b_2 - p_{2,2} & 1 - a_2 - b_2 + p_{2,2} \\ \hline b_2 & 1 - b_2 \\ \hline \end{array}
 \begin{array}{|c|} \hline a_2 \\ \hline 1 - a_2 \\ \hline \end{array}
 \end{array}
 \tag{18}$$

are compatible with the connections

$$\begin{array}{c}
 \begin{array}{|c|} \hline \mathbf{A}_{1,1} = +1 \\ \hline \mathbf{A}_{1,1} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{A}_{1,2} = +1 & \mathbf{A}_{1,2} = -1 \\ \hline a_1 - \alpha_1 & \alpha_1 \\ \hline \alpha_1 & 1 - a_1 - \alpha_1 \\ \hline a_1 & 1 - a_1 \\ \hline \end{array}
 \begin{array}{|c|} \hline a_1 \\ \hline 1 - a_1 \\ \hline \end{array}
 \qquad
 \begin{array}{|c|} \hline \mathbf{B}_{1,1} = +1 \\ \hline \mathbf{B}_{1,1} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{B}_{2,1} = +1 & \mathbf{B}_{2,1} = -1 \\ \hline b_1 - \beta_1 & \beta_1 \\ \hline \beta_1 & 1 - b_1 - \beta_1 \\ \hline b_1 & 1 - b_1 \\ \hline \end{array}
 \begin{array}{|c|} \hline b_1 \\ \hline 1 - b_1 \\ \hline \end{array}
 \\
 \\
 \begin{array}{|c|} \hline \mathbf{A}_{2,1} = +1 \\ \hline \mathbf{A}_{2,1} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{A}_{2,2} = +1 & \mathbf{A}_{2,2} = -1 \\ \hline a_2 - \alpha_2 & \alpha_2 \\ \hline \alpha_2 & 1 - a_2 - \alpha_2 \\ \hline a_2 & 1 - a_2 \\ \hline \end{array}
 \begin{array}{|c|} \hline a_2 \\ \hline 1 - a_2 \\ \hline \end{array}
 \qquad
 \begin{array}{|c|} \hline \mathbf{B}_{1,2} = +1 \\ \hline \mathbf{B}_{1,2} = -1 \\ \hline \end{array}
 \begin{array}{|c|c|} \hline \mathbf{B}_{2,2} = +1 & \mathbf{B}_{2,2} = -1 \\ \hline b_2 - \beta_2 & \beta_2 \\ \hline \beta_2 & 1 - b_2 - \beta_2 \\ \hline b_2 & 1 - b_2 \\ \hline \end{array}
 \begin{array}{|c|} \hline b_2 \\ \hline 1 - b_2 \\ \hline \end{array}
 \end{array}
 \tag{19}$$

if and only if

$$\begin{aligned}
 & s_u(\langle \mathbf{A}_{1,1} \mathbf{B}_{1,1} \rangle, \langle \mathbf{A}_{1,2} \mathbf{B}_{1,2} \rangle, \langle \mathbf{A}_{2,1} \mathbf{B}_{2,1} \rangle, \langle \mathbf{A}_{2,2} \mathbf{B}_{2,2} \rangle) \\
 & \leq 6 - s_{1-u}(\langle \mathbf{A}_{1,1} \mathbf{A}_{1,2} \rangle, \langle \mathbf{A}_{2,1} \mathbf{A}_{2,2} \rangle, \langle \mathbf{B}_{1,1} \mathbf{B}_{2,1} \rangle, \langle \mathbf{B}_{1,2} \mathbf{B}_{2,2} \rangle),
 \end{aligned}
 \tag{20}$$

where  $u$  stands for 0 or 1, and

$$\begin{aligned}
 s_0(x_1, \dots, x_n) &= \max\{\pm x_1 \pm \dots \pm x_n : \text{even \# of -'s}\}, \\
 s_1(x_1, \dots, x_n) &= \max\{\pm x_1 \pm \dots \pm x_n : \text{odd \# of -'s}\}.
 \end{aligned}$$

Here we use the parameterization by the 12 expectation variables defined as (with  $i, j \in \{1, 2\}$ )

$$\langle \mathbf{A}_{i,j} \mathbf{B}_{i,j} \rangle = (4p_{ij} - 1) - (2a_i - 1) - (2b_j - 1),
 \tag{21}$$

$$\langle \mathbf{A}_{i,1} \mathbf{A}_{i,2} \rangle = 1 - 4\alpha_i = 1 - 2 \Pr [\mathbf{A}_{i,1} \neq \mathbf{A}_{i,2}], \quad (22)$$

$$\langle \mathbf{B}_{1,j} \mathbf{B}_{2,j} \rangle = 1 - 4\beta_j = 1 - 2 \Pr [\mathbf{B}_{1,j} \neq \mathbf{B}_{2,j}], \quad (23)$$

$$\langle \mathbf{A}_i \rangle = 2a_i - 1, \quad \langle \mathbf{B}_j \rangle = 2b_j - 1. \quad (24)$$

Writing the inequality (20) in terms of these expectations rather than in terms of probabilities is the most economic way of presenting the 128 non-trivial inequalities of the system, as the marginal probabilities  $a_1, a_2, b_1, b_2$  (or expectations  $\langle \mathbf{A}_1 \rangle, \langle \mathbf{A}_2 \rangle, \langle \mathbf{B}_1 \rangle, \langle \mathbf{B}_2 \rangle$ ) vanish in this form. However, it should be noted that in addition to these 128 inequalities, the form of the observed distributions and connections itself imposes further 28 trivial constraints on the 12 expectation variables of the system: the probabilities within each  $2 \times 2$  matrix in (18) and (19) should be nonnegative and sum to one. 16 of these trivial constraints pertain to the observed distributions and 12 to the connections. In terms of the expectations, these trivial constraints correspond to

$$-1 + |\langle \mathbf{A} \rangle + \langle \mathbf{B} \rangle| \leq \langle \mathbf{AB} \rangle \leq 1 - |\langle \mathbf{A} \rangle - \langle \mathbf{B} \rangle|, \quad (25)$$

for given marginals for each pair  $(\mathbf{A}, \mathbf{B})$  of random variables in (21)–(23). This expands to four inequalities for each of the observed distributions and to three inequalities for each of the connections (the two upper bounds in (25) coincide when  $\langle \mathbf{A} \rangle = \langle \mathbf{B} \rangle$ ). Although these trivial constraints can usually be assumed implicitly, it is important to keep them explicitly in the system for the next step.

Adding the equation

$$\begin{aligned} \Delta &= \Pr [\mathbf{A}_{1,1} \neq \mathbf{A}_{1,2}] + \Pr [\mathbf{A}_{2,1} \neq \mathbf{A}_{2,2}] + \Pr [\mathbf{B}_{1,1} \neq \mathbf{B}_{2,1}] + \Pr [\mathbf{B}_{1,2} \neq \mathbf{B}_{2,2}] \\ &= 2 - \frac{1}{2} (\langle \mathbf{A}_{1,1} \mathbf{A}_{1,2} \rangle + \langle \mathbf{A}_{2,1} \mathbf{A}_{2,2} \rangle + \langle \mathbf{B}_{1,1} \mathbf{B}_{2,1} \rangle + \langle \mathbf{B}_{1,2} \mathbf{B}_{2,2} \rangle) \end{aligned}$$

to the system and then eliminating the connection correlations  $\langle \mathbf{A}_{1,1} \mathbf{A}_{1,2} \rangle, \langle \mathbf{A}_{2,1} \mathbf{A}_{2,2} \rangle, \langle \mathbf{B}_{1,1} \mathbf{B}_{2,1} \rangle, \langle \mathbf{B}_{1,2} \mathbf{B}_{2,2} \rangle$  from the system using the Fourier–Motzkin elimination algorithm, we obtain the system

$$-1 + \frac{1}{2} S_{CHSH} \leq \Delta \leq 4 - \left[ -1 + \frac{1}{2} S_{CHSH} \right], \quad (26)$$

$$0 \leq \Delta \leq 4 - (|\langle \mathbf{A}_1 \rangle| + |\langle \mathbf{A}_2 \rangle| + |\langle \mathbf{B}_1 \rangle| + |\langle \mathbf{B}_2 \rangle|), \quad (27)$$

where we denote

$$S_{CHSH} = s_1(\langle \mathbf{A}_{1,1} \mathbf{B}_{1,1} \rangle, \langle \mathbf{A}_{1,2} \mathbf{B}_{1,2} \rangle, \langle \mathbf{A}_{2,1} \mathbf{B}_{2,1} \rangle, \langle \mathbf{A}_{2,2} \mathbf{B}_{2,2} \rangle) \quad (28)$$

as in the main text. This means that  $\Delta$  is compatible with the given observed probabilities if and only if the above inequalities are satisfied. Since the set of possible values of  $\Delta$  constrained by (26) and (27) is known to be nonempty, it follows that the minimum value of  $\Delta$  is always given by

$$\Delta_{\min} = \max \left\{ 0, \frac{1}{2} S_{CHSH} - 1 \right\}.$$

### A.2 EPR-Bell: Negative probabilities

The analogous result for the negative probabilities approach is that the observable distributions (18) are obtained as the marginals of some negative probability joint of  $\mathbf{A}_1 = \mathbf{A}_{1,1} = \mathbf{A}_{1,2}$ ,  $\mathbf{A}_2 = \mathbf{A}_{2,1} = \mathbf{A}_{2,2}$ ,  $\mathbf{B}_1 = \mathbf{B}_{1,1} = \mathbf{B}_{2,1}$ , and  $\mathbf{B}_2 = \mathbf{B}_{1,2} = \mathbf{B}_{2,2}$  given by

$$\Pr[\mathbf{A}_1 = a'_1, \mathbf{A}_2 = a'_2, \mathbf{B}_1 = b'_1, \mathbf{B}_2 = b'_2] = p^+(a'_1, a'_2, b'_1, b'_2) - p^-(a'_1, a'_2, b'_1, b'_2),$$

$a'_1, a'_2, b'_1, b'_2 \in \{1, -1\}$ , for some nonnegative functions  $p^+$  and  $p^-$  having a total probability mass value

$$M = \sum_{a'_1, a'_2, b'_1, b'_2} p^+(a'_1, a'_2, b'_1, b'_2) + p^-(a'_1, a'_2, b'_1, b'_2)$$

if and only if  $M \geq 1 + \Gamma_{\min}$ , where

$$\Gamma_{\min} = \max \left\{ 0, \frac{1}{2} S_{CHSH} - 1 \right\}.$$

The computations are similar to those of the CbD approach, but there are two general differences. First, in the CbD approach, the convex range of the possible observed and connection expectations (21)–(24) over the convex polytope of all possible joints is obtained by looking at these expectations at the  $2^8$  vertices defining the polytope of all joints and then applying a computer algorithm to find the set of inequalities delineating the extreme values of the expectations at these vertices. However, in the negative probabilities approach, the joint is represented by the  $2^4$  differences of the positive and negative components of the distribution and so, although these  $2 \cdot 2^4$  components are nonnegative as in the CbD approach, they do not need to sum to one. Hence, the joint is represented by a convex cone rather than a bounded polytope. Still, a convex cone is a special case of a general polytope and can be handled by the same algorithms that we have used in the CbD approach.

Second, we do not need to apply the Fourier–Motzkin elimination algorithm here as we have defined  $M$  directly by the representation of the joint so there are no extra variables we would need to eliminate. This difference, however, is not really a difference between the two approaches, as we could have done the same in the CbD approach as well: we could have defined  $\Delta$  directly based on the joint of all eight variables without explicitly defining the connection correlations (22)–(23), and then we would have obtained the result directly from the half-space representation, as we do in the negative probabilities approach.

### A.3 Leggett–Garg: Contextuality-by-Default

The results for Leggett–Garg  $\mathbf{Q}_1, \mathbf{Q}_2, \mathbf{Q}_3$  can be obtained in the same way as for the EPR-Bell systems. In the CbD approach, the observed correlations  $\langle \mathbf{Q}_{1,2} \mathbf{Q}_{2,1} \rangle$ ,  $\langle \mathbf{Q}_{1,3} \mathbf{Q}_{3,1} \rangle$ ,  $\langle \mathbf{Q}_{2,3} \mathbf{Q}_{3,2} \rangle$ ,  $\langle \mathbf{Q}_1 \rangle = \langle \mathbf{Q}_{1,2} \rangle = \langle \mathbf{Q}_{1,3} \rangle$ ,  $\langle \mathbf{Q}_2 \rangle = \langle \mathbf{Q}_{2,1} \rangle =$

$\langle \mathbf{Q}_{3,2} \rangle, \langle \mathbf{Q}_3 \rangle = \langle \mathbf{Q}_{3,1} \rangle = \langle \mathbf{Q}_{3,2} \rangle$  are consistent with the connection correlations  $\langle \mathbf{Q}_{1,2} \mathbf{Q}_{1,3} \rangle, \langle \mathbf{Q}_{2,1} \mathbf{Q}_{2,3} \rangle, \langle \mathbf{Q}_{3,1} \mathbf{Q}_{3,2} \rangle$  if and only if these connection correlations are realizable with the given marginals (i.e., each correlation  $\langle \mathbf{AB} \rangle$  has to satisfy  $-1 + |\langle \mathbf{A} \rangle + \langle \mathbf{B} \rangle| \leq \langle \mathbf{AB} \rangle \leq 1 - |\langle \mathbf{A} \rangle - \langle \mathbf{B} \rangle|$  as discussed in the EPR-Bell case above) and satisfy

$$s_u (\langle \mathbf{Q}_{1,2} \mathbf{Q}_{2,1} \rangle, \langle \mathbf{Q}_{1,3} \mathbf{Q}_{3,1} \rangle, \langle \mathbf{Q}_{2,3} \mathbf{Q}_{3,2} \rangle) + s_{1-u} (\langle \mathbf{Q}_{1,2} \mathbf{Q}_{1,3} \rangle, \langle \mathbf{Q}_{2,1} \mathbf{Q}_{2,3} \rangle, \langle \mathbf{Q}_{3,1} \mathbf{Q}_{3,2} \rangle) \leq 4, \quad (29)$$

These two inequalities (with  $u$  standing for 0 or 1) expand to 32 linear inequalities and there are 21 trivial constraints.

Denoting

$$\begin{aligned} \Delta &= \Pr [\mathbf{Q}_{1,2} \neq \mathbf{Q}_{1,3}] + \Pr [\mathbf{Q}_{2,1} \neq \mathbf{Q}_{2,3}] + \Pr [\mathbf{Q}_{3,1} \neq \mathbf{Q}_{3,2}] \\ &= \frac{3}{2} - \frac{1}{2} (\langle \mathbf{Q}_{1,2} \mathbf{Q}_{1,3} \rangle + \langle \mathbf{Q}_{2,1} \mathbf{Q}_{2,3} \rangle + \langle \mathbf{Q}_{3,1} \mathbf{Q}_{3,2} \rangle) \end{aligned}$$

and eliminating the connection correlations from the system using the Fourier–Motzkin algorithm, we obtain the system

$$-\frac{1}{2} + \frac{1}{2} S_{LG} \leq \Delta \leq 3 - \left[ -\frac{1}{2} + \frac{1}{2} S_{LG}^0 \right], \quad (30)$$

$$0 \leq \Delta \leq 3 - |\langle \mathbf{Q}_1 \rangle| - |\langle \mathbf{Q}_2 \rangle| - |\langle \mathbf{Q}_3 \rangle|, \quad (31)$$

where we denote

$$\begin{aligned} S_{LG} &= s_1 (\langle \mathbf{Q}_{1,2} \mathbf{Q}_{2,1} \rangle, \langle \mathbf{Q}_{1,3} \mathbf{Q}_{3,1} \rangle, \langle \mathbf{Q}_{2,3} \mathbf{Q}_{3,2} \rangle), \\ S_{LG}^0 &= s_0 (\langle \mathbf{Q}_{1,2} \mathbf{Q}_{2,1} \rangle, \langle \mathbf{Q}_{1,3} \mathbf{Q}_{3,1} \rangle, \langle \mathbf{Q}_{2,3} \mathbf{Q}_{3,2} \rangle). \end{aligned}$$

That is,  $\Delta$  is consistent with the observed probabilities if and only if the above inequalities are satisfied. It follows that the minimum value of  $\Delta$  is given by

$$\Delta_{\min} = \max \left\{ 0, \frac{1}{2} S_{LG} - \frac{1}{2} \right\}.$$

#### A.4 Leggett–Garg: Negative probabilities

With the same additional comments as in the negative probability calculations for the EPR-Bell case, our calculations show that the observable probabilities  $\Pr [\mathbf{Q}_{12} = q_1, \mathbf{Q}_{21} = q_2], \Pr [\mathbf{Q}_{13} = q_1, \mathbf{Q}_{31} = q_3], \Pr [\mathbf{Q}_{23} = q_2, \mathbf{Q}_{32} = q_3], q_1, q_2, q_3 \in \{0, 1\}$ , can be obtained as the marginals of a negative probability jpd of  $\mathbf{Q}_1 = \mathbf{Q}_{1,2} = \mathbf{Q}_{1,3}, \mathbf{Q}_2 = \mathbf{Q}_{2,1} = \mathbf{Q}_{2,3}$ , and  $\mathbf{Q}_3 = \mathbf{Q}_{3,1} = \mathbf{Q}_{3,2}$  with the total probability mass of  $M$  if and only if  $M \geq 1 + \Gamma_{\min}$ , where

$$\Gamma_{\min} = \max \left\{ 0, \frac{1}{2} S_{LG} - \frac{1}{2} \right\}.$$

## References

1. Abramsky, S., Brandenburger, A.: The sheaf-theoretic structure of non-locality and contextuality. *New J. Phys.* 13, 113036 (2011).
2. Abramsky, S., Brandenburger, A.: An operational interpretation of negative probabilities and no-signalling models. In *Horizons of the Mind. A Tribute to Prakash Panangaden*, Lecture Notes in Computer Science No. 8464, edited by F. van Breugel, E. Kashefi, C. Palamidessi, and J. Rutten (Springer Int. Pub., 2014) pp. 59–75.
3. Al-Safi, S., Short, A.: Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities. *Phys. Rev. Lett.* 111, 170403 (2013).
4. Bacciagaluppi, G.: Leggett-Garg inequalities, pilot waves and contextuality. *Int. J. Quant. Found.*, 1(1), 1-17 (2015).
5. Brunner, N., Scarani, V., Gisin, N.: Bell-type inequalities for non-local resources. *J. Math. Phys.* 47, 112101 (2006).
6. Chaves, R., Fritz, T.: Entropic approach to local realism and noncontextuality. *Phys. Rev. A* 85, 032113 (2012).
7. de Barros, J.A.: Decision making for inconsistent expert judgments using negative probabilities. *Lecture Notes in Computer Science* 8369, 257 (2014).
8. de Barros, J.A., Suppes, P.: Inequalities for dealing with detector inefficiencies in Greenberger-Horne-Zeilinger type experiments. *Phys. Rev. Lett.* 84, 793 (2000).
9. de Barros, J.A., Suppes, P.: Probabilistic results for six detectors in a three-particle GHZ experiment. In *Chance in Physics, Lecture Notes in Physics*, Vol. 574, edited by J. Bricmont, D. Durr, M. C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi (2001) p. 213.
10. de Barros, J. A., Oas, G., Suppes, P.: Negative probabilities and counterfactual reasoning on the double-slit experiment. In J.-Y. Beziau, D. Krause, & J. B. Arenhart (Eds.), *Conceptual Clarification: Tributes to Patrick Suppes (1992-2014)*. London: College Publications (2015).
11. de Barros, J.A., Oas, G.: Negative probabilities and counter-factual reasoning in quantum cognition. *Phys. Scripta* T163, 014008 (2014).
12. Dzhafarov, E., Kujala, J.: All-Possible-Couplings approach to measuring probabilistic context. *PLoS ONE* 8, e61712 (2013).
13. Dzhafarov, E., Kujala, J.: No-Forcing and No-Matching theorems for classical probability applied to quantum mechanics. *Found. Phys.*, 44, 248-265 (2014).
14. Dzhafarov, E., Kujala, J.: Embedding quantum into classical: Contextualization vs conditionalization. *PLoS One* 9, e92818 (2014).
15. Dzhafarov, E., Kujala, J.: contextuality is about identity of random variables. *Phys. Scripta* T163, 014009 (2014).
16. Dzhafarov, E., Kujala, J.: random variables recorded under mutually exclusive conditions: Contextuality-by-Default. In H. Liljenström (Ed.) *Advances in Cognitive Neurodynamics IV* (pp. 405-410) (2015).
17. Dzhafarov, E., Kujala, J.: A qualified kolmogorovian account of probabilistic contextuality. *Lecture Notes in Computer Science* 8369, 201 (2014).
18. Dzhafarov, E., Kujala, J.: probabilistic contextuality in EPR/Bohm-type systems with signaling allowed. *arXiv:1406.0243 [quant-ph, q-bio]* (2014).
19. Dzhafarov, E., Kujala, J.: generalizing Bell-type and Leggett-Garg-type inequalities to systems with signaling. *arXiv:1407.2886 [quant-ph]* (2014).
20. Dzhafarov, E., Kujala, J., Larsson, J.-Å.: contextuality in three types of quantum-mechanical systems. *Found. Phys.* DOI 10.1007/s10701-015-9882-9 (2014).

21. Dzhafarov, E., Kujala, J., Larsson, J.-Å.: Contextuality-by-Default: A brief overview of ideas, concepts, and terminology. arXiv:1504.00530 [quant-ph] (2015).
22. Fine, A.: Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* 48, 291 (1982).
23. Grudka, A., Horodecki, K., Horodecki, M., Horodecki, P., Horodecki, R., Joshi, P., Kłobus, W., Wójcik, A.: Quantifying contextuality. *Phys. Rev. Lett.* 112, 120401 (2014).
24. Halmos, P.: *Measure Theory*. Springer-Verlag, New York, NY (1974).
25. Kleinmann, M., Gühne, O., Portillo, J., Larsson, J.-Å., Cabello, A.: Memory cost of quantum contextuality. *New J. Phys.* 13, 113011 (2011).
26. Khrennikov, A. (2005). The principle of supplementarity: A contextual probabilistic viewpoint to complementarity, the interference of probabilities, and the incompatibility of variables in quantum mechanics. *Found. Phys.* 35, 1655-1693.
27. Khrennikov, A. (2008). Bell-Boole inequality: Nonlocality or probabilistic incompatibility of random variables? *Entropy* 10, 19-32.
28. Khrennikov, A. (2009). *Contextual approach to quantum formalism*. Springer: Berlin.
29. Kurzyński, P., Ramanathan, R., Kaszlikowski, D.: entropic test of quantum contextuality. *Phys. Rev. Lett.* 109, 020404 (2012).
30. Kujala, J., Dzhafarov, E., J.-Å. Larsson, J.-Å.: Necessary and Sufficient Conditions for Maximal Noncontextuality in a Broad Class of Quantum Mechanical Systems. *Phys. Rev. Lett.* (in press).
31. Kujala, J., Dzhafarov, E. Proof of a conjecture on contextuality in cyclic systems with binary variables. arXiv:1503.02181 [quant-ph] (2015).
32. Larsson, J.-Å.: A Kochen-Specker inequality. *EPL (Europhysics Letters)* 58, 799 (2002).
33. Leggett, A., Garg, A.: Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Phys. Rev. Lett.* 54, 857 (1985).
34. Oas, G., de Barros, J.A. Carvalhaes, C.: Exploring non-signalling polytopes with negative probability. *Phys. Scripta T163*, 014034 (2014).
35. Popescu, S., Rohrlich, D.: Nonlocality as an axiom. *Found. Phys.* 24, 379 (1994).
36. Simon, C., Brukner, Č., Zeilinger, A.: Hidden-variable theorems for real experiments. *Phys. Rev. Lett.*, 86(20):4427–4430 (2001).
37. Specker, E.P.: The logic of propositions which are not simultaneously decidable. In *The Logico-Algebraic Approach to Quantum Mechanics*, The University of Western Ontario Series in Philosophy of Science No. 5a, edited by C. A. Hooker (Springer Netherlands, 1975) pp. 135–140.
38. Spekkens, R.W.: Negativity and contextuality are equivalent notions of nonclassicality. *Phys. Rev. Lett.* 101, 020401 (2008).
39. Suppes, P., Zanotti, M.: When are probabilistic explanations possible? *Synthese* 48, 191 (1981).
40. Svozil, K.: How much contextuality? *Natural Computing* 11, 261 (2012).
41. Winter, A.: What does an experimental test of quantum contextuality prove or disprove? *J. Phys. A: Math. Theor.*, 47, 424031 (2014).