

Original Paper

The Role of (Non)contextuality in Bell's Theorems from the Perspective of an Operational Modeling Framework

Michael L. Ulrey

4 Longvue Ct, Mt Vernon, Ohio 43050, USA.

<https://orcid.org/0000-0002-5143-973X>

E-mail: mulrey@hotmail.com

Received: 10 March 2022 / Accepted: 21 March 2022 / Published: 29 March 2022

Abstract: A novel approach for analyzing "classical" alternatives to quantum mechanics for explaining the statistical results of an EPRB-like experiment is proposed. This perspective is top-down instead of bottom-up. Rather than beginning with an inequality derivation, a hierarchy of model types is constructed, each distinguished by appropriately parameterized conditional probabilities. This hierarchy ranks the "classical" model types in terms of their ability to reproduce QM statistics or not. The analysis goes beyond the usual consideration of model types that "fall short" (i.e., satisfy all of the CHSH inequalities) to ones that are "excessive" (i.e., not only violate CHSH but even exceed a Tsirelson bound). This approach clearly shows that noncontextuality is the most general property of an operational model that blocks replication of at least some QM statistical predictions. Factorizability is naturally revealed to be a special case of noncontextuality. The same is true for the combination of remote context independence and outcome determinism (RCI+OD). It is noncontextuality that determines the dividing line between "classical" model instances that satisfy the CHSH inequalities and those that don't. Outcome deterministic operational models are revealed to be the "building blocks" of all the rest, including quantum mechanical, noncontextual, and contextual ones. The set of noncontextual model instances is exactly the convex hull of all 16 RCI+OD model instances, and furthermore, the set of all model instances, including all QM ones, is equal to the convex hull of the 256 OD model instances. It is shown that, under a mild assumption, the construction of convex hulls of finite ensembles of OD model instances is (mathematically) equivalent to the traditional hidden variables approach. Via the introduction of

operational models that possess outcome and measurement "predictability", a new perspective is gained on the impossibility of faster-than-light transfer of information in an EPRB experiment. Finally, many plots and figures, some of which appear to be new, provide visual affirmation of many of the results.

Keywords: Bell's theorem(s); Bell-Kochen-Specker theorem; hidden variables; convex hull; outcome determinism; locality; non-contextuality; predictability; correlation plots; modeling hierarchy

1. Introduction

Consider Fig. 1, which shows sets of 3-tuples of correlations attached to different probabilistic models of an EPRB experiment (the fourth correlation is fixed at -1 , $-\frac{1}{2}$, 0 , $\frac{1}{2}$, or 1 , from left to right in each row). The first four rows show correlations corresponding to the assumptions, respectively, of Hilbert-space quantum mechanics (QM), remote context independence together with outcome determinism (RCI+OD), factorizability (or Bell locality), and noncontextuality. The remaining rows show pairwise comparisons of selected plots.

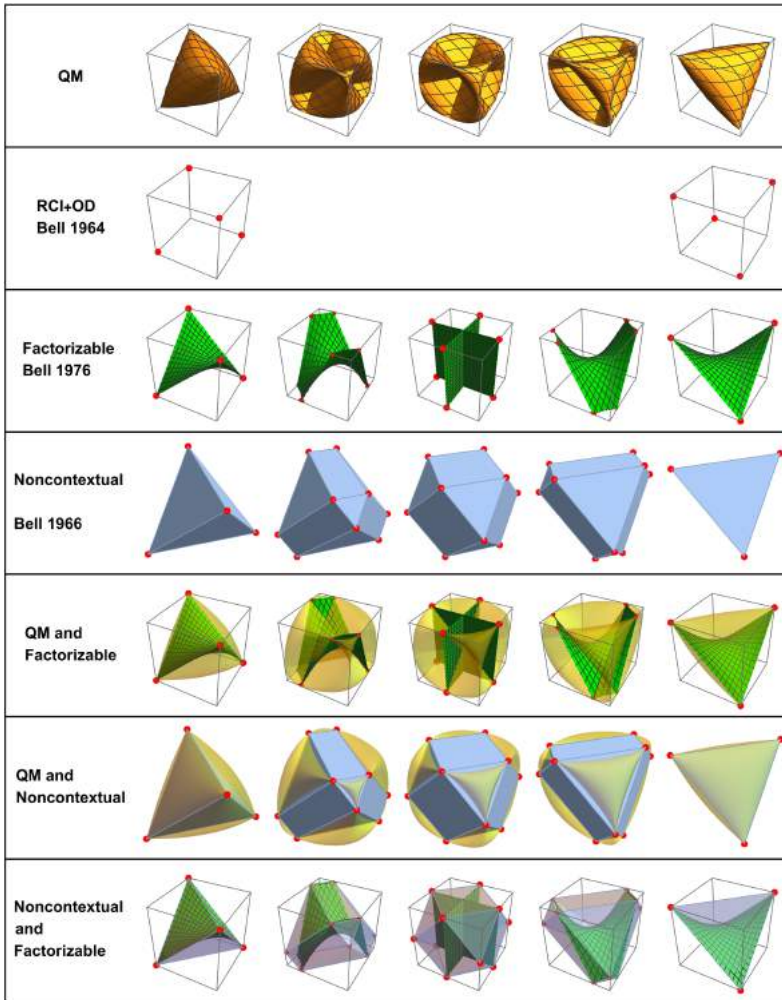
Rows 2-4 and the last row strongly suggest that RCI+OD and factorizability are simply stronger notions of noncontextuality. Notice how the factorizable correlations (green surfaces) fit entirely inside the noncontextual correlations (blue solids) and the RCI+OD correlations (red dots) are "vertices" of both of them (in the first and last columns only, where the correlation in the 4th dimension is fixed at -1 and $+1$, respectively). In other words, in terms of sets of correlations,

$$RCI + OD \subset Factorizable \subset Noncontextual. \quad (1)$$

Furthermore, all of the correlations in each of these categories satisfy all of the CHSH inequalities.¹

¹ Clauser, Horne, Shimony, and Holt. See [1] and [2].

Figure 1. 3D slices through sets of 4D correlation vectors corresponding to various probabilistic assumptions, and combinations thereof. In each row, exactly one correlation is fixed in each of the five plots at the values $-1, -\frac{1}{2}, 0, \frac{1}{2},$ and 1 , from left to right. The plots in the noncontextual row (4) correspond to the polyhedra defined by the CHSH inequalities (see Corollary 3). The frame for each plot is the "correlation cube" $[-1, 1] \times [-1, 1] \times [-1, 1]$. See Appendix A.2 for more details.



Wiseman [3] points out that many of the debates surrounding "Bell's theorem" stem from the fact that there are really two Bell's theorems. In [4], Bell showed that RCI+OD were sufficient to derive his inequality, and in [5], he showed that factorizability is also sufficient to derive a similar inequality. Both inequalities are inconsistent with at least some statistical predictions of QM.

Mermin [6] compared Bell's 1964 paper [4] to the earlier so-called

"Bell-Kochen-Specker" (BKS) paper [7]². In the BKS paper, Bell dismantled von Neumann's "silly" argument against a hidden variables theory and replaced it with his own argument, which he then self-criticized because of the tacit assumption that the measurement of a given observable should yield the same value, regardless of what other measurements are made simultaneously. Although he did not name it at the time, this assumption later came to be called "noncontextuality" – again see [6] and also the noncontextuality discussion in [9].

The increasing level of generality of Eq. 1 thus reflects the content of Bell's three papers [4,5,7], mathematically if not chronologically (also noted in Fig. 1 in the descriptions to the left of rows 2, 3, and 4). They each indicate Bell's ideas at the time about what kind of hidden variables theory is blocked from explaining at least some QM statistical predictions. This culminates in the BKS paper [7], which effectively shows there can be no noncontextual (NC) hidden variables theory that is consistent with all QM statistical predictions.

The central importance of noncontextuality was already evident in the pioneering work of Fine [10,11]. For example in [10] he said: *"Finally, I believe that Proposition (1) – conjoined with the other two – shows what hidden variables and the Bell inequalities are all about; namely, imposing requirements to make well-defined precisely those probability distributions for noncommuting observables whose rejection is the very essence of quantum mechanics."*

Abramsky and Bradenburger [12], exploring nonlocality and contextuality in the very general framework of sheaf theory, echo this conclusion by saying: *"We show that contextuality, and non-locality as a special case, correspond exactly to obstructions to global sections."*

In Held [13], the Kochen-Specker theorem and noncontextuality are linked by the statement: *"The second important no-go theorem against HV theories is the theorem of Kochen and Specker (KS) which states that, given a premise of noncontextuality (to be explained presently) certain sets of QM observables cannot consistently be assigned values at all (even before the question of their statistical distributions arises)."*

Klyachko, et. al. [14] emphasize the important implications for quantum computing, given that the contextuality of QM is more fundamental than nonlocality: *"However, for quantum computation the magic ability of entanglement to bypass constraints imposed by the so-called classical realism is far more important. The latter is understood here as the existence of hidden parameters, or equivalently a joint probability distribution of all involved quantum observables."* See also [15] for a specific example of better-than-classical

² Due to a publication delay, this paper did not appear until 1966. Kochen and Specker independently proved the same result by different means in [8].

performance of quantum protocols using polarization qubits, enabled by preparation contextuality.

For deep connections of noncontextuality to graph theoretic concepts, see for example [16, 17], and for a surprising connection to relational database theory, see [18].

In this paper, within the context of a Bell/Aspect-like experiment, many of the concepts surrounding Bell's theorems are organized into a hierarchy of parameterized operational model types, together with sets of their instantiations, that represent various concepts, such as remote context independence, remote outcome independence, factorizability, outcome determinism, noncontextuality, and combinations thereof. In this parameterized modeling framework, it is possible to "rank" the different model types in their ability to replicate QM predictions while retaining certain "classical" features (or not). This top-down approach provides a broader perspective than mere manipulation of inequalities. It lends intuition into why certain known results are true, reveals a few underappreciated facts, renders many proofs almost trivial, leads to a more intuitive, simpler formulation of "hidden variables", provides a different insight into why the experiment cannot be used for faster-than-light communication, and provides many relevant plots and figures, some of which appear to be new.

In particular, this parameterized operational modeling paradigm can be used to confirm the suggestiveness of Fig. 1, namely that the hierarchy of noncontextual correlations in Eq. 1 reflects a similar relationship among the model types representing these properties. Both RCI+OD and factorizability are merely special cases of noncontextuality. If one assumes "locality", either as RCI in the combination RCI+OD, or as factorizability, one is implicitly assuming noncontextuality. Noncontextuality is the most general "classical" notion among these three that blocks complete duplication of QM statistical behaviors. In fact, the 8 CHSH inequalities become a "witness" for noncontextuality. That is, if noncontextuality is assumed, all 8 CHSH inequalities hold and (almost) conversely, if a model instance (not necessarily noncontextual) has correlations that satisfy all of the CHSH inequalities, *there exists* a (possibly different) noncontextual model instance with those same correlations.

2. The operational model types

The basic motivation behind a "hidden variables" theory is to posit an alternative explanation of quantum mechanics that explains its behaviors in a "classical" way. Some "hidden variables" theories can be eliminated straight away if they lead to mathematical models that produce statistical predictions that are inconsistent with QM. The strategy in this paper is to consider a specific class of statistical models, which will be called "operational" models, and determine if certain properties that they possess are consistent (or not) with QM.

An operational model type is characterized by a set of 16 conditional probabilities

$$P(A = s, B = t | M_A = u, M_B = v)$$

for Alice and Bob's outcomes A and B given their measurement choices (context) M_A and M_B , where $s, t = \pm 1$ and $u, v = 1$ or 2 . The novel approach in this paper is to represent these conditional probabilities with symbolic parameters

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_{16}),$$

and then suitably "re-parameterize" this *generic* representation to characterize various *specialized* model types.

2.1. Generic

The generic representation of an *operational* (that is, statistical) model of an EPRB experiment is shown in Table 1.

Table 1. The **generic** model type, where the γ_k 's represent the 16 conditional probabilities $P(A = s, B = t | M_A = u, M_B = v)$, where $s, t = \pm 1$ denote Alice and Bob's measurement outcomes and $u, v = 1$ or 2 indicate their measurement choices (context), respectively. Each row consists of nonnegative real numbers summing to 1.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	γ_1	γ_2	γ_3	γ_4
$(1, 2)$	γ_5	γ_6	γ_7	γ_8
$(2, 1)$	γ_9	γ_{10}	γ_{11}	γ_{12}
$(2, 2)$	γ_{13}	γ_{14}	γ_{15}	γ_{16}

In this table, each γ_k represents a conditional probability, that is,

$$\gamma_k = P(A = s, B = t | M_A = u, M_B = v)$$

for appropriate $s, t = \pm 1, u, v = 1, 2$, and $k = 1, 2, \dots, 16$. For example,

$$\gamma_7 = P(A = 1, B = -1 | M_A = 1, M_B = 2),$$

and similarly for the other 15 entries. Table 1 corresponds to what Abramsky and

Bradenburger [12] call an *empirical* model, or in this case, a $(2, 2, 2)$ *Bell-type scenario*³. This is just one example in their comprehensive analysis of non-locality and contextuality in the language of sheaf theory. Whereas [12] encompasses general (n, k, l) Bell-type scenarios (and more), the focus here is on the $(2, 2, 2)$ case only.

This model type is called "generic" because the γ_k 's are unspecified, except for the following restrictions, which apply to each and every model type in this paper, whether generic or specialized.

Definition 1. *The measurement context constraints say that*

$$\begin{aligned}\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 &= 1, \\ \gamma_5 + \gamma_6 + \gamma_7 + \gamma_8 &= 1, \\ \gamma_9 + \gamma_{10} + \gamma_{11} + \gamma_{12} &= 1, \\ \gamma_{13} + \gamma_{14} + \gamma_{15} + \gamma_{16} &= 1.\end{aligned}\tag{2}$$

Definition 2. *The term model type refers to any table like Table 1, which includes a specification of the outcomes, context, and conditional probabilities $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$. These conditional probabilities might be written in terms of a different set of symbolic parameters.*

Definition 3. *A model instance is a model type in which the symbolic parameters have been assigned specific numeric values.*

Definition 4. *A conditional probability vector, or cpv, is an ordered set of 16 nonnegative real numbers $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ that satisfy the measurement context constraints of Def. 1.*

In this paper, the terms "model type" and "cpv" will be used interchangeably, when they involve symbolic parameters, and "model instance" and "cpv" used interchangeably, when they involve numeric parameters.

2.2. Correlations, s -functions, and the CHSH inequalities

Before proceeding with the definitions of the more specialized model types, it is convenient to define the correlations, s -functions, and CHSH inequalities in terms of the generic model type parameters. These definitions will then apply *mutatis mutandis* to any other specialized model type by a suitable change of parameters according to the table that defines the relevant model type. The results for all model types are summarized in Appendix A.

³ 2 observers, 2 measurements per observer, 2 outcomes per measurement.

Definition 5. Define the four **correlations** (w, x, y, z) of any generic model type (and hence any other model type in this paper) in terms of its parameters $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$:

$$\begin{aligned} w &= \gamma_1 - \gamma_2 - \gamma_3 + \gamma_4, \\ x &= \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8, \\ y &= \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12}, \\ z &= \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}. \end{aligned} \tag{3}$$

Now define the four **s-functions**⁴ of the correlations (w, x, y, z) :

$$\begin{aligned} s_1 &= -w + x + y + z, \\ s_2 &= w - x + y + z, \\ s_3 &= w + x - y + z, \\ s_4 &= w + x + y - z. \end{aligned} \tag{4}$$

Finally the 8 **CHSH inequalities** are:

$$-2 \leq s_1, s_2, s_3, s_4 \leq 2. \tag{5}$$

The following two matrices will come in handy for reducing lengthy lists of equations/inequalities to a more compact matrix format.

Definition 6. The **correlation matrix** C turns a set of generic parameters $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ into a set of correlations $(w, x, y, z)^T = C\gamma^T$. See Eq. 3.

$$C = \begin{pmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{pmatrix} \tag{6}$$

Definition 7. The matrix S turns a set of correlations (w, x, y, z) into the corresponding

⁴ The use of the letter s for these expressions is taken from Aspect [2].

s -functions $(s_1, s_2, s_3, s_4)^T = S(w, x, y, z)^T$. See Eq. 4.

$$S = \begin{pmatrix} -1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{pmatrix}. \tag{7}$$

The CHSH inequalities (Eq. 5) can now be written in terms of the generic model type parameters $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ as

$$-2 \leq SC\gamma^T \leq 2. \tag{8}$$

Inequalities like these are meant to be interpreted as applying componentwise to the vector in the middle. One of the main tasks is to characterize the sets of generic parameters $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ that either satisfy all or violate at least one of the inequalities represented by Eq. 8.

Table 2. The correlations and s -functions for the generic type. Not only is it possible to violate a CHSH inequality, it is possible to exceed a Tsirelson bound $(\pm 2\sqrt{2})$.

w	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4$
x	$\gamma_5 - \gamma_6 - \gamma_7 + \gamma_8$
y	$\gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12}$
z	$\gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_1	$-\gamma_1 + \gamma_2 + \gamma_3 - \gamma_4 + \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 + \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} + \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_2	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 - \gamma_5 + \gamma_6 + \gamma_7 - \gamma_8 + \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} + \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_3	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 + \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 - \gamma_9 + \gamma_{10} + \gamma_{11} - \gamma_{12} + \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_4	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 + \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 + \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} - \gamma_{13} + \gamma_{14} + \gamma_{15} - \gamma_{16}$
Min and Max	-4 and 4

It should be clear from Table 2 that, since the γ_k 's are nonnegative real numbers that sum to 4, the s -functions of the generic model type can be as high as +4 or as low as -4.

2.3. QM

Table 3 shows the model type derived from standard Hilbert-space quantum mechanics

(QM)⁵. It is clear that this is an example of the generic model type, since the sum of row entries is 1 for all four rows.

Table 3. The **QM** model type is defined by the following 16 expressions in terms of the parameters θ_k for $k = 1, 2, 3, 4$. They represent the conditional probabilities $P(A = s, B = t | M_A = u, M_B = v)$ of Alice and Bob’s outcomes $s, t = \pm 1$ given their measurement choices (context) $u, v = 1$ or 2 , respectively. Note the sum of values in each of the four rows is 1.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\frac{1}{2} \cos^2 \theta_1$	$\frac{1}{2} \sin^2 \theta_1$	$\frac{1}{2} \sin^2 \theta_1$	$\frac{1}{2} \cos^2 \theta_1$
$(1, 2)$	$\frac{1}{2} \cos^2 \theta_2$	$\frac{1}{2} \sin^2 \theta_2$	$\frac{1}{2} \sin^2 \theta_2$	$\frac{1}{2} \cos^2 \theta_2$
$(2, 1)$	$\frac{1}{2} \cos^2 \theta_3$	$\frac{1}{2} \sin^2 \theta_3$	$\frac{1}{2} \sin^2 \theta_3$	$\frac{1}{2} \cos^2 \theta_3$
$(2, 2)$	$\frac{1}{2} \cos^2 \theta_4$	$\frac{1}{2} \sin^2 \theta_4$	$\frac{1}{2} \sin^2 \theta_4$	$\frac{1}{2} \cos^2 \theta_4$

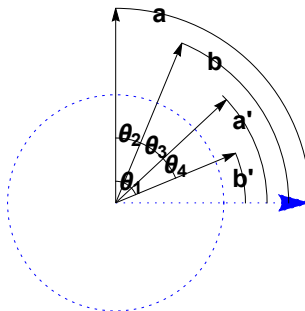
Comparing to Table 1, for example, the entry in row 2, column 3 is

$$\gamma_7 = P(A = 1, B = -1 | M_A = 1, M_B = 2) = \frac{1}{2} \sin^2 \theta_2.$$

The θ_k ’s represent the measurement *difference* angles as shown in Eq. 9 and Fig. 2, where a, a' and b, b' represent Alice and Bob’s photon polarization analyzer orientation choices, respectively. This notation is based on (not completely equivalent to) Aspect’s account [2].

$$\theta_1 = a - b', \theta_2 = a - b, \theta_3 = b - a', \theta_4 = a' - b'. \tag{9}$$

Figure 2. The relationships among the measurement difference angles θ_k .



It is clear from this figure that the measurement difference angles satisfy the constraint in Eq. 10. If the names or orientations are different, it will be assumed that they can always

⁵ These probabilities correspond to the photon-polarization version of the experiment, *not* the spin- $\frac{1}{2}$ singlet pair version. This assumption holds throughout this paper.

be renamed so as to conform to Fig. 2.

$$\theta_1 = \theta_2 + \theta_3 + \theta_4. \tag{10}$$

The correlations and s -functions for the QM model type are shown in Table 4. These can be calculated from Eq.'s 3 and 4 by substituting the QM conditional probabilities from Table 3 for the corresponding γ_k 's in Table 1.

Table 4. The correlations and s -functions for the QM model type. The proof that the s -function bounds are $\pm 2\sqrt{2}$ can be found in Appendix B.

w	$\cos 2\theta_1$
x	$\cos 2\theta_2$
y	$\cos 2\theta_3$
z	$\cos 2\theta_4$
s_1	$-\cos 2\theta_1 + \cos 2\theta_2 + \cos 2\theta_3 + \cos 2\theta_4$
s_2	$\cos 2\theta_1 - \cos 2\theta_2 + \cos 2\theta_3 + \cos 2\theta_4$
s_3	$\cos 2\theta_1 + \cos 2\theta_2 - \cos 2\theta_3 + \cos 2\theta_4$
s_4	$\cos 2\theta_1 + \cos 2\theta_2 + \cos 2\theta_3 - \cos 2\theta_4$
Min and Max	$-2\sqrt{2}$ and $2\sqrt{2}$

2.4. Noncontextual

It is important to note that the probabilistic structure postulated in Tables 1 or 3 consists of four probability mass functions (expressed as conditional probabilities) on four distinct probability spaces, each associated with a different measurement context.

Definition 8. A *probability mass function (pmf)* is a finite ordered set of real numbers $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_n)$ where all $\pi_k \geq 0$ and $\sum_k \pi_k = 1$.

So far there are no assumptions concerning an overall joint pmf from which all 16 marginal pmfs can be derived. This should be evident from the fact that the QM probabilities from Table 3 are a special case of the ones in Table 1, and it is well known that there does not always exist an overall joint distribution from which all QM probabilities can be derived as marginals.

If one *does* make such an assumption, however, then it is easy to show, by marginalization, that the resulting four sets of marginal probabilities must be of the *form* shown in Table 5. If these are viewed as double conditional probabilities, they define the noncontextual (NC) model type. It is assumed that $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_{16})$ is a pmf, that is, $0 \leq \rho_k \leq 1$ for all $k = 1, 2, \dots, 16$, and $\sum_{k=1}^{16} \rho_k = 1$. Note that the ρ_k 's do not *individually* denote conditional probabilities, only appropriate sums of four as indicated. In terms of generic parameters, for

example,

$$\gamma_7 = P(A = 1, B = -1 | M_A = 1, M_B = 2) = \rho_3 + \rho_4 + \rho_{11} + \rho_{12}$$

is the conditional probability in row 2, column 3.

Table 5. The **noncontextual (NC)** model type. Each entry represents a conditional probability $P(A = s, B = t | M_A = u, M_B = v)$ where $s, t = \pm 1$ denote Alice and Bob’s measurement outcomes and $u, v = 1$ or 2 indicate their measurement choices (context), respectively. Since the ρ_k ’s form a pmf, the conditional probabilities in each row sum to $\sum_{k=1}^{16} \rho_k = 1$.

Context	Outcome (s, t)			
	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\rho_1 + \rho_2 + \rho_5 + \rho_6$	$\rho_9 + \rho_{10} + \rho_{13} + \rho_{14}$	$\rho_3 + \rho_4 + \rho_7 + \rho_8$	$\rho_{11} + \rho_{12} + \rho_{15} + \rho_{16}$
$(1, 2)$	$\rho_1 + \rho_2 + \rho_9 + \rho_{10}$	$\rho_5 + \rho_6 + \rho_{13} + \rho_{14}$	$\rho_3 + \rho_4 + \rho_{11} + \rho_{12}$	$\rho_7 + \rho_8 + \rho_{15} + \rho_{16}$
$(2, 1)$	$\rho_1 + \rho_3 + \rho_5 + \rho_7$	$\rho_9 + \rho_{11} + \rho_{13} + \rho_{15}$	$\rho_2 + \rho_4 + \rho_6 + \rho_8$	$\rho_{10} + \rho_{12} + \rho_{14} + \rho_{16}$
$(2, 2)$	$\rho_1 + \rho_3 + \rho_9 + \rho_{11}$	$\rho_5 + \rho_7 + \rho_{13} + \rho_{15}$	$\rho_2 + \rho_4 + \rho_{10} + \rho_{12}$	$\rho_6 + \rho_8 + \rho_{14} + \rho_{16}$

The correlations and s -functions for the noncontextual model type are shown in Table 6.

Table 6. The correlations and s -functions for the noncontextual type. The fact that $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ is a pmf guarantees that all s -functions lie between -2 and $+2$. Hence all 8 CHSH inequalities are satisfied.

w	$\rho_1 + \rho_2 - \rho_3 - \rho_4 + \rho_5 + \rho_6 - \rho_7 - \rho_8 - \rho_9 - \rho_{10} + \rho_{11} + \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16}$
x	$\rho_1 + \rho_2 - \rho_3 - \rho_4 - \rho_5 - \rho_6 + \rho_7 + \rho_8 + \rho_9 + \rho_{10} - \rho_{11} - \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16}$
y	$\rho_1 - \rho_2 + \rho_3 - \rho_4 + \rho_5 - \rho_6 + \rho_7 - \rho_8 - \rho_9 + \rho_{10} - \rho_{11} + \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16}$
z	$\rho_1 - \rho_2 + \rho_3 - \rho_4 - \rho_5 + \rho_6 - \rho_7 + \rho_8 + \rho_9 - \rho_{10} + \rho_{11} - \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16}$
s_1	$2(\rho_1 - \rho_2 + \rho_3 - \rho_4 - \rho_5 - \rho_6 + \rho_7 + \rho_8 + \rho_9 + \rho_{10} - \rho_{11} - \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16})$
s_2	$2(\rho_1 - \rho_2 + \rho_3 - \rho_4 + \rho_5 + \rho_6 - \rho_7 - \rho_8 - \rho_9 - \rho_{10} + \rho_{11} + \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16})$
s_3	$2(\rho_1 + \rho_2 - \rho_3 - \rho_4 - \rho_5 + \rho_6 - \rho_7 + \rho_8 + \rho_9 - \rho_{10} + \rho_{11} - \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16})$
s_4	$2(\rho_1 + \rho_2 - \rho_3 - \rho_4 + \rho_5 - \rho_6 + \rho_7 - \rho_8 - \rho_9 + \rho_{10} - \rho_{11} + \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16})$
Min and Max	-2 and 2

Note how the parameterized operational modeling approach makes the following

statement obvious.

Theorem 1. *All 8 CHSH inequalities are satisfied for every NC model instance.*

Proof. The range of each s -function is restricted to $[-2, 2]$, since $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ is a pmf by definition of the NC model type. \square

2.5. Factorizable

The model types RCI+OD and factorizable are relevant to Bell's original derivations of his inequalities [4,5]. It turns out that both of these are special cases of noncontextuality. RCI+OD is defined in Sec. 2.7. Factorizability is defined here.

Definition 9. *A model instance will be called **factorizable** or **Bell local** if*

$$\begin{aligned} P(A = s, B = t | M_A = u, M_B = v) = \\ P(A = s | M_A = u) \times P(B = t | M_B = v). \end{aligned} \quad (11)$$

for all $s, t = \pm 1$ and $u, v = 1, 2$. It is sometimes said that the **factorization condition (FC)** is satisfied. Both [3] and [9] use the term **Bell local** to describe models that satisfy the factorization condition.

Factorizability is closely related to local causality. Bell developed the idea of local causality in [5]. Also see [3] for a discussion of this topic. It was used by Bell to capture the idea that anything outside of Alice's past light cone, such as Bob's settings or outcomes, are statistically irrelevant to her outcome, and similarly for Bob's outcome with regard to Alice's actions or observations.

Definition 10. Local causality or LC:

$$\begin{aligned} P(A = s | B = t, M_A = u, M_B = v) = P(A = s | M_A = u) \text{ and} \\ P(B = t | A = s, M_A = u, M_B = v) = P(B = t | M_B = v), \end{aligned} \quad (12)$$

for all $s, t = \pm 1$ and $u, v = 1, 2$, but only if all of the conditional probabilities involved are well-defined.

The reason for caution in the definition of LC is that the conditional probabilities on the left may be undefined, because of possible zero conditioning probabilities. However, if all conditional probabilities are well-defined, it is equivalent to factorizability. See Sec. 4.1.2 for details. Based on these definitions, the factorizable model type is defined by Table 7.

Table 7. The conditional probabilities $P(A = s, B = t|M_A = u, M_B = v)$ (where $s, t = \pm 1$ and $u, v = 1, 2$) for the **factorizable** (or **Bell local**) model type are characterized by 8 parameters α_k and β_k for $k = 1, 2, 3, 4$. These conditional probabilities satisfy the measurement context constraints (each row sums to 1) because of the factorization constraints in Eq. 14.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\alpha_1\beta_1$	$\alpha_1\beta_3$	$\alpha_3\beta_1$	$\alpha_3\beta_3$
$(1, 2)$	$\alpha_1\beta_2$	$\alpha_1\beta_4$	$\alpha_3\beta_2$	$\alpha_3\beta_4$
$(2, 1)$	$\alpha_2\beta_1$	$\alpha_2\beta_3$	$\alpha_4\beta_1$	$\alpha_4\beta_3$
$(2, 2)$	$\alpha_2\beta_2$	$\alpha_2\beta_4$	$\alpha_4\beta_2$	$\alpha_4\beta_4$

As an example of the correspondence to the generic parameters, the entry in row 2, column 3 is

$$\gamma_7 = P(A = 1, B = -1|M_A = 1, M_B = 2) = \alpha_3\beta_2.$$

Definition 11. The factorization parameters are defined by

$$\begin{aligned} \alpha_1 &= P(A = -1|M_A = 1), & \beta_1 &= P(B = -1|M_B = 1), \\ \alpha_2 &= P(A = -1|M_A = 2), & \beta_2 &= P(B = -1|M_B = 2), \\ \alpha_3 &= P(A = +1|M_A = 1), & \beta_3 &= P(B = +1|M_B = 1), \\ \alpha_4 &= P(A = +1|M_A = 2), & \beta_4 &= P(B = +1|M_B = 2). \end{aligned} \tag{13}$$

Definition 12. The factorization parameters satisfy the factorization constraints given by

$$\alpha_1 + \alpha_3 = \alpha_2 + \alpha_4 = 1, \quad \beta_1 + \beta_3 = \beta_2 + \beta_4 = 1. \tag{14}$$

2.6. Factorization implies noncontextuality

Theorem 2. The factorizable model type is a special case of the noncontextual model type.

Proof. Assume a set of factorizable conditional probabilities (Table 7) based on parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and $\beta_1, \beta_2, \beta_3, \beta_4$. To construct the corresponding noncontextual conditional probabilities (Table 5), simply assign the parameters $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ according to Eq. 15. Then use the factorization constraints of Eq. 14 to show that ρ is in fact a pmf, and that if one substitutes these values for the ρ_k 's into Table 5 one gets Table 7.

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \rho_4 \\ \rho_5 \\ \rho_6 \\ \rho_7 \\ \rho_8 \\ \rho_9 \\ \rho_{10} \\ \rho_{11} \\ \rho_{12} \\ \rho_{13} \\ \rho_{14} \\ \rho_{15} \\ \rho_{16} \end{pmatrix} = \begin{pmatrix} \alpha_1 \alpha_2 \beta_1 \beta_2 \\ \alpha_1 \alpha_4 \beta_1 \beta_2 \\ \alpha_2 \alpha_3 \beta_1 \beta_2 \\ \alpha_3 \alpha_4 \beta_1 \beta_2 \\ \alpha_1 \alpha_2 \beta_1 \beta_4 \\ \alpha_1 \alpha_4 \beta_1 \beta_4 \\ \alpha_2 \alpha_3 \beta_1 \beta_4 \\ \alpha_3 \alpha_4 \beta_1 \beta_4 \\ \alpha_1 \alpha_2 \beta_2 \beta_3 \\ \alpha_1 \alpha_4 \beta_2 \beta_3 \\ \alpha_2 \alpha_3 \beta_2 \beta_3 \\ \alpha_3 \alpha_4 \beta_2 \beta_3 \\ \alpha_1 \alpha_2 \beta_3 \beta_4 \\ \alpha_1 \alpha_4 \beta_3 \beta_4 \\ \alpha_2 \alpha_3 \beta_3 \beta_4 \\ \alpha_3 \alpha_4 \beta_3 \beta_4 \end{pmatrix} \tag{15}$$

□

This statement can be strengthened to say that factorizability is a *strictly stronger* classical notion than noncontextuality, meaning that there is at least one noncontextual cpv that is not also factorizable (in fact infinitely many). See Example 1 in Sec. 2.10.

The correlations and *s*-functions for the factorizable type are shown in Table 8. Being a special case of the noncontextual type, all of its *s*-functions also satisfy all 8 CHSH inequalities. Note that since the single conditional expectations are given by

$$\begin{aligned} E[A|M_A = 1] &= \alpha_3 - \alpha_1 & E[B|M_B = 1] &= \beta_3 - \beta_1, \\ E[A|M_A = 2] &= \alpha_4 - \alpha_2 & E[B|M_B = 2] &= \beta_4 - \beta_2, \end{aligned} \tag{16}$$

it follows that each correlation (conditional expectation of the product) is the product of the corresponding single conditional expectations.

$$\begin{aligned} w &= E[AB|M_A = 1, M_B = 1] = E[A|M_A = 1] \times E[B|M_B = 1], \\ x &= E[AB|M_A = 1, M_B = 2] = E[A|M_A = 1] \times E[B|M_B = 2], \\ y &= E[AB|M_A = 2, M_B = 1] = E[A|M_A = 2] \times E[B|M_B = 1], \\ z &= E[AB|M_A = 2, M_B = 2] = E[A|M_A = 2] \times E[B|M_B = 2]. \end{aligned} \tag{17}$$

Table 8. The correlations and s -functions for the factorizable (Bell local) type. As a special case of the NC model type (Thm. 2), all factorizable model instances satisfy all CHSH inequalities.

w	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1)$
x	$(\alpha_3 - \alpha_1)(\beta_4 - \beta_2)$
y	$(\alpha_4 - \alpha_2)(\beta_3 - \beta_1)$
z	$(\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_1	$-(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) + (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) + (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) + (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_2	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) - (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) + (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) + (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_3	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) + (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) - (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) + (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_4	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) + (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) + (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) - (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
Min and Max	-2 and 2

From either Eq. 17 or Table 8, it is easy to see that for the factorizable correlations (w, x, y, z) ,

$$wz = xy,$$

and therefore

$$\begin{cases} z = w^{-1}xy & \text{if } w \neq 0 \\ xy = 0 & \text{if } w = 0. \end{cases} \tag{18}$$

This explains the shapes of the correlation plots in the factorizable row 3 in Fig. 1. At the far left ($w = -1$) and far right ($w = 1$), the surfaces are hyperbolic paraboloids. In the middle ($w = 0$), $xy = 0$ determines two intersecting planes.

2.7. Outcome determinism

Definition 13. A generic model instance will be called **outcome deterministic (OD)** if $\gamma_k = 0$ or 1 for all k .

An example is shown in Table 9. There are 256 model instances like this (4 ways to place exactly one 1 in each of the four rows, hence a total of $4^4 = 256$ such arrangements of 1's).

Table 9. Example of an **outcome deterministic (OD)** model instance. There are exactly 256 such model instances (including this one). Obviously each row sums to 1 in all cases.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	1	0	0	0
$(1, 2)$	1	0	0	0
$(2, 1)$	1	0	0	0
$(2, 2)$	1	0	0	0

These turn out to be the fundamental building blocks for all other model types (including QM), via convex combinations. In Sec. 3, the traditional hidden variables approach is reformulated in terms of convex combinations of finite ensembles of OD model instances.

It is fairly obvious that all of the correlations of any of the 256 OD model instances must be ± 1 . See Def. 13 and Eq. 3. This property defines an even larger (in fact infinite) class of model instances as follows.

Definition 14. A model instance will be called **perfectly correlated (PC)** if its correlations are all ± 1 . In other words, in terms of generic parameters,

$$\begin{aligned}
 w &= \gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 = \pm 1, \\
 x &= \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 = \pm 1, \\
 y &= \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} = \pm 1, \\
 z &= \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16} = \pm 1.
 \end{aligned}
 \tag{19}$$

Since the subsets of γ_k 's that go into defining each of the correlations are mutually disjoint, it is clear that (w, x, y, z) can be any of 16 possible sequences of ± 1 's. The s -functions are determined by the patterns of the ± 1 's according to Table 10. Although the arrangement of numbers in any row of the last four columns might be different, the numbers that appear will be the same. Table 10 shows that perfectly correlated model instances, including the 256 OD model instances, are inconsistent with the predictions of QM, either by having s -functions that are "too small" or "too big".

Table 10. All possible values of s -functions for a perfectly correlated model instance. Note there are cases where the CHSH inequalities are satisfied and others where even a Tsirelson bound is exceeded.

Perfect correlations		Typical s -functions			
No. -1's	No. +1's	s_1	s_2	s_3	s_4
4	0	-2	-2	-2	-2
3	1	0	0	0	-4
2	2	2	2	-2	-2
1	3	4	0	0	0
0	4	2	2	2	2

The instances that have s -functions which exceed one of the Tsirelson bounds come from a set of correlations with either one +1 and three -1's or a set with one -1 and three +1's (table 10). Consider the case with one -1 and three +1's. There are 4 ways to choose which of w, x, y, z will be the -1. Within that correlation definition, there are just two γ_k choices to set to 1 in order for that to happen. In the remaining three correlations, there are two choices for which γ_k to set to 1 as well. This yields $4 \times 2 \times 2 \times 2 = 64$ ways to get one -1 and three +1's. The same argument works for one +1 and three -1's. Hence a total of $64 + 64 = 128$ sets of correlations with either one -1 and three +1's or one +1 and three -1's, which are the ones with an s -function that violates a CHSH inequality. This leaves $256 - 128 = 128$ sets of correlations with either zero, two, or four +1's (equivalent to four, two, or zero -1's). These are the cases that satisfy CHSH.

2.8. RCI+OD

Only 16 of 256 OD model instances satisfy RCI. Their cpvs are shown as columns in Table 11.⁶

Definition 15. A model instance satisfies **remote context independence (RCI)** if:

$$\begin{aligned}
 P(A = s | M_A = u, M_B = v) &= P(A = s | M_A = u) \text{ and} \\
 P(B = t | M_A = u, M_B = v) &= P(B = t | M_B = v)
 \end{aligned}
 \tag{20}$$

for all $s, t = \pm 1$ and $u, v = 1, 2$. That is, Alice's outcome probability conditioned on her measurement choice is independent of Bob's measurement choice, and vice-versa. RCI is sometimes called **parameter independence**.

⁶ This format was chosen because 16 individual tables take up too much space.

Table 11. The 16 **RCI+OD** model instances (shown as columns). They are also factorizable and noncontextual. Any NC (including factorizable) model instance can be written as a convex combination of these 16. The first column (1) corresponds to the model instance shown in Table 9.

<i>s</i>	<i>t</i>	<i>u</i>	<i>v</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-1	-1	1	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0
-1	1	1	1	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0
1	-1	1	1	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1
-1	-1	1	2	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
-1	1	1	2	0	0	0	0	1	1	0	0	0	0	0	0	1	1	0	0
1	-1	1	2	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0
1	1	1	2	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1
-1	-1	2	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
-1	1	2	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1	0
1	-1	2	1	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0
1	1	2	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	1
-1	-1	2	2	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0	0
-1	1	2	2	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0
1	-1	2	2	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	0
1	1	2	2	0	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1

*Mathematica*⁷ was used to identify these 16 special model instances. In fact, *Mathematica* was used to classify all OD model instances according to certain properties. Fig. 3 shows the finite set of OD model instances embedded in the infinite set of perfectly correlated model instances. Subsets that satisfy various combinations of properties are indicated, together with their numbers, as follows:

- RCI+OD (16)
- CHSH+OD (128)
- CHSH+OD but do not satisfy RCI (112)
- ROI+OD (256)
- ROI+OD but do not satisfy CHSH (128)

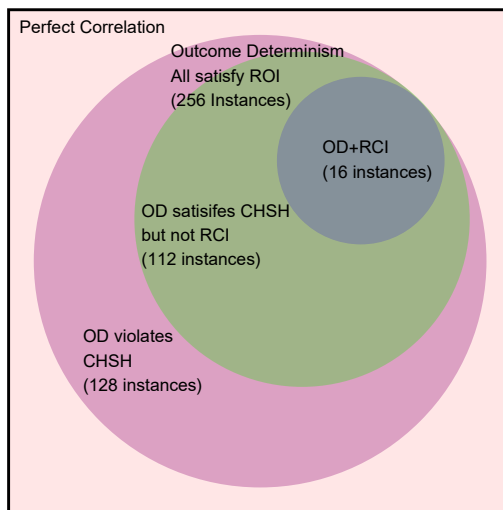
⁷ The computational workhorse behind everything in this paper is the Wolfram language platform *Mathematica*, including the initial idea exploration and feasibility analysis, the heavy-duty matrix generation and calculations, function optimization and equation solving, the computation of expectations, marginal and conditional distributions, and especially the generation of the plots and visualizations.

Definition 16. A model instance satisfies **remote outcome independence (ROI)** if:

$$P(A = s, B = t | M_A = u, M_B = v) = P(A = s | M_A = u, M_B = v) \times P(B = t | M_A = u, M_B = v). \quad (21)$$

for all $s, t = \pm 1$ and $u, v = 1, 2$. That is, the outcomes A and B are **conditionally independent** given the measurement choice pair (M_A, M_B) .

Figure 3. Relationships among classes of OD model instances. All 256 of them satisfy ROI. Only 16 satisfy RCI. 112 others satisfy CHSH but do not satisfy RCI, hence are neither noncontextual or factorizable. The remaining 128 not only violate CHSH but exceed a Tsirelson bound.



Here are examples from each of the four regions shown in Fig. 3.

1. PC but not OD. Choose generic parameters

$$\gamma_1 = \gamma_4 = \gamma_6 = \gamma_7 = \gamma_{10} = \gamma_{11} = \gamma_{13} = \gamma_{16} = \frac{1}{2} \text{ and all other } \gamma_k = 0.$$

Then the correlations are (see Eq. 3)

$$(w, x, y, z) = (1, -1, -1, 1).$$

2. OD but does not satisfy CHSH. Choose generic parameters

$$\gamma_2 = \gamma_6 = \gamma_{10} = \gamma_{13} = 1 \text{ and all other } \gamma_k = 0.$$

Then the correlations are (see Eq. 3)

$$(w, x, y, z) = (-1, -1, -1, 1)$$

and the s -functions are (see Eq. 4)

$$(s_1, s_2, s_3, s_4) = (0, 0, 0, -4).$$

3. OD and satisfies CHSH but is not RCI. Choose generic parameters

$$\gamma_1 = \gamma_7 = \gamma_9 = \gamma_{14} = 1 \text{ and all other } \gamma_k = 0.$$

Then the correlations are (see Eq. 3)

$$(w, x, y, z) = (1, -1, 1, -1)$$

and the s -functions are (see Eq. 4)

$$(s_1, s_2, s_3, s_4) = (-2, 2, -2, 2),$$

but this is not one of the sixteen RCI+OD instances in Table 11.

4. OD and satisfies RCI. See any one of the 16 examples in Table 11.

2.9. Noncontextuality built up from RCI+OD model instances

In this section, it is shown that the convex hull of the RCI+OD model instances is precisely the set of NC model instances. In Appendix C, a more general result is proved, namely that *any* generic model instance, including *all* QM ones, can be written as a convex combination of 256 (or fewer) OD model instances. The process of constructing interesting families of cpvs from *finite* ensembles of more basic cpvs via convex combinations adds some intuition to the traditional hidden variables representation. In Sec. 3, the (almost) equivalence of the two approaches is demonstrated.

Theorem 3. *Any NC model instance can be written as a convex combination of the 16 RCI+OD model instances (cpvs) shown as columns in Table 11.*

Proof. Define the matrix R as follows and consider Eq. 22.

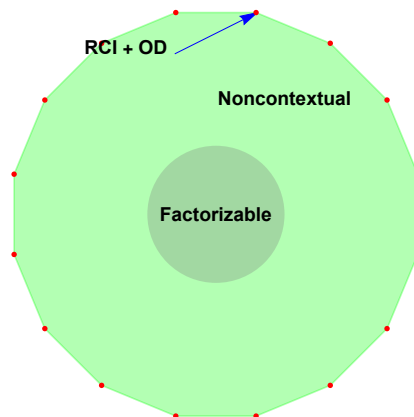
Definition 17. *The matrix R is the 16×16 matrix consisting of the last 16 columns of Table 11. R turns a pmf $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ into a NC conditional probability vector (see Eq. 22). R is an example of a **Bell-type scenario incidence matrix** in [12].*

$$R\rho^T = \begin{pmatrix} \rho_1 + \rho_2 + \rho_5 + \rho_6 \\ \rho_9 + \rho_{10} + \rho_{13} + \rho_{14} \\ \rho_3 + \rho_4 + \rho_7 + \rho_8 \\ \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16} \\ \rho_1 + \rho_2 + \rho_9 + \rho_{10} \\ \rho_5 + \rho_6 + \rho_{13} + \rho_{14} \\ \rho_3 + \rho_4 + \rho_{11} + \rho_{12} \\ \rho_7 + \rho_8 + \rho_{15} + \rho_{16} \\ \rho_1 + \rho_3 + \rho_5 + \rho_7 \\ \rho_9 + \rho_{11} + \rho_{13} + \rho_{15} \\ \rho_2 + \rho_4 + \rho_6 + \rho_8 \\ \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16} \\ \rho_1 + \rho_3 + \rho_9 + \rho_{11} \\ \rho_5 + \rho_7 + \rho_{13} + \rho_{15} \\ \rho_2 + \rho_4 + \rho_{10} + \rho_{12} \\ \rho_6 + \rho_8 + \rho_{14} + \rho_{16} \end{pmatrix} \tag{22}$$

Since $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ is a pmf, the expression on the left is a convex combination of the columns of R . The column vector on the right constitutes the conditional probabilities of the NC model type – see Table 5 for comparison. \square

Fig. 4 illustrates Thm. 3. This is a schematic representation of a polytope in a 9-dimensional subspace of \mathbb{R}^{16} . The rank of R was computed using *Mathematica*'s `MatrixRank` function.

Figure 4. The convex hull of the 16 RCI+OD cpvs (red dots) equals the full set of NC cpvs, including the factorizable (Bell local) ones.



Corollary 1. All of the 16 RCI+OD model instances of Table 11 are noncontextual.

Proof. In Eq. 22, for any $k = 1, 2, \dots, 16$, set $\rho_k = 1$ and $\rho_j = 0$ for all $j \neq k$. This

produces the k th RCI+OD cpv, that is, the k th column of Table 11 (also the k th column of R , by definition). □

Corollary 2. All of the 16 RCI+OD model instances of Table 11 are factorizable.

Proof. As an example, it is proved that the first cpv (column 1) in Table 11 is factorizable. The proofs for the other 15 columns are similar. The proof of Corollary 1 showed that the first RCI+OD cpv is noncontextual – simply set $\rho_1 = 1$ and $\rho_j = 0$ for all $j \neq 1$. Then Eq. 15 makes it obvious what to do. Set $\alpha_1 = \alpha_2 = 1$, $\beta_1 = \beta_2 = 1$ and $\alpha_j = \beta_j = 0$ for all $j \neq 1$ or 2. Hence the first RCI+OD model instance is a legitimate instance of the factorizable model type shown in Table 7. □

The correlations and s -functions for the 16 RCI+OD model instances are shown in Table 12. Clearly all of the CHSH inequalities are satisfied; in fact the s -functions attain these extreme values exactly as functions of the correlations of the RCI+OD cpvs.

Table 12. The correlations and s -functions for the 16 RCI+OD model instances. See Table 11 for the corresponding cpv’s.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
w	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1
x	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1
y	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
z	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1
s_1	2	-2	2	-2	-2	-2	2	2	2	2	-2	-2	-2	2	-2	2
s_2	2	-2	2	-2	2	2	-2	-2	-2	-2	2	2	-2	2	-2	2
s_3	2	2	-2	-2	-2	2	-2	2	2	-2	2	-2	-2	-2	2	2
s_4	2	2	-2	-2	2	-2	2	-2	-2	2	-2	2	-2	-2	2	2
Min and Max	-2 and 2															

It is hard to overemphasize the pivotal role played by the NC model type.

- The correlations of any NC model instance satisfy all 8 CHSH inequalities.
- Any factorizable model instance and any outcome deterministic model instance that also satisfies remote context independence are special cases of the NC model type.
- Noncontextuality is the most general classical notion (among RCI+OD, factorizability, and noncontextuality) that blocks replication of at least some QM statistical predictions. Both factorization and RCI+OD are *sufficient* to derive the CHSH inequalities, but neither is *necessary* (see sec. 2.10).
- In other words, if one assumes RCI+OD or one assumes factorizability, one has implicitly assumed noncontextuality (in the mathematical sense).

2.10. Noncontextuality does not imply factorizability

To show that the set of factorizable (Bell local) model instances is a *proper* subset of the set of noncontextual model instances, it is sufficient to provide an example of a NC model instance that cannot be factorizable. First make the simple observation that if two model instances have the same conditional probabilities, then they have the same correlations (see Eq. 3). Hence, to produce a counterexample, it is sufficient to find a NC model instance whose correlations cannot be consistent with *any* factorizable model instance.

Example 1. *The NC model instance with parameters*

$$\rho = (0, \frac{1}{4}, 0, 0, 0, 0, 0, 0, \frac{1}{4}, \frac{1}{4}, 0, 0, 0, 0, 0, 0, \frac{1}{4}, 0)$$

leads to the generic representation

$$\gamma = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0, 0, 0, \frac{1}{2}, 0, \frac{1}{2}, \frac{1}{2}, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}),$$

which has correlations $(w, x, y, z) = (0, 1, -1, 0)$. However, $(0, 1, -1, 0)$ cannot be achieved by any assignment of values to the parameters of a factorizable model.

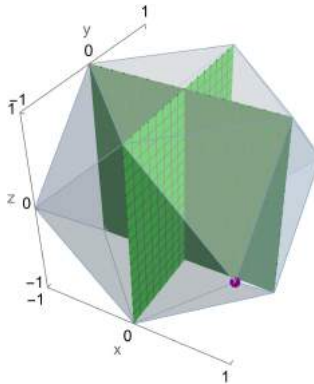
Consider the equations that must be solved.

$$\begin{pmatrix} (\alpha_1 - \alpha_3)(\beta_1 - \beta_3) \\ (\alpha_1 - \alpha_3)(\beta_2 - \beta_4) \\ (\alpha_2 - \alpha_4)(\beta_1 - \beta_3) \\ (\alpha_2 - \alpha_4)(\beta_2 - \beta_4) \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \tag{23}$$

The expressions on the left represent the four correlations (w, x, y, z) for the factorizable model type (see Table 8 for the factorizable correlations). To make the first term zero, either $\alpha_1 = \alpha_3$ or $\beta_1 = \beta_3$. But then either the second or third term (or both) is (are) zero, making one (or both) of the second and third equations false. So it is impossible for a factorizable model instance to achieve this particular set of correlations.

As an illustration of what has just been proved, see Fig. 5 (see Appendix A.2 for details on correlation plots). It shows a 3D slice of all correlations $(0, x, y, z)$ (corresponding to $w = 0$) for the NC region (blue solid) and the factorizable surfaces (green planes). The point $(x, y, z) = (1, -1, 0)$ (purple dot) is a vertex of the NC region (blue), but it is way off the factorizable contour (green), i.e., the intersecting planes $x = 0$ and $y = 0$. Note that, since this is the 3D-slice corresponding to $w = 0$, the 4-tuple of correlations in question is the one specified above, namely $(w, x, y, z) = (0, 1, -1, 0)$.

Figure 5. $(w, x, y, z) = (0, 1, -1, 0)$ is an example of a correlation vector consistent with a NC model instance but not consistent with any factorizable model instance. The triple $(x, y, z) = (1, -1, 0)$ is shown (purple dot) where $w = 0$ is assumed to be fixed.



At first glance, this counterexample seems to be in conflict with Proposition 2 in Fine [10], since he proves that "There exists a factorizable stochastic hidden-variables model for a correlation experiment if and only if there exists a deterministic hidden variables model for the experiment". First let's translate this statement into the language of convex combinations of model instances, namely, "A model instance is expressible as a convex combination of factorizable model instances iff it is expressible as a convex combination of RCI+OD model instances". Here is a proof:

Proof. (\Leftarrow) Suppose a model instance is a convex combination of the 16 RCI+OD model instances. Since all of these are factorizable (Corollary 1), this is also a convex combination of factorizable model instances.

(\Rightarrow) Suppose a model instance is a convex combination of factorizable model instances. But each of these factorizable model instances is also NC by Thm. 2, and thus expressible as a convex combination of RCI+OD model instances by Thm. 3. Hence the given model instance is expressible as a convex combination of RCI+OD model instances since a convex combination of convex combinations is again a convex combination. \square

So how does Example 1 escape being factorizable? After all, it can be expressed as a convex combination of RCI+OD model instances, since it is NC (Thm. 3). Hence it is expressible as a convex combination of factorizable model instances. However, there is nothing to guarantee that such a convex combination is also factorizable. That is, the factorization property does not generally "survive" the process of constructing convex combinations, or in the traditional language of hidden variables, does not "survive" the process of integration with respect to a measure over the hidden variables. This is a simple consequence of the fact that the integral of a product of functions is usually not the product of the integrals of the functions.

2.11. CHSH almost implies noncontextuality

As an example of a (possibly) underappreciated result, consider one of the propositions from Fine’s pioneering paper from 1982 [10], where among other things, he linked the existence of a joint distribution on all observables (commuting or not) to satisfaction of the Bell/CH inequalities. Although he was careful to say *there exists* a factorizable model, or *there exists* a deterministic hidden variables model throughout, it may be not be well-appreciated by everyone that there exist model instances that satisfy all 8 CHSH inequalities, but which are *not* noncontextual (hence also not factorizable). Here is an example of a model instance whose correlations satisfy all of the CHSH inequalities but it is not NC. Note this is one of the 112 OD model instances with this property (see Fig. 3).

Example 2. Consider the OD model instance with parameters

$$\gamma = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1).$$

The corresponding correlations are $(w, x, y, z) = (1, 1, 1, 1)$ and the s -functions are $(s_1, s_2, s_3, s_4) = (2, 2, 2, 2)$. However, it is fairly obvious that there is no pmf $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ such that

$$\begin{pmatrix} \rho_1 + \rho_2 + \rho_5 + \rho_6 \\ \rho_3 + \rho_4 + \rho_7 + \rho_8 \\ \rho_9 + \rho_{10} + \rho_{13} + \rho_{14} \\ \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16} \\ \rho_1 + \rho_2 + \rho_9 + \rho_{10} \\ \rho_3 + \rho_4 + \rho_{11} + \rho_{12} \\ \rho_5 + \rho_6 + \rho_{13} + \rho_{14} \\ \rho_7 + \rho_8 + \rho_{15} + \rho_{16} \\ \rho_1 + \rho_3 + \rho_5 + \rho_7 \\ \rho_2 + \rho_4 + \rho_6 + \rho_8 \\ \rho_9 + \rho_{11} + \rho_{13} + \rho_{15} \\ \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16} \\ \rho_1 + \rho_3 + \rho_9 + \rho_{11} \\ \rho_2 + \rho_4 + \rho_{10} + \rho_{12} \\ \rho_5 + \rho_7 + \rho_{13} + \rho_{15} \\ \rho_6 + \rho_8 + \rho_{14} + \rho_{16} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Just start knocking off all of the ρ_k ’s on the left in rows where 0’s appear in the column vector on the right, and pretty soon there is nothing left to make 1’s in rows 1,5,9, or 16.

Thus it is not true that satisfaction of the CHSH inequalities implies noncontextuality.

What is true is a slightly weaker statement, namely, if a model instance (noncontextual or not) has correlations that satisfy all 8 CHSH inequalities, then *there exists* a noncontextual model instance with those same correlations. A constructive proof of this result is given, based on convex geometry, which may lend some intuition and emphasis to Fine’s earlier result. It will be called "Fine’s Theorem" here, even though it is just part of what he did. See Halliwell [19] for some interesting alternative proofs of this result. The underlying takeaway, although it should be obvious from Eq. 3, is that the mapping from cpvs to correlations is many-to-one.

The proof given here is grounded in the geometry of Fig. 6, which may add some intuition compared to the original proof in [10]. These five regions represent 3D slices (x, y, z) through 4D sets of correlations (w, x, y, z) that satisfy the CHSH inequalities, for a given fixed value for w in each case, where $w = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$, from left to right. Since these are closed and bounded convex regions, any such correlation can be written as a convex combination of the vertices (red dots). To get the parameters ρ of a NC model instance which has the given correlations, one merely computes the same convex combination of a special set of probability vectors.

First note that in each case, the coordinates for the vertices (red dots) can be written as the columns of the last three rows of the following matrices.

Definition 18. *The vertex matrix V is defined in three different ways, depending on the value of w .*

$$V = \begin{pmatrix} w & w & w & w & w & w & w & w & w & w & w & w \\ w & -1 & 1 & -w & w & 1 & -1 & -w & 1 & 1 & -1 & -1 \\ 1 & -w & w & -1 & -1 & -w & w & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -w & w & -w & w \end{pmatrix}$$

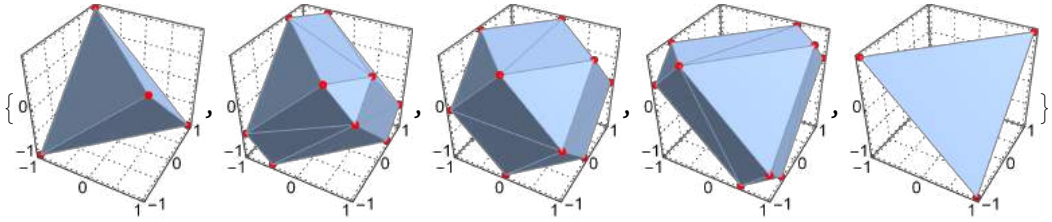
if $-1 < w < 1$,

$$V = \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \text{ if } w = -1,$$

$$V = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix} \text{ if } w = +1.$$

Then the following lemma basically says that any correlation vector $\mathbf{p} = (w, x, y, z)$

Figure 6. Five regions where the CHSH inequalities are satisfied. $w = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$ from left to right.



which satisfies all eight CHSH inequalities can be written as a convex combination of the columns of V .

Lemma 1. Pick an arbitrary set of correlations $\mathbf{p} = (w, x, y, z)$ that satisfy CHSH. Then there exists a pmf $\boldsymbol{\pi}$ such that $\mathbf{p}^T = (w, x, y, z)^T = V\boldsymbol{\pi}^T$. If V has 12 rows (i.e., $-1 < w < 1$), $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4, \dots, \pi_{12})$, and if it has 4 rows (i.e., $w = \pm 1$), $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$.

Proof. Start by ignoring the first row of V and consider the last three rows only. The columns of the resulting sub-matrix represent the coordinates of the vertices of certain 3D volumes defined by the CHSH inequalities. Some examples are shown in Fig. 6. Since all of these volumes, for any $-1 \leq w \leq 1$, are formed by the intersection of a finite number of closed half-spaces, the result is a closed convex region. The particular (hyper)planes represented by the CHSH inequalities also make the region bounded. Invoking a theorem of Minkowski, an arbitrary point $\mathbf{p}_w = (x, y, z)$ inside any of these bounded, closed, convex regions can be written as a convex combination of the extreme points of the region, in this case, the vertices. That is, $\mathbf{p}_w^T = (x, y, z)^T = V(2, 3, 4)\boldsymbol{\pi}^T$, for some appropriate vertex matrix V (where $V(2, 3, 4)$ denotes the last three rows of V) and pmf $\boldsymbol{\pi}$. Since $\boldsymbol{\pi}$ is a pmf, $w = (w, w, \dots, w)\boldsymbol{\pi}^T$, so by inserting all w 's as the first row of $V(2, 3, 4)$ to make V , it is easily seen that $\mathbf{p}^T = (w, x, y, z)^T = V\boldsymbol{\pi}^T$ as well. \square

The next matrix was discovered by using *Mathematica* to find instances of solutions M to the matrix equation shown in Lemma 2. That it worked so well earned it the name "magic".

Definition 19. The *magic matrix* M is defined in three different ways, depending on the value of w . M is a (right) stochastic matrix, since the rows consist of non-negative numbers that sum to 1.

$$M = \begin{pmatrix} \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+w}{2} & 0 & 0 & \frac{1-w}{2} & 0 & 0 \\ \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1+w}{2} & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 \\ 0 & \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+w}{2} & \frac{1-w}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1+w}{2} & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+w}{2} & 0 & \frac{1-w}{2} & 0 & 0 \\ \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1+w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1-w}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+w}{2} & 0 & \frac{1-w}{2} & 0 & 0 & 0 \end{pmatrix}$$

if $-1 < w < 1$,

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ if } w = -1,$$

$$M = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \text{ if } w = +1.$$

Lemma 2. $CRM^T = V$.

Proof. This is a straightforward but tedious matrix multiplication, so it is left to the reader. (The definitions of the C and R matrices can be found in Def.'s 6 and 17, respectively.) There are three cases to consider depending on the value of w . *Mathematica* or other computational software platform is helpful here! □

Finally Thm. 4 can be stated and proved. The proof may be of interest due to its compactness and its basis in the geometry exemplified by the polyhedra shown in Fig. 6.

Theorem 4. (Fine’s Theorem) Given an arbitrary correlation vector $\mathbf{p} = (w, x, y, z)$ that satisfies the CHSH inequalities. Let $\boldsymbol{\pi}$ be a pmf such that $\mathbf{p}^T = V\boldsymbol{\pi}^T$ as guaranteed by Lemma 1. Define

- $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_{16}) = \boldsymbol{\pi}M$,
- $\boldsymbol{\gamma}^T = (\gamma_1, \gamma_2, \dots, \gamma_{16})^T = R\boldsymbol{\rho}^T = RM^T\boldsymbol{\pi}^T$.

Then

- ρ is a pmf and therefore γ is a set of NC conditional probabilities.
- The NC conditional probabilities γ are consistent with the correlations \mathbf{p} .

Proof. For the first conclusion, note that ρ is a pmf because the rows of M are pmf's, π is a pmf, therefore $\rho = \pi M$, being a convex combination of pmf's, is also a pmf. Thus $\gamma^T = R\rho^T$ is a set of NC conditional probabilities by the very definition of R (Def. 17).

For the second conclusion, invoke Lemma 2 to write

$$\mathbf{p}^T = V\pi^T = CRM^T\pi^T = C\gamma^T, \tag{24}$$

that is, the initial arbitrary correlation vector \mathbf{p} , that was assumed to satisfy the CHSH inequalities, is equal to the computed correlations $C\gamma^T$ based on the NC conditional probabilities γ . □

An important consequence of Thm. 4 is that the polyhedra shown in the "Noncontextual" (fourth) row of Fig. 1 correspond to satisfaction of the CHSH inequalities. Here is a formal proof. Assume $-1 \leq w \leq 1$ is fixed and define two polyhedrons as follows:

$$\mathcal{P}_1 = \{(x, y, z) : (w, x, y, z) \text{ satisfy all of the CHSH inequalities}\}, \tag{25}$$

and (see Table 6 for the NC correlations)

$$\begin{aligned} \mathcal{P}_2 = \{ & (x, y, z) : \text{there exists a pmf } \rho = (\rho_1, \rho_2, \dots, \rho_{16}) \text{ such that} \\ & \rho_1 + \rho_2 - \rho_3 - \rho_4 + \rho_5 + \rho_6 - \rho_7 - \rho_8 - \\ & \rho_9 - \rho_{10} + \rho_{11} + \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16} = w, \\ & \text{and where } x, y, z \text{ are defined by} \\ & x = \rho_1 + \rho_2 - \rho_3 - \rho_4 - \rho_5 - \rho_6 + \rho_7 + \rho_8 + \\ & \rho_9 + \rho_{10} - \rho_{11} - \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16}, \\ & y = \rho_1 - \rho_2 + \rho_3 - \rho_4 + \rho_5 - \rho_6 + \rho_7 - \rho_8 - \\ & \rho_9 + \rho_{10} - \rho_{11} + \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16}, \text{ and} \\ & z = \rho_1 - \rho_2 + \rho_3 - \rho_4 - \rho_5 + \rho_6 - \rho_7 + \rho_8 + \\ & \rho_9 - \rho_{10} + \rho_{11} - \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16}\}. \end{aligned} \tag{26}$$

Corollary 3. $\mathcal{P}_1 = \mathcal{P}_2$.

Proof. $\mathcal{P}_2 \subseteq \mathcal{P}_1$ since the correlations of any NC model instance satisfy all of the CHSH inequalities. Conversely, $\mathcal{P}_1 \subseteq \mathcal{P}_2$ by Thm. 4, since if $(x, y, z) \in \mathcal{P}_1$, then *there exists*

a NC cpv (i.e., based on the parameters of some pmf $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$) that has the correlations (w, x, y, z) , hence $(x, y, z) \in \mathcal{P}_2$. \square

However, recall that there are cpvs whose correlations satisfy all CHSH inequalities but are *not* NC, for example the 112 OD+CHSH cpvs (see Sec. 2.8). The set of cpvs whose correlations satisfy all of the CHSH inequalities is *strictly larger* than the set of NC cpvs. Corollary 3 only tells us this: *For any given polyhedron defined by the CHSH inequalities (such as in row 4 of Fig. 1), it is possible to associate to any point (x, y, z) in the polyhedron some NC model instance that has correlations (w, x, y, z) for the appropriate w .*

2.12. Fine’s Theorem in action

Figure 7. Mathematica interactive demo for finding a set of NC model instance parameters ρ consistent with a randomly chosen set of correlations \mathbf{p} (purple dot) that satisfy CHSH. $w = 0$ case is shown. Other cases are obtained by moving the slider in the control area at the top.

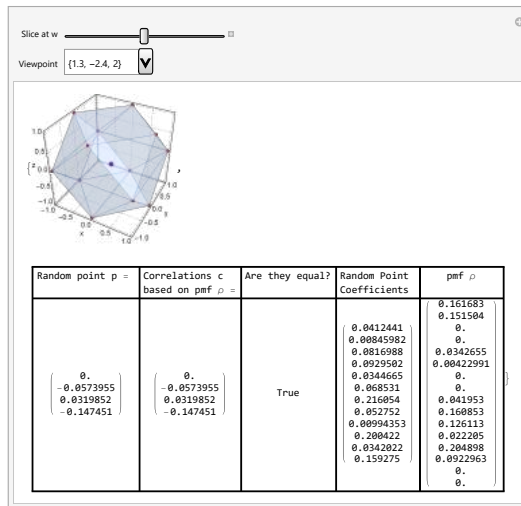


Fig. 7 shows example output from a Mathematica interactive program that finds the NC parameters ρ that are consistent with a random set of correlations \mathbf{p} . Moving the slider selects different values for the first correlation w , and a triple of the remaining correlations (x, y, z) is chosen randomly (constrained so that (w, x, y, z) satisfies the CHSH inequalities, of course). Both the appropriate CHSH inequality region shape (as a 3D slice) and the random point (x, y, z) are then displayed, along with tables of values showing, respectively from left to right, the coordinates of the random point \mathbf{p} , the computed correlations based on the NC parameters ρ , whether or not they are equal, the coefficients of the point \mathbf{p} as a convex combination of the vertices of the polyhedron, and finally the NC model instance parameters ρ that were generated from \mathbf{p} ’s coefficients.

2.13. No faster-than-light communication

Despite the fact that QM model instances are not factorizable⁸ (i.e., are "nonlocal"), it is important to point out that the EPRB experiment cannot be used for faster-than-light transmission of information between Alice and Bob.⁹ Of course this is well known, but the parameterized operational modeling approach provides a different viewpoint. Instead of focusing on proving the "no-communication" properties of QM itself, generic operational models are constructed that *do* allow Bob to send messages to Alice, and then show that no QM model instance could possibly have this property.

To demonstrate this perspective, consider what it takes for Bob to communicate with Alice. There must be something that Bob *controls* that is simultaneously something Alice can *deduce*. Table 13 shows an example of how this can be done. This will be called the **measurement predictable** model type. (See Sec. 5.4 for more details.) The following argument is predicated on the assumption that Alice knows in advance that Table 13 specifies the conditional probabilities for the experiment. That is, she knows the experimental arrangement, which is not unreasonable to assume. Furthermore, it is assumed that these probabilities hold for every iteration of the experiment.

Referring then to Table 13, suppose, for example, that Bob chooses measurement $v = 1$. Then whether Alice chooses her measurement $u = 1$ or $u = 2$, she gets the outcome $s = -1$ corresponding to the only (possibly) nonzero probabilities $\gamma_1, \gamma_2, \gamma_9, \gamma_{10}$. Hence she *deduces* that Bob's measurement choice must be $v = 1$.

On the other hand, suppose that Bob chooses measurement $v = 2$. Then whether Alice chooses her measurement $u = 1$ or $u = 2$, she gets the outcome $s = 1$ corresponding to the only (possibly) nonzero probabilities $\gamma_7, \gamma_8, \gamma_{15}, \gamma_{16}$. Hence she *deduces* that Bob's measurement choice must be $v = 2$.

Table 13. Example of a **measurement predictable (MP)** model type, in which Alice can reliably predict (with probability 1) Bob's measurement. Some of the γ_k 's have been set to 0 *a priori*, but the measurement context constraints still hold. There are exactly 4 such templates for measurement predictability.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	γ_1	γ_2	0	0
$(1, 2)$	0	0	γ_7	γ_8
$(2, 1)$	γ_9	γ_{10}	0	0
$(2, 2)$	0	0	γ_{15}	γ_{16}

⁸ Except for the trivial one. See Table 19.

⁹ Assuming QM rules, that is.

Hence Bob can purposely send a message encoded as a sequence of measurement choices, which Alice can then decode. Of course, there is no QM model instance that can be measurement predictable. Just compare this template to the QM model type of Table 3, in which each row has the pattern "abba". This is inconsistent with each row of Table 13. The same is true of the other three measurement predictable model templates (see Sec. 5.4). The proof of Thm. 12 shows that no noncontextual model instance can be measurement predictable, either (hence eliminating any factorizable model instances too, of course).

The advantage of this viewpoint is that the set of model instances that allow faster-than-light communication are specifically identified. They are inside the complement of the so-called "no-signaling" model instances (i.e., those that satisfy RCI), of course, but more than that, they constitute a *proper* subset of this complement, thus identifying more precisely the boundary between FTL "signaling" and "no-signaling".

This same technique can be used to produce an "outcome predictable" model type, in which Alice can reliably predict Bob's outcome (see Sec. 5.1). This analysis confirms the fact that only certain QM model instances can be outcome predictable (these are the cases that Bell compared to "Bertlmann's socks" [20]). Note that outcome predictability is *not sufficient* for Bob to send messages to Alice, since he does not *control* his outcomes.

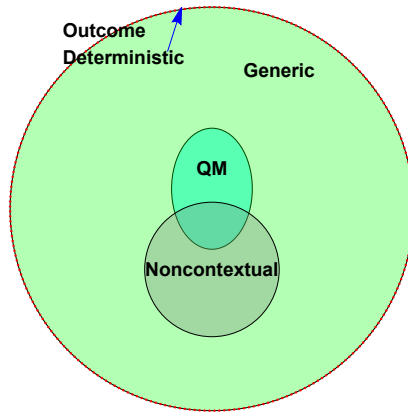
3. Hidden variables and convex combinations of OD cpvs are (almost) equivalent

In this paper, finite ensembles of OD cpvs (model instances) are used to build up more complicated cpvs (model instances) – see Thm.'s 3 and 14. This is an alternative perspective to the traditional hidden variables approach involving an unspecified measure on a (possibly) infinite probability space of hidden variables. In this section, the (mathematical) equivalence of the two representations is demonstrated, under the assumption of integrability of certain random variables with respect to the hidden variables measure. But first it is helpful to summarize the convex combination approach and to review standard hidden variables, rephrasing it slightly to conform to the notation in this paper.

3.1. Convex combinations of OD model instances

In Sec. 2.9 (Thm. 3) it was shown that the convex of hull the 16 RCI+OD model instances equals the set of noncontextual model instances. This naturally leads to: What is the convex hull of all 256 OD model instances? The proof of Thm. 14 in Appendix C shows that it equals the set of *all* generic model instances, including QM, factorizable, and noncontextual ones. See Fig. 8 and compare to Fig. 4 in Sec. 2.9.

Figure 8. The convex hull of all 256 OD conditional probability vectors (red dots on the boundary) equals the full set of generic conditional probability vectors, including the QM and NC ones. See Thm. 14 in Appendix C for proof.



In Fig. 8, the 256 OD conditional probability vectors are the extreme points on the boundary, shown as red dots (must look closely). The figure is a schematic representation of a polytope in a 13-dimensional subspace of \mathbb{R}^{16} . This is because the 16×256 matrix with the OD conditional probability vectors as columns has rank 13.¹⁰

3.2. Review of hidden variables

The "hidden variables" idea is to construct a statistical model of the experiment by combining more basic models, each of which is associated with a "hidden variable" which explains its behavior. These elementary models are then combined via integration over the space of hidden variables to produce the overall final model for the statistics of the experiment. Formally,

1. There is a probability space $(\Lambda, \mathcal{F}, \mu)$ where Λ is a set of "hidden variables" (or "complete states"), \mathcal{F} is a suitable σ -algebra of subsets of Λ , and μ is a probability measure on \mathcal{F} .
2. For each $\lambda \in \Lambda$, there exists a cpv

$$\gamma_\lambda = (\gamma_1(\lambda), \gamma_2(\lambda), \dots, \gamma_{16}(\lambda))$$

where it is assumed that each γ_k is μ -integrable.

3. The final overall cpv describing the experimental statistics is constructed by integrating

¹⁰ *Mathematica's* MatrixRank function was used to compute this.

over Λ (integration is performed componentwise):

$$\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_{16}) = \int_{\Lambda} \boldsymbol{\gamma}_{\lambda} \mu(d\lambda).$$

Note that $\boldsymbol{\gamma}$ is a cpv since the measurement context constraints are satisfied. For example

$$\begin{aligned} \sum_{k=1}^4 \gamma_k &= \\ \sum_{k=1}^4 \int_{\Lambda} \gamma_k(\lambda) \mu(d\lambda) &= \\ \int_{\Lambda} \sum_{k=1}^4 \gamma_k(\lambda) \mu(d\lambda) &= \\ \int_{\Lambda} 1 \mu(d\lambda) &= 1, \end{aligned}$$

and similarly for the other three measurement context constraints.

3.3. Proof of (almost) equivalence of convex combinations and hidden variables

In the proof of Thm. 5, it is shown that any generic model instance written as a convex combination of OD model instances can be turned into a *finite* hidden variables *model*. No attempt is made to determine what sort of hidden variables *theory* might give rise to this model.

In the proof of Thm. 6 it is shown, that with an extra "reasonable" assumption, the converse is true, namely a hidden variables model can always be turned into a representation as a convex combination of OD cpvs.

Theorem 5. Assume a generic cpv $\boldsymbol{\gamma}$ written as

$$\boldsymbol{\gamma} = \sum_{k=1}^{256} \pi_k \boldsymbol{\gamma}_k \tag{27}$$

for some pmf $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_{256})$ and where $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \dots, \boldsymbol{\gamma}_{256})$ denotes the set of 256 OD cpvs in some order. Then Eq. 27 can be written in the form of a hidden variables model as

$$\boldsymbol{\gamma} = \int_{\Lambda} \boldsymbol{\gamma}_{\lambda} \mu(d\lambda)$$

over a space Λ of hidden variables, where each $\boldsymbol{\gamma}_{\lambda}$ is a cpv.

Proof. Assume $\boldsymbol{\gamma} = \sum_{k=1}^{256} \pi_k \boldsymbol{\gamma}_k$. Define $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{256}\}$, $\mathcal{F} = 2^{\Lambda}$, and $\mu(\{\lambda_k\}) =$

π_k for all k . Extend μ to all of \mathcal{F} in the natural way. Define

$$\gamma_\lambda = \gamma_k$$

whenever $\lambda = \lambda_k$ for all $k = 1, 2, \dots, 256$. Then summation and integration become the same, and

$$\gamma = \sum_{k=1}^{256} \pi_k \gamma_k = \int_{\Lambda} \gamma_\lambda \mu(d\lambda). \tag{28}$$

□

Theorem 6. Assume a generic cpv can be written in the form

$$\gamma = \int_{\Lambda} \gamma_\lambda \mu(d\lambda) \tag{29}$$

over a space Λ of hidden variables, where γ_λ is a cpv for each λ . By Thm. 14, for each $\lambda \in \Lambda$, there exists a pmf $(\eta_1(\lambda), \eta_2(\lambda), \dots, \eta_{256}(\lambda))$ such that

$$\gamma_\lambda = \sum_{k=1}^{256} \eta_k(\lambda) \gamma_k,$$

where $(\gamma_1, \gamma_2, \dots, \gamma_{256})$ denotes the set of 256 OD cpvs in some order. It is assumed that each η_k is μ -measurable.¹¹ Then

$$\pi_k \doteq \int_{\Lambda} \eta_k(\lambda) \mu(d\lambda)$$

is well-defined for each $k = 1, 2, \dots, 256$. Finally Eq. 29 can be written as a convex combination of cpvs as follows:

$$\gamma = \sum_{k=1}^{256} \pi_k \gamma_k.$$

Proof. Assume $\gamma = \int_{\Lambda} \gamma_\lambda \mu(d\lambda)$. It was assumed that each random variable $\eta_k(\cdot) : \Lambda \rightarrow [0, 1]$ is μ -measurable. Since a bounded, Lebesgue-measurable function defined on a domain of finite measure is Lebesgue-integrable, this yields μ -integrability for each η_k . Hence the

¹¹ This is the "reasonable" assumption. It is certainly true, for example, if Λ is finite (and \mathcal{F} is the power set of Λ).

π_k 's are well-defined. Also, since $(\eta_1(\lambda), \eta_2(\lambda), \dots, \eta_{256}(\lambda))$ is a pmf for each λ ,

$$\begin{aligned} \sum_{k=1}^{256} \pi_k &= \sum_{k=1}^{256} \int_{\Lambda} \eta_k(\lambda) \mu(d\lambda) = \\ &\int_{\Lambda} \sum_{k=1}^{256} \eta_k(\lambda) \mu(d\lambda) = \\ &\int_{\Lambda} 1 \mu(d\lambda) = 1, \end{aligned} \tag{30}$$

hence $(\pi_1, \pi_2, \dots, \pi_{256})$ is a pmf as well. Finally,

$$\begin{aligned} \gamma &= \int_{\Lambda} \sum_{k=1}^{256} \eta_k(\lambda) \gamma_k \mu(d\lambda) = \\ &\sum_{k=1}^{256} \gamma_k \int_{\Lambda} \eta_k(\lambda) \mu(d\lambda) = \\ &\sum_{k=1}^{256} \pi_k \gamma_k. \end{aligned} \tag{31}$$

□

4. QM model instances can be both contextual and noncontextual

In this section, examples are given showing how to write a noncontextual QM model instance and a contextual QM model instance explicitly as a convex combination of OD model instances. Following these examples, the essential difference between noncontextuality and contextuality is discussed using a fable.

4.1. QM model instance that is also noncontextual

Table 14 shows a model instance that is both QM and NC. Table 15 shows how to write it as a convex combination of the 4 RCI+OD model instances in columns 2,8,9, and 15 from Table 11.

Table 14. Example of QM model instance that is also NC. The QM parameters are $\theta_1 = 135^\circ, \theta_2 = 0^\circ, \theta_3 = 90^\circ, \theta_4 = 45^\circ$, and the NC parameters are $\rho_k = \frac{1}{4}$ for $k = 2, 8, 9, 15$, and $\rho_k = 0$ otherwise.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$(1, 2)$	$\frac{1}{2}$	0	0	$\frac{1}{2}$
$(2, 1)$	0	$\frac{1}{2}$	$\frac{1}{2}$	0
$(2, 2)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Table 15. The model instance in Table 14 can be written as the convex combination of the RCI+OD model instances in columns 1-4 below, with weights $\pi_k = \frac{1}{4}$ for $k = 1, 2, 3, 4$. These conditional probability vectors are columns 2,8,9,15 from Table 11.

s	t	u	v	1	2	3	4
-1	-1	1	1	1	0	0	0
-1	1	1	1	0	0	1	0
1	-1	1	1	0	1	0	0
1	1	1	1	0	0	0	1
-1	-1	1	2	1	0	1	0
-1	1	1	2	0	0	0	0
1	-1	1	2	0	0	0	0
1	1	1	2	0	1	0	1
-1	-1	2	1	0	0	0	0
-1	1	2	1	0	0	1	1
1	-1	2	1	1	1	0	0
1	1	2	1	0	0	0	0
-1	-1	2	2	0	0	1	0
-1	1	2	2	0	0	0	1
1	-1	2	2	1	0	0	0
1	1	2	2	0	1	0	0
Weights				$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

4.2. QM model instance that is also contextual

Consider the infamous QM model instance with an s_1 -function value of $2\sqrt{2}$. This is obviously not NC because one of the CHSH inequalities is violated. It has QM parameters $\theta_1 = 67.5^\circ$ and $\theta_k = 22.5^\circ$ for $k = 2, 3, 4$. Table 16 shows how to write this as a convex

combination of 13 OD model instances.¹² None of the OD model instances satisfy RCI, 8 come from the subset of 112 OD model instances that satisfy CHSH but do not satisfy RCI, and 5 come from the subset of 128 model instances that have an s -function that exceeds a Tsirelson bound (see Fig. 3). The corresponding weights π_k in the bottom row are defined in Eq. 32.

Table 16. The QM model instance with $\theta_1 = 67.5^\circ$ and $\theta_k = 22.5^\circ$ for $k = 2, 3, 4$ can be written as a convex combination of these 13 OD model instances. The weights π_k are defined in Eq. 32.

s	t	u	v	1	2	3	4	5	6	7	8	9	10	11	12	13
-1	-1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
-1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0
1	-1	1	1	0	0	0	0	0	0	0	1	1	1	1	1	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1
-1	-1	1	2	1	1	0	0	0	0	0	1	0	0	0	0	1
-1	1	1	2	0	0	0	0	0	0	0	0	1	1	0	0	0
1	-1	1	2	0	0	0	1	1	0	0	0	0	0	1	1	0
1	1	1	2	0	0	1	0	0	1	1	0	0	0	0	0	0
-1	-1	2	1	1	0	1	0	0	0	0	1	0	0	0	0	1
-1	1	2	1	0	0	0	0	0	0	0	0	1	0	1	0	0
1	-1	2	1	0	0	0	1	0	1	0	0	0	1	0	1	0
1	1	2	1	0	1	0	0	1	0	1	0	0	0	0	0	0
-1	-1	2	2	0	1	1	0	0	0	0	1	0	0	0	0	1
-1	1	2	2	0	0	0	0	0	0	0	0	1	0	0	1	0
1	-1	2	2	0	0	0	0	1	1	0	0	0	1	1	0	0
1	1	2	2	1	0	0	1	0	0	1	0	0	0	0	0	0
Weights				π_1	π_2	π_3	π_4	π_5	π_6	π_7	π_8	π_9	π_{10}	π_{11}	π_{12}	π_{13}

¹² *Mathematica* was used to find this solution.

The weights $\pi = (\pi_1, \pi_2, \dots, \pi_{13})$ form a pmf, where

$$\begin{aligned}
 \pi_k &= \frac{1}{6} \cos^2 67.5^\circ \text{ for } k = 1, 2, 3, 10, 11, 12, \\
 \pi_k &= \frac{1}{12} \cos^2 67.5^\circ \text{ for } k = 4, 5, 6, \\
 \pi_7 &= \frac{1}{8} (1 - 3 \cos(2 \times 67.5^\circ)), \\
 \pi_8 &= -\frac{1}{6} (1 + 4 \cos(2 \times 67.5^\circ)), \\
 \pi_9 &= \frac{1}{3} \cos^2 67.5^\circ, \\
 \pi_{13} &= \frac{1}{2} \cos^2 67.5^\circ.
 \end{aligned} \tag{32}$$

It is straightforward to check that the correlations and s -functions for this model instance are given by, respectively,

$$\begin{aligned}
 (w, x, y, z) &= \frac{1}{2} \sqrt{2} (-1, 1, 1, 1), \text{ and} \\
 (s_1, s_2, s_3, s_4) &= (2\sqrt{2}, 0, 0, 0).
 \end{aligned}$$

As was shown in general in Sec. 3.3, this construction can be turned into a hidden variables *model* $(\Lambda, \mathcal{F}, \mu)$ (no claims of a *theory*) by setting

$$\begin{aligned}
 \Lambda &= \{\lambda_1, \lambda_2, \dots, \lambda_{13}\}, \mathcal{F} = 2^\Lambda, \\
 \mu(\{\lambda_k\}) &= \pi_k, \text{ (extend to } \mathcal{F} \text{ in the natural way), and} \\
 \gamma_\lambda &= \gamma_k \text{ whenever } \lambda = \lambda_k,
 \end{aligned} \tag{33}$$

where γ_k is the k th column of Table 16 for $k = 1, 2, \dots, 13$. This is a perfectly legitimate *outcome deterministic* hidden variables model, but it is not *local*, either in the sense of RCI or factorizability. It is interesting to note here that

- Each of the 13 individual components of the construction are OD but do not satisfy RCI.
- The QM model instance that is the culmination of the convex combination (equivalent to a hidden variables' construction) is *not* OD but *does* satisfy RCI.

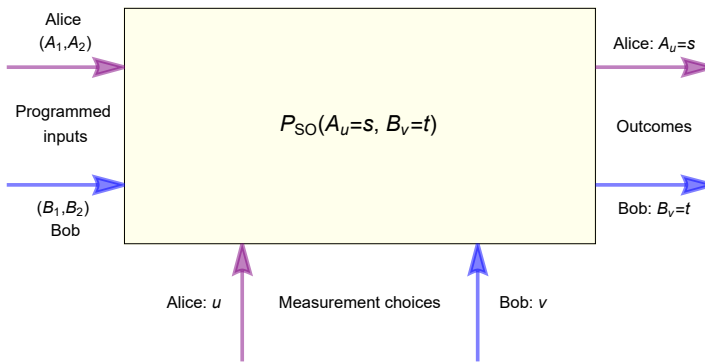
4.3. A fable illustrating the difference between contextuality and noncontextuality

Intuitively, what is the difference between the *noncontextual* QM model instance in Sec. 4.1 and the corresponding *contextual* QM model instance in Sec. 4.2? After all, in both cases, the QM model instance can be written as a convex combination of OD model

instances. Therefore, it is not unreasonable to hope that, in both cases, one could reproduce the statistics of the target QM model instance in a "classical" way. For example, imagine a preparation process that, on each iteration of the experiment, produces one of the basic OD model instances with a probability equal to its weight in the convex combination. The long-term statistics should then approach those of the target QM model instance.

Putting this into the form of a fable, imagine that Eve (Alice and Bob’s friend) has been asked to try and fool Alice and Bob into thinking they are getting data from a real EPRB experiment. She sends out four values (A_1, A_2, B_1, B_2) on each iteration of the experiment. (A_1, A_2) go to Alice and (B_1, B_2) go to Bob. Also suppose that if Alice chooses her first measurement option $M_A = 1$, she gets the result $A = A_1$ and if she chooses her second $M_A = 2$, she gets $A = A_2$. Similarly for Bob’s result B with respect to his own measurement options. Fig. 9 illustrates this setup. Borrowing suggestive terminology from Fine’s excellent (and more general) paper [11] on joint distributions, the random variables (A_1, A_2, B_1, B_2) will be called "statistical observables".

Figure 9. The imaginary scenario in which Eve attempts to fool Alice and Bob into thinking they are observing data from an EPRB experiment. Eve is in control of the 4-tuple of inputs (A_1, A_2, B_1, B_2) .



Recall that in the actual EPRB experiment, Alice and Bob can only (jointly) record statistics on *compatible pairs* (A_u, B_v) for $u, v = 1$ or 2 . It is not possible to perform simultaneous measurements of non-commuting observables. Therefore, Eve’s task is to send out the statistical observables (A_1, A_2, B_1, B_2) in such a way that the resulting *occurrence frequencies of compatible pairs* $(A_1, B_1), (A_1, B_2), (A_2, B_1), (A_2, B_2)$ approach the corresponding expected frequencies for the target QM model instance. Here are two strategies she could try.

1. Strategy 1. Eve adopts a full joint probability distribution

$$P_{SO}(A_1 = s, A_2 = s', B_1 = t, B_2 = t')$$

on (A_1, A_2, B_1, B_2) . See Table 17. The long-term frequency with which she sends out

a given 4-tuple (s, s', t, t') is determined entirely by the associated ρ_k . For example,

$$P_{SO}(A_1 = 1, A_2 = -1, B_1 = -1, B_2 = 1) = \rho_7,$$

so in the long-term, Alice and Bob will be presented with the 4-tuple $(1, -1, -1, 1)$ approximately $100\rho_7\%$ of the time, from which they choose a compatible pair, based on their measurement choices.

Table 17. A full joint pmf $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$ for the imaginary scenario. Each ρ_k corresponds to a full joint probability $P_{SO}(A_1 = s, A_2 = s', B_1 = t, B_2 = t')$. These determine the approximate long-term frequency with which Eve sends out any given 4-tuple (s, s', t, t') .

	(s, s')			
(t, t')	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(-1, -1)$	ρ_1	ρ_2	ρ_3	ρ_4
$(-1, 1)$	ρ_5	ρ_6	ρ_7	ρ_8
$(1, -1)$	ρ_9	ρ_{10}	ρ_{11}	ρ_{12}
$(1, 1)$	ρ_{13}	ρ_{14}	ρ_{15}	ρ_{16}

Then a straightforward marginal probability calculation yields the compatible double marginals in Table 18. Obviously these double marginals are exactly the same form as the noncontextual double conditional probabilities of Table 5.

Table 18. The statistical observables' double marginals $P_{SO}(A_u = s, B_v = t)$. These are identical to the NC double conditional probabilities of Table 5.

Indices	Outcomes (s, t)			
	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\rho_1 + \rho_2 + \rho_5 + \rho_6$	$\rho_9 + \rho_{10} + \rho_{13} + \rho_{14}$	$\rho_3 + \rho_4 + \rho_7 + \rho_8$	$\rho_{11} + \rho_{12} + \rho_{15} + \rho_{16}$
$(1, 2)$	$\rho_1 + \rho_2 + \rho_9 + \rho_{10}$	$\rho_5 + \rho_6 + \rho_{13} + \rho_{14}$	$\rho_3 + \rho_4 + \rho_{11} + \rho_{12}$	$\rho_7 + \rho_8 + \rho_{15} + \rho_{16}$
$(2, 1)$	$\rho_1 + \rho_3 + \rho_5 + \rho_7$	$\rho_9 + \rho_{11} + \rho_{13} + \rho_{15}$	$\rho_2 + \rho_4 + \rho_6 + \rho_8$	$\rho_{10} + \rho_{12} + \rho_{14} + \rho_{16}$
$(2, 2)$	$\rho_1 + \rho_3 + \rho_9 + \rho_{11}$	$\rho_5 + \rho_7 + \rho_{13} + \rho_{15}$	$\rho_2 + \rho_4 + \rho_{10} + \rho_{12}$	$\rho_6 + \rho_8 + \rho_{14} + \rho_{16}$

Therefore, Eve can mimic the statistics of the QM model instance in Sec. 4.1 but *not* the one in Sec. 4.2.

2. Strategy 2. On each iteration of the experiment, Eve *guesses* what Alice and Bob's joint measurement choice will be, then sends out a 4-tuple (s, s', t, t') with a probability based on that assumption. This strategy is illustrated for an arbitrary QM model instance in the case where she guesses that both Alice and Bob will make their *first* measurement choice. In this case, Eve sends out (where " * " denotes "don't

care")

$$\begin{aligned}
 &(-1, *, -1, *) \text{ with probability } \frac{1}{2} \cos^2 \theta_1, \\
 &(-1, *, 1, *) \text{ with probability } \frac{1}{2} \sin^2 \theta_1, \\
 &(1, *, -1, *) \text{ with probability } \frac{1}{2} \sin^2 \theta_1, \\
 &(1, *, 1, *) \text{ with probability } \frac{1}{2} \cos^2 \theta_1.
 \end{aligned} \tag{34}$$

Then if Alice and Bob actually make their first measurement choices, thus choosing the values in the first (Alice’s) and third (Bob’s) places in each of these 4-tuples, in the long-term they will get the values shown with a frequency close to the those predicted by the QM model type as shown in Table 3. Similarly for the other three measurement choice combinations for Alice and Bob. The problem with this strategy of course is that Eve will often be wrong in her guess about Alice and Bob’s upcoming measurement choices, so the values she sends them will occur with the "wrong" frequencies. In fact, she can only reliably make this strategy work if the QM model instance is also the trivial model instance. See Table 19. Since this is the completely uniform, all zero-correlation situation, it is not of much value or interest.

Table 19. The trivial model instance is the only one that is simultaneously QM ($\theta_1 = 135^\circ, \theta_k = 45^\circ$ for $k = 2, 3, 4$), factorizable ($\alpha_k = \beta_k = \frac{1}{2}$ for all $k = 1, 2, 3, 4$), and noncontextual ($\rho_k = \frac{1}{16}$ for all $k = 1, 2, \dots, 16$). All of its correlations and s -functions are 0.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$(1, 2)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$(2, 1)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$(2, 2)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

As a last hope, what about sending out the 13 OD model instances of Table 16 with frequencies given by the weights in Eq. 32? The problem with this idea is that at least one (in fact all) of the OD model instances in Table 16 fail to satisfy RCI, hence cannot be noncontextual. In other words, for every one of these OD model instance, there is no full joint probability distribution on (A_1, A_2, B_1, B_2) whose marginals match the double conditionals of that OD model instance.

It is interesting to pause and contemplate what it means that neither of these strategies always works. In strategy 1, the overall joint probability distribution (for both compatible and incompatible statistical observables) basically *encodes* the desired correlations into

the experimental results, *regardless of future measurement choices*. Strategy 2 *requires* foreknowledge of the measurement choices. In other words, in a QM world, in which the behavior can be contextual, it is not always possible to fake the results in a way that ignores the future measurements.

5. Predictability

Correlations are one thing, actual transfer of information quite another. The parameterized operational modeling framework is very useful in exploring this concept. "Predictability" is all about Alice's ability "predict" or "know" something about Bob's wing of the experiment. It comes in two different flavors. The first, "outcome predictability" (OP), indicates that Alice can reliably predict (i.e., with probability 1) what Bob's outcome will be, based solely on her measurement choice and outcome. "Measurement predictability" (MP) indicates that Alice can reliably deduce what Bob's measurement choice is, again based solely on her measurement choice and outcome. No *direct* knowledge of Bob's measurement choice is required in either case. In the "measurement predictability" case, Bob can send information to Alice through his choice of measurement.

In Sec. 5.2, it will be shown that *some* QM model instances are outcome predictable. But in Sec. 5.5, it will be shown that *no* QM model instance is measurement predictable – alas no faster-than-light communication in a Bell/Aspect experiment under QM rules. On the other hand, it will be shown that there are infinitely many generic model instances which *are* outcome predictable and/or measurement predictable.

The ability of Alice to deduce either Bob's outcome or measurement choice is predicated on the assumptions

- The generic parameters $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ are fixed throughout the experiment,
- The generic parameters γ conform to one of several special "patterns" of zero and non-zero entries,
- Alice knows this pattern.

These assumptions merely state that Alice knows the experimental setup, and that it is fixed for all iterations of the experiment.

5.1. Outcome predictability

OP implies that every Alice outcome-measurement choice pair $(A, M_A) = (s, u)$ is mapped to only one of Bob's outcomes $B = t$. Table 20 shows the 16 possible "patterns" for generic parameters that fulfill this condition.

Table 20. The 16 **outcome predictable (OP)** model type patterns. Alice can always deduce what Bob’s outcome is, given her own measurement choice and outcome. This is predicated on the assumption that Alice knows in advance which conditional probabilities are zero.

s	t	u	v	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
-1	-1	1	1	γ_1	γ_1	γ_1	γ_1	γ_1	γ_1	γ_1	γ_1	0	0	0	0	0	0	0	0
-1	1	1	1	0	0	0	0	0	0	0	0	γ_2	γ_2	γ_2	γ_2	γ_2	γ_2	γ_2	γ_2
1	-1	1	1	γ_3	γ_3	γ_3	γ_3	0	0	0	0	γ_3	γ_3	γ_3	γ_3	0	0	0	0
1	1	1	1	0	0	0	0	γ_4	γ_4	γ_4	γ_4	0	0	0	0	γ_4	γ_4	γ_4	γ_4
-1	-1	1	2	γ_5	γ_5	γ_5	γ_5	γ_5	γ_5	γ_5	γ_5	0	0	0	0	0	0	0	0
-1	1	1	2	0	0	0	0	0	0	0	0	γ_6	γ_6	γ_6	γ_6	γ_6	γ_6	γ_6	γ_6
1	-1	1	2	γ_7	γ_7	γ_7	γ_7	0	0	0	0	γ_7	γ_7	γ_7	γ_7	0	0	0	0
1	1	1	2	0	0	0	0	γ_8	γ_8	γ_8	γ_8	0	0	0	0	γ_8	γ_8	γ_8	γ_8
-1	-1	2	1	γ_9	γ_9	0	0	γ_9	γ_9	0	0	γ_9	γ_9	0	0	γ_9	γ_9	0	0
-1	1	2	1	0	0	γ_{10}	γ_{10}	0	0	γ_{10}	γ_{10}	0	0	γ_{10}	γ_{10}	0	0	γ_{10}	γ_{10}
1	-1	2	1	γ_{11}	0	γ_{11}	0	γ_{11}	0	γ_{11}	0	γ_{11}	0	γ_{11}	0	γ_{11}	0	γ_{11}	0
1	1	2	1	0	γ_{12}	0	γ_{12}	0	γ_{12}	0	γ_{12}	0	γ_{12}	0	γ_{12}	0	γ_{12}	0	γ_{12}
-1	-1	2	2	γ_{13}	γ_{13}	0	0	γ_{13}	γ_{13}	0	0	γ_{13}	γ_{13}	0	0	γ_{13}	γ_{13}	0	0
-1	1	2	2	0	0	γ_{14}	γ_{14}	0	0	γ_{14}	γ_{14}	0	0	γ_{14}	γ_{14}	0	0	γ_{14}	γ_{14}
1	-1	2	2	γ_{15}	0	γ_{15}	0	γ_{15}	0	γ_{15}	0	γ_{15}	0	γ_{15}	0	γ_{15}	0	γ_{15}	0
1	1	2	2	0	γ_{16}	0	γ_{16}	0	γ_{16}	0	γ_{16}	0	γ_{16}	0	γ_{16}	0	γ_{16}	0	γ_{16}

To see this, note that every one of the 16 patterns in Table 20 is like a standard set of generic parameters, except that certain γ_k ’s are set to zero *a priori*. It is assumed that the measurement context constraints (Eq. 2) still hold.

Consider the first model instance in the column marked "1". Suppose Alice chooses her measurement $M_A = u = 1$ and gets outcome $A = s = -1$. This can only happen in rows 1 and/or 5, since the other possibilities are associated with $\gamma_2 = 0$ and $\gamma_6 = 0$. And in these cases (where at least one of γ_1 and γ_5 must be nonzero), Alice can see that Bob’s outcome is always $B = t = -1$. Similarly suppose Alice chooses her measurement $M_A = u = 1$ and gets outcome $A = s = +1$. This can only happen in rows 3 and/or 7, since the other possibilities are associated with $\gamma_4 = 0$ and $\gamma_8 = 0$. And in these cases (where at least one of γ_3 and γ_7 must be nonzero), Alice can see that Bob’s outcome is again always $B = t = -1$. A similar argument can be made for the cases where $(A, M_A) = (s, u) = (-1, 2)$ or $(1, 2)$. Extend the same arguments to all of the other columns.

In other words, Alice can always deduce Bob’s outcome because any ambiguities were preemptively removed by strategically zeroing out certain γ_k ’s while retaining the measurement context constraints.

5.2. Can a QM model instance be outcome predictable?

Theorem 7. *There are only four OP model instances that are also QM model instances.*

Proof. The (generic) parameters for any QM model instance must satisfy the two conditions (see for example Table 3):

- Each block of four parameters must have the pattern "abba", and
- All parameters must be between 0 and 1/2.

The only OP model instances in Table 20 that could possibly satisfy the first condition are numbers 6,7,10, and 11. Combining the second condition with the usual measurement context constraints (see Eq. 2) implies that all nonzero γ_k 's must be equal to 1/2. Therefore the only such cases are as shown in Table 21. □

Examples of QM parameter assignments that yield instances 1,2,3, and 4 of Table 21, respectively, are

$$\begin{aligned}
 1. & \theta_1 = \theta_2 = \theta_3 = \theta_4 = 0^\circ, \\
 2. & \theta_1 = 0^\circ, \theta_2 = 0^\circ, \theta_3 = 90^\circ, \theta_4 = -90^\circ, \\
 3. & \theta_1 = 90^\circ, \theta_2 = 90^\circ, \theta_3 = 0^\circ, \theta_4 = 0^\circ, \\
 4. & \theta_1 = 270^\circ, \theta_2 = 90^\circ, \theta_3 = 90^\circ, \theta_4 = 90^\circ.
 \end{aligned}
 \tag{35}$$

These QM model instances are all perfectly correlated, as their sets of correlations are, respectively

$$\begin{aligned}
 1. & (1, 1, 1, 1), \\
 2. & (1, 1, -1, -1), \\
 3. & (-1, -1, 1, 1), \\
 4. & (-1, -1, -1, -1).
 \end{aligned}
 \tag{36}$$

One suspects there are RCI+OD model instances that can achieve the same correlations. Table 22 shows a comparison of an RCI+OD model instance ("OD" column) and a QM model instance ("QM" column), both of which achieve the perfect correlations

$$(w, x, y, z) = (-1, -1, -1, -1)$$

with s -functions

$$(s_1, s_2, s_3, s_4) = (-2, -2, -2, -2).$$

Parameter assignments which produce the two model instances in Table 22 are:

$$\begin{aligned}
 \text{OD: } & \gamma_2 = \gamma_6 = \gamma_{10} = \gamma_{14} = 1 \text{ and all other } \gamma_k = 0, \\
 \text{QM: } & \theta_1 = 270^\circ, \theta_2 = \theta_3 = \theta_4 = 90^\circ.
 \end{aligned}
 \tag{37}$$

Once again, this illustrates the fact that the mapping from conditional probabilities to correlations is "many-to-one".

Table 21. The four QM model instances that are also outcome predictable.

s	t	u	v	1	2	3	4
-1	-1	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
-1	1	1	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	1	1	0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
-1	-1	1	2	$\frac{1}{2}$	$\frac{1}{2}$	0	0
-1	1	1	2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	1	2	0	0	$\frac{1}{2}$	$\frac{1}{2}$
1	1	1	2	$\frac{1}{2}$	$\frac{1}{2}$	0	0
-1	-1	2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0
-1	1	2	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	-1	2	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	1	2	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0
-1	-1	2	2	$\frac{1}{2}$	0	$\frac{1}{2}$	0
-1	1	2	2	0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	-1	2	2	0	$\frac{1}{2}$	0	$\frac{1}{2}$
1	1	2	2	$\frac{1}{2}$	0	$\frac{1}{2}$	0

Table 22. RCI+OD and QM model instances which both achieve the correlations $(w, x, y, z) = (-1, -1, -1, -1)$.

s	t	u	v	OD	QM
-1	-1	1	1	0	0
-1	1	1	1	1	$\frac{1}{2}$
1	-1	1	1	0	$\frac{1}{2}$
1	1	1	1	0	0
-1	-1	1	2	0	0
-1	1	1	2	1	$\frac{1}{2}$
1	-1	1	2	0	$\frac{1}{2}$
1	1	1	2	0	0
-1	-1	2	1	0	0
-1	1	2	1	1	$\frac{1}{2}$
1	-1	2	1	0	$\frac{1}{2}$
1	1	2	1	0	0
-1	-1	2	2	0	0
-1	1	2	2	1	$\frac{1}{2}$
1	-1	2	2	0	$\frac{1}{2}$
1	1	2	2	0	0

5.3. Outcome predictable instances that are not QM

Inspection of Table 20 yields the following three theorems.

Theorem 8. *There are an infinite number of outcome predictable model instances that are also NC.*

Proof. Many examples can be generated from the 16 patterns in Table 20. Consider model instance pattern number 6, for example. Then for the corresponding NC model parameters, set

$$\rho_k = 0 \text{ for all } k \text{ except } 1 > \rho_1 > 0 \text{ and } \rho_{16} = 1 - \rho_1.$$

Now plug these into the NC parameter list in Table 5. This produces a set of generic parameters that match the pattern of the 6th column of Table 20, namely

$$(\gamma_1, 0, 0, \gamma_4, \gamma_5, 0, 0, \gamma_8, \gamma_9, 0, 0, \gamma_{12}, \gamma_{13}, 0, 0, \gamma_{16}) =$$

$$(\rho_1, 0, 0, \rho_{16}, \rho_1, 0, 0, \rho_{16}, \rho_1, 0, 0, \rho_{16}, \rho_1, 0, 0, \rho_{16}).$$

□

Theorem 9. *There are an infinite number of outcome predictable model instances that satisfy CHSH.*

Proof. The example from Thm. 8 is NC, hence satisfies all CHSH inequalities. Obviously there are an infinite number of choices for the two defining parameters (ρ_1, ρ_{16}) . As another example, consider model instance number 7 in Table 20. The correlations are $(w, x, y, z) = (\gamma_1 + \gamma_4, \gamma_5 + \gamma_8, -\gamma_{10} - \gamma_{11}, -\gamma_{14} - \gamma_{15}) = (1, 1, -1, -1)$. Hence the s -functions are $(-2, -2, 2, 2)$, all of which satisfy CHSH. Obviously infinitely many values can be assigned to the non-zero γ_k 's which yield these same s -function values. \square

Theorem 10. *There are an infinite number of outcome predictable model instances that exceed at least one Tsirelson bound.*

Proof. Consider model instance pattern number 1 in Table 20. The correlations are $(w, x, y, z) = (\gamma_1 - \gamma_3, \gamma_5 - \gamma_7, \gamma_9 - \gamma_{11}, \gamma_{13} - \gamma_{15})$. One possible assignment is $\gamma_1 = 0, \gamma_3 = 1, \gamma_5 = 1, \gamma_7 = 0, \gamma_9 = 1, \gamma_{11} = 0, \gamma_{13} = 1, \gamma_{15} = 0$, which yields the correlations $(-1, 1, 1, 1)$. This results in the s -functions $(4, 0, 0, 0)$, so one s -function exceeds a Tsirelson bound. To obtain infinitely many assignments of values to the γ_k 's so that at least one s -functions exceeds a Tsirelson bound, introduce an $0 < \epsilon < 1$ to write the γ_k 's as follows: $\gamma_1 = \epsilon, \gamma_3 = 1 - \epsilon, \gamma_5 = 1, \gamma_7 = 0, \gamma_9 = 1, \gamma_{11} = 0, \gamma_{13} = 1, \gamma_{15} = 0$. Then the correlations become

$$(w, x, y, z) = (\gamma_1 - \gamma_3, \gamma_5 - \gamma_7, \gamma_9 - \gamma_{11}, \gamma_{13} - \gamma_{15}) = (2\epsilon - 1, 1, 1, 1) \tag{38}$$

which yields s -functions

$$(4 - 2\epsilon, 2\epsilon, 2\epsilon, 2\epsilon). \tag{39}$$

The first s -function exceeds a Tsirelson bound if $\epsilon < 2 - \sqrt{2} \approx 0.59$. For example, if $\epsilon = 0.5$, then the s -functions are $(3, 1, 1, 1)$, and the first one exceeds the upper Tsirelson bound, i.e. $3 > 2\sqrt{2}$. \square

5.4. Measurement predictability

MP implies that every Alice outcome-measurement choice pair $(A, M_A) = (s, u)$ is mapped to only one of Bob's measurements $M_B = v$. Table 23 shows the 4 possible "patterns" for generic parameters that fulfill this condition.

Table 23. The 4 measurement predictable (MP) model type patterns. Alice can always deduce what Bob’s measurement is, given her own measurement choice and outcome. This is predicated on the assumption that Alice knows in advance which conditional probabilities are zero.

s	t	u	v	1	2	3	4
-1	-1	1	1	γ_1	γ_1	0	0
-1	1	1	1	γ_2	γ_2	0	0
1	-1	1	1	0	0	γ_3	γ_3
1	1	1	1	0	0	γ_4	γ_4
-1	-1	1	2	0	0	γ_5	γ_5
-1	1	1	2	0	0	γ_6	γ_6
1	-1	1	2	γ_7	γ_7	0	0
1	1	1	2	γ_8	γ_8	0	0
-1	-1	2	1	γ_9	0	γ_9	0
-1	1	2	1	γ_{10}	0	γ_{10}	0
1	-1	2	1	0	γ_{11}	0	γ_{11}
1	1	2	1	0	γ_{12}	0	γ_{12}
-1	-1	2	2	0	γ_{13}	0	γ_{13}
-1	1	2	2	0	γ_{14}	0	γ_{14}
1	-1	2	2	γ_{15}	0	γ_{15}	0
1	1	2	2	γ_{16}	0	γ_{16}	0

To see this, note that every one of the 4 patterns in Table 23 is like a standard set of generic parameters, except that certain γ_k ’s are set to zero *a priori*. It is assumed that the measurement context constraints (Eq. 2) still hold. Consider the first model instance in the column marked "1". Suppose Alice chooses her measurement $M_A = u = 1$ and gets outcome $A = s = -1$. This can only happen in rows 1 and/or 2, since the other possibilities are associated with $\gamma_5 = 0$ and $\gamma_6 = 0$. And in the cases where either or both γ_1 and γ_2 are nonzero, Alice can see that Bob’s measurement is always $M_B = v = 1$. Now suppose Alice chooses her measurement $M_A = u = 1$ and gets outcome $A = s = +1$. This can only happen in rows 7 and/or 8, since the other possibilities are associated with $\gamma_3 = 0$ and $\gamma_4 = 0$. And in the cases where either or both γ_7 and γ_8 are nonzero, Alice can see that Bob’s measurement is always $M_B = v = 2$. A similar argument can be made for the cases where $(A, M_A) = (s, u) = (-1, 2)$ or $(1, 2)$. Extend the same arguments to all of the other columns.

Alice can always deduce Bob’s measurement because any ambiguities were preemptively removed by strategically zeroing out certain γ_k ’s. There are 16 cpvs that have this property, but after the zeroing process, 12 fail to satisfy the measurement context constraints, so only 4 are left.

5.5. The no-communication theorem for QM

It would be a surprise if any of these four model instance patterns could also be QM. The next theorem says they cannot, and therefore confirms the no-communication theorem for QM (in the context of a Bell/Aspect-like experiment, that is).

Theorem 11. *No QM model instance can also be measurement predictable.*

Proof. Consider the first four entries $(\gamma_1, \gamma_2, 0, 0)$ in the first measurement predictable pattern of Table 23. Recall these 4 numbers must sum to 1, so necessarily $\gamma_1 + \gamma_2 = 1$. In order to be a QM model instance, the pattern of these first four entries must be "abba" (see Table 3). This means $\gamma_1 = 0$. Then $\gamma_2 = 1$. But then the second and third entries 1 and 0 do not match, so this cannot be a QM model instance. A similar argument can be used to show the other three patterns could not possibly conform to the QM pattern either. \square

In the literature, the model instances that satisfy RCI are often called "no-signaling", presumably implying that any "signaling" (i.e. faster-than-light communication capable) model instances must lie in the set that do not satisfy RCI. The set of measurement predictable model instances are inside this RCI complement, of course, but more than that, they form a *proper* subset. That is, measurement predictability defines more precisely the boundary between "no-signaling" and "signaling". Consider the following OD model instance:¹³

Example 3. $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ where $\gamma_k = 1$ for $k = 1, 5, 9, 16$ and $\gamma_k = 0$ otherwise.

This OD model instance does not satisfy RCI (it is not one of the 16 listed in Table 11). It also cannot be measurement predictable. Consider Table 23, and simply check that there is no way to assign 1's and 0's to the non-zero γ_k 's in any of the 4 columns to produce Ex. 3.¹⁴

5.6. Measurement predictability and noncontextuality

Thm. 11 shows that no QM model instance can be MP. The next theorem shows a similar result for NC model instances.

Theorem 12. *There are no measurement predictable model instances that are NC, hence none that are factorizable either.*

Proof. Consider the table of MP instances in Table 23. The proof consists of a straightforward check that no instance in that table can be NC. That is, in Eq. 40, can

¹³ This is also Ex. 2 in Sec. 2.10, repeated here for convenience.

¹⁴ This is true for Bob sending messages to Alice. The corresponding table for Alice sending messages to Bob is different, and must be checked as well (exercise for reader).

the column vector on the left be equal to any of the four column vectors on the right?

$$\begin{pmatrix} \rho_1 + \rho_2 + \rho_5 + \rho_6 \\ \rho_9 + \rho_{10} + \rho_{13} + \rho_{14} \\ \rho_3 + \rho_4 + \rho_7 + \rho_8 \\ \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16} \\ \rho_1 + \rho_2 + \rho_9 + \rho_{10} \\ \rho_5 + \rho_6 + \rho_{13} + \rho_{14} \\ \rho_3 + \rho_4 + \rho_{11} + \rho_{12} \\ \rho_7 + \rho_8 + \rho_{15} + \rho_{16} \\ \rho_1 + \rho_3 + \rho_5 + \rho_7 \\ \rho_9 + \rho_{11} + \rho_{13} + \rho_{15} \\ \rho_2 + \rho_4 + \rho_6 + \rho_8 \\ \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16} \\ \rho_1 + \rho_3 + \rho_9 + \rho_{11} \\ \rho_5 + \rho_7 + \rho_{13} + \rho_{15} \\ \rho_2 + \rho_4 + \rho_{10} + \rho_{12} \\ \rho_6 + \rho_8 + \rho_{14} + \rho_{16} \end{pmatrix} = \begin{pmatrix} \gamma_1 & \gamma_1 & 0 & 0 \\ \gamma_2 & \gamma_2 & 0 & 0 \\ 0 & 0 & \gamma_3 & \gamma_3 \\ 0 & 0 & \gamma_4 & \gamma_4 \\ 0 & 0 & \gamma_5 & \gamma_5 \\ 0 & 0 & \gamma_6 & \gamma_6 \\ \gamma_7 & \gamma_7 & 0 & 0 \\ \gamma_8 & \gamma_8 & 0 & 0 \\ \gamma_9 & 0 & \gamma_9 & 0 \\ \gamma_{10} & 0 & \gamma_{10} & 0 \\ 0 & \gamma_{11} & 0 & \gamma_{11} \\ 0 & \gamma_{12} & 0 & \gamma_{12} \\ 0 & \gamma_{13} & 0 & \gamma_{13} \\ 0 & \gamma_{14} & 0 & \gamma_{14} \\ \gamma_{15} & 0 & \gamma_{15} & 0 \\ \gamma_{16} & 0 & \gamma_{16} & 0 \end{pmatrix} \tag{40}$$

Pick the first column on the right. Just start knocking off all of the ρ_k 's on the left in rows where 0's appear in the column vector on the right, and pretty soon there is nothing left to make *any* non-zero γ_k 's in rows 1,2,7,8,9,10,15, or 16. There can be no such factorizable instances either, since every factorizable model instance is also NC (see Thm. 2). A similar argument works for the other three column vectors on the right. \square

5.7. Measurement predictability, CHSH and the Tsirelson bounds

In spite of Thm. 12, there are infinitely many MP model instances that satisfy CHSH. These instances show that even NC is not *necessary* to derive the CHSH inequalities.

Theorem 13. *There are an infinite number of measurement predictable model instances that satisfy all CHSH inequalities and an infinite number that violate at least one CHSH inequality. In fact there are instances that exceed a Tsirelson bound.*

Proof. Consider the first MP model pattern from Table 23. An easy calculation shows the correlations to be

$$\begin{aligned} (w, x, y, z) &= (\gamma_1 - \gamma_2, \gamma_8 - \gamma_7, \gamma_9 - \gamma_{10}, \gamma_{16} - \gamma_{15}) = \\ &(2\gamma_1 - 1, 2\gamma_8 - 1, 2\gamma_9 - 1, 2\gamma_{16} - 1), \end{aligned} \tag{41}$$

with s -functions given by

$$(s_1, s_2, s_3, s_4) = (2(-1 - \gamma_1 + \gamma_8 + \gamma_9 + \gamma_{16}), 2(-1 + \gamma_1 - \gamma_8 + \gamma_9 + \gamma_{16}), 2(-1 + \gamma_1 + \gamma_8 - \gamma_9 + \gamma_{16}), 2(-1 + \gamma_1 + \gamma_8 + \gamma_9 - \gamma_{16})), \quad (42)$$

where $\gamma_1, \gamma_8, \gamma_9, \gamma_{16}$ are independently adjustable parameters between 0 and 1. Here are some examples for obtaining specific (s_1, s_2, s_3, s_4) values:

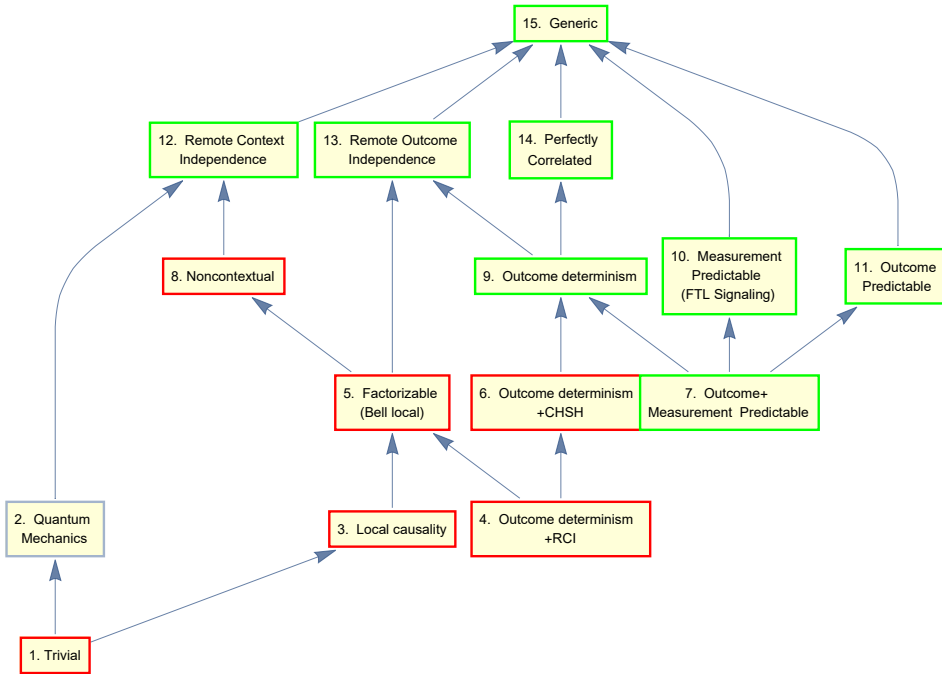
- $(-4, 0, 0, 0)$ by setting $\gamma_1 = 1, \gamma_8 = \gamma_9 = \gamma_{16} = 0$.
- $(-2, 2, -2, 2)$ by setting $\gamma_1 = 1, \gamma_8 = 0, \gamma_9 = 1, \gamma_{16} = 0$.
- $(-1, -3, 1, 1)$ by setting $\gamma_1 = \frac{1}{2}, \gamma_8 = 1, \gamma_9 = 0, \gamma_{16} = 0$.
- $(0, 0, 0, 0)$ by setting $\gamma_1 = \gamma_8 = \gamma_9 = \gamma_{16} = \frac{1}{2}$.

□

6. Summary and discussion

The original motivation for this paper was to understand why there continue to be so many papers concerning the meaning and significance of Bell's theorems even after several decades. The thinking was that ambiguous everyday words, such as "locality", "determinism", and "realism" must be represented by assumptions inside of precise mathematical models, and the originally intended meanings may get lost in translation, either going into or coming out of the model. The operational modeling viewpoint, as represented by Tables 1, 3, 5, 7, and 19, represents properties such as "noncontextual", "factorizable", "QM", "outcome determinism", etc. in a uniform way that enables direct comparisons. This led ultimately to the taxonomy of notions and concepts (not necessarily complete) shown in Fig. 10.

Figure 10. Relationships among sets of model instances for selected types. Arrows indicate set inclusion. Any box outlined in red indicates that all such model instances satisfy all of the CHSH inequalities. A box outlined in green indicates that at least some model instances may not only violate a CHSH inequality, but even exceed a Tsirelson bound.

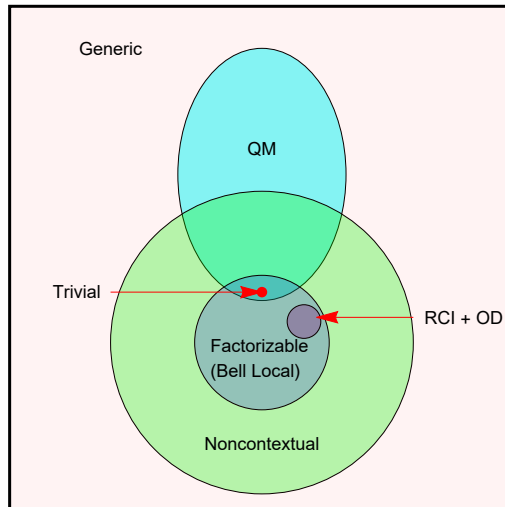


Each node of the graph consists of a set of conditional probability vectors (aka model instances) which satisfy a precise mathematical definition that is intended to capture the notion embodied in the name of the node. For example, box 5, labeled "Factorizable", represents the set of all factorizable model instances of the form shown in Table 7. The arrows represent set inclusion.

Every model instance in any box outlined in red satisfies all of the CHSH inequalities in Eq. 5. Some model instances in any box outlined in green not only violate at least one CHSH inequality, but also exceed a Tsirelson bound ($\pm 2\sqrt{2}$). In other words, none of the concepts represented by any of these red or green boxes are completely consistent with standard quantum mechanics. Put another way, hypothetical alternatives to quantum mechanics have been considered that not only "fall short" (i.e. satisfy CHSH) but are "excessive", i.e. have correlations that exceed even what quantum mechanics predicts.

Fig. 11 illustrates some of the relationships among the major model types that may not be completely evident from Fig. 10, such as the fact that the QM and noncontextual model instances have a nonempty intersection. The "boundary" between the noncontextual and contextual QM model instances is defined by those that satisfy all CHSH inequalities and those that do not.

Figure 11. Relationships among sets of instances of the main operational model types. The sets of QM and NC model instances have a nonempty intersection, but neither is a subset of the other. The only QM model instance that is also factorizable is the trivial one (Table 19).



The top-down, operational modeling perspective developed in this paper unifies many known facts in an intuitive way, suggests more compact or revealing proofs of some of these facts, highlights some facts which are perhaps underappreciated, and lends itself to the production of unique figures and plots that illustrate them. Some new insights resulting from this approach include

- The 256 outcome deterministic model instances can be classified according to whether or not they satisfy RCI, are factorizable, satisfy all CHSH inequalities, violate at least one CHSH inequality, or even exceed a Tsirelson bound.
- Convex combinations of subsets of the 256 OD model instances can be translated into a traditional hidden variables model (not necessarily a hidden variables theory) for any generic model instance, including all QM ones.
- The convex hull of the 16 RCI+OD model instances is equal to the set of noncontextual model instances, including the factorizable ones. This implies the existence of a local hidden variables model for any noncontextual model instance (including some QM ones).
- The convex hull of all 256 OD model instances equals the set of all generic model instances, including all QM ones. From this construction, a hidden variables model can be derived for any QM model instance, whether contextual or not. Of course for a contextual QM model instance, at least some of the OD model instances in the convex combination cannot satisfy RCI or factorizability.

- It seems that at least some of the plots have not appeared in the literature before. See Fig.'s 1, 12, and 13. One usually sees the polyhedral plots corresponding to the CHSH inequalities (the same as the noncontextual plots). But the QM or factorizable (Bell local) plots (and their various intersections) appear to be new. The QM plots are interesting in that their contours roughly "mimic" the shapes of the solid noncontextual plots, especially on the ends, where $w = \pm 1$. The factorizable plots are also not solids, but contours such as hyperbolic paraboloids (on the ends where $w = \pm 1$), which happen to fit neatly inside the noncontextual solids, visually affirming that factorizability is a special case of noncontextuality.

As for interesting proofs of known facts or particular emphasis on perhaps underappreciated facts, here are some examples.

- The fact that every factorizable model instance and every RCI+OD model instance is noncontextual is easily proved by "matching" the parameters in the symbolic forms of their cpvs, as in Thm. 2 and Corollary 1.
- A compact, geometrically-motivated proof of one of Fine's theorems was given that emphasizes that fact that if a model instance has correlations that satisfy all 8 CHSH inequalities, then *there exists* a NC model instance with those same correlations. The original model instance may or may not be NC. In other words, just because a model instance satisfies all CHSH inequalities does not necessarily mean it is noncontextual (hence not necessarily factorizable or RCI+OD, either). Many examples of model instances that satisfy all CHSH inequalities but are not noncontextual have been given (for example, there are 112 such OD model instances).
- Operational model types that allow Alice to deduce Bob's outcomes or that allow Alice to deduce Bob's measurements have been constructed. It then becomes trivial to show that some QM model instances can be outcome predictable, but none can be measurement predictable. The latter result means that the EPRB experimental setup cannot be used for faster-than-light communication (assuming the world is governed by QM rules, that is). These are well-known results, but not only does the operational modeling perspective make these proofs very intuitive, the method naturally extends to all model types. In fact it was shown that there are no noncontextual measurement predictable model instances either, but there is an infinite number of contextual (but not QM!) model instances that are not only measurement predictable, but also satisfy all of the CHSH inequalities.

6.1. Final thoughts on (non)contextuality

It is important to point out that the use of the name "noncontextual" for the operational model type in Table 5 was intended to be suggestive but is in fact quite arbitrary. It could have

been called "Model Type 3", and none of the subsequent proofs, examples, or correlation plots would change (except for the names). It would still be the case that all of the model instances belonging to the "Model Type 3" class would have correlations that satisfy all of the CHSH inequalities. So why give it the name "noncontextual" in the first place?

As discussed in Sec. 4.3, one way to arrive at the NC model type is to assume definite, preexisting values A_1, A_2, B_1, B_2 , just waiting to be revealed by Alice and Bob's measurements, and equipped with a full joint pmf on all 4 r.v.'s. This leads directly to Table 5 through marginalization. Then it is clear, for example, that if Alice chooses her first measurement $M_A = 1$ she gets $A = A_1$ regardless of Bob's measurement choice. Similarly for Alice's other measurement choice and both of Bob's. In other words, the measurement of each of their observables is independent of context, which is the standard definition of noncontextuality.

Bell himself argued in [7] against such an assumption. His line of reasoning is also described in [9], where it is noted that assumptions like this would later be dubbed "noncontextual". Of course, Bell did show that the assumption of "locality" as embedded in the mathematical form RCI+OD leads to a Bell inequality [4] and later showed that "locality" in the mathematical form of factorization also leads to a Bell inequality [5]. And as has been shown in this paper, these are both special cases of noncontextuality, in the *mathematical* sense. So what gives? Did Bell *implicitly* assume noncontextuality (which he himself rejected as a legitimate starting point), thus effectively assuming what he was trying to prove?

He had legitimate physical reasons to make these assumptions. The *mathematical* assumption of RCI, i.e. the probability of Alice's outcome should depend only on her own (local) measurement (and similarly for Bob), followed from the *physical* assumption of separability. Likewise, the *mathematical* assumption of factorizability was derived from the *physical* assumption of local causality, namely that, given space-like separation, the probability of Alice's outcome should be independent of *both* Bob's measurement choice and outcome, and similarly for Bob with respect to Alice's wing of the experiment. In other words, Bell's physical assumptions were not explicitly "noncontextual". Nevertheless, once these are translated into a *mathematical* model with conditional probabilities as in Tables 7 and 11, it is clear that these are special cases of the noncontextual model type of Table 5 (see Thm. 2 and Corollary 1 of Thm. 3). Whatever one's opinion about what Bell explicitly or implicitly assumed, the fact remains that noncontextuality is at the *mathematical* core of these notions of locality.

Funding This research received no external funding.

Acknowledgments Thanks to the math/physics discussion group at Bellevue Community College for suffering through my first talks on this topic, and in particular to Victor Polinger, who believed that I had something to say. I would like to thank Doug Stoll for reading early

drafts and providing helpful comments, Benjamin Schumacher for comments, suggestions, and encouragement, and Howard Wiseman for being a sponsor to get early versions on the arXiv. Long discussions with Frank Lad and Karl Hess were very valuable in refining my thoughts on noncontextuality vs. locality in Bell's theorems. It should be noted that neither of these two agrees entirely with the views expressed in this paper, but I am grateful for their critiques, because they forced careful rethinking at several junctures, that wouldn't have happened otherwise.

Conflicts of interest The author declares no conflict of interest.

Appendices

A. Correlations, s -functions, and plots

In this section correlations and s -functions tables are displayed for several different model types, including QM, together with plots of correlations (Appendices A.2 and A.3). These plots provide visual affirmation (not necessarily proof in all cases) that

- The intersection of the sets of QM and NC model instances is non-empty, but neither is a subset of the other.
- Every RCI+OD model instances and every Bell local (factorizable) model instance is also NC.
- There are NC model instances that are not factorizable.

The advantage of showing sets of correlations (dimension 4) rather than sets of the original conditional probabilities (dimension 16) is that only one dimension needs to be fixed in order to produce human-understandable 3D plots. The disadvantage is that the "mapping" from generic parameters to correlations is "many-to-one" (see Eq. 3), and one must be careful when interpreting the correlation visualizations. In Appendix A.2 individual examples are shown, and in the following Appendix A.3, selected comparisons.

A.1. Correlations and s -functions

1.1.1. Generic

See Table 24, which shows the correlations and s -functions of the generic model type of Table 1. Since the γ_k 's are nonnegative and sum to 4, the s -functions are obviously between

-4 and +4. There exist generic model instances that not only violate at least one of the CHSH inequalities, but exceed one of the Tsirelson bounds as well.

Table 24. The correlations and s -functions for the generic type.

w	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4$
x	$\gamma_5 - \gamma_6 - \gamma_7 + \gamma_8$
y	$\gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12}$
z	$\gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_1	$-\gamma_1 + \gamma_2 + \gamma_3 - \gamma_4 + \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 + \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} + \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_2	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 - \gamma_5 + \gamma_6 + \gamma_7 - \gamma_8 + \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} + \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_3	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 + \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 - \gamma_9 + \gamma_{10} + \gamma_{11} - \gamma_{12} + \gamma_{13} - \gamma_{14} - \gamma_{15} + \gamma_{16}$
s_4	$\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4 + \gamma_5 - \gamma_6 - \gamma_7 + \gamma_8 + \gamma_9 - \gamma_{10} - \gamma_{11} + \gamma_{12} - \gamma_{13} + \gamma_{14} + \gamma_{15} - \gamma_{16}$
Min and Max	-4 and 4

1.1.2. QM

See Table 25, which shows the correlations and s -functions of the QM type of Table 3. Not all QM model instances satisfy all the CHSH inequalities. The parameterization of a QM model that maximally violates one CHSH inequality is given by

$$\theta_1 = 67.5^\circ \text{ and } \theta_k = 22.5^\circ \text{ for } k = 2, 3, 4. \tag{43}$$

This QM model instance has correlations

$$(w, x, y, z) = \frac{1}{2}\sqrt{2}(-1, 1, 1, 1) \tag{44}$$

and s -functions

$$(s_1, s_2, s_3, s_4) = (2\sqrt{2}, 0, 0, 0). \tag{45}$$

Obviously the first one (s_1) does not satisfy one of the CHSH inequalities.

Table 25. The correlations and s -functions for the QM type.

w	$\cos 2\theta_1$
x	$\cos 2\theta_2$
y	$\cos 2\theta_3$
z	$\cos 2\theta_4$
s_1	$-\cos 2\theta_1 + \cos 2\theta_2 + \cos 2\theta_3 + \cos 2\theta_4$
s_2	$\cos 2\theta_1 - \cos 2\theta_2 + \cos 2\theta_3 + \cos 2\theta_4$
s_3	$\cos 2\theta_1 + \cos 2\theta_2 - \cos 2\theta_3 + \cos 2\theta_4$
s_4	$\cos 2\theta_1 + \cos 2\theta_2 + \cos 2\theta_3 - \cos 2\theta_4$
Min and Max	$-2\sqrt{2}$ and $2\sqrt{2}$

1.1.3. Noncontextual

See Table 26, which shows the correlations and s -functions of the NC model type of Table 5. Since the ρ_k 's form a pmf, it is obvious that all of the s -functions must be between -2 and 2. Note how the parameterized modeling framework makes derivation of the CHSH inequalities trivial.

Table 26. The correlations and s -functions for the NC type.

w	$\rho_1 + \rho_2 - \rho_3 - \rho_4 + \rho_5 + \rho_6 - \rho_7 - \rho_8 - \rho_9 - \rho_{10} + \rho_{11} + \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16}$
x	$\rho_1 + \rho_2 - \rho_3 - \rho_4 - \rho_5 - \rho_6 + \rho_7 + \rho_8 + \rho_9 + \rho_{10} - \rho_{11} - \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16}$
y	$\rho_1 - \rho_2 + \rho_3 - \rho_4 + \rho_5 - \rho_6 + \rho_7 - \rho_8 - \rho_9 + \rho_{10} - \rho_{11} + \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16}$
z	$\rho_1 - \rho_2 + \rho_3 - \rho_4 - \rho_5 + \rho_6 - \rho_7 + \rho_8 + \rho_9 - \rho_{10} + \rho_{11} - \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16}$
s_1	$2(\rho_1 - \rho_2 + \rho_3 - \rho_4 - \rho_5 - \rho_6 + \rho_7 + \rho_8 + \rho_9 + \rho_{10} - \rho_{11} - \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16})$
s_2	$2(\rho_1 - \rho_2 + \rho_3 - \rho_4 + \rho_5 + \rho_6 - \rho_7 - \rho_8 - \rho_9 - \rho_{10} + \rho_{11} + \rho_{12} - \rho_{13} + \rho_{14} - \rho_{15} + \rho_{16})$
s_3	$2(\rho_1 + \rho_2 - \rho_3 - \rho_4 - \rho_5 + \rho_6 - \rho_7 + \rho_8 + \rho_9 - \rho_{10} + \rho_{11} - \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16})$
s_4	$2(\rho_1 + \rho_2 - \rho_3 - \rho_4 + \rho_5 - \rho_6 + \rho_7 - \rho_8 - \rho_9 + \rho_{10} - \rho_{11} + \rho_{12} - \rho_{13} - \rho_{14} + \rho_{15} + \rho_{16})$
Min and Max	-2 and 2

1.1.4. Factorizable (Bell local)

See Table 27, which shows the correlations and s -functions of the factorizable model type of Table 7. As a special case of the NC model type (Thm. 2 in Sec. 2), all factorizable (Bell local) model instances also satisfy all CHSH inequalities.

Table 27. The correlations and s -functions for the factorizable (Bell local) type.

w	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1)$
x	$(\alpha_3 - \alpha_1)(\beta_4 - \beta_2)$
y	$(\alpha_4 - \alpha_2)(\beta_3 - \beta_1)$
z	$(\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_1	$-(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) + (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) + (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) + (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_2	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) - (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) + (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) + (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_3	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) + (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) - (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) + (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
s_4	$(\alpha_3 - \alpha_1)(\beta_3 - \beta_1) + (\alpha_3 - \alpha_1)(\beta_4 - \beta_2) + (\alpha_4 - \alpha_2)(\beta_3 - \beta_1) - (\alpha_4 - \alpha_2)(\beta_4 - \beta_2)$
Min and Max	-2 and 2

Note that since the single conditional expectations are given by

$$\begin{aligned}
 E[A|M_A = 1] &= \alpha_3 - \alpha_1 & E[B|M_B = 1] &= \beta_3 - \beta_1, \\
 E[A|M_A = 2] &= \alpha_4 - \alpha_2 & E[B|M_B = 2] &= \beta_4 - \beta_2,
 \end{aligned}
 \tag{46}$$

each correlation (conditional expectation of the product) is the product of the corresponding single conditional expectations.

$$\begin{aligned}
 w &= E[AB|M_A = 1, M_B = 1] = E[A|M_A = 1] \times E[B|M_B = 1], \\
 x &= E[AB|M_A = 1, M_B = 2] = E[A|M_A = 1] \times E[B|M_B = 2], \\
 y &= E[AB|M_A = 2, M_B = 1] = E[A|M_A = 2] \times E[B|M_B = 1], \\
 z &= E[AB|M_A = 2, M_B = 2] = E[A|M_A = 2] \times E[B|M_B = 2].
 \end{aligned}
 \tag{47}$$

From either Eq. 47 or Table 27 it is easy to see that $wz = xy$ and hence, for any fixed w , the following equation determines the shape of the sets of factorizable correlations in 3-space. See Fig. 1 or 12.

$$\begin{cases} z = w^{-1}xy & \text{if } w \neq 0 \\ xy = 0 & \text{if } w = 0. \end{cases}
 \tag{48}$$

However, the *unconditional* correlation is equal to the product of the individual unconditional expectations, that is,

$$E[AB] = E[A] \times E[B]
 \tag{49}$$

only if all joint parameter setting probabilities are equal, that is,

$$\begin{aligned}
 P(M_A = 1, M_B = 1) &= P(M_A = 1, M_B = 2) = \\
 P(M_A = 2, M_B = 1) &= P(M_A = 2, M_B = 2) = \frac{1}{4}.
 \end{aligned}
 \tag{50}$$

1.1.5. Remote context independence plus outcome determinism

See Table 28, which shows the correlations and s -functions of the 16 OD model instances from Table 11. It is clear that they satisfy all 8 CHSH inequalities. It was shown in Sec. 1 that these model instances also satisfy RCI. Note the CHSH inequality bounds are tight in this case.

Table 28. The correlations and s -functions for the 16 RCI+OD model instances. See Table 11 for the underlying cpvs.

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
w	1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	1
x	1	1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1
y	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1
z	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1
s_1	2	-2	2	-2	-2	-2	2	2	2	2	-2	-2	-2	2	-2	2
s_2	2	-2	2	-2	2	2	-2	-2	-2	-2	2	2	-2	2	-2	2
s_3	2	2	-2	-2	-2	2	-2	2	2	-2	2	-2	-2	-2	2	2
s_4	2	2	-2	-2	2	-2	2	-2	-2	2	-2	2	-2	-2	2	2
Min and Max	-2 and 2															

1.1.6. Trivial

See Table 29. It is obvious that all CHSH inequalities are satisfied. This is the completely uniform, no-information case.

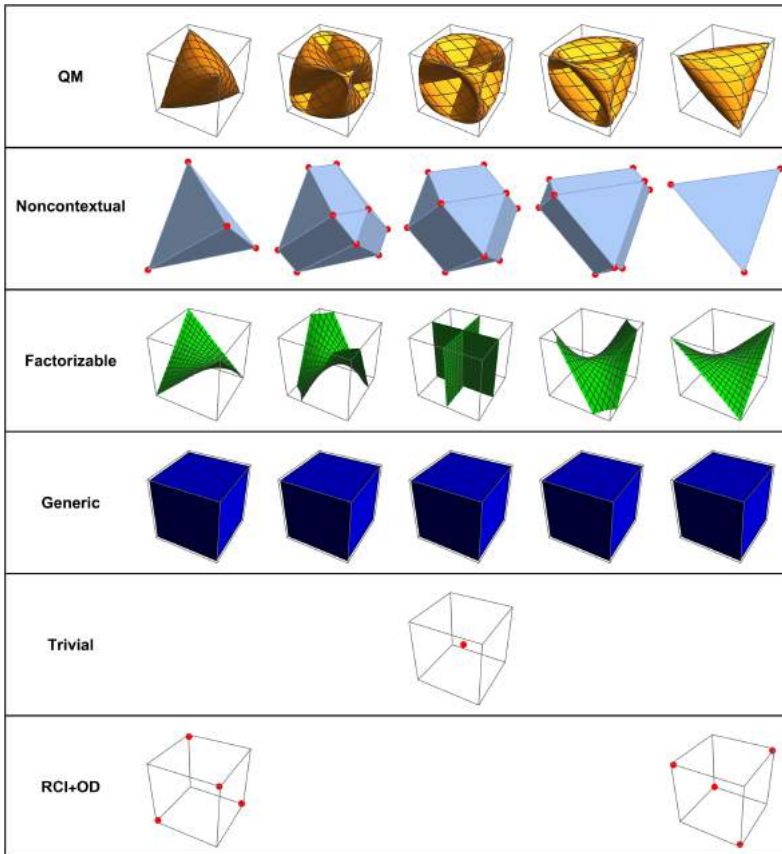
Table 29. The correlations and s -functions for the trivial type.

w	0
x	0
y	0
z	0
s_1	0
s_2	0
s_3	0
s_4	0
Min and Max	0 and 0

A.2. Correlation plots

Fig. 12 shows different sets of correlations corresponding to six different model types. Each of the five (or fewer) pictures in any row is a 3D-slice (x, y, z) through the set of possible correlations (w, x, y, z) for the given model type, where w is fixed in each case. In these figures, $w = -1, -\frac{1}{2}, 0, \frac{1}{2}, 1$ from left to right across the 5 columns.

Figure 12. Sets of correlations (x, y, z) of some model types. w is fixed at either $-1, -\frac{1}{2}, 0, \frac{1}{2},$ or 1 from left to right in each row.



QM (row 1): Each contour represents the set of points

$$\begin{aligned}
 (x, y, z) &= (\cos 2\theta_2, \cos 2\theta_3, \cos 2\theta_4) \text{ where } \theta_1 = \theta_2 + \theta_3 + \theta_4, \\
 \theta_1 &= \frac{1}{2} \arccos w, \text{ and } -90^\circ \leq \theta_2, \theta_3, \theta_4 \leq 90^\circ.
 \end{aligned}
 \tag{51}$$

Noncontextuality (row 2): Each of the five figures is a representation of all possible correlations (x, y, z) for a NC model type with w fixed. The figures at the far left ($w = -1$) and the far right ($w = 1$) are tetrahedrons with 4 vertices and 4 equilateral triangular faces.

The polyhedra in between have 12 vertices and 6 rectangular faces and 8 triangular faces. The one in the middle ($w = 0$) is called a "cuboctahedron", a semi-regular polyhedron where the rectangles are congruent squares and the triangles are congruent equilateral triangles. The points on the boundary and in the interior of each of these polyhedra represent those correlations (w, x, y, z) that satisfy the CHSH inequalities for a given fixed value of w . These pictures can be viewed as more detailed representations of the "correlation polytope" in Fig. 3 of Pitowski [21]. For more detailed connections between the Bell inequalities and the Platonic and Archimedean solids, see [22,23].

Factorizable (row 3): At the far left ($w = -1$) and far right ($w = 1$), the surfaces are hyperbolic paraboloids. All 5 surfaces can be represented by the equation

$$\begin{cases} z = w^{-1}xy & \text{if } w \neq 0 \\ xy = 0 & \text{if } w = 0. \end{cases} \quad (52)$$

Generic (row 4): These fill the entire "correlation cube" $([-1, 1] \times [-1, 1] \times [-1, 1])$. This is because each of the four correlations is determined by a set of four parameters completely independent of the others, hence they all can independently take any value between -1 and 1.

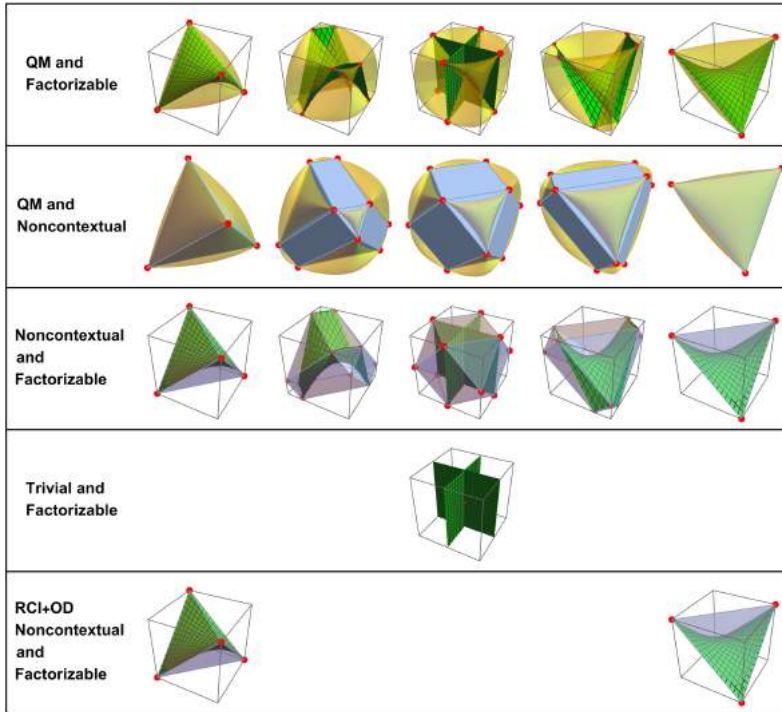
Trivial (row 5): This can only appear in the middle ($w = 0$) because all of its correlations are 0.

RCI+OD (row 6): These can only appear on the far left ($w = -1$) or far right ($w = 1$) because any OD model instance has perfect correlations (± 1) .

A.3. Correlation plots compared

Now let's do some comparisons. See Fig. 13. In the first and second rows, QM correlations are compared to factorizable and NC ones. In both cases, there is a non-empty intersection, but neither is a subset of the other. In the end cases ($w = \pm 1$), the surfaces (or region in the NC case) stretch out to the RCI+OD vertices shown as red dots.

Figure 13. Sets of correlations (x, y, z) of some pairs of model instances. Fixed $w = -1, -\frac{1}{2}, 0, \frac{1}{2},$ or 1 from left to right in each row.



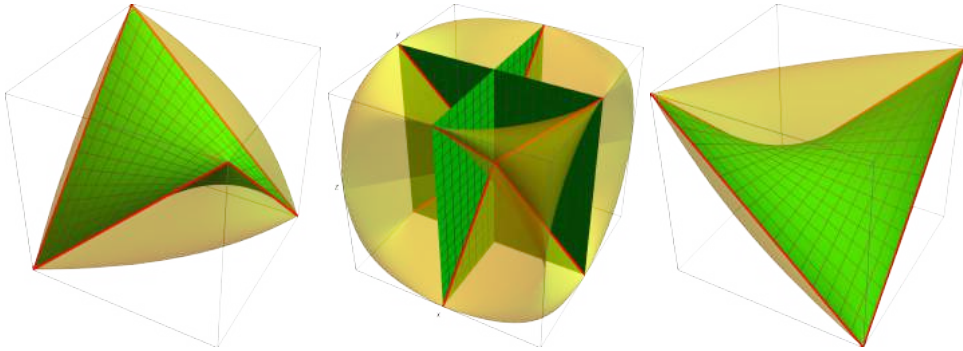
In the third row (noncontextuality and factorization) it is clear that in every case, the Bell local (factorizable) correlations are completely contained in the corresponding NC correlations. This is a direct consequence of the fact that every factorizable model instance is a special case of the NC model type (Thm. 2 in Sec. 1).

The first and last plots ($w = \pm 1$) in rows 3 and 5 show that the perfect correlations (red dots) of the OD model types can be achieved by some NC and Bell local (factorizable) model instances. This reflects the fact that every RCI+OD model instance is also factorizable and NC.

Fig. 14 shows the first, middle and last plot of row 1 in Fig. 13 "blown up". The red lines have been added to highlight the intersections of the QM and factorizable correlation plots. How does this square with the fact that the *only* model instance that is simultaneously QM and factorizable is the trivial one (corresponding to the point at the very center of the middle plot, at the intersection of the diagonals)? This is a dramatic demonstration of the fact that the mapping from model instance parameters to correlations is many-to-one. That is, every point on the each of the red lines (i.e., a set of correlations) is achievable by *some* QM model instance, and it is achievable by *some* factorizable model instance, but in all but one case, the conditional probabilities (generic parameters) of these two model instances are *different*.

B. QM model s-function max and min

Figure 14. Red lines indicate correlations (x, y, z) mutually achievable by different QM and factorizable model instances. Fixed $w = -1, 0, 1$ from left to right.



Define a function $\bar{s}_1 : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \rightarrow R$ where

$$\bar{s}_1(\theta_2, \theta_3, \theta_4) = -\cos 2(\theta_2 + \theta_3 + \theta_4) + \cos 2\theta_2 + \cos 2\theta_3 + \cos 2\theta_4.$$

\bar{s}_1 is related to s_1 by $\bar{s}_1(\theta_2, \theta_3, \theta_4) = s_1(\cos 2\theta_1, \cos 2\theta_2, \cos 2\theta_3, \cos 2\theta_4)$ where $\theta_1 = \theta_2 + \theta_3 + \theta_4$.

Tables 30, 31, and 32 show the results of using Mathematica to search for stationary points. Of the 20 stationary points, two are local maxima, two are local minima, and the other sixteen are neither.

Table 30. Critical point angles.

Pt	θ_1	θ_2	θ_3	θ_4
1	$\pm 67.5^\circ$	$\pm 22.5^\circ$	$\pm 22.5^\circ$	$\pm 22.5^\circ$
2	$\pm 202.5^\circ$	$\pm 67.5^\circ$	$\pm 67.5^\circ$	$\pm 67.5^\circ$
3	0°	0°	0°	0°
4	$\pm 90^\circ$	0°	0°	$\pm 90^\circ$
5	$\pm 90^\circ$	0°	$\pm 90^\circ$	0°
6	$\pm 90^\circ$	$\pm 90^\circ$	0°	0°
7	$\pm 180^\circ$	0°	$\pm 90^\circ$	$\pm 90^\circ$
8	$\pm 180^\circ$	$\pm 90^\circ$	0°	$\pm 90^\circ$
9	$\pm 180^\circ$	$\pm 90^\circ$	$\pm 90^\circ$	$\pm 0^\circ$
10	$\pm 270^\circ$	$\pm 90^\circ$	$\pm 90^\circ$	$\pm 90^\circ$

Table 31. Critical point correlations.

Pt	w	x	y	z
1	$-\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$	$\frac{1}{2}\sqrt{2}$
2	$\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}\sqrt{2}$	$-\frac{1}{2}\sqrt{2}$
3	1	1	1	1
4	-1	1	1	-1
5	-1	1	-1	1
6	-1	-1	1	1
7	1	1	-1	-1
8	1	-1	1	-1
9	1	-1	-1	1
10	-1	-1	-1	-1

Table 32. Critical point value, gradient, and Hessian.

Pt	Value	Gradient	Hessian
1	$2\sqrt{2}$	(0, 0, 0)	Negative Definite
2	$-2\sqrt{2}$	(0, 0, 0)	Positive Definite
3	2	(0, 0, 0)	Indefinite
4	2	(0, 0, 0)	Indefinite
5	2	(0, 0, 0)	Indefinite
6	2	(0, 0, 0)	Indefinite
7	-2	(0, 0, 0)	Indefinite
8	-2	(0, 0, 0)	Indefinite
9	-2	(0, 0, 0)	Indefinite
10	-2	(0, 0, 0)	Indefinite

Figure 15. Maxima and minima for the QM \bar{s}_1 function.

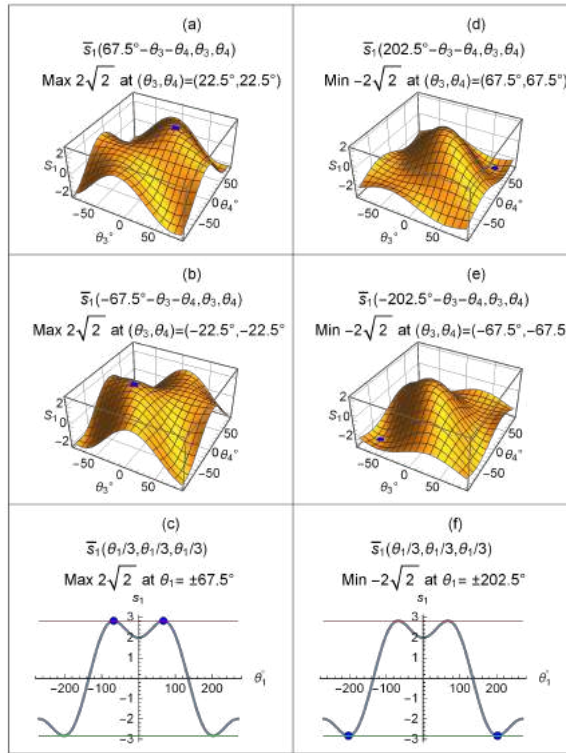
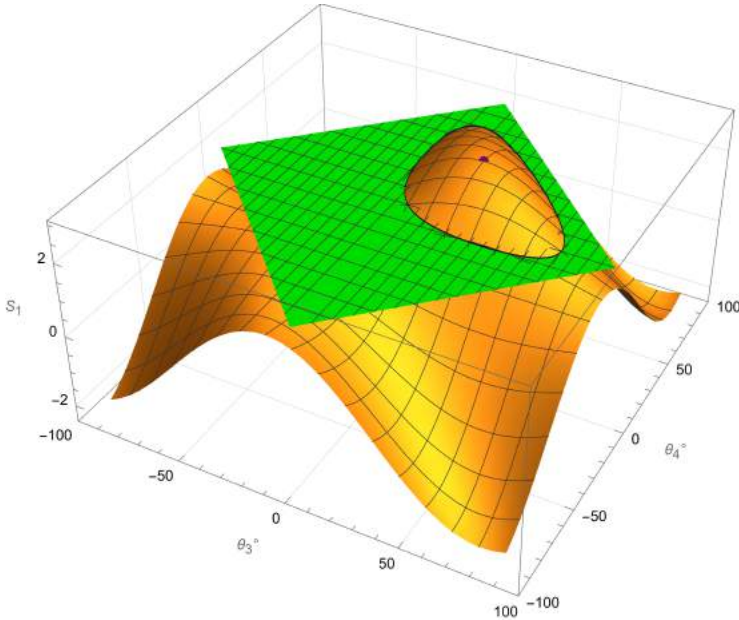


Fig. 15 illustrates the local maxima and minima indicated by rows 1 and 2 of Tables 30, 31, and 32. The first column (panels (a), (b), and (c)) indicate the maxima and the second column (panels (d), (e), and (f)) indicate the minima (both shown as blue dots). In panels (a) and (b), θ_1 is held fixed at $\pm 67.5^\circ$, respectively, and θ_3 and θ_4 are free parameters. In (c), only θ_1 is free, and the maxima are seen to coincide at $(\theta_1, \theta_2, \theta_3, \theta_4) = (\pm 67.5^\circ, \pm 22.5^\circ, \pm 22.5^\circ, \pm 22.5^\circ)$. In panels (d) and (e), θ_1 is held fixed at $\pm 202.5^\circ$, respectively, and θ_3 and θ_4 are free parameters. In (f), only θ_1 is free, and the minima are seen to coincide at $(\theta_1, \theta_2, \theta_3, \theta_4) = (\pm 202.5^\circ, \pm 67.5^\circ, \pm 67.5^\circ, \pm 67.5^\circ)$.

B.1. Bell's theorem in a nutshell

Fig. 16 is an illustration of the relationship between the first QM s -function, as a function of measurement difference angles (θ_3, θ_4) and one of the CHSH bounds. The green square represents the CHSH inequality bound of 2, and the yellow surface represents the QM s_1 -function, as a function of the two free parameters θ_3 and θ_4 , based on the two constraints $\theta_1 = \theta_2 + \theta_3 + \theta_4$ and $\theta_2 = 22.5^\circ$. The small hill above the green square indicates where the CHSH inequality is violated. The purple dot at the top of the hill shows the maximum of $2\sqrt{2}$ when $\theta_2 = \theta_3 = \theta_4 = 22.5^\circ$ and $\theta_1 = 67.5^\circ$.

Figure 16. Illustration of the violation of one of the CHSH inequalities by the s_1 -function of QM correlations ($\theta_2 = 22.5^\circ$ is held fixed).



B.2. The Tsirelson bounds

The QM s -function max and min of $\pm 2\sqrt{2}$ are known as the *Tsirelson bound(s)*. If all four of the parameters $\theta_1, \theta_2, \theta_3$, and θ_4 were free, then the max and min of the s -functions would be $+4$ and -4 , just like the generic model – see Table 1. It is the simple geometric relationship among the four measurement difference angles that imposes a linear constraint among these angles (see Fig. 2). Note that these would not be *completely* generic models, however, since the form of the QM conditional probabilities imposes 16 extra constraints on the generic parameters $\gamma_1, \gamma_2, \dots, \gamma_{16}$, such as $\gamma_1 + \gamma_2 = \frac{1}{2} \sin^2(\theta_1) + \frac{1}{2} \cos^2(\theta_1) = \frac{1}{2}$.

C. Convex hull of OD model instances equals the set of all generic model instances

The proof of Thm. 3 shows that the convex hull of the 16 RCI+OD model instances equals the class of noncontextual model instances. This motivates the question: What is the convex hull of all 256 OD conditional probability vectors? The next theorem answers this question.

Theorem 14. *Let G be the set of generic conditional probability vectors and K the set of 256 OD conditional probability vectors. Then any $\gamma \in G$ can be written as a convex combination of the members of K . See Fig. 8.*

Proof. A theorem of Minkowski states that any convex, closed, and bounded set in \mathbb{R}^n is equal to the convex hull of its extreme points. Therefore the plan is to show that G is convex, closed, and bounded, and that the members of K comprise all of the extreme points of G .

- G is convex because if $\gamma = \pi\gamma_1 + (1 - \pi)\gamma_2$ for any $\gamma_1, \gamma_2 \in G$ and $0 \leq \pi \leq 1$, it is clear that all of the entries of γ are nonnegative and satisfy the measurement context constraints, hence $\gamma \in G$.
- G is closed since for any sequence $\gamma_n = (\gamma_1^{(n)}, \gamma_2^{(n)}, \dots, \gamma_{16}^{(n)}) \in G$ converging to $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16})$,

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = \lim_{n \rightarrow \infty} (\gamma_1^{(n)} + \gamma_2^{(n)} + \gamma_3^{(n)} + \gamma_4^{(n)}) = \lim_{n \rightarrow \infty} 1 = 1,$$

since each $\gamma_n \in G$. Similarly for the other successive blocks of four entries in γ . Hence $\gamma \in G$ and thus G is closed.

- G is bounded because if $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16}) \in G$, then $\|\gamma\| = \sqrt{\sum \gamma_k^2} \leq \sqrt{\sum \gamma_k} = \sqrt{4} = 2$, so G is contained in a sphere of radius 2 centered at the origin in \mathbb{R}^{16} .
- Now it is shown that K is the set of all extreme points of G . It is obvious that $K \subseteq G$ since every OD model instance has nonnegative entries that satisfy the measurement context constraints. Next it is demonstrated that every member of K is an extreme point of G . To simplify notation, the proof is restricted to showing a particular member of K is an extreme point of G . It should be clear that the argument is not affected by the exact positions of the four 1's in any of the other 255 OD conditional probability vectors. Thus, consider this representative member of K :

$$o = (1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0).$$

Suppose $o = \frac{1}{2}(\gamma + \gamma')$ where $\gamma, \gamma' \in G$. Since the entries of γ, γ' are nonnegative, it is pretty clear that all entries in each, except for positions 1, 5, 9, 13, must be 0, and that leaves only the choice of 1 in those positions for both γ, γ' . In other words, o cannot be the midpoint of a non-trivial line in G , hence it must be an extreme point.

Now it is proved that any $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_{16}) \in G \setminus K$ cannot be an extreme point of G . In one of the successive blocks of four entries, there must be one with at least two entries not equal to either 0 or 1. Wlog, assume that $0 < \gamma_1 \leq \gamma_2 < 1$. Choose a real x such that $0 < x < \gamma_1$, and define

$$\begin{aligned} \gamma_1 &= (2\gamma_1 - x, \gamma_2 - \gamma_1 + x, \gamma_3, \dots, \gamma_{16}) \text{ and} \\ \gamma_2 &= (x, \gamma_1 + \gamma_2 - x, \gamma_3, \dots, \gamma_{16}). \end{aligned}$$

Then it is not too hard to show that $\gamma_1, \gamma_2 \in G \setminus K$ are distinct and therefore, since G is convex,

$$\pi\gamma_1 + (1 - \pi)\gamma_2$$

for $0 \leq \pi \leq 1$ is a non-trivial line segment completely contained in G . In other words,

$$\gamma = \frac{1}{2}(\gamma_1 + \gamma_2)$$

is the midpoint of a non-trivial line segment entirely inside G , so it cannot be an extreme point.

Finally the since set of all extreme points of G is exactly K , the set G of generic model instances is equal to the convex hull of K , the 256 OD model instances. \square

D. Filling in the gaps

This appendix has been reserved for proofs and counterexamples related to some of the details of Fig. 10 (which shows the relationships among the different model types and conditions) that haven't been proved or illustrated elsewhere. This was done in order not to interrupt the flow in the main body of the paper.

The first order of business is to address the issue of possibly undefined conditional probabilities, which affects both the underlying definitions of model types in general and the LC model type in particular.

D.1. Undefined conditional probabilities

In the definition of conditional probabilities, for example,

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

for events E and F , it is assumed that $P(F) > 0$, otherwise the conditional probability is undefined. In this paper, this tradition is respected, which requires mention of two situations where the issue of zero denominators might arise.

4.1.1. Measurement choice probabilities

Consider Table 33, which shows the *unconditional* probabilities

$$P(A = s, B = t, M_A = u, M_B = v) = P(A = s, B = t | M_A = u, M_B = v) \delta_{uv}$$

for the generic model type, where the joint measurement choice probabilities are given by

$$\delta_{uv} = P(M_A = u, M_B = v) \text{ for } u, v = 1, 2. \tag{53}$$

Table 33. PMF for the generic model type. These are the unconditional probabilities based on measurement choice probabilities (assumed frequencies).

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	$\gamma_1\delta_{11}$	$\gamma_2\delta_{11}$	$\gamma_3\delta_{11}$	$\gamma_4\delta_{11}$
$(1, 2)$	$\gamma_5\delta_{12}$	$\gamma_6\delta_{12}$	$\gamma_7\delta_{12}$	$\gamma_8\delta_{12}$
$(2, 1)$	$\gamma_9\delta_{21}$	$\gamma_{10}\delta_{21}$	$\gamma_{11}\delta_{21}$	$\gamma_{12}\delta_{21}$
$(2, 2)$	$\gamma_{13}\delta_{22}$	$\gamma_{14}\delta_{22}$	$\gamma_{15}\delta_{22}$	$\gamma_{16}\delta_{22}$

If the joint measurement choice probabilities satisfy $0 < \delta_{11}, \delta_{12}, \delta_{21}, \delta_{22} < 1$ strictly, then the parameters $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_{16})$ can be identified with the conditional probabilities $P(A, B|M_A, M_B)$, e.g.,

$$P(A = -1, B = -1|M_A = 1, M_B = 1) = \frac{\gamma_1\delta_{11}}{(\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4)\delta_{11}} = \gamma_1,$$

since $\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 1$ by the measurement context constraint. In this example, if $\delta_{11} = 0$, then the conditional probability on the LHS is undefined, whereas γ_1 presumably is defined. For this reason, it will be assumed that the measurement choice probabilities are all strictly between 0 and 1 unless explicitly stated otherwise. The implication for any actual experiment is that every one of the 4 possible joint measurement combinations by Alice and Bob must occur at least once.

4.1.2. Local causality conditionals

Recall the definitions of various forms of locality (RCI, factorizability, ROI, LC) in Sec. 1. Note that the first three involve conditioning on M_A and/or M_B only, whereas LC involves conditioning on either (A, M_A, M_B) or (B, M_A, M_B) . For example, in terms of the generic parameters, conditional probabilities such as the following can be computed.

$$P(A = -1|B = -1, M_A = 1, M_B = 1) = \frac{\gamma_1}{\gamma_1 + \gamma_3} \tag{54}$$

$$P(B = -1|A = -1, M_A = 1, M_B = 1) = \frac{\gamma_1}{\gamma_1 + \gamma_2}.$$

There are 32 conditional probabilities like this that will be undefined unless the corresponding denominators are nonzero. It turns out that a *necessary* condition for being well-defined is that at least 12 γ_k 's must be nonzero. This is the reason that "Local causality" (box 3) is shown below "Factorizable (Bell local)" (box 5) in Fig. 10. That is, the LC model instances form a subset of factorizable model instances that just happen to have 12 or more non-zero generic parameters.

D.2. Examples and counterexamples

In this section, examples and counterexamples are constructed that justify some of the relationships among model types and properties shown in Fig. 10.

4.2.1. QM and NC but not factorizable

Table 34 shows parameterizations for a model instance that is both NC and QM, but it is not Bell local (factorizable). It shows that a model instance does not have to be factorizable to satisfy all of the CHSH inequalities.

Table 34. Example of a model instance that is QM and NC but not factorizable (Bell local).

Model type	Parameter assignment
Generic	$\gamma_k = \frac{1}{4}$ for $k = 1, 2, 3, 4, 13, 14, 15, 16,$ $\gamma_k = \frac{1}{2}$ for $k = 5, 8, 10, 11,$ 0 otherwise
QM	$\theta_1 = 135^\circ, \theta_2 = 0^\circ, \theta_3 = 90^\circ, \theta_4 = 45^\circ$
NC	$\rho_k = \frac{1}{4}$ for $k = 2, 8, 9, 15$ and $\rho_k = 0$ otherwise .

4.2.2. NC but neither QM nor factorizable

Table 35 shows a NC model instance that does not belong to either the set of QM model instances or Bell local (factorizable) model instances.

Table 35. Example of a model instance that is NC but not QM or factorizable (Bell local).

Model type	Parameter assignment
Generic	$\gamma_3 = 1,$ $\gamma_k = \frac{1}{3}$ for $k = 8, 9, 13, 15, 16,$ $\gamma_k = \frac{2}{3}$ for $k = 7, 11,$ 0 otherwise
NC	$\rho_k = \frac{1}{3}$ for $k = 3, 4, 8$ and $\rho_k = 0$ otherwise

4.2.3. QM but not NC

Table 36 shows a QM model instance that is not NC, hence also not factorizable.

Table 36. Example of a model instance that is QM but not NC.

Model type	Parameter assignment
Generic	$\gamma_k = \frac{1}{2} \cos^2 67.5^\circ$ for $k = 1, 4, 6, 7, 10, 11, 14, 15,$ and $\gamma_k = \frac{1}{2} \sin^2 67.5^\circ$ for $k = 2, 3, 5, 8, 9, 12, 13, 16.$
QM	$\theta_1 = 67.5^\circ$ and $\theta_k = 22.5^\circ$ for $k = 2, 3, 4.$

4.2.4. The trivial model instance

Table 37 shows that there are parameterizations for the generic, QM, NC, and factorizable model types that produce the trivial model instance.

Table 37. The only model instance achievable by all five main operational model types is the trivial one.

Model type	Parameter assignment
Generic	$\gamma_k = \frac{1}{4}$ for all k
QM	$\theta_1 = 135^\circ$ and $\theta_k = 45^\circ$ for $k = 2, 3, 4$
NC	$\rho_k = \frac{1}{16}$ for $k = 1, 2, \dots, 16$
Factorizable	$\alpha_k = \beta_k = \frac{1}{2}$ for $k = 1, 2, 3, 4$
Trivial	All entries = $\frac{1}{4}$

D.3. All QM model instances satisfy RCI but only one satisfies ROI

Theorem 15. Every QM model instance satisfies RCI but only one satisfies ROI (the trivial model instance).

Proof. See Table 38 for the proof that QM model instances satisfy RCI. Since all of the conditional probabilities are equal to $\frac{1}{2}$, it is obvious that Alice’s outcome probabilities do not depend on Bob’s measurement, and similarly Bob’s outcome probabilities do not depend on Alice’s measurement.

Table 38. The QM model type satisfies RCI. Both Alice’s conditionals (next-to-last column) and Bob’s conditionals (last column) are obviously independent of the other’s measurement choice.

s	t	u	v	$P(A = s M_A = u, M_B = v)$	$P(B = t M_A = u, M_B = v)$
-1	-1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	-1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
1	1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
-1	-1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
1	1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	-1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
1	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$

As for satisfying ROI, consult the definition of ROI (Def. 16) and then refer to Tables 3 and 38. These imply that

$$\frac{1}{2} \cos^2 \theta_k = \frac{1}{2} \sin^2 \theta_k = \frac{1}{4}$$

for all $k = 1, 2, 3, 4$. This works only if the measurement difference angles are all of the form

$$\theta_k = (2j_k + 1) \frac{\pi}{4}$$

for integers j_k . This is the definition of the trivial model type (Table 19). The only QM model instance that satisfies ROI is the trivial one. The QM model type is highly "nonlocal" in this sense. □

D.4. All NC model instances satisfy RCI but only the factorizable ones satisfy ROI

Theorem 16. *Every NC model instance satisfies RCI, but not all satisfy ROI.*

Proof. See Tables 39 and 40. Note the probabilities in each pair of rows are equal whenever (s, u) is the same (for Alice), or whenever (t, v) is the same (for Bob), no matter the value of the other’s measurement. That is, the measurement in either wing of the experiment does not affect the outcome probability in the other.

Table 39. The NC model type satisfies RCI. Alice’s conditionals are shown. The probabilities in each pair of rows in which (s, u) is the same are equal.

s	u	v	$P(A = s M_A = u, M_B = v)$
-1	1	1	$\rho_1 + \rho_2 + \rho_5 + \rho_6 + \rho_9 + \rho_{10} + \rho_{13} + \rho_{14}$
-1	1	2	$\rho_1 + \rho_2 + \rho_5 + \rho_6 + \rho_9 + \rho_{10} + \rho_{13} + \rho_{14}$
-1	2	1	$\rho_1 + \rho_3 + \rho_5 + \rho_7 + \rho_9 + \rho_{11} + \rho_{13} + \rho_{15}$
-1	2	2	$\rho_1 + \rho_3 + \rho_5 + \rho_7 + \rho_9 + \rho_{11} + \rho_{13} + \rho_{15}$
1	1	1	$\rho_3 + \rho_4 + \rho_7 + \rho_8 + \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16}$
1	1	2	$\rho_3 + \rho_4 + \rho_7 + \rho_8 + \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16}$
1	2	1	$\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16}$
1	2	2	$\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16}$

Table 40. The NC model type satisfies RCI. Bob’s conditionals are shown. The probabilities in each pair of rows in which (t, v) is the same are equal.

t	u	v	$P(B = t M_A = u, M_B = v)$
-1	1	1	$\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5 + \rho_6 + \rho_7 + \rho_8$
-1	2	1	$\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5 + \rho_6 + \rho_7 + \rho_8$
-1	1	2	$\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_9 + \rho_{10} + \rho_{11} + \rho_{12}$
-1	2	2	$\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_9 + \rho_{10} + \rho_{11} + \rho_{12}$
1	1	1	$\rho_9 + \rho_{10} + \rho_{11} + \rho_{12} + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}$
1	2	1	$\rho_9 + \rho_{10} + \rho_{11} + \rho_{12} + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}$
1	1	2	$\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}$
1	2	2	$\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}$

To see that a NC model instance does not satisfy ROI in general, see Table 41. These represent the equations that define ROI for the NC model type (see Def. 16 and Tables 5, 39, and 40). There are 16 nonlinear equations that must be solved for $\rho = (\rho_1, \rho_2, \dots, \rho_{16})$, where of course ρ must be a pmf.

It turns out that these are exactly the equations which must be solved in order for a NC model instance to be factorizable. This should not be a surprise. Since every NC model instance is already RCI, finding a solution to these equations would make it ROI as well, hence factorizable (Thm. 17).

Table 41. A NC model instance satisfies ROI only if these 16 equations are satisfied.

$(\rho_1 + \rho_2 + \rho_5 + \rho_6 + \rho_9 + \rho_{10} + \rho_{13} + \rho_{14})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5 + \rho_6 + \rho_7 + \rho_8) =$	$\rho_1 + \rho_2 + \rho_5 + \rho_6$
$(\rho_1 + \rho_2 + \rho_5 + \rho_6 + \rho_9 + \rho_{10} + \rho_{13} + \rho_{14})$ $\times(\rho_9 + \rho_{10} + \rho_{11} + \rho_{12} + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_9 + \rho_{10} + \rho_{13} + \rho_{14}$
$(\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16})$ $\times(\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_6 + \rho_8 + \rho_{14} + \rho_{16}$
$(\rho_3 + \rho_4 + \rho_7 + \rho_8 + \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5 + \rho_6 + \rho_7 + \rho_8) =$	$\rho_3 + \rho_4 + \rho_7 + \rho_8$
$(\rho_3 + \rho_4 + \rho_7 + \rho_8 + \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16})$ $\times(\rho_9 + \rho_{10} + \rho_{11} + \rho_{12} + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_{11} + \rho_{12} + \rho_{15} + \rho_{16}$
$(\rho_1 + \rho_2 + \rho_5 + \rho_6 + \rho_9 + \rho_{10} + \rho_{13} + \rho_{14})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_9 + \rho_{10} + \rho_{11} + \rho_{12}) =$	$\rho_1 + \rho_2 + \rho_9 + \rho_{10}$
$(\rho_1 + \rho_2 + \rho_5 + \rho_6 + \rho_9 + \rho_{10} + \rho_{13} + \rho_{14})$ $\times(\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_5 + \rho_6 + \rho_{13} + \rho_{14}$
$(\rho_3 + \rho_4 + \rho_7 + \rho_8 + \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_9 + \rho_{10} + \rho_{11} + \rho_{12}) =$	$\rho_3 + \rho_4 + \rho_{11} + \rho_{12}$
$(\rho_3 + \rho_4 + \rho_7 + \rho_8 + \rho_{11} + \rho_{12} + \rho_{15} + \rho_{16})$ $\times(\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_7 + \rho_8 + \rho_{15} + \rho_{16}$
$(\rho_1 + \rho_3 + \rho_5 + \rho_7 + \rho_9 + \rho_{11} + \rho_{13} + \rho_{15})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5 + \rho_6 + \rho_7 + \rho_8) =$	$\rho_1 + \rho_3 + \rho_5 + \rho_7$
$(\rho_1 + \rho_3 + \rho_5 + \rho_7 + \rho_9 + \rho_{11} + \rho_{13} + \rho_{15})$ $\times(\rho_9 + \rho_{10} + \rho_{11} + \rho_{12} + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_9 + \rho_{11} + \rho_{13} + \rho_{15}$
$(\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_5 + \rho_6 + \rho_7 + \rho_8) =$	$\rho_2 + \rho_4 + \rho_6 + \rho_8$
$(\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16})$ $\times(\rho_9 + \rho_{10} + \rho_{11} + \rho_{12} + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_{10} + \rho_{12} + \rho_{14} + \rho_{16}$
$(\rho_1 + \rho_3 + \rho_5 + \rho_7 + \rho_9 + \rho_{11} + \rho_{13} + \rho_{15})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_9 + \rho_{10} + \rho_{11} + \rho_{12}) =$	$\rho_1 + \rho_3 + \rho_9 + \rho_{11}$
$(\rho_1 + \rho_3 + \rho_5 + \rho_7 + \rho_9 + \rho_{11} + \rho_{13} + \rho_{15})$ $\times(\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_5 + \rho_7 + \rho_{13} + \rho_{15}$
$(\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16})$ $\times(\rho_1 + \rho_2 + \rho_3 + \rho_4 + \rho_9 + \rho_{10} + \rho_{11} + \rho_{12}) =$	$\rho_2 + \rho_4 + \rho_{10} + \rho_{12}$
$(\rho_2 + \rho_4 + \rho_6 + \rho_8 + \rho_{10} + \rho_{12} + \rho_{14} + \rho_{16})$ $\times(\rho_5 + \rho_6 + \rho_7 + \rho_8 + \rho_{13} + \rho_{14} + \rho_{15} + \rho_{16}) =$	$\rho_6 + \rho_8 + \rho_{14} + \rho_{16}$

□

D.5. Factorization is equivalent to RCI + ROI

In the interest of reducing clutter, in this subsection and Appendix D.6, the *random*

variables will be suppressed, and only the corresponding values will be shown. For example,

$$P(A = s, B = t | M_A = u, M_B = v) \text{ will be shortened to } P(s, t | u, v)$$

for $s, t = \pm 1$ and $u, v = 1, 2$. The following fact was proved in Jarrett [24].

Theorem 17. *A model instance is factorizable iff it satisfies both the remote context independent (RCI) condition and the remote outcome independence (ROI) condition.*

Proof. (\Rightarrow) Assume factorizability. Then

$$\begin{aligned} P(s|u, v) &= \sum_t P(s, t|u, v) = \sum_t P(s|u)P(t|v) = P(s|u) \text{ and} \\ P(t|u, v) &= \sum_s P(s, t|u, v) = \sum_s P(s|u)P(t|v) = P(t|v). \end{aligned} \tag{55}$$

In each case, the first equality is a marginal probability calculation, the second equality follows from factorizability, and the third is just the fact that a pmf sums to 1. This establishes that factorizability implies RCI. As for establishing that factorizability implies ROI, note

$$P(s, t|u, v) = P(s|u)P(t|v) = P(s|u, v)P(t|u, v). \tag{56}$$

The first equality is the definition of factorizability and the second equation follows from RCI, which was proved in Eq. 55.

(\Leftarrow) Assume RCI and ROI. Then

$$P(s, t|u, v) = P(s|u, v)P(t|u, v) = P(s|u)P(t|v). \tag{57}$$

The first equality follows from ROI and the second equality follows from RCI. □

D.6. LC implies factorizability

Recall that, by the definition of LC, conditional probabilities such as $P(s|t, u, v)$, $P(t|s, u, v)$, must be well-defined, that is, $P(s, u, v) > 0$ and $P(t, u, v) > 0$ (Def. 10). Also $P(u, v) = P(M_A = u, M_B = v) > 0$ is a given – see Sec. 4.1.1.

Theorem 18. *If a model instance is locally causal (LC), then it is also factorizable.*

Proof. First show that LC implies RCI. Assume LC.

$$P(s|u, v) = \sum_t P(s, t|u, v) = \sum_t P(s|t, u, v)P(t) = \sum_t P(s|u)P(t) = P(s|u)$$

and (58)

$$P(t|u, v) = \sum_s P(s, t|u, v) = \sum_s P(t|s, u, v)P(s) = \sum_s P(t|v)P(s) = P(t|v).$$

In each line the third equality invokes LC. The other three equalities follow by ordinary marginal, conditional, and total probability manipulations.

Then factorizability follows from

$$\begin{aligned} P(s, t|u, v) &= \frac{P(s, t, u, v)}{P(u, v)} = \frac{P(s|t, u, v)P(t, u, v)}{P(u, v)} = \\ &= \frac{P(s|t, u, v)P(t|u, v)P(u, v)}{P(u, v)} = P(s|u)P(t|u, v) = P(s|u)P(t|v). \end{aligned} \quad (59)$$

where the first three equalities are conditional probability manipulations and the last two equalities invoke a combination of LC and RCI. □

D.7. RCI model instances that are not NC

Consider the generic model instance with

$$\gamma_k = \frac{1}{2} \text{ for } k = 2, 3, 5, 8, 9, 12, 13, 16 \text{ and } \gamma_k = 0 \text{ otherwise.}$$

Table 42 clearly shows that RCI is satisfied, both on Alice and Bob's sides.

Table 42. Proof that the example satisfies RCI: Each of Alice and Bob’s outcomes depends only on her/his own measurement choice.

s	t	u	v	$P(A = s M_A = u, M_B = v)$	$P(B = t M_A = u, M_B = v)$
-1	-1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
1	1	1	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	-1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
1	1	1	2	$\frac{1}{2}$	$\frac{1}{2}$
-1	-1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
1	1	2	1	$\frac{1}{2}$	$\frac{1}{2}$
-1	-1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
-1	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
1	-1	2	2	$\frac{1}{2}$	$\frac{1}{2}$
1	1	2	2	$\frac{1}{2}$	$\frac{1}{2}$

The correlations and s -functions for this example are

$$(w, x, y, z) = (-1, 1, 1, 1)$$

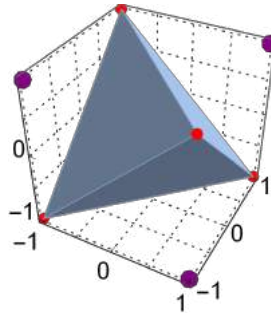
$$(s_1, s_2, s_3, s_4) = (4, 0, 0, 0).$$

Hence one of the CHSH inequalities is violated, so this example cannot be NC. Because it is not NC, it also cannot be factorizable or LC by Thm.’s 2 and 18.

See Thm. 16 in Appendix D.4 for a proof that every NC model instance satisfies RCI. Together with this example, it explains the positioning of "Remote Context Independence" (box 12) strictly above "NC" (box 8) in Fig. 10. It also explains why box 12 labeled "Remote Context Independence" is outlined in green, since this example has an s -function that exceeds a Tsirelson bound.

One can find more examples like these which are RCI but whose correlations have a pattern either of three -1’s and one +1 or three +1’s and one -1. Consider Fig. 17, which shows the NC region for $w = -1$, a tetrahedron. The vertices of the tetrahedron (red dots) correspond to model instances with correlations consisting of zero, two, or four +1’s (or alternatively four, two, or zero -1’s), hence have s -functions that satisfy CHSH. The purple dots not in the NC region, but on the corners of the correlation cube, correspond to model instances that have correlations consisting of one -1 and three +1’s or three -1’s and one +1, hence have an s -function that exceeds a Tsirelson bound. The same holds for the NC region

Figure 17. Perfect correlations (x, y, z) on (red) and off (purple) the NC region, where $w = -1$ is fixed. Red dots at the corners correspond to correlations that satisfy CHSH. Purple dots off of the tetrahedron correspond to correlations that violate CHSH. (The purple dot corresponding to the point $(-1, 1, -1)$ is hidden.)



for $w = 1$, also a tetrahedron, but in which the red dots and the purple dots are switched. See Fig.'s 12 and 13.

D.8. ROI model instance that is not NC

Table 43 shows an assignment of generic parameters $\gamma_1, \gamma_2, \dots, \gamma_{16}$, and the associated correlations and s -functions are shown in Table 44. Note that $s_3 = -2.6108$ exceeds the lower bound of -2 of the CHSH inequalities. Hence the model instance cannot be NC, factorizable, or LC (by Thm. 18).

Table 43. Example that is ROI but does not satisfy CHSH.

Context	Outcome (s, t)			
(u, v)	$(-1, -1)$	$(-1, 1)$	$(1, -1)$	$(1, 1)$
$(1, 1)$	0.0435	0.9540	0.0001	0.0023
$(1, 2)$	0.2198	0.7777	0.0006	0.0020
$(2, 1)$	0.7469	0.0180	0.2296	0.0055
$(2, 2)$	0.1428	0.8142	0.0064	0.0366

Table 44. Proof that example does not satisfy CHSH.

w	-0.9083
x	-0.5565
y	0.5048
z	-0.6412
s_1	0.2154
s_2	-0.4882
s_3	-2.6108
s_4	-0.3188
Min and Max	-2.6108 and 0.2154

However, Table 45 shows that this model instance satisfies ROI. Note that the columns labeled $P(s, t|u, v)$ and $P(s|u, v) \times P(t|u, v)$ are equal, which shows that Alice and Bob's outcomes are conditionally independent, given their joint measurement choices. This example explains the positioning of "Remote Outcome Independence" (box 13) strictly above "Bell local (Factorization)" (box 5) in Fig. 10.

Table 45. Proof that example satisfies ROI.

s	t	u	v	$P(s, t u, v)$	$P(s u, v) \times P(t u, v)$
-1	-1	1	1	0.0435	0.0435
-1	1	1	1	0.9541	0.9541
1	-1	1	1	0.0001	0.0001
1	1	1	1	0.0023	0.0023
-1	-1	1	2	0.2198	0.2198
-1	1	1	2	0.7776	0.7776
1	-1	1	2	0.0006	0.0006
1	1	1	2	0.0020	0.0020
-1	-1	2	1	0.7469	0.7469
-1	1	2	1	0.0180	0.0180
1	-1	2	1	0.2296	0.2296
1	1	2	1	0.0055	0.0055
-1	-1	2	2	0.1428	0.1428
-1	1	2	2	0.8142	0.8142
1	-1	2	2	0.0064	0.0064
1	1	2	2	0.0366	0.0366

D.9. RCI and ROI are distinct

In Sec.'s D.7 and D.8 example model instances were shown that are, respectively,

- RCI but not NC and

- ROI but not NC.

These instances are obviously different, demonstrating that neither the set of RCI model instances nor the set of ROI model instances can contain the other. To see this recall that $RCI \cap ROI = FC \subset NC$ and both of these instances are outside NC . This fact is reflected by the positioning of RCI (box 12) and ROI (box 13) on different paths in Fig. 10.

E. CHSH inequalities survive convex combinations

The proof of Lemma 4 shows that convex combinations of model instances that satisfy CHSH also satisfy all the CHSH inequalities. The definitions of the correlation matrix C and the s -function matrix S can be found in Sec. 2.2.

Lemma 3. *Let $\gamma_1, \gamma_2, \dots, \gamma_n$ be n generic conditional probability vectors, and let $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ be a pmf. Define*

$$\gamma = \sum_k \pi_k \gamma_k.$$

Then

$$SC\gamma^T = \sum_k \pi_k SC\gamma_k^T.$$

That is, the s -functions of a convex combination of conditional probability vectors equal the same convex combination of the s -functions of the underlying cpvs.

Proof. This follows immediately from the linearity of matrix multiplication. □

Lemma 4. *Given the same assumptions as in Lemma 3. Then if the s -functions associated with each γ_k satisfy all of the CHSH inequalities, so do the s -functions associated with γ .*

Proof. From Eq. 8,

$$-2 \leq SC\gamma_k^T \leq 2$$

for all k and since the π_k 's are nonnegative, therefore

$$-2\pi_k \leq \pi_k SC\gamma_k^T \leq 2\pi_k$$

for all k as well. By Lemma 3 and the fact that $\sum_k \pi_k = 1$, adding these n inequalities yields

$$-2 \leq SC\gamma^T \leq 2.$$

□

References

1. J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt. *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett., 23 (15): 880–4, Bibcode:1969PhRvL..23..880C, doi:10.1103/PhysRevLett.23.880), 1969.
<https://doi.org/10.1103/PhysRevLett.23.880>
2. Alain Aspect. *Bell's Theorem: The Naive View of an Experimentalist*, "Quantum [Un]speakables – From Bell to Quantum information", edited by R. A. Bertlmann and A. Zeilinger, Springer, 2002.
https://dx.doi.org/10.1007/978-3-662-05032-3_9
3. Howard Wiseman. *The Two Bell's Theorems of John Bell*, Journal of Physics A: Mathematical and Theoretical, Volume 47, Number 42, Oct 8, 2014.
<https://doi.org/10.1088/1751-8113/47/42/424001>
4. John S. Bell. *On the Einstein, Podolsky, Rosen Paradox*, Physics Vol. 1, No. 3, pp. 195-290, 1964. (Reproduced in [25] [26].)
<https://dx.doi.org/10.1103/PhysicsPhysiqueFizika.1.195>
5. John S. Bell. *The Theory of Local Beables*, Epistemological Letters 9, 1976; Dialectica 39 1985. (Reproduced in [25] [26].)
https://doi.org/10.1142/9789812386540_0008
6. N.D.Mermin. *Hidden variables and the two theorems of John Bell*, Rev. Mod. Phys., 65, 803–815, 1993.
<https://doi.org/10.1103/RevModPhys.65.803>
7. John S. Bell. *On the Problem of Hidden Variables in Quantum Mechanics* Review of Modern Physics 38, 1966 (Reproduced in [25] [26].)
<https://dx.doi.org/10.1103/RevModPhys.38.447>
8. S. Kochen, E. Specker. *The Problem of Hidden Variables in Quantum Mechanics*, Journal of Mathematics and Mechanics, 17: 59–87, 1967.
<http://dx.doi.org/10.1512/iumj.1968.17.17004>
9. W. Myrvold, M. Genovese, A. Shimony. *Bell's Theorem*, The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), Fall 2021 Edition. First published Wed Jul 21, 2004; substantive revision Wed Mar 13, 2019.
<https://plato.stanford.edu/archives/fall2021/entries/bell-theorem/>
10. Arthur Fine. *Hidden Variables, Joint Probability, and the Bell Inequalities*, Phys. Rev. Lett. 48, 291, February 1982.
<https://doi.org/10.1103/PhysRevLett.48.291>
11. Arthur Fine. *Joint distributions, quantum correlations, and commuting observables*, J. Math. Phys., 23, 1306, 1982.
<https://doi.org/10.1063/1.525514>

12. S. Abramsky, A. Brandenburger. *The sheaf-theoretic structure of non-locality and contextuality*, IOP Publishing and Deutsche Physikalische Gesellschaft, New Journal of Physics, Volume 13, November 2011.
<https://dx.doi.org/10.1088/1367-2630/13/11/113036>
13. C. Held. *The Kochen-Specker Theorem*, The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), Spring 2018 Edition.
<https://plato.stanford.edu/archives/spr2018/entries/kochen-specker/>
14. A.A. Klyachko, M.A. Can, S. Binicioglu, A.S. Shumovsky. *Simple Test for Hidden Variables in Spin-1 Systems*, Phys. Rev. Lett., 101, 020403, July 2008
<https://doi.org/10.1103/PhysRevLett.101.020403>
15. R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, G. J. Pryde. *Preparation Contextuality Powers Parity-Oblivious Multiplexing*, Phys. Rev. Letters, 102, 010401, Jan, 2009.
<https://dx.doi.org/10.1103/PhysRevLett.102.010401>
16. A. Cabello, S. Severini, A. Winter. *(Non)-Contextuality of Physical Theories as an Axiom*, arXiv:1010.2163v1 [quant-ph], Oct 1, 2010.
<https://arxiv.org/abs/1010.2163>
17. K. Bharti, A.S. Arora, L.C. Kwek, J. Roland. *Uniqueness of all fundamental noncontextuality inequalities*, Phys. Rev. Research, 2, 033010, 2020.
<https://doi.org/10.1103/PhysRevResearch.2.033010>
18. S. Abramsky. *Relational Databases and Bell's Theorem*, In Search of Elegance in the Theory and Practice of Computation: Essays dedicated to Peter Buneman, ed. V. Tannen, L. Wong, L. Libkin, W. Fan, W.C. Tan and M. Fourman, Springer, pages 13-35, 2013.
<http://arxiv.org/abs/1208.6416>
19. J.J.Halliwell. *Two Proofs of Fine's Theorem*, Physics Letters A, Vol. 378, Issue 40, Aug 14 2014, pp. 2945-2950.
<https://doi.org/10.1016/j.physleta.2014.08.012>
20. John S. Bell. *Bertlmann's socks and the nature of reality*, Journal de Physique Colloques, 42 (C2), 1981. (Reproduced in [25] [26].)
<https://doi.org/10.1051/jphyscol:1981202>
21. I. Pitowski. *Quantum Mechanics as a Theory of Probability*, Oct 2005.
<https://arxiv.org/abs/quant-ph/0510095v1>
22. K. Bolonek-Lason, P. Kosinski. *Groups, Platonic Solids and Bell Inequalities*, Quantum, Vol. 5, 593, Nov 2021.
<https://doi.org/10.22331/q-2021-11-29-593>
23. A. Tavakoli, N. Gisin. *The Platonic solids and fundamental tests of quantum mechanics*, Quantum, Vol. 4, 293, July 2020.

<https://doi.org/10.22331/q-2020-07-09-293>

24. J. P. Jarrett. *On the Physical Significance of the Locality Conditions in the Bell Arguments*, *Nous*, Vol. 18 No. 4, Special Issue on the Foundations of Quantum Mechanics, pp. 569-589, Nov 1984.

<https://doi.org/10.2307/2214878>

25. John S. Bell. *Speakable and unspeakable in quantum mechanics*, Cambridge University Press, 1987 (also 1988 and 1989 paperback).

<https://doi.org/10.1017/CBO9780511815676>

26. John S. Bell. *John S. Bell on the Foundations of Quantum Mechanics*, M. Bell, K. Gottfried, M. Veltman, eds., World Scientific, Singapore, 2001.

<https://doi.org/10.1142/4757>

Copyright © 2022 by Michael L. Ulrey. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.