All-Possible-Couplings Approach to Measuring Probabilistic Context

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

1 Department of Psychological Sciences, Purdue University, USA
2 Department of Mathematical Information Technology, University of Jyväskylä, Finland
∗ E-mail: ehtibar@purdue.edu
This work was supported by NSF grant SES-1155956.

Abstract

From behavioral sciences to biology to quantum mechanics, one encounters situations where (i) a system outputs several random variables in response to several inputs, (ii) for each of these responses only some of the inputs may "directly" influence them, but (iii) other inputs provide a "context" for this response by influencing its probabilistic relations to other responses. These contextual influences are very different, say, in classical kinetic theory and in the entanglement paradigm of quantum mechanics, which are traditionally interpreted as representing different forms of physical determinism. One can mathematically construct systems with other types of contextuality, whether or not empirically realizable: those that form special cases of the classical type, those that fall between the classical and quantum ones, and those that violate the quantum type. We show how one can quantify and classify all logically possible contextual influences by studying various sets of probabilistic couplings, i.e., sets of joint distributions imposed on random outputs recorded at different (mutually incompatible) values of inputs.

Introduction

Consider a system with two inputs, α, β, and two random outputs, A, B, about which it is assumed that A is *not* influenced by β, nor B by α. A necessary condition for this selectivity of influences is *marginal selectivity* [*1*]: changes in the values of β do not influence the distribution of A, and analogously for α and B. Let, for example, both inputs and outputs be binary: $\alpha = \{\alpha_1, \alpha_2\}$, $\beta = \{\beta_1, \beta_2\}$, and A, B attain values $+1$ and -1 each. Denoting by A_{ij} and B_{ij} the two outputs conditioned on $\alpha = \alpha_i, \beta = \beta_j$ ($i, j \in \{1, 2\}$), the distribution of (A_{ij}, B_{ij}) is described by the joint probabilities $p_{ij}, q_{ij}, r_{ij}, s_{ij}$ (summing to 1) in the matrix

α_i, β_j	$B_{ij} = +1$	$B_{ij} = -1$
$A_{ij} = +1$	p_{ij}	q_{ij}
$A_{ij} = -1$	r_{ij}	s_{ij}

. (1)

Assuming all four combinations $\{\alpha_1, \alpha_2\} \times \{\beta_1, \beta_2\}$ are possible, marginal selectivity in this example means

$$\begin{array}{c} p_{i1} + q_{i1} = p_{i2} + q_{i2} = \Pr[A_{ij} = +1], \\ p_{1j} + r_{1j} = p_{2j} + r_{2j} = \Pr[B_{ij} = +1], \end{array} \quad (2)$$

for all $i, j \in \{1, 2\}$.

The assumption of *selective influences*, however, is stronger. It requires that the joint distribution of the two outputs satisfies, for all $i, j \in \{1, 2\}$,

$$(A_{ij}, B_{ij}) \sim (f(R, \alpha_i), g(R, \beta_j)) \tag{3}$$

where $\sim$ stands for "has the same distribution as," f, g are some functions, and R is a source of randomness that does not depend on α, β [*2-8*]. In our example (1) this means

$$\begin{array}{c} p_{ij} = \Pr[f(R, \alpha_i) = +1, g(R, \beta_j) = +1], \\ r_{ij} = \Pr[f(R, \alpha_i) = +1, g(R, \beta_j) = -1], \\ \text{etc.} \end{array} \tag{4}$$

In the quantum mechanical context (see below) R is interpreted as "hidden variables." Such a representation may or may not exist when marginal selectivity is satisfied. For instance, the latter is satisfied in the following four distributions,

α_1, β_1	$B_{11} = +1$	$B_{11} = -1$
$A_{11} = +1$	$1/4$	0
$A_{11} = -1$	0	$3/4$

α_1, β_2	$B_{12} = +1$	$B_{12} = -1$
$A_{12} = +1$	0	$1/4$
$A_{12} = -1$	$1/2$	$1/4$

(5)

α_2, β_1	$B_{21} = +1$	$B_{21} = -1$
$A_{21} = +1$	0	$1/2$
$A_{21} = -1$	$1/4$	$1/4$

α_2, β_2	$B_{22} = +1$	$B_{22} = -1$
$A_{22} = +1$	0	$1/2$
$A_{22} = -1$	$1/2$	0

It can be shown, however, that no representation (3) here is possible as the joint probabilities violate the Bell/CHSH inequalities considered below (Section 1 of Theory and Text S1). At the same time, a representation in the form of (3) is possible for the similar distributions

α_1, β_1	$B_{11} = +1$	$B_{11} = -1$
$A_{11} = +1$	$1/4$	0
$A_{11} = -1$	0	$3/4$

α_1, β_2	$B_{12} = +1$	$B_{12} = -1$
$A_{12} = +1$	$1/4$	0
$A_{12} = -1$	$1/4$	$1/2$

(6)

α_2, β_1	$B_{21} = +1$	$B_{21} = -1$
$A_{21} = +1$	0	$1/2$
$A_{21} = -1$	$1/4$	$1/4$

α_2, β_2	$B_{22} = +1$	$B_{22} = -1$
$A_{22} = +1$	0	$1/2$
$A_{22} = -1$	$1/2$	0

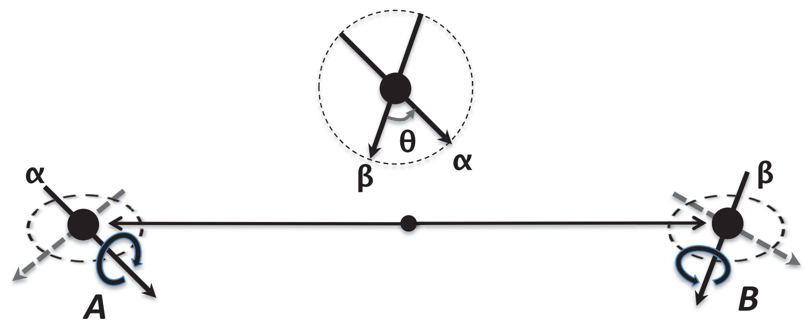

Figure 1. Entanglement paradigm. Schematic representation of two spin-$1/2$ particles, e.g., electrons, in the singlet state (represented by $\langle\uparrow| \otimes \langle\downarrow| - |\downarrow\rangle \otimes |\uparrow\rangle$ in quantum-mechanical notation) running away from each other. The directions α and β are detector settings for spin measurements (in our language, inputs). The measured spins A and B (outputs) in these directions are shown by rotation arrows: one direction of rotation (say, clockwise) represents "spin-up"$= +1$ in one particle and "spin-down"$= -1$ in the other. By the quantum theory, for any α, β, $\Pr[A = +1, B = +1] = 1/2 \cos^2 \theta/2$ (equivalently, expected value of AB is $\cos\theta$). The two measurements are made simultaneously (in some inertial frame of reference).

One can think of α and β in (5) and (6) as being involved in different kinds of *probabilistic context* for the "direct" dependence of, respectively, B on β and A on α.

We propose a principled way of quantifying and classifying conceivable contextual influences, whether within or outside the scope of (3). Our approach is neutral with respect to such issues as causality or what distinguishes direct influences from contextual. We merely accept as a given a diagram of direct input-output correspondences (e.g., $A \leftarrow \alpha, B \leftarrow \beta$) and study the joint distribution of the outputs at all possible values of the inputs. The interpretation of the diagram is irrelevant insofar as it is compatible with the observed pattern of marginal selectivity: as α changes while β remains fixed, the distribution of B does not change, and as β changes while α remains fixed, the distribution of A does not change. Note that the distribution of A may but does not have to change in response to changes in α, and analogously for B and β.

Our approach is maximally general in the sense of applying to arbitrary sets of inputs and outputs (see Section 5 of Theory). To demonstrate it by detailed computations, however, we focus primarily on binary α, β influencing binary A, B; and even more narrowly, on the "homogeneous" case with the two values of both A and B equiprobable at all values of the inputs α_i, β_j $(i, j \in \{1, 2\})$,

$$\Pr[A_{ij} = +1] = \Pr[B_{ij} = +1] = 1/2. \tag{7}$$

Marginal selectivity then is satisfied trivially (because all marginal distributions are fixed).

The example focal for this paper is Bohm's version of the Einstein-Podolsky-Rosen paradigm (EPR/B) [*9*]: a quantum mechanical system consisting (in the simplest case) of two entangled spin-$1/2$ particles separated by a space-like interval (see Fig. 1). The two inputs here are spin measurements on these particles: input α has two values corresponding to spin axes α_1, α_2 chosen for one particle, and input β has two values corresponding to spin axes β_1, β_2 for another particle. The two outputs are spin values recorded: having chosen axes α_i and β_j, $i, j \in \{1, 2\}$, one records A_{ij} for the first particle and B_{ij} for the second, each being a random variable with values $+1$ and -1. (Note that the spins of a given particle along two different axes are *noncommuting* (see Text S2), because of which if one spin value is determined precisely, $+1$ or -1, the other one has a nonzero uncertainty. This means that α_1, α_2 considered as measurements yielding precise values of spins are mutually exclusive, and this is the reason α_1, α_2 can be viewed as values of a single input α; and analogously for β_1, β_2 [*10,11*].) Marginal selectivity (2) in this context is known under a variety of other names, such as "parameter independence" and "physical locality" [*12*]. We confine ourselves to the case (7), with the two spin values $+1$ and -1 being equiprobable for both A_{ij} and B_{ij}.

Formally equivalent situations are abundant in behavioral and social sciences [*8,13-17*], where the issue of selective influences was initially introduced in [*18,19*], in the context of information processing architectures. An example of a system here (from our laboratory) can be a human observer who adjusts a visual stimulus until it matches in appearance another, "target" visual stimulus. Let the latter be characterized by two properties, α and β (e.g., amplitudes of two Fourier-components), each varying on two levels, α_1, α_2 and β_1, β_2. Denoting by S^1_{ij} and S^2_{ij} the corresponding properties (amplitudes) of the adjusted stimulus in response to α_i, β_j, we define a binary random output A_{ij} as having the value "high"$= +1$ or "low"$= -1$ according as the variable S^1_{ij} is above or below the median of its distribution; output B_{ij} is defined from S^2_{ij} analogously. Marginal selectivity in the form (7) is ensured here by construction.

In an example from a biological domain S^1_{ij} and S^2_{ij} could be activity levels of two neurons tuned to two stimulus properties, α and β, respectively. Making α and β vary on two levels each and defining A_{ij}, B_{ij} with respect to the medians of S^1_{ij}, S^2_{ij} by the same rule as above, we get precisely the same mathematical formulation.

The formal equivalence of these three examples should by no means be interpreted as a hint at their physical affinity. Unlike in the EPR/Bohm paradigm, no physical laws prohibit the activity level A of a neuron tuned to stimulus property α from being affected by stimulus property β. Similarly, the amplitude A of the first Fourier component of the adjusted stimulus in the second example may very well be affected by the amplitude β of the second Fourier component of the target stimulus. Our only claim is that if these "secondary" influences do not change the marginal distributions of A and B (which in the two examples in question is ensured by the definition of A and B), they can be viewed within the framework of a formal treatment that also includes the (physically very different) case of entangled particles.

Theory

1 Forms of context (determinism)

In the following, symbols i, j, k (possibly with primes) always take on values $1, 2$ each, and each of the outputs A_{ij}, B_{ij} takes on values $+1, -1$ with equal probabilities. Representation (3) is equivalent to the existence of a jointly distributed system

$$H = \left(H^1_1, H^1_2, H^2_1, H^2_2\right), \tag{8}$$

such that every output pair A_{ij}, B_{ij} is distributed as H^1_i, H^2_j; in symbols,

$$\left(H^1_i, H^2_j\right) \sim \left(A_{ij}, B_{ij}\right). \tag{9}$$

As this entails

$$H^1_i \sim A_{ij},\ H^2_j \sim B_{ij},$$

all components of H are random variables with equiprobable +1/-1, and (9) reduces to

$$\begin{array}{l}\Pr\left[A_{ij}=+1,B_{ij}=+1\right]\\ =\Pr\left[H_i^1=+1,H_j^2=+1\right].\end{array} \tag{10}$$

The existence of H in (8) satisfying (9) is known as (a special case of) the *Joint Distribution Criterion* (JDC) [*6,7,14,20,21*]. It follows from (3) by

$$H_i^1=f\left(R,\alpha_i\right),H_j^2=g\left(R,\beta_j\right). \tag{11}$$

Conversely, if (9) holds for some H, then one can put $R=H$ and

$$\begin{array}{l}f\left(H,\alpha_i\right)=\mathrm{Proj}_i\left(H_1^1,H_2^1,H_1^2,H_2^2\right),\\ g\left(H,\beta_j\right)=\mathrm{Proj}_{2+j}\left(H_1^1,H_2^1,H_1^2,H_2^2\right),\end{array} \tag{12}$$

where Proj_k stands for the "kth member" (in the list of arguments). The JDC is a deep criterion that provides a probabilistic foundation for our understanding of the classical (non)contextuality (or classical determinism in physics). In particular, it immediately follows from the JDC that if representation (3) for (A_{ij},B_{ij}) exists, the "hidden variables" R can always be reduced to a single discrete random variable with 2^4 possible values (corresponding to the possible values of H).

Using the same notation as above,

$$p_{ij}=\Pr\left[A_{ij}=+1,B_{ij}=+1\right], \tag{13}$$

the JDC in our case (two binary inputs and two binary outputs with equiprobable values) is equivalent to four double-inequalities

$$0\le p_{ij}+p_{ij'}+p_{i'j'}-p_{i'j}\le 1 \tag{14}$$

with $i\neq i'$, $j\neq j'$ [*6,7*]. (See Text S1 for a derivation.) They are often referred to as *the Bell/CHSH inequalities* (in the homogeneous form), CHSH acronymizing the authors of [*4*], although the first appearance of these inequalities dates to [*5*].

The theory of the EPR/B paradigm predicts and experimental data confirm violations of the Bell/CHSH inequalities [*22,23*], but quantum mechanics imposes its own constraint on the same linear combinations of probabilities :

$$\frac{1-\sqrt{2}}{2}\le p_{ij}+p_{ij'}+p_{i'j'}-p_{i'j}\le\frac{1+\sqrt{2}}{2}. \tag{15}$$

This constraint is known as the Cirel'son inequalities [24, 25] (see Text S2 for a derivation). Since the class of vectors $(p_{11},p_{12},p_{21},p_{22})$ that satisfy these double-inequalities include those allowed by (14) as a proper subset, it is natural to expect that (15) represents some relaxation, or generalization of the JDC. No such generalization, however, has been previously proposed. Developing one is the main goal of this paper.

This generalization is not confined to quantum mechanical systems. In other (e.g., behavioral) applications, one cannot exclude a priori the possibility of the bounds m and M in

$$m\le p_{ij}+p_{ij'}+p_{i'j'}-p_{i'j}\le M \tag{16}$$

being wider than in (15), or falling between the bounds in (14) and (15), or being more narrow than in (14). One can think of all kinds of other constraints imposed on the possible values of $(p_{11},p_{12},p_{21},p_{22})$, from confining this vector to one specific value to allowing it to vary freely. The latter ("complete chaos") is represented by the "no-constraint" constraint

$$-\,{}^1\!/\!_2\le p_{ij}+p_{ij'}+p_{i'j'}-p_{i'j}\le{}^3\!/\!_2 \tag{17}$$

with $m = -1/2$ attained if one of $p_{11}, p_{12}, p_{21}, p_{22}$ is $1/2$ and the rest are zero, and $M = 3/2$ attained if three of $p_{11}, p_{12}, p_{21}, p_{22}$ are $1/2$ and the remaining one is zero. Recall that we only consider the outputs with equiprobable outcomes, so

$$0 \le p_{ij} \le 1/2. \tag{18}$$

All these conceivable constraints on the possible values of $(p_{11}, p_{12}, p_{21}, p_{22})$ represent different forms and degrees of contextual influences. It would be unsatisfactory if all these possibilities, whether or not empirically realizable, could not be treated within a unified probabilistic framework including JDC as a special case. We construct such a framework, based on the classical (Kolmogorov's) theory of probability and the probabilistic coupling theory [*26*].

2 Connections

It is easy to see that for any vector of probabilities $p = (p_{11}, p_{12}, p_{21}, p_{22})$ one can find a jointly distributed system of +1/-1 variables

$$H = \left(H^1_{11}, H^2_{11}, H^1_{12}, H^2_{12}, H^1_{21}, H^2_{21}, H^1_{22}, H^2_{22}\right) \tag{19}$$

such that

$$\left[\begin{array}{c} \left(H^1_{ij}, H^2_{ij}\right) \sim (A_{ij}, B_{ij}) \\ \text{i.e.} \\ \Pr\left[H^1_{ij} = +1, H^2_{ij} = +1\right] = p_{ij} \end{array}\right] \tag{20}$$

for all i, j. The JDC then amounts to additionally assuming that among all such vectors H there is one with

$$\begin{array}{c} \Pr\left[H^1_{i1} \neq H^1_{i2}\right] = 0, \\ \Pr\left[H^2_{1j} \neq H^2_{2j}\right] = 0, \end{array} \tag{21}$$

and this is the assumption that is rejected by quantum theory in the EPR/B paradigm. Once (21) is explicitly formulated, however, it becomes clear that it is not the only way of thinking of H. Since A_{i1} and A_{i2} occur under mutually exclusive conditions, one cannot identify the distribution of $\left(H^1_{i1}, H^1_{i2}\right)$ with that of (A_{i1}, A_{i2}). The latter does not exist as a pair of jointly distributed random variables. There is therefore no privileged pairing scheme for realizations of H^1_{i1} and H^1_{i2},and zero values for $\Pr\left[H^1_{i1} \neq H^1_{i2}\right], \Pr\left[H^2_{1j} \neq H^2_{2j}\right]$ are as acceptable a priori as any other. Analogous considerations apply to $\left(H^2_{1j}, H^2_{2j}\right)$ and (B_{1j}, B_{2j}).

Our approach consists in replacing (21) with more general

$$\begin{array}{c} \Pr\left[H^1_{i1} \neq H^1_{i2}\right] = 2\varepsilon^1_i \in [0,1], \\ \Pr\left[H^2_{1j} \neq H^2_{2j}\right] = 2\varepsilon^2_j \in [0,1], \end{array} \tag{22}$$

and characterizing the dependence of (A, B) on (α, β) by properties of the set of all 4-vectors $\varepsilon = \left(\varepsilon^1_1, \varepsilon^1_2, \varepsilon^2_1, \varepsilon^2_2\right)$ that are compatible with or imply certain constraints imposed on the vectors $p = (p_{11}, p_{12}, p_{21}, p_{22})$. Having adopted a particular diagram of input-output correspondences (in our case, $A \leftarrow \alpha, B \leftarrow \beta$), we can also say that these sets of ε characterize the contextual role of α, β for B and A, respectively.

We call ε a vector of *connection probabilities*. The connection probabilities are of a principally non-empirical nature: they are joint probabilities of events that can never co-occur. By contrast, due to (20) the components of p are joint probabilities of events that do co-occur, and by observing these co-occurrences the probabilities in p can be estimated. To emphasize this distinction we refer to p as a vector of *empirical probabilities*.

To distinguish our approach from other forms and meanings of probabilistic contextualism, e.g., [*27,28,29*], we dub it the *"all-possible-couplings"* approach. The term "coupling" refers to imposing a joint distribution (say, that of H^1_{11}, H^1_{12}) on random variables that otherwise are not jointly distributed (A_{11} and A_{12}). For a rigorous and general discussion of couplings and connections see Section 5.

3 Extended Linear Feasibility Polytope (ELFP)

ELFP is the set of all possible (p, ε) for which there exists a vector H in (19) with jointly distributed components H_{ij}^k such that (20) holds, and, in accordance with (22),

$$\begin{array}{l} \Pr\left[H_{i1}^1 = +1, H_{i2}^1 = +1\right] = \varepsilon_i^1, \\ \Pr\left[H_{1j}^2 = +1, H_{2j}^2 = +1\right] = \varepsilon_j^2, \end{array} \tag{23}$$

for all i, j. The existence of such an H means the existence of a probability vector Q consisting of the 2^8 joint probabilities

$$\Pr\left[H_{11}^1 = h_{11}^1, H_{11}^2 = h_{11}^2, \ldots, H_{22}^2 = h_{22}^2\right], \tag{24}$$

$h_{ij}^1, h_{ij}^2 \in \{+1, -1\}$. Let P denote the 2^5-component vector consisting of 2^4 empirical probabilities

$$\Pr\left[A_{ij} = a_{ij}, B_{ij} = b_{ij}\right] \tag{25}$$

and 2^4 connection probabilities

$$\begin{array}{l} \Pr\left[A_{i1} = a_{i1}, A_{i2} = a_{i2}\right], \\ \Pr\left[B_{1j} = b_{1j}, B_{2j} = b_{2j}\right], \end{array} \tag{26}$$

$a_{ij}, b_{ij} \in \{+1, -1\}$.

Define a $2^5 \times 2^8$ Boolean matrix M whose rows are enumerated in accordance with components of P (i.e., by equalities $[A_{ij} = a_{ij}, B_{ij} = b_{ij}]$, $[A_{i1} = a_{i1}, A_{i2} = a_{i2}]$, or $[B_{1j} = b_{1j}, B_{2j} = b_{2j}]$) and columns in accordance with components of Q (i.e., by equalities $\left[H_{11}^1 = h_{11}^1, H_{11}^2 = h_{11}^2, \ldots, H_{22}^2 = h_{22}^2\right]$). An entry of M contains 1 if and only if the corresponding random variables in the enumerations of its row and its column have the same values: e.g., if a row is enumerated by $[B_{12} = b_{12}, B_{22} = b_{22}]$ and a column by $\left[H_{11}^1 = h_{11}^1, \ldots, H_{12}^2 = h_{12}^2, \ldots, H_{22}^2 = h_{22}^2\right]$, then their intersection contains 1 if and only if $h_{12}^2 = b_{12}, h_{22}^2 = b_{22}$.

It is easy to see that H exists if and only if

$$MQ = P \tag{27}$$

for some vector $Q \geq 0$ (componentwise) of probabilities. The vectors P for which such a Q exists are exactly those within the polytope whose vertices are the columns of the matrix M. The term ELFP is due to this construction extending that of the linear feasibility test in [*10*]. This test, among other applications, is the most general way of extending the Bell/CHSH criterion to an arbitrary number of particles, spin axes, and spin quantum numbers [*10,11,30-32*]. Its application to binary inputs/outputs (not necessarily with equiprobable outcomes) is shown in Text S1.

To describe ELFP by inequalities on (p, ε), we introduce the 16-component sets

$$\begin{array}{c} \mathsf{S}p = \left\{ \begin{array}{l} \pm(p_{11} - 1/4) \pm (p_{12} - 1/4) \\ \pm(p_{21} - 1/4) \pm (p_{22} - 1/4) : \\ \text{each } \pm \text{ is } + \text{ or } - \end{array} \right\}, \\ \\ \mathsf{S}\varepsilon = \left\{ \begin{array}{l} \pm\left(\varepsilon_1^1 - 1/4\right) \pm \left(\varepsilon_1^2 - 1/4\right) \\ \pm\left(\varepsilon_2^1 - 1/4\right) \pm \left(\varepsilon_2^2 - 1/4\right) : \\ \text{each } \pm \text{ is } + \text{ or } - \end{array} \right\}. \end{array} \tag{28}$$

$\mathsf{S}_0 p$ and $\mathsf{S}_1 p$ denote the subsets of $\mathsf{S}p$ with, respectively, even (0,2, or 4) and odd (1 or 3) number of + signs; $\mathsf{S}_0\varepsilon$ and $\mathsf{S}_1\varepsilon$ are defined analogously. ELFP is described by

$$\max \left(\begin{array}{c} \max \mathsf{S}_0 p + \max \mathsf{S}_1 \varepsilon, \\ \max \mathsf{S}_1 p + \max \mathsf{S}_0 \varepsilon \end{array} \right) \leq 3/2 \tag{29}$$

(see Text S3).

4 All, Fit, Force, and Equi sets

Let constr (p) denote any constraint (e.g., inequalities) imposed on p. Our approach consists in characterizing this constraint by solving the following four problems:

1. Find the set $\text{All}_{\text{constr}}$ of all $(p, \varepsilon) \in [0, 1/2]^8$ with p subject to constr (p): i.e., $(p, \varepsilon) \in \text{All}_{\text{constr}}$ if and only if
$$\text{constr}(p) \text{ and } (p, \varepsilon) \in \text{ELFP}. \quad (30)$$

2. Find the set $\text{Fit}_{\text{constr}}$ of connection vectors $\varepsilon \in [0, 1/2]^4$ that fit (are compatible with) all empirical probability vectors p satisfying constr: i.e., $\varepsilon \in \text{Fit}_{\text{const}}$ if and only if
$$\text{constr}(p) \Longrightarrow (p, \varepsilon) \in \text{ELFP}. \quad (31)$$

3. Find the set $\text{Force}_{\text{constr}}$ of $\varepsilon \in [0, 1/2]^4$ that force all compatible empirical probability vectors p to satisfy constr: i.e., $\varepsilon \in \text{Force}_{\text{constr}}$ if and only if
$$(p, \varepsilon) \in \text{ELFP} \Longrightarrow \text{constr}(p) \quad (32)$$

4. Find the set $\text{Equi}_{\text{constr}}$ of $\varepsilon \in [0, 1/2]^4$ for which an empirical probability vector p satisfies constr if and only if (p, ε) is in the ELFP set: i.e., $\varepsilon \in \text{Equi}_{\text{constr}}$if and only if
$$\text{constr}(p) \Longleftrightarrow (p, \varepsilon) \in \text{ELFP}. \quad (33)$$

Clearly, $\text{Equi}_{\text{constr}} = \text{Force}_{\text{constr}} \cap \text{Fit}_{\text{constr}}$.

To illustrate, we focus on the following four benchmark constraints. The no-constraint, or "complete chaos" situation is given by
$$\text{chaos(p)} \Longleftrightarrow p \in [0, 1/2]^4, \quad (34)$$
equivalent to (17) . The quantum mechanical constraint is given by
$$\text{quant}(p) \Longleftrightarrow \max \mathsf{S}_1 p \leq \sqrt{2}/2, \quad (35)$$
equivalent to (15) . The "classical" constraint is given by
$$\text{class}(p) \Longleftrightarrow \max \mathsf{S}_1 p \leq 1/2, \quad (36)$$
equivalent to the Bell/CHSH inequalities (14). Finally, we consider the constraint
$$\text{fix}(p) \Longleftrightarrow p = \text{specific vector}. \quad (37)$$
For all constraints except for fix (p) the sets All, Fit, Force, and Equi are as shown in Table 1 (for derivations see Text S4).

Thus, $\text{Fit}_{\text{chaos}}$ is the set of all ε such that $\max \mathsf{S}\varepsilon \leq 1/2$: if an ε is in this set, then any p (with no constraints) is compatible with it. $\text{Force}_{\text{quant}}$ is characterized by $\max \mathsf{S}_0\varepsilon \geq \frac{3-\sqrt{2}}{2}$: if an ε is in this set, then all compatible with it p satisfy quant(p). $\text{Equi}_{\text{class}}$ is the set of all ε such that $\mathsf{S}_0\varepsilon$ contains 1: for any such an ε, a p is compatible with it if and only if it satisfies class (p).

For each of these sets we compute Vol^d, its volume normalized by that of $[0, 1/2]^d$, with d being the dimensionality of the set (Fig. 2). Thus, the defining property of $\text{Force}_{\text{class}}$, $1 \in \mathsf{S}_0\varepsilon$, is satisfied if and only if either all ε_i^k are 0, or they all are 1/2, or two of them are 0 and two 1/2. Hence $\text{Vol}^4 (\text{Force}_{\text{class}}) = 0$. For

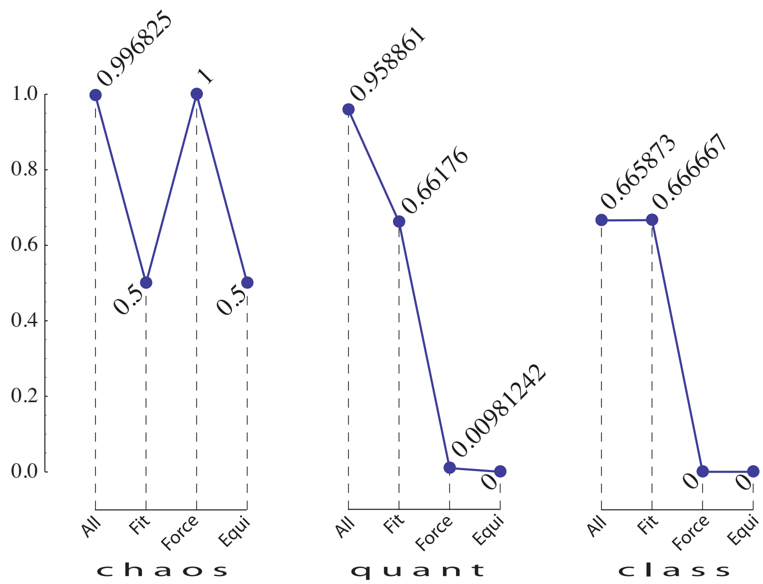

Figure 2. Volume profiles under constraints. Profiles $\mathrm{Vol}^8\left(\mathrm{All}_{\mathrm{constr}}\right) \to \mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{constr}}\right) \to \mathrm{Vol}^4\left(\mathrm{Force}_{\mathrm{constr}}\right) \to \mathrm{Vol}^4\left(\mathrm{Equi}_{\mathrm{constr}}\right)$ for constraints chaos, quant, and class.

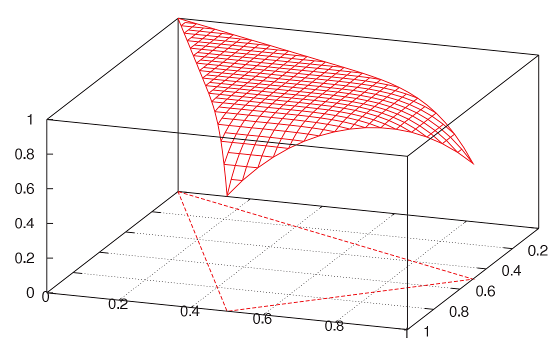

Figure 3. Fit-set volumes for fixed probabilities. $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right)$ is shown as a function of $x = \max \mathsf{S}_0 p$ and $y = \max \mathsf{S}_1 p$. The possible (x, y)-pairs form the triangle $((0,0), (1/2, 1), (1, 1/2))$, and $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right) = 1 + \frac{\rho(x)}{3}\left(-1 + 8x - 24x^2 + 32x^3 - 16x^4\right) + \frac{\rho(y)}{3}\left(-1 + 8y - 24y^2 + 32y^3 - 16y^4\right)$, where $\rho(z) = 1$ if $z \geq 1/2$ and $\rho(z) = 0$ otherwise.

Table 1. Characterizations of the sets of four different types (columns) subject to three constrains (rows). In all cells, $\varepsilon \in [0, 1/2]^4$ and $p \in [0, 1/2]^4$.

	All (p, ε)	Fit (ε)	Force (ε)	Equi (ε)
chaos	$(p, \varepsilon) \in$ ELFP	$\max \mathsf{S}\varepsilon \leq 1/2$	arbitrary	$\max \mathsf{S}\varepsilon \leq 1/2$
quant	$\max \mathsf{S}_1 p \leq \sqrt{2}/2$ & $(p, \varepsilon) \in$ ELFP	$\max \mathsf{S}_0\varepsilon \leq \frac{3-\sqrt{2}}{2}$, $\max \mathsf{S}_1\varepsilon \leq 1/2$	$\max \mathsf{S}_0\varepsilon \geq \frac{3-\sqrt{2}}{2}$	$\frac{3-\sqrt{2}}{2} \in \mathsf{S}_0\varepsilon$, $\max \mathsf{S}_1\varepsilon \leq 1/2$
class	$\max \mathsf{S}_1 p \leq 1/2$ & $(p, \varepsilon) \in$ ELFP	$\max \mathsf{S}_1\varepsilon \leq 1/2$	$1 \in \mathsf{S}_0\varepsilon$	$1 \in \mathsf{S}_0\varepsilon$

nonzero volumes, the derivation is described in Text S4. Each panel of Fig. 2 can be viewed as a "profile" of the corresponding constraint. Each of the first three volumes in a panel can be viewed as characterizing the "strictness" of a constraint, in three different meanings. The intuition of a stricter constraint is that it corresponds to a smaller $\mathrm{Vol}^8\left(\mathrm{All}_{\mathrm{constr}}\right)$, larger $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{constr}}\right)$, and smaller $\mathrm{Vol}^4\left(\mathrm{Force}_{\mathrm{constr}}\right)$. Characterizing constraints imposed on empirical probabilities by multidimensional volumes is not a new idea [33], but our computations are different: they are aimed at sets of nonempirical connection probabilities in relation to constraints imposed on empirical probabilities.

The constraint $\mathrm{fix}\,(p)$ has to be handled separately. Clearly, $\mathrm{Vol}^8\left(\mathrm{All}_{\mathrm{fix(p)}}\right) = 0$. $\mathrm{Fit}_{\mathrm{fix}(p)}$ is described by

$$\begin{aligned} \max \mathsf{S}_1\varepsilon &\leq 3/2 - \max \mathsf{S}_0 p, \\ \max \mathsf{S}_0\varepsilon &\leq 3/2 - \max \mathsf{S}_1 p, \end{aligned} \tag{38}$$

and $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right)$ is a polynomial function of $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$, these two quantities forming the triangle $((0,0),(1/2,1),(1,1/2))$. The polynomial and its values are shown in Fig. 3 (see Text S5, for computational details). $\mathrm{Force}_{\mathrm{fix}(p)}$ is clearly empty, hence so is $\mathrm{Equi}_{\mathrm{fix}(p)}$.

5 All-possible-couplings approach on the general level

We show here how the approach presented so far generalizes to arbitrary sets of inputs and random outputs. We use the term *sequence* to refer to any indexed family (a function from an index set into a set), with index sets not necessarily countable. We present sequences in the form $(x^y : y \in Y)$, $(x_z : z \in Z)$, or $(x_z^y : y \in Y, z \in Z)$. A random variable is understood most broadly, as a measurable mapping between any two probability spaces. In particular, any sequence of jointly distributed random variables is a random variable. For brevity, we omit an explicit presentation of probability spaces and distributions. In all other respects the notation and terminology closely follow [*15,11*].

An *input* is a set of elements called *input values*. Let $\alpha = \left(\alpha^k : k \in K\right)$ be a sequence of inputs. A *treatment* is a sequence $\phi = (x^k : k \in K)$ that belongs to a nonempty set $\Phi \subset \prod_{k\in K} \alpha^k$ (so that $x^k \in \alpha^k$ for all $k \in K$). If $\phi \in \Phi$, $k \in K$, and $I \subset K$, then $\phi(k) = x^k \in \alpha^k$ and $\phi|I$ is the restriction of ϕ to I, i.e., the sequence $(x^k : k \in I)$.

An *output* is a random variable. Let $\left(A_\phi^k : k \in K, \phi \in \Phi\right)$ be a sequence of outputs such that

1. $A_\phi = \left(A_\phi^k : k \in K\right)$ is a random variable for every $\phi \in \Phi$, i.e., the random variables A_ϕ^k across all possible k possess a joint distribution;

2. if $\phi, \phi' \in \Phi$, $I \subset K$, and $\phi|I = \phi'|I$, then $\left(A_\phi^k : k \in I\right) \sim \left(A_{\phi'}^k : k \in I\right)$.

Property 2 is *(complete) marginal selectivity* [8]. A_ϕ is called an *empirical random variable*, and $A = (A_\phi : \phi \in \Phi)$ is the *sequence of empirical random variables*.

Remark 1. The interpretation is that for every ϕ, each α^k may "directly" influence A_ϕ^k but no other output in A_ϕ. The fact that inputs in $\alpha = \left(\alpha^k : k \in K\right)$ and outputs in an *empirical random variable* $A_\phi = \left(A_\phi^k : k \in K\right)$ are in a bijective correspondence is not restrictive: this can always be achieved by an appropriate grouping of inputs and (re)definition of treatments ϕ [10].

Remark 2. The special case considered in the previous sections corresponds to $K = \{1, 2\}$,

$$\begin{gathered} \alpha = \left(\alpha^1, \alpha^2\right) \\ \text{with} \\ \alpha^k = \left\{\alpha_1^k, \alpha_2^k\right\} \text{ for } k \in \{1, 2\}, \end{gathered} \tag{39}$$

$$\begin{gathered} \Phi = \{\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}\} \\ \text{with} \\ \phi_{ij} = \left(\alpha_i^1, \alpha_j^2\right) \text{ for } i, j \in \{1, 2\}, \end{gathered} \tag{40}$$

and (abbreviating $A_{\phi_{ij}}$ as A_{ij} and $A_{\phi_{ij}}^k$ as A_{ij}^k)

$$\begin{gathered} A = (A_{11}, A_{12}, A_{21}, A_{22}), \\ \text{with} \\ A_{ij} = \left(A_i^1, A_j^2\right) \text{ for } i, j \in \{1, 2\}, \end{gathered} \tag{41}$$

where each A_{ij}^k is a binary random variable with $\Pr\left[A_{ij}^k = a_1^k\right] = \Pr\left[A_{ij}^k = a_2^k\right] = 1/2$.

Given a sequence of *empirical random variables* $A = (A_\phi : \phi \in \Phi)$, a sequence of random variables

$$C_A = \left(C_\tau^I : \tau \in \prod_{k \in I} \alpha^k, I \in 2^K - \{\emptyset, K\}\right) \tag{42}$$

(not necessarily jointly distributed) is called a *connecting set* for A if each C_τ^I is a coupling for

$$A_\tau^I = \left(A_\phi^I : \phi \in \Phi, \phi|I = \tau\right), \tag{43}$$

where $A_\phi^I = \left(A_\phi^k : k \in I\right)$. This means that C_τ^I is a random variable of the form

$$C_\tau^I = \left(C_{\tau,\phi}^I : \phi \in \Phi, \phi|I = \tau\right) \tag{44}$$

with

$$C_{\tau,\phi}^I \sim A_\phi^I \tag{45}$$

for all $\phi \in \Phi$ such that $\phi|I = \tau$. C_τ^I is called an (I, τ)*-connection*. The indexation in $C_{\tau,\phi}^I$ is to ensure that if $(I, \tau) \neq (I', \tau')$, then C_τ^I and $C_{\tau'}^{I'}$ are stochastically unrelated. An *identity* (I, τ)*-connection* C_τ^I is one with $\Pr\left[C_{\tau,\phi}^I = C_{\tau,\phi'}^I\right] = 1$ for any $\phi, \phi' \in \Phi$.

Remark 3. It is generally convenient not to distinguish identically distributed connections. By abuse of language, the distribution of C_τ^I (or some characterization thereof) can also be called (I, τ)-connection. We used this language in the previous sections when we represented $\left(\{k\}, k \mapsto \alpha_i^k\right)$-connections (without introducing them explicitly) by probabilities ε_i^k and called ε a connection vector. See Remark 4.

A jointly distributed sequence

$$H = \left(H_{\phi}^{k} : k \in K, \phi \in \Phi\right) \tag{46}$$

is called an *Extended Joint Distribution Sequence* (EJDS) for (A, C_A) if for any $I \in 2^K - \{\emptyset, K\}$ and any $\tau \in \prod_{k \in I} \alpha^k$,

$$H_{\tau}^{I} = \left(H_{\phi}^{I} : \phi \in \Phi, \phi|I = \tau\right) \sim C_{\tau}^{I}, \tag{47}$$

where $H_{\phi}^{I} = \left(H_{\phi}^{k} : k \in I\right)$, and

$$H_{\phi}^{K} = \left(H_{\phi}^{k} : k \in K\right) \sim A_{\phi} \tag{48}$$

for any $\phi \in \Phi$.

Remark 4. For the special case considered in the previous sections, a connecting set for A is (conveniently replacing $C_{\phi_{ij}}^{\{k\}}$, $C_{\phi_{ij}|\{1\}}^{\{1\}}$, and $C_{\phi_{ij}|\{2\}}^{\{2\}}$ with C_{ij}^{k} , C_i^1, and C_j^2, respectively)

$$\begin{gathered} C_A = \left(C_1^1, C_2^1, C_1^2, C_2^2\right) \\ \text{with} \\ C_i^1 = \left(C_{i,i1}^1, C_{i,i2}^1\right), \ C_j^2 = \left(C_{j,1j}^2, C_{j,2j}^2\right), \end{gathered} \tag{49}$$

such that

$$C_{i,ij}^1 \sim A_{ij}^1, \ C_{j,ij}^2 \sim A_{ij}^2 \tag{50}$$

for $i, j \in \{1, 2\}$. An EJDS for (A, C_A) is a random variable (using analogous abbreviations)

$$H = \left(H_{11}^1, H_{11}^2, H_{12}^1, H_{12}^2, H_{21}^1, H_{21}^2, H_{22}^1, H_{22}^2\right) \tag{51}$$

such that

$$\left(H_{i1}^1, H_{i2}^1\right) \sim C_i^1, \ \left(H_{1j}^2, H_{2j}^2\right) \sim C_j^2 \tag{52}$$

and

$$H_{ij}^{12} = \left(H_{ij}^1, H_{ij}^2\right) \sim A_{ij} = \left(A_{ij}^1, A_{ij}^2\right) \tag{53}$$

for $i, j \in \{1, 2\}$. In the previous sections each C_i^k was represented by ε_i^k and each H_{ij}^{12} by p_{ij}.

An EJDS for (A, C_A) reduces to the Joint Distribution Criterion set (JDC set) of the theory of selective influences [*11, 14*] if all connections in C_A are identity ones. Note that no connection has an empirical meaning: for distinct $\phi, \phi' \in \Phi$, the variables A_{ϕ}^{I} and $A_{\phi'}^{I}$ corresponding to $C_{\tau,\phi}^{I}$ and $C_{\tau,\phi'}^{I}$ do not have an empirically observable (or theoretically privileged) pairing scheme.

Let X be any set whose elements are sequences of *empirical random variables* $A = (A_{\phi} : \phi \in \Phi)$. X can be viewed as the set of all possible *empirical random variables* satisfying certain constraints. We define the sets All_X, Fit_X, Force_X, and Equi_X as follows:

1. All_X is the set of all pairs (A, C_A) such that

$$\begin{gathered} A \in X \\ \text{and} \\ \text{there exists an EJDS } H \text{ for } (A, C_A). \end{gathered} \tag{54}$$

2. Fit_X is the set of all C_A such that

$$\begin{gathered} A \in X \\ \Downarrow \\ \text{there exists an EJDS } H \text{ for } (A, C_A). \end{gathered} \tag{55}$$

3. Force_X is the set of all C_A such that

$$\begin{array}{c}\text{there exists an EJDS } H \text{ for } (A, C_A)\\ \Downarrow\\ A \in X.\end{array} \tag{56}$$

4. $\mathrm{Equi}_X = \mathrm{Force}_X \cap \mathrm{Fit}_X$, that is, $C_A \in \mathrm{Equi}_X$ if and only if

$$\begin{array}{c}A \in X\\ \Updownarrow\\ \text{there exists an EJDS } H \text{ for } (A, C_A)\,.\end{array} \tag{57}$$

The all-possible-couplings approach in the general case consists in characterizing any X (interpreted as a type of contextuality or determinism) by All_X, Fit_X, Force_X, and Equi_X. A straightforward generalization of this approach that might be useful in some applications is to replace C_A in all definitions with a subset of C_A, or several subsets of C_A tried in turn. Thus one might consider connections involving only particular $I \subset K$ (e.g., only singletons), or one might require that some of the connections are identity ones.

Conclusion

The essence of the proposed mathematical framework is as follows. We consider all possible couplings for empirically observed vectors of random outputs. In the case of two binary inputs/outputs these vectors are pairs

$$\begin{array}{c}(A_{11}, B_{11})\,, (A_{12}, B_{12})\,,\\ (A_{21}, B_{21})\,, (A_{22}, B_{22})\,,\end{array} \tag{58}$$

the couplings H for them have the form (19), with the coupling relation (20). We assume that the joint distributions (in our case described by pairwise joint probabilities) of the empirically observed (A_{ij}, B_{ij}) are subject to a certain constraint, given to us by substantive considerations outside the scope of our approach: for instance, if a system consists of entangled particles, a constraint, say (15), is derived from the quantum theory. Due to (20), the constraint is imposed on

$$\begin{array}{c}\left(H_{11}^{1}, H_{11}^{2}\right), \left(H_{12}^{1}, H_{12}^{2}\right),\\ \left(H_{21}^{1}, H_{21}^{2}\right), \left(H_{22}^{1}, H_{22}^{2}\right).\end{array} \tag{59}$$

We investigate then the unobservable "connections", the subvectors of the components of H that correspond to outputs obtained at mutually exclusive values of the inputs (i.e., never co-occurring). In our case these are the pairs

$$\begin{array}{c}\left(H_{11}^{1}, H_{12}^{1}\right), \left(H_{21}^{1}, H_{22}^{1}\right),\\ \left(H_{11}^{2}, H_{21}^{2}\right), \left(H_{12}^{2}, H_{22}^{2}\right)\end{array} \tag{60}$$

corresponding to, respectively,

$$\begin{array}{c}(A_{11}, A_{12})\,, (A_{21}, A_{22})\,,\\ (B_{11}, B_{21})\,, (B_{12}, B_{22})\,.\end{array} \tag{61}$$

We then characterize the constraint imposed on the empirical pairs (59) by describing the "fitting" or "forcing" (or both "fitting and forcing") distributions of the unobservable connections (60). By fitting distributions of (60) we mean those that are compatible with any (59) subject to the constraint in question, the compatibility meaning that all these eight pairs can be embedded into a single H (with jointly distributed components). By forcing distributions of (60) we mean those that are compatible with (59) only if the latter are subject to the given constraint.

The value of this approach is in providing a unified language for speaking of probabilistic contextuality. At the cost of greater computational complexity but with no conceptual complications the computations involved in our demonstration of the all-possible-couplings approach can be extended to more general cases: arbitrary marginal probabilities (satisfying marginal selectivity), nonlinear constraints, and greater numbers of inputs, outputs, and their possible values. The language for a completely general theory, involving unrestricted (not necessarily finite) sets of inputs, outputs, and their values, is presented in Section 5 of Theory.

Supporting Information

Text S1 Derivation of the Bell/CHSH bounds

Text S2 Derivation of the Cirel'son bounds

Text S3 Computations for ELFP

Text S4 Computations for chaos(p), quant(p), and class(p) constraints

Text S5 Computations for $\mathrm{Fit}_{\mathrm{fix}(p)}$ constraint

REFERENCES

1. Townsend JT, Schweickert R (1989) Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *J. Math. Psych.* **33**: 309-327.

2. Bell J (1964) On the Einstein-Podolsky-Rosen paradox. *Physics* **1**: 195-200.

3. Bell J (1966) On the problem of hidden variables in quantum mechanics. *Rev. Modern Phys.* **38**: 447-453.

4. Clauser JF, Horne MA, Shimony A, Holt RA (1969) Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**: 880-884.

5. Clauser JF, Horne MA (1974) Experimental consequences of objective local theories. *Phys. Rev. D* **10**: 526-535.

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

7. Fine A (1982) Hidden variables, joint probability, and the Bell inequalities. *Phys. Rev. Lett.* **48**: 291-295.

8. Dzhafarov EN (2003) Selective influence through conditional independence. *Psychom.* **68**: 7-26.

9. Bohm D, Aharonov Y (1957) Discussion of experimental proof for the paradox of Einstein, Rosen and Podolski. *Phys. Rev.* **108**: 1070-1076.

10. Dzhafarov EN, Kujala JV (2012) Selectivity in probabilistic causality: Where psychology runs into quantum physics. *J. Math. Psych.* **56**: 54-63.

11. Dzhafarov EN, Kujala JV (2012) Quantum entanglement and the issue of selective influences in psychology: An overview. *Lect. Notes in Comp. Sci.* 7620: 184-195.

12. Cereceda J (2000) Quantum mechanical probabilities and general probabilistic constraints for Einstein–Podolsky–Rosen–Bohm experiments. *Found. Phys. Lett.* **13**: 427-442

13. Kujala JV, Dzhafarov EN (2008) Testing for selectivity in the dependence of random variables on external factors. *J. Math. Psych.* **52**: 128-144.

14. Dzhafarov EN, Kujala JV (2010) The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. *Front. Quant. Psych. Meas.* **1:151** doi: 10.3389/fpsyg.2010.00151.

15. Dzhafarov EN, Kujala JV (in press) Order-distance and other metric-like functions on jointly distributed random variables. *Proc. Amer. Math. Soc.* (available as arXiv:1110.1228v3 [math.PR])

16. Dzhafarov EN, Schweickert R, Sung K (2004) Mental architectures with selectively influenced but stochastically interdependent components. *J. Math. Psych.* **48**: 51-64.

17. Schweickert R, Fisher DL, Goldstein WM (2010) Additive factors and stages of mental processes in task networks. *J. Math. Psych.* **54**: 405-414.

18. Sternberg S (1969) The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), *Attention and Performance II. Acta Psych.* **30**: 276–315.

19. Townsend JT (1984) Uncovering mental processes with factorial experiments. *J. Math. Psych.* **28**: 363–400.

20. Suppes P, Zanotti M (1981) When are probabilistic explanations possible? *Synthese* **48**: 191-199.

21. Suppes P (2002) Representation *and Invariance of Scientific Structures* (CSLI, Stanford) pp. 332-351.

22. Aspect A, Grangier P, Roger G (1981) Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.* **47**: 460–463.

23. Aspect A, Grangier P, Roger G (1982) Experimental realization of Einstein-Podolsky-Rosen-Bohm gedankenexperiment: A new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**: 91–94.

24. Cirel'son BS (1980) Quantum generalizations of Bell's inequality. *Lett. Math. Phys.* **4**: 93–100.

25. Landau LJ (1987) On the violation of Bell's inequality in quantum theory. *Phys. Lett. A* **120:** 54–56.

26. Thorisson H (2000) *Coupling, Stationarity, and Regeneration* (New York, Springer).

27. Laudisa F (1997) Contextualism and nonlocality in the algebra of EPR observables. *Phil. Sci.* **64**: 478-496.

28. Khrennikov AYu (2008) EPR–Bohm experiment and Bell's inequality: Quantum physics meets probability theory. *Theor. Math. Phys.* **157**: 1448–1460.

29. Khrennikov AYu (2009) *Contextual Approach to Quantum Formalism. Fundamental Theories of Physics* **160** (Dordrecht, Springer).

30. Pitowsky I *(*1989) *Quantum Probability–Quantum Logic. Lect. Notes in Phys.* **321** (Heidelberg, Springer).

31. Peres A (1999) All the Bell inequalities. *Found. Phys.* **29** :589-614.

32. Basoalto RM, Percival IC (2003) BellTest and CHSH experiments with more than two settings. *J. Phys. A: Math. & Gen.* **36**: 7411–7423.

33. Cabello A (2005) How much larger quantum correlations are than classical ones. Phys. Rev. A **72**: 12113 1-5.

34. Landau LJ (1988) Empirical two-point correlation functions. *Found. Phys.* **18**: 449–460.

35. Masanes Ll. (2003) Necessary and sufficient condition for quantum-generated correlations. *arXiv:quant-ph/0309137*.

S1 Derivation of the Bell/CHSH bounds

A representation (8)-(9) exists if and only if the 2^4 possible values $\left(h_1^1, h_2^1, h_1^2, h_2^2\right)$ of H $(h_i^k \in \{+1, -1\}, i, k \in \{1, 2\})$ can be assigned probabilities

$$p\left(h_1^1, h_2^1, h_1^2, h_2^2\right) = \Pr\left[H_1^1 = h_1^1, H_2^1 = h_2^1, H_1^2 = h_1^2, H_2^2 = h_2^2\right], \tag{S1.1}$$

so that, for all $a_{ij}, b_{ij} \in \{+1, -1\}$, $i, j \in \{1, 2\}$,

$$\sum_{h_1^1, h_2^1, h_1^2, h_2^2} \chi\left(h_i^1 = a_{ij} \wedge h_j^2 = b_{ij}\right) p\left(h_1^1, h_2^1, h_1^2, h_2^2\right) = \Pr\left[A_{ij} = a_{ij}, B_{ij} = b_{ij}\right], \tag{S1.2}$$

where $\chi(...)$ indicates the truth value (1 or 0) of the statement within the parentheses. It is easy to see that this system of linear equations can be written as

$$MQ = P, \tag{S1.3}$$

where P is the 16-vector of probabilities $\Pr\left[A_{ij} = a_{ij}, B_{ij} = b_{ij}\right]$ indexed (together with the columns of matrix M) by (i, j, a_{ij}, b_{ij})-values, say, lexicographically; Q is the 16-vector of unknown probabilities $p\left(h_1^1, h_2^1, h_1^2, h_2^2\right)$ indexed (together with the rows of M) by $\left(h_1^1, h_2^1, h_1^2, h_2^2\right)$-values in some order; and the cells of M indexed by $\left((i, j, a_{ij}, b_{ij}), \left(h_1^1, h_2^1, h_1^2, h_2^2\right)\right)$ contain $\chi\left(h_i^1 = a_{ij} \wedge h_j^2 = b_{ij}\right)$. We conclude that a representation (8)-(9) exists if and only if

$$B(M, P) = 1, \tag{S1.4}$$

where $B(M, P)$ is a Boolean function equal to 1 if (S1.3) has at least one solution with nonnegative components of Q. It is easy to show (see [10] for details) that solutions Q of (S1.3) always have the property

$$\sum_{h_1^1, h_2^1, h_1^2, h_2^2} p\left(h_1^1, h_2^1, h_1^2, h_2^2\right) = 1. \tag{S1.5}$$

It is known from the linear programming theory that $B(M, P)$ is always computable. A standard facet enumeration algorithm allows one to obtain the system of all linear inequalities and equations imposed on P that are equivalent to (S1.4). This system turns out to consist of the equalities (2) representing marginal selectivity, and inequalities that can be written as

$$-2 \leq \mathrm{E}_{ij} + \mathrm{E}_{i'j} + \mathrm{E}_{i'j'} - \mathrm{E}_{ij'} \leq 2, \tag{S1.6}$$

where, in reference to (1), $\mathrm{E}_{ij} = p_{ij} + s_{ij} - q_{ij} - r_{ij}$ is the expected value of $A_{ij}B_{ij}$. When marginal probabilities are all $1/2$, these inequalities reduce to (14), using $p_{ij} = \left(\mathrm{E}_{ij} + 1\right)/4$.

Remark 5. It would be a mistake to consider this proof "computer-assisted" because it mentions a facet enumeration algorithm. The latter is merely a long chain of trivial algebraic transformations, that can always be written out *in extenso* if needed.

S2 Derivation of the Cirel'son bounds

The following is a modification of the derivation given in [25]. Let a, a', b, b' be the Hermitian operators in complex Hilbert space corresponding to, respectively, outputs $A_{1j}, A_{2j}, B_{i1}, B_{i2}$ (where i and j are irrelevant, i.e., a represents both A_{11} and A_{12}, b both B_{11} and B_{21}, etc.). Denoting by E expected value and by Tr trace, we have, for any state (density operator) W,

$$\begin{array}{c} 4p_{11} - 1 = \mathrm{E}\left[A_{11}B_{11}\right] = \mathrm{Tr}\left(Wab\right), \\ 4p_{12} - 1 = \mathrm{E}\left[A_{12}B_{12}\right] = \mathrm{Tr}\left(Wab'\right), \\ etc. \end{array} \tag{S2.1}$$

where either of a and a' commutes with either of b and b'. Inequalities (15) to be demonstrated are equivalent to

$$\begin{array}{c} R_1 = \left|\mathrm{Tr}\left(Wab\right) + \mathrm{Tr}\left(Wab'\right) + \mathrm{Tr}\left(Wa'b\right) - \mathrm{Tr}\left(Wa'b'\right)\right| = \left|\mathrm{Tr}\left(Ws_1\right)\right| \leq 2\sqrt{2}, \\ R_2 = \left|\mathrm{Tr}\left(Wab\right) + \mathrm{Tr}\left(Wab'\right) - \mathrm{Tr}\left(Wa'b\right) + \mathrm{Tr}\left(Wa'b'\right)\right| = \left|\mathrm{Tr}\left(Ws_2\right)\right| \leq 2\sqrt{2}, \\ etc. \end{array} \tag{S2.2}$$

where

$$\begin{array}{c} s_1 = ab + ab' + a'b - a'b' = a\left(b + b'\right) + a'\left(b - b'\right), \\ s_2 = ab + ab' - a'b + a'b' = a\left(b + b'\right) - a'\left(b - b'\right), \\ etc. \end{array} \tag{S2.3}$$

Since the values of the outputs, +1/-1, are the eigenvalues of the corresponding operators, it can easily be seen (e.g., by spectral decomposition, squaring, and then multiplication by an arbitrary vector) that

$$a^2 = b^2 = a'^2 = b'^2 = I, \tag{S2.4}$$

where I is the identity operator. Using this we show by straightforward if somewhat tedious algebra that

$$\begin{array}{c} s_1^2 = s_4^2 = 4I - \left(aa' - a'a\right)\left(bb' - b'b\right), \\ s_2^2 = s_3^2 = 4I + \left(aa' - a'a\right)\left(bb' - b'b\right), \end{array} \tag{S2.5}$$

whence, using the conventional notation for commutators, $[x, y] = xy - yx$,

$$\begin{array}{c} \mathrm{Tr}\left(Ws_1^2\right) = \mathrm{Tr}\left(Ws_4^2\right) = 4 - \mathrm{Tr}\left(W\left[a, a'\right]\left[b, b'\right]\right), \\ \mathrm{Tr}\left(Ws_2^2\right) = \mathrm{Tr}\left(Ws_3^2\right) = 4 + \mathrm{Tr}\left(W\left[a, a'\right]\left[b, b'\right]\right). \end{array} \tag{S2.6}$$

For $k = 1, 2, 3, 4$, since s_k is a Hermitian operator (as the sum of products of commuting Hermitian operators), we know that

$$0 \leq \left(\mathrm{Tr}\left(Ws_k\right)\right)^2 \leq \mathrm{Tr}\left(Ws_k^2\right). \tag{S2.7}$$

It follows from (S2.6) then that

$$\left|\mathrm{Tr}\left(W\left[a, a'\right]\left[b, b'\right]\right)\right| \leq 4 \tag{S2.8}$$

and

$$\mathrm{Tr}\left(Ws_k^2\right) \leq 8. \tag{S2.9}$$

But then

$$R_k^2 = \left(\mathrm{Tr}\left(Ws_k\right)\right)^2 \leq 8. \tag{S2.10}$$

This implies (S2.2) and (15).

That the value $2\sqrt{2}$ in (S2.2) can be attained is easy to show using the EPR/B paradigm: if $\alpha_1 = 0$, $\alpha_2 = \pi/2$, $\beta_1 = \pi/4$, $\beta_2 = -\pi/4$, then

$$R_1 = \cos\left(\alpha_1 - \beta_1\right) + \cos\left(\alpha_1 - \beta_2\right) + \cos\left(\alpha_2 - \beta_1\right) - \cos\left(\alpha_2 - \beta_2\right) = 2\sqrt{2}. \tag{S2.11}$$

Remark 6. It is instructive to see that if the operators a, a' (or b, b') commute, (S2.6) leads to $R_k^2 \leq 4$, which, in view of (S2.1), is equivalent to (14). It is tempting therefore to consider (14) as merely a special (commutative) case of the construction used above to prove (15). Notice however that this view cannot be accepted without additional arguments: the proof of (14) makes no use of the assumption that the outputs are eigenvalues of Hermitian operators in a Hilbert space.

Remark 7. It is known from [6, 7] that if a vector $(p_{11}, p_{12}, p_{21}, p_{22})$ satisfies (14), then this vector can be generated by a system with binary inputs and equiprobable binary outputs that satisfies (3), that is, is explainable by classical (non)contextuality. By contrast, if a vector $(p_{11}, p_{12}, p_{21}, p_{22})$ satisfies (15), it is not known to us whether this vector can be generated by a quantum mechanical system with binary inputs and equiprobable binary outputs. In this sense our characterization of quantum contextuality is improvable. The issue of conditions that are both necessary and sufficient for quantum contextuality has been addressed [33, 34, 35], but only in terms of the existence of *some* quantum systems, not necessarily those with binary inputs and outputs.

S3 Computations for ELFP

A convex bounded polytope can be equivalently defined either as the convex hull of a set of points (V-representation) or as the intersection of half-spaces (H-representation). For our purposes, a V-representation of a convex polytope in d-space is given by a set of points $x_1, \ldots, x_n \in \mathbb{R}^d$. The polytope consists of all convex combinations of these points: $\lambda_1 x_1 + \cdots + \lambda_n x_n$, for all $\lambda_1, \ldots, \lambda_n \geq 0$, $\lambda_1 + \cdots + \lambda_n = 1$. It is possible that the polytope is of lower dimension than the space $\mathbb{R}^d$ in which it is defined if all the points x_i reside in a lower dimensional affine subspace of $\mathbb{R}^d$. A minimal V-representation (including only extreme points, i.e., points that are vertices of the polytope) is unique. The H-representation of a convex polytope is given by vectors $a_1, \ldots, a_m \in \mathbb{R}^d$ and a vector $b \in \mathbb{R}^m$. The polytope consists of the points $x \in \mathbb{R}^d$ satisfying $a_i^T x \leq b_i$ for all $i = 1, \ldots, m$. A lower-dimensional convex polytope can be represented by including inequalities of the forms $a^T x \leq b$ and $(-a)^T x \leq -b$ for some a and b or by explicitly specifying certain constraints as equations in the representation. For a full-dimensional convex polytope, the minimal H-representation is unique. However, for a lower-dimensional polytope, the equation constraints can be specified in many equivalent ways and the inequality constraints can look different depending on which of the linearly related coordinates are used to specify them.

There exist algorithms for converting between the two representations of a convex polytope in exact rational arithmetic. We have used our own program for these conversions but other programs, such as *lrs* (http://cgm.cs.mcgill.ca/~avis/C/lrs.html), can do the same. The conversion between the two representation is computationally demanding, the algorithms generally requiring superpolynomial time in the size of the input.

A computationally simpler problem is eliminating redundant points (those that are not vertices of the polytope) from a V-representation or eliminating redundant equations or inequalities from an H-representation. This problem can be solved by linear programming and the algorithm is implemented in the *redund* program that comes with *lrs*. However, the *redund* program is not sufficient for putting an H-representation to a minimal form as it cannot convert sets of inequalities into equivalent equations (e.g., the three inequalities $x \geq 0$, $y \geq 0$, $x + y \leq 0$ should be minimally represented as the two equations $x = 0$, $y = 0$). To find the minimal H-representation, for every constraint $a_i^T x \leq b_i$ or $a_i^T x = b_i$ in turn, one can find the upper and lower bounds u and l by maximizing and minimizing the expression $a_i^T x$ given the other constraints, and apply the following rules:

1. if this is an equation constraint (i.e., $a_i^T x = b_i$) and $u = l = b_i$, then the constraint is redundant and can be eliminated;

2. if this is an inequality constraint (i.e., $a_i^T x \leq b_i$) and $u \leq b_i$, then the inequality is redundant and can be eliminated. Otherwise, if $l = b_i$, then the constraint should be converted to an equation.

The dimension of a polytope can be determined from a minimal H-representation. It is the dimension of the space minus the number of equation constraints in the minimal representation. Given a full-dimensional polytope, its volume can be computed using the *lrs* program alongside the conversion from a V-representation to an H-representation. If the polytope is given as an H-representation, then it has to be converted to a V-representation first to compute its volume using *lrs*. To compute the volume of a lower-dimensional polytope, we first move to a lower-dimensional parameterization that spans the affine subspace where the polytope resides.

To compute ELFP, we begin by formulating the linear programming problem $MQ = P$ subject to $Q \geq 0$, as described in the main text (M being $2^5 \times 2^8$, P having 2^5 components). M defines the V-representation for ELFP, and Vol^8 for ELFP is computed directly from it. Applying an algorithm to find an equivalent H-representation we obtain a system of 160 inequalities and 16 equations. We can then substitute the expressions in the above matrices into this system and reduce any redundant inequalities and equations. The resulting system has 144 nonredundant inequalities and no equations with the $p_{11}, p_{12}, p_{21}, p_{22}, \varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^2, \varepsilon_2^2$ variables. Then (dropping the implicit $\varepsilon \in [0, 1/2]^4$ and $p \in [0, 1/2]^4$ constraints), we algebraically simplify the list of 144 inequalities, first into

$$\begin{aligned} -\Gamma \leq -p_{11} + p_{21} + p_{12} + p_{22} &\leq 1 + \Gamma, \\ -\Gamma \leq p_{11} - p_{21} + p_{12} + p_{22} &\leq 1 + \Gamma, \\ -\Gamma \leq p_{11} + p_{21} - p_{12} + p_{22} &\leq 1 + \Gamma, \\ -\Gamma \leq p_{11} + p_{21} + p_{12} - p_{22} &\leq 1 + \Gamma, \end{aligned} \tag{S3.1}$$

$$-\Lambda \leq p_{11} + p_{21} + p_{12} + p_{22} \leq 2 + \Lambda, \tag{S3.2}$$

$$\begin{aligned} |-p_{11} - p_{21} + p_{12} + p_{22}| &\leq 1 + \Lambda, \\ |-p_{11} + p_{21} - p_{12} + p_{22}| &\leq 1 + \Lambda, \\ |-p_{11} + p_{21} + p_{12} - p_{22}| &\leq 1 + \Lambda, \end{aligned} \tag{S3.3}$$

where

$$\begin{aligned} \Gamma = \min\{ \quad & 1 - \varepsilon_1^1 - \varepsilon_1^2 + \varepsilon_2^1 + \varepsilon_2^2, \\ & 1 - \varepsilon_1^1 + \varepsilon_1^2 - \varepsilon_2^1 + \varepsilon_2^2, \\ & 1 - \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_2^1 - \varepsilon_2^2, \\ & 1 + \varepsilon_1^1 - \varepsilon_1^2 - \varepsilon_2^1 + \varepsilon_2^2, \\ & 1 + \varepsilon_1^1 - \varepsilon_1^2 + \varepsilon_2^1 - \varepsilon_2^2, \\ & 1 + \varepsilon_1^1 + \varepsilon_1^2 - \varepsilon_2^1 - \varepsilon_2^2, \\ & \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_2^1 + \varepsilon_2^2, \\ & 2 - \varepsilon_1^1 - \varepsilon_1^2 - \varepsilon_2^1 - \varepsilon_2^2 \}, \end{aligned} \tag{S3.4}$$

$$\begin{aligned} \Lambda = \min\{ - \quad & \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_2^1 + \varepsilon_2^2, \\ & \varepsilon_1^1 - \varepsilon_1^2 + \varepsilon_2^1 + \varepsilon_2^2, \\ & \varepsilon_1^1 + \varepsilon_1^2 - \varepsilon_2^1 + \varepsilon_2^2, \\ & \varepsilon_1^1 + \varepsilon_1^2 + \varepsilon_2^1 - \varepsilon_2^2, \\ & 1 - \varepsilon_1^1 - \varepsilon_1^2 - \varepsilon_2^1 + \varepsilon_2^2, \\ & 1 - \varepsilon_1^1 - \varepsilon_1^2 + \varepsilon_2^1 - \varepsilon_2^2, \\ & 1 - \varepsilon_1^1 + \varepsilon_1^2 - \varepsilon_2^1 - \varepsilon_2^2, \\ & 1 + \varepsilon_1^1 - \varepsilon_1^2 - \varepsilon_2^1 - \varepsilon_2^2 \}, \end{aligned} \tag{S3.5}$$

and then, by noticing regularities, into the compact inequality (29).

Remark 8. Changing $\varepsilon_j^i \to 1/2 - \varepsilon_j^i$ leads to (denoting the new ε-vector by ε')

$$\max \mathsf{S}_1 \varepsilon' = \max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_0 \varepsilon' = \max \mathsf{S}_1 \varepsilon. \tag{S3.6}$$

Analogously for $p_{ij} \to 1/2 - p_{ij}$,

$$\max \mathsf{S}_1 p' = \max \mathsf{S}_0 p, \max \mathsf{S}_0 p' = \max \mathsf{S}_1 p. \tag{S3.7}$$

It follows that we cannot without loss of generality confine all components of ε or p to $[0, 1/4]$. But ELFP does not change if the transformation $x \to 1/2 - x$ is applied to an even number of the components of (p, ε).

S4 Computations for chaos(p), quant(p), and class(p) constraints

The $\text{All}_{\text{constr}}$ polytopes for the three constraints are obtained by concatenating the ELFP equations and inequalities with the constraint inequalities. Then, the volumes are computed by using the *lrs* program as described above.

For $\text{Fit}_{\text{constr}}$ polytopes, we observe first that they are convex. This follows from

$$\begin{aligned}\text{Fit}_{\text{constr}} &= \{\varepsilon : \forall i = 1, \ldots, n : \left(p_{(i)}, \varepsilon\right) \in \text{ELFP}\,\} \\ &= \text{ELFP}_{p_{(1)}} \cap \cdots \cap \text{ELFP}_{p_{(n)}},\end{aligned} \tag{S4.1}$$

where $p_{(i)}$, $i = 1, \ldots, n$, denote the vertices of the 4D convex polytope defined by constr and $\text{ELFP}_{p_{(i)}}$ denotes the (convex) cross-section of the ELFP set formed with $p = p_{(i)}$. It follows that $\text{Fit}_{\text{constr}}$ is convex as the intersection of convex sets. Following the logic of this observation, we have implemented a general program for eliminating variables from a system of linear equations and inequalities so that the resulting system is satisfied for exactly those values for which there exist such values of the eliminated variables for which the original system is satisfied. This program together with steps to ensure that the resulting representation is minimal was used to find all the Fit sets shown in the main text.

Finding the forcing sets is more difficult as they are generally not convex. We characterize them using the equation

$$\begin{aligned}&\text{Force}_{\text{chaos}} - \text{Force}_{\text{constr}} \\ &= \{\varepsilon : (\exists p : (p, \varepsilon) \in \text{ELFP} \wedge \neg\text{constr}\,(p))\}\,.\end{aligned} \tag{S4.2}$$

This equation provides an algorithm: for each inequality in constr, form the conjunction of the ELFP inequalities with the negation of the inequality. Then project this conjunction to the ε 4-space. The union of these projections over all inequalities in constr is the set $\text{Force}_{\text{chaos}} - \text{Force}_{\text{constr}}$. We have implemented a general program that takes as input a representation of a polytope, a list of additional constraints, and a list of variables to eliminate. It then outputs a representation of the difference of the polytope and the set represented by the additional constraints projected to the remaining (not eliminated) variables. This representation consists of a list of linear systems whose disjunction characterizes the resulting set. In all our computations it turned out that all the linear systems in the disjunction were the same, and so the sets $\text{Force}_{\text{chaos}} - \text{Force}_{\text{constr}}$ are in fact convex in these cases.

The computations of Equi sets require no elaboration.

Remark 9. There is the practical problem that the negation of a $\leq$ -inequality is a $>$ -inequality while standard algorithms only accept closed convex polytopes. To cope with this problem, we approximated $a > b$ by $a \geq b +$ (very small number). We also used a rational approximation to $\sqrt{2}$ in the quant constraints. In both cases, we have repeated the computations with decreasing values of "very small number" until it was obvious where the results converged.

S5 Computations for $\text{Fit}_{\text{fix}(p)}$ constraint

That $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$ are contained in and completely fill the triangle $\{(0, 0), (1/2, 1), (1, 1/2)\}$ can be verified by splitting (38) into 64 component cases according as which of the values of $\mathsf{S}_0 p$ and $\mathsf{S}_1 p$ are the maxima, finding the vertices of each component system, and drawing the union of these components in $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$ coordinates. The triangle is described by

$$\begin{aligned} &2\max \mathsf{S}_0 p - \max \mathsf{S}_1 p \geq 0, \\ &2\max \mathsf{S}_1 p - \max \mathsf{S}_0 p \geq 0, \\ &\max \mathsf{S}_0 p + \max \mathsf{S}_1 p \leq 3/2. \end{aligned} \tag{S5.1}$$

Adding these inequalities to the representation of (38) as linear inequalities according to the definitions of $\max \mathsf{S}_0 \varepsilon$ and $\max \mathsf{S}_1 \varepsilon$, we obtain a 6D polytope $P^{(6)}$ in $(\varepsilon, \max \mathsf{S}_0 p, \max \mathsf{S}_1 p)$-coordinates. In the V-representation of $P^{(6)}$, all vertices have values of $\max \mathsf{S}_0 p$ and $\max \mathsf{S}_1 p$ in the set

$$\{(0,0), (1/4, 1/2), (1/2, 1/4), (1/2, 1), (1, 1/2)\}\,. \tag{S5.2}$$

It follows that every edge of the polytope projects to one of these 5 points or to a line connecting two of them. Consequently, as $(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon)$ changes within any triangle T formed by these lines, the cross-section $P^{(4)}_{(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon)}$ of $P^{(6)}$ retains its structure (face lattice) while its coordinates change as affine functions of $(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon) \in T$. It follows that the volume of $P^{(4)}_{(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon)}$ is a polynomial of $(\max \mathsf{S}_0 \varepsilon, \max \mathsf{S}_1 \varepsilon) \in T$ of at most degree four. The coefficients of these polynomials were obtained by fitting unconstrained degree 4 polynomials to the exact volumes $\mathrm{Vol}^4\left(\mathrm{Fit}_{\mathrm{fix}(p)}\right)$ for $(\max \mathsf{S}_0 p, \max \mathsf{S}_1 p) \in \{0, .01, .02, \ldots, 1\}^2$. It turns out that the coefficients change only if either of the differences $\max \mathsf{S}_0 p - 1/2$ and $\max \mathsf{S}_1 p - 1/2$ changes its sign. In all cases the fit is perfect for the number of points far exceeding the number of coefficients, confirming that the computations are correct.