A Qualified Kolmogorovian Account of Probabilistic Contextuality

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

[1]Purdue University
ehtibar@purdue.edu
[2]University of Jyväskylä
jvk@iki.fi

Abstract. We describe a mathematical language for determining all possible patterns of contextuality in the dependence of stochastic outputs of a system on its deterministic inputs. The central principle (contextuality-by-default) is that the outputs indexed by mutually incompatible values of inputs are stochastically unrelated; but they can be coupled (imposed a joint distribution on) in a variety of ways. A system is characterized by a pattern of which outputs can be "directly influenced" by which inputs (a primitive relation, hypothetical or normative), and by certain constraints imposed on the outputs (such as Bell-type inequalities or their quantum analogues). The set of couplings compatible with these constraints determines the form of contextuality in the dependence of outputs on inputs.

Keywords: Bell-type inequalities, Cirelson inequalties, context, contextuality-by-default, coupling, direct influences, determinism, EPR paradigm, marginal selectivity, sample spaces, stochastically unrelated variables.

1 Introduction

In this paper we describe a language for analyzing dependence of *stochastic outputs* of a system on *deterministic inputs*. This language applies to systems of all imaginable kinds: quantum physical, macroscopic physical, biological, psychological, and even purely mathematical, created on paper. The notion of "dependence," as well as related to it notions of "influence," "causality," and "context" may have different meanings in different areas. Even if not, we do not know how to define them. We circumvent the necessity of designing these definitions by simply accepting that some inputs are connected to some outputs by arrows called *direct influences*. We ignore the question of how these direct influences are determined, except for a certain necessary condition they must satisfy (*marginal selectivity*). A system is also characterized by certain *constraints* imposed on the joint distribution of its outputs across different inputs. A prominent example when both direct influences and constraints are justified by a well-developed

---

[*] Corresponding author.

theory is the EPR paradigm in quantum physics, where it is assumed that measurement settings for a given particle directly affect measurement outcomes in that particle only, and the joint distributions of the measurement outcomes on different particles satisfy certain inequalities or parametric equalities. If these constraints can be accounted for entirely in terms of the posited direct influences, the system can be viewed as "contextless." If this is not the case, we characterize *probabilistic contexts* by studying the deviations from the contextless behavior exhibited by the system.

Whether one deals with quantum contextuality or thinks of contextuality beyond even quantum bounds, our approach does squarely remains within the domain of the classical probability theory, which we refer to as *Kolmogorovian*. A caveat for using this attribution is that we do not mean the "naive" Kolmogorovian theory in which all random variables are thought of as defined on a single sample space (equivalently, as functions of a single random variable). Such a notion is no more tenable than the "set of all sets" of the naive set theory. The qualified Kolmogorovian approach we adopt is based on the principle of *contextuality-by-default:*

> *any two random variables recorded under mutually exclusive conditions are stochastically unrelated, defined on different sample spaces.*

This is a radical version of views previously expressed in the literature, e.g., in Khrennikov, 2008a-b, where it is traced back to Andrei Kolmogorov himself and even to George Boole. Our emphasis, however, is on the fact that any set of stochastically unrelated variables (but never "all of them") can be *coupled*, or imposed a joint distribution upon, in many different ways (Thorisson, 2000). In particular, the identity coupling is sometimes (but not always) possible, in which the two random variables defined under mutually exclusive conditions and "automatically" (by default) labeled as different and stochastically unrelated, merge into one and the same random variable.

The basics of this approach are presented in Section 2. In Sections 3 and 4 we use it to investigate contextual influences with respect to a given pattern of direct influences. The theory and notation there closely follows Dzhafarov and Kujala (2013a). The departure point is that since different treatments (combinations of input values) are mutually exclusive, the joint distributions of the outputs corresponding to them, according to the principle of contextuality-by-default, are stochastically unrelated. We then consider all possible ways of coupling them across different treatments. From each such a coupling we extract stochastic relations that are "hidden," principally unobservable, because they correspond to outputs obtained under different treatments. We focus on the special kind of these hidden relations, those between random variables that share the same pattern of direct influences. We call these hidden relations *connections*. Given a certain constraint imposed on the system by a theory or empirical observations, we pose the question of what connections imply (or force) this constraint and what connections are implied by (or compatible with) it. Taken over all possible couplings, these relations between connections and constraints characterize the

type of contextuality exhibited by the system. This view of contextuality is different from the existing approaches (Khrennikov, 2009; Laudisa, 1997).

2   Probability Theory: Multiple Sample Spaces

Given two probability spaces, (S, Σ, p) and (S_A, Σ_A, p_A), with standard meaning of the terms, a *random variable* is defined as a (Σ, Σ_A)-measurable function $A : S \to S_A$ subject to

$$p_A(X) = p\left(A^{-1}(X)\right), \tag{1}$$

for any $X \in \Sigma_A$. The probability space (S, Σ, p) is usually called a *sample space*, and we will refer to (S_A, Σ_A, p_A) as the *distribution* of A. The sample space itself is a distribution of the random variable R (let us call it a *basic variable*) which is the (Σ, Σ)-measurable identity function, $x \mapsto x$, $x \in \Sigma$. Any random variable A defined on this sample space can also be presented as a function $A = f(R)$, and (1) can be written as

$$p_A(X) = \Pr[A \in X] = \Pr\left[R \in f^{-1}(X)\right], \tag{2}$$

for any $X \in \Sigma_A$.

Let $\left(A^k = f_k(R) : k \in K\right)$ be a sequence[1] of random variables, all functions of one and the same basic variable R, with A^k distributed as $\left(S^k, \Sigma^k, p_k\right)$. Then $A = \left(A^k : k \in K\right) = f(R)$ too is a random variable that is a function of R, with the distribution

$$\left(S_A = \prod_{k \in K} S^k, \Sigma_A = \bigotimes_{k \in K} \Sigma^k, p_A\right). \tag{3}$$

Here, $\bigotimes_{k \in K} \Sigma^k$ is the minimal sigma-algebra containing sets of the form $X^k \times \prod_{i \in K-\{k\}} S^i$ for all $X^k \in \Sigma^k$, and p_A is defined by (2), with

$$f^{-1}(X) = \{x \in S : (f_k(x) : k \in K) \in X\}. \tag{4}$$

The distribution of A can also be given by (3) with no reference to its sample space, or basic variable. It can be viewed as a *joint distribution* of the components of a sequence $A = \left(A^k : k \in K\right)$, such that, for any nonempty $K' \subset K$, the subsequence $A' = \left(A^k : k \in K'\right)$ is a random variable distributed as

$$\left(S_{A'} = \prod_{k \in K'} S^k, \Sigma_{A'} = \bigotimes_{k \in K'} \Sigma^k, p_{A'}\right), \tag{5}$$

with

$$p_{A'}(X) = p_A\left(X \times \prod_{k \in K-K'} S^k\right), \tag{6}$$

---

[1] The term *sequence* in this paper is used in the generalized meaning, as any indexed family, a function from an index set into a set. Index sets need not be countable.

for any $X \in \Sigma_{A'}$. The distribution $\left(S^k, \Sigma^k, p_k\right)$ of a single A^k is determined by that of the one-element subsequence $\left(A^k\right)$ in the obvious way. All the random variables A^k obtained in this way from A can be viewed as functions on one and the same basic variable, e.g., $R = A$ itself.

We see that the relation "are jointly distributed" is synonymous to the relation "are functions of one and the same basic variable." But clearly there cannot be a single basic variable of which all imaginable random variables are functions. This is obvious from the cardinality considerations alone, as random variables may have arbitrarily large sets of possible values. But this is true even if one confines consideration to all imaginable random variables with any given distribution, provided it is not concentrated at a point. Let, e.g., $\mathcal{B}$ be a class (not necessarily a set) of all functions of R that are Bernoulli $(0/1)$ variables with equiprobable values. That is, each $B \in \mathcal{B}$ is a function $f(R)$ with $f : S \to \{0, 1\}$, such that $\Pr\left(R \in f^{-1}\left(\{0\}\right)\right) = 1/2$. Consider a Bernoulli variable B^* with equiprobable values such that for any $B \in \mathcal{B}$,

$$\Pr\left(B = 0, B^* = 0\right) = 1/4. \tag{7}$$

Then B^* cannot be a function of R because it is independent of (hence is not the same as) any of the elements of $\mathcal{B}$. If needed, however, one can redefine the basic variable, e.g., as $R^* = (R, B^*)$, with independent R and B^*, so that all elements of $\mathcal{B} \cup \{B^*\}$ become functions of R^*.

This simple demonstration shows that the Kolmogorovian approach to probability is not represented by a single sample space with measurable functions on it. Rather the true picture is an "open-ended" class (definitely not a set) of basic variables that are *stochastically unrelated* to each other, each with its own class of random variables defined as its functions: schematically,

$$\ldots \qquad R^1 \qquad \ldots \qquad R^2 \qquad \ldots \tag{8}$$
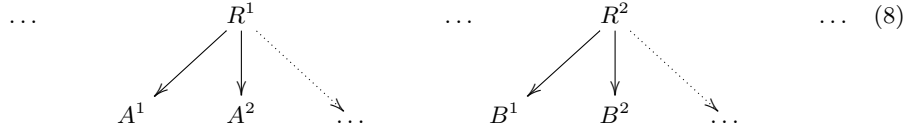
If necessary, using some *coupling scheme* as discussed below, any sequence of stochastically unrelated basic variables $\left(R^k : k \in K\right)$ can be redefined into a random variable $H = \left(H^k : k \in K\right)$ such that H^k and R^k are identically distributed for all k. This amounts to considering all individual R^k, as well as their functions, as functions of H. But this procedure is not unique, and it cannot be performed for "all random variables."

The contextuality-by-default principle requires that any two random variables conditioned upon mutually exclusive values of some third variable are stochastically unrelated. Indeed, there is never a unique way for coupling their realizations. A simple example: I flip a coin and depending on the outcome weigh one of two lumps of clay, lump 1 (if "heads") or lump 2 (if "tails"). The random

variables $A =$"weight reading for lump 1" and $B =$"weight reading for lump 2" do not a priori possess a joint distribution because there is no privileged way of deciding whether a given value of A *co-occurs* with a given value of B. If necessary, however, such a co-occurrence (or coupling) scheme can always be constructed. For instance, one can list the values of A and B chronologically and then couple the nth realization of A with the nth realization of B ($n = 1, 2, \ldots$). Or one could rank-order the values of A and B and couple the realizations of the same quantile rank (this would create positive correlation between the variables) or of the complementary ranks (negative correlation). One cannot say that one way of paring is better justified than another, each one represents "a point of view" and creates its own joint distribution of A and B.

3   All Possible Couplings Approach

Consider a sequence of random variables $A = (A_\phi : \phi \in \Phi)$. The elements of Φ are called *(allowable) treatments*. Two distinct treatments ϕ, ϕ' are mutually exclusive, so A_ϕ and $A_{\phi'}$ are stochastically unrelated. This means that A is not a random variable.

Let there be a sequence of nonempty sets $\alpha = (\alpha^k : k \in K)$ such that $\Phi \subset \prod_{k \in K} \alpha^k$. This means that every treatment is a sequence $\phi = (x^k : k \in K)$, with $x^k \in \alpha^k$. The sets α^k are called *inputs*, and their elements x^k *input values*. Note that generally $\Phi \neq \prod_{k \in K} \alpha^k$, that is, not all possible combinations of input values form treatments (hence the adjective "allowable").

For every treatment ϕ, let the random variable A_ϕ be a sequence of jointly distributed random variables $A_\phi = \left(A_\phi^\ell : \ell \in L \right)$. For each ℓ, the sequence $A^\ell = \left(A_\phi^\ell : \phi \in \Phi \right)$ is called an *output*. Its element A_ϕ^ℓ can then be referred to as *output A^ℓ at treatment ϕ* (or simply output A_ϕ^ℓ, when this does not create confusion). Note that A^ℓ is not a random variable, because its components are stochastically unrelated.

We postulate that, for every input α^k and every output A^ℓ, either α^k *directly influences* A^ℓ, and we write $A^\ell \leftarrow \alpha^k$, or this is not true, $A^\ell \nleftarrow \alpha^k$. This relation is treated as primitive. Its intuitive meaning can be different in different applications. The only constraint imposed on this relation, *(complete) marginal selectivity*, is as follows (Dzhafarov, 2003). Let index subsets $I \subset K$ and $J \subset L$ be such that if $A^\ell \leftarrow \alpha^k$ for some $\ell \in J$ then $k \in I$. That is, no input belonging to $(\alpha^k : k \in K - I)$ directly influences any output belonging to $(A^\ell : \ell \in J)$. Let $\phi = (x^k : k \in K)$ and $\phi' = (y^k : k \in K)$ be any allowable treatments such that

$$\phi | I = (x^k : k \in I) = (y^k : k \in I) = \phi' | I. \tag{9}$$

The slash here indicates restriction of a function (sequence) on a subset of arguments (indices). Marginal selectivity means that under these assumptions

$$\left(A_\phi^k : k \in J \right) \sim \left(A_{\phi'}^k : k \in J \right), \tag{10}$$

where $\sim$ means "has the same distribution as." In other words, the joint distribution of a subset of outputs does not depend on inputs that do not directly influence any of these outputs. This does not mean, however, that these inputs, $\left(\alpha^k : k \in K - I\right)$, can be ignored altogether when dealing with $\left(A^\ell : \ell \in J\right)$: generally, this will not allow one to account for its stochastic relation to other outputs, $\left(A^\ell : \ell \in L - J\right)$.

By appropriately redefining the inputs the relation of "being directly influenced by" can always be made bijective: each output is directly influenced by one and only one input. The procedure is easier to illustrate on an example. Let the diagram of direct influences be

$$
\begin{array}{cccc}
\alpha^1 & \alpha^2 & \alpha^3 & \\
& & & \\
A^1 & A^2 & A^3 & A^4 \quad A^5
\end{array}
\tag{11}
$$
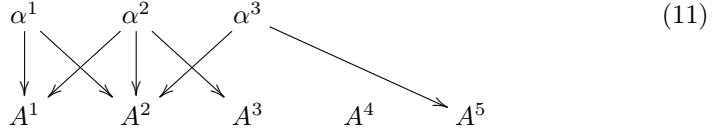
Assume, for simplicity, that all combinations of input values are allowable, $\Phi = \alpha^1 \times \alpha^2 \times \alpha^3$. Then the redefined inputs are as shown:

$$
\begin{array}{ccccc}
\beta^1 = \alpha^1 \times \alpha^2 & \beta^2 = \alpha^1 \times \alpha^2 \times \alpha^3 & \beta^3 = \alpha^2 & \beta^4 = \{.\} & \beta^5 = \alpha^3 \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
A^1 & A^2 & A^3 & A^4 & A^5
\end{array}
\tag{12}
$$

The set $\{.\}$ represents a dummy (single-valued) input, it should be paired with any output that is not directly influenced by any inputs. The rest of the redefinition should be clear. The set of allowable treatments is redefined into a new set Ψ, which is not the Cartesian product of the new inputs but rather a proper subsequence thereof: e.g., if β^2 attains the value $\left(x^1, x^2, x^3\right)$, then the only treatment allowable is

$$
\left(\left(x^1, x^2\right), \left(x^1, x^2, x^3\right), x^2, ., x^3\right).
\tag{13}
$$

We assume from now on that the direct influences are defined in a bijective form: $\alpha = \left(\alpha^k : k \in K\right)$, $\Phi \subset \prod_{k \in K} \alpha^k$, $A_\phi = \left(A_\phi^k : k \in K\right)$, $A^k \leftarrow \alpha^k$ for every $k \in K$, and there are no other direct influences.

Let us return to the sequence of random variables[2]

$$
A = (A_\phi : \phi \in \Phi) = \left(A_\phi^k : k \in K, \phi \in \Phi\right),
\tag{14}
$$

with stochastically unrelated components. Consider a *complete coupling* for A,

$$
H = \left(H_\phi^k : k \in K, \phi \in \Phi\right),
\tag{15}
$$

---

[2] In (14) and subsequently we are conveniently confusing differently grouped subsequences, such as (A, B, C), $((A, B), C)$, $(A, (B, C))$.

a random variable (that is, its components are jointly distributed) such that

$$H_\phi = \left(H_\phi^k : k \in K\right) \sim \left(A_\phi^k : k \in K\right) = A_\phi. \tag{16}$$

Such a random variable H always exists. It suffices, e.g., to consider every element of H_ϕ to be stochastically independent of every element in $H_{\phi'}$, for all $\phi \neq \phi'$. But generally, the complete couplings H for a given A can be chosen arbitrarily, except for the defining requirement (16).

Our approach consists in thinking of H, in addition to (16), in terms of "connections" it contains, by which we understand couplings for sequences of random variables that are indexed by different treatments sharing the same pattern of direct influences. Consider, e.g., the components A_ϕ^k for all ϕ whose kth element equals a given value $\phi(k) = x$. This subsequence can be written as

$$A_x^k = \left(A_\phi^k : \phi \in \Phi, \phi(k) = x\right). \tag{17}$$

Since $A^k \leftarrow \alpha^k$ only, all random variables A_ϕ^k are directly influenced by the same input value. Let

$$C_x^k = \left(C_\phi^k : \phi \in \Phi, \phi(k) = x\right) \tag{18}$$

be a coupling for $A_{x^k}^k$. This means that if $\phi(k) = x$,

$$C_\phi^k \sim A_\phi^k, \tag{19}$$

and it follows from the marginal selectivity property that the distribution of C_ϕ^k across all ϕ with $\phi(k) = x$ remains unchanged (and equal to the distribution of A_ϕ^k). There can be many joint distributions of (18) with this property. One possibility is that C_x^k is an *identity coupling*, meaning that for any two $C_\phi^k, C_{\phi'}^k$ in (18),

$$\Pr\left(C_\phi^k = C_{\phi'}^k\right) = 1. \tag{20}$$

If this is assumed for all $k \in K$ and $x \in \alpha^k$, then the complete coupling H in (15) can be written as the *reduced coupling*

$$R = \left(R_x^k : k \in K, x \in \alpha^k\right), \tag{21}$$

such that

$$R_\phi = \left(R_x^k : k \in K, \phi(k) = x\right) \sim A_\phi. \tag{22}$$

The existence of such a reduced coupling for a given A is the central theme of the theory of selective influences (Dzhafarov, 2003; Dzhafarov & Kujala, 2010, 2012a-b, 2013b, in press; Kujala & Dzhafarov, 2008; Schweickert, Fisher, & Sung, 2012, Ch. 10), which includes the Bell-type theorems as special cases. Using the language of the present paper, if R exists, one can say that each A^k is influenced only by the input α^k that directly influences it. In other words, there are no influences that are not direct ("no context"). Other examples from behavioral sciences involve recent work on combination of concepts (Aerts, Gabora, & Sozzo, in press; Bruza, Kitto, Ramm, & Sitbon, 2013; for a critical overview see Wang,

Busemeyer, Atmanspacher, & Pothos, in press, and Dzhafarov & Kujala, in press). In quantum physics the existence of the reduced coupling represents classical, pre-quantum determinism; it is the foundation of all Bell-type theorems (Basoalto & Percival, 2003; Dzhafarov & Kujala, 2012a).

We know, however, that Bell-type inequalities are violated in quantum physics. This leads us to explore alternatives to the assumption (20) and to the ensuing existence of a reduced coupling. This can be done by allowing C_x^k in (18) to be different from an identity coupling. The random variable C_x^k is called a *connection*. If its distribution is posited, we constrain the complete coupling (15) not just by (16), but also by its consistency with this connection:

$$H_x^k = \left(H_\phi^k : \phi \in \Phi, \phi\left(k\right) = x \right) \sim C_x^k. \tag{23}$$

With this additional constraint, the coupling H need not exist.

Generalizing, let I be a subset of K other than empty set and K itself. Then the (I, τ)-*connection* is defined as a random variable

$$C_\tau^I = \left(C_\phi^I : \phi \in \Phi, \phi | I = \tau \right) \tag{24}$$

such that for $\phi | I = \tau$,

$$C_\phi^I \sim A_\phi^I = \left(A_\phi^k : k \in I \right). \tag{25}$$

Recall that $\phi | I = \tau$ is the restriction of the treatment on a subset of its indices.[3] Note that the components of a given C_τ^I are jointly distributed, but if $(I, \tau) \neq (I', \tau')$, C_τ^I and $C_{\tau'}^{I'}$ are stochastically unrelated.

Given a sequence of outputs A in (14), denote the sequence of the connections C_τ^I for all I and τ by C_A (not a random variable). Assume that the distributions of all these connections are known. Then one can ask whether a complete coupling H for A is consistent with all connections in C_A, that is, whether in addition to (16) H also satisfies, for any $I \in 2^K - \{\emptyset, K\}$ and any $\tau \in \prod_{k \in I} \alpha^k$,

$$H_\tau^I = \left(H_\phi^I : \phi \in \Phi, \phi | I = \tau \right) \sim C_\tau^I, \tag{26}$$

where

$$H_\phi^I = \left(H_\phi^k : k \in I \right). \tag{27}$$

If this is true, then H is called an *Extended Joint Distribution Sequence* (EJDS) for (A, C_A). This notion is a generalization of the Joint Distribution Sequence (or "Joint Distribution Criterion set") that coincides with the reduced coupling (21) in the theory of selective influences (Dzhafarov & Kujala, 2010, 2012a, 2013b). It is obtained from EJDS by requiring that all connections be identity ones, that is, for any ϕ, ϕ' in (24),

$$\Pr\left(C_\phi^I = C_{\phi'}^I \right) = 1. \tag{28}$$

---

[3] Strictly speaking, this notation makes the upper index I in C_τ^I redundant. But it is convenient as it allows one to abridge the presentation of τ. Thus, if $K = \{1, 2, 3\}$, $I = \{1, 3\}$, $\phi(1) = x$, $\phi(3) = y$, then a strict reading of C_τ^I is $C_{\{(1,x),(3,y)\}}^{\{1,3\}}$, but it is naturally abridged into $C_{x,y}^{1,3}$, which seems more convenient than $C_{\{(1,x),(3,y)\}}$. Note that our opening example of a connection, C_x^k, is an abridged form of $C_{\{(k,x)\}}^{\{k\}}$.

4   Characterizing Contextuality

The notion of an EJDS can be used to characterize contextuality in relation to constraints imposed on the outputs of a system. Suppose that it is known that the outputs A taken across all allowable treatments in (14) satisfy a certain property $\mathcal{P}(A)$. This property may be described by certain equations and inequalities relating to each other parameters of the outputs, such as Bell-type inequalities, or Cirelson-Landau's quantum inequalities (see below). One should investigate then the set of possible C_A in relation to this property $\mathcal{P}(A)$.

To understand this better, let us consider a simple example of A. Let K be $\{1,2\}$, the sequence of inputs $\left(\alpha^k : k \in K\right)$ be $\left(\alpha^1 = \{1,2\}, \alpha^2 = \{1,2\}\right)$, the sequence of allowable treatments be $\Phi = \alpha^1 \times \alpha^2$, and the sequence of outputs be $A = \left(\left(A_{ij}^1, A_{ij}^2\right) : i,j \in \{1,2\}\right)$ (where each subscript ij represents the treatment (i,j)). The diagram of direct influences is assumed to be

$$\begin{array}{cc} \alpha^1 & \alpha^2 \\ \downarrow & \downarrow \\ A^1 & A^2 \end{array} \qquad (29)$$

The only choices of $I \subset K$ here other than $\emptyset$ and K are the singletons $\{1\}$ and $\{2\}$, so the only four connections are, for $i \in \{1,2\}$,

$$C_i^1 = \left(C_{i1}^1, C_{i2}^1\right), C_i^2 = \left(C_{1i}^2, C_{2i}^2\right), \qquad (30)$$

where $C_{ij}^k \sim A_{ij}^k$ for all $i,j,k \in \{1,2\}$. Recall that the logic of forming $C_i^1 = \left(C_{i1}^1, C_{i2}^1\right)$ is that A_{i1}^1 and A_{i2}^1, while they are recorded at different treatments, $(i,1)$ and $(i,2)$, share the same pattern of direct influences, namely, both are directly influenced by the value i of α^1 (in our general notation, $\phi|\{1\} = (i)$). So if their joint distribution is described by anything other than $\Pr\left(C_{i1}^1 = C_{i2}^1\right) = 1$, we can speak of indirect, contextual influences. The situation with C_i^2 is analogous. The complete coupling for A here is the 8-vector

$$H = \left(H_{ij}^1, H_{ij}^2 : i,j \in \{1,2\}\right). \qquad (31)$$

Assume that each A_{ij}^k (hence also H_{ij}^k in the complete coupling, $i,j,k \in \{1,2\}$) is a binary random variable with equiprobable outcomes $+1$ and -1. Then A is represented by four probabilities $p = (p_{11}, p_{12}, p_{21}, p_{22})$, where

$$p_{ij} = \Pr\left[A_{ij}^1 = +1, A_{ij}^2 = +1\right] = \Pr\left[H_{ij}^1 = +1, H_{ij}^2 = +1\right]. \qquad (32)$$

One prominent situation encompassed by this example is the Bohmian version of the EPR paradigm involving two spin-$1/2$ particles with two settings (spatial directions) per particle. As examples of a constraint $\mathcal{P}(A)$ consider the Bell/CH/Fine inequalities (Bell, 1964; Clauser & Horn, 1974; Fine, 1982)

$$0 \leq p_{ij} + p_{ij'} + p_{i'j'} - p_{i'j} \leq 1 \qquad (33)$$

and Cirel'son's (1980) inequalities

$$\frac{1-\sqrt{2}}{2} \le p_{ij} + p_{ij'} + p_{i'j'} - p_{i'j} \le \frac{1+\sqrt{2}}{2}, \qquad (34)$$

where $i, j \in \{1, 2\}$, $i' = 3 - i$, $j' = 3 - j$ (so each expression contains four double-inequalities). The Bell/CH/Fine inequalities are known to be necessary and sufficient for the existence of a classical explanation for the EPR paradigm in question (Fine, 1982), whereas the Cirel'son inequalities are necessary for the existence of a quantum mechanical explanation (Landau, 1987).

One question to pose about the connections is: what is the set of all C_A such that whenever $\mathcal{P}(A)$ is satisfied, an EJDS for (A, C_A) exists? We call any connection belonging to this C_A a *fitting connection* for $\mathcal{P}(A)$. A question can also be posed about the opposite implication: what is the set of all C_A such that whenever an EJDS for (A, C_A) exists, $\mathcal{P}(A)$ is satisfied? We call any connection in this C_A a *forcing connection* for $\mathcal{P}(A)$. In our example, C_A is the sequence of four connections C_i^k in (30), and they are uniquely characterized by the 4-vector $\varepsilon = \left(\varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^2, \varepsilon_2^2 \right)$, where

$$\varepsilon_i^1 = \Pr\left[C_{i1}^1 = +1, C_{i2}^1 = +1 \right], \varepsilon_i^2 = \Pr\left[C_{1i}^2 = +1, C_{2i}^2 = +1 \right]. \qquad (35)$$

Hence the complete coupling H, in order to be an EJDS for (A, C_A), should satisfy not only (32), but also

$$\Pr\left[H_{i1}^1 = +1, H_{i2}^1 = +1 \right] = \varepsilon_i^1, \Pr\left[C_{1i}^2 = +1, C_{2i}^2 = +1 \right] = \varepsilon_i^2, \qquad (36)$$

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

To describe the fitting and forcing connections for our example, it is convenient to introduce the following abbreviations:

$$s_0 = \max\left\{ \pm \left(\varepsilon_1^1 - 1/4 \right) \pm \left(\varepsilon_1^2 - 1/4 \right) \pm \left(\varepsilon_2^1 - 1/4 \right) \pm \left(\varepsilon_2^2 - 1/4 \right) : \# \text{ of } + \text{ signs is even} \right\},$$
$$s_1 = \max\left\{ \pm \left(\varepsilon_1^1 - 1/4 \right) \pm \left(\varepsilon_1^2 - 1/4 \right) \pm \left(\varepsilon_2^1 - 1/4 \right) \pm \left(\varepsilon_2^2 - 1/4 \right) : \# \text{ of } + \text{ signs is odd} \right\}. \qquad (37)$$

It turns out (Dzhafarov and Kujala, 2013a) that the sets of fitting connections for the Bell/CH/Fine and Cirel'son inequalities are described by, respectively,

$$s_1 \le 1/2, \qquad (38)$$

and

$$s_0 \le \frac{3 - \sqrt{2}}{2}, s_1 \le 1/2. \qquad (39)$$

This means that if p satisfies (33), then any ε with $s_1 \le 1/2$ is compatible with it, that is, this p and this ε can be embedded in the same EJDS H. If p satisfies (34), the set of ε compatible with it is more narrow: they should additionally satisfy $s_0 \le \frac{3-\sqrt{2}}{2}$. Both sets include, of course, the vector $\varepsilon = (0, 0, 0, 0)$, which represents no-contextuality and corresponds to the reduced coupling R in (21).

The sets of forcing connections for the Bell/CH/Fine and Cirel'son inequalities are described by, respectively,

$$s_0 = 1, \tag{40}$$

and

$$s_0 \geq \frac{3 - \sqrt{2}}{2}. \tag{41}$$

The set of ε such that $s_0 = 1$ consists of $\varepsilon = (0, 0, 0, 0)$, $\varepsilon = (1/2, 1/2, 1/2, 1/2)$, and vectors with two zeros and two $1/2$'s. All of them represent no-contextuality, with +1 and -1 interpreted differently in different connections. Only if ε is one of these vectors, p must satisfy the Bell/CH/Fine inequalities in order to be compatible with it. In other words, such an ε and no other "forces" p to satisfy these inequalities. The class of ε that force p to satisfy the Cirel'son inequalities should include these ε because every p satisfying (33) also satisfies (34). But there are other ε, all those with $s_0 \geq \frac{3-\sqrt{2}}{2}$, that too are compatible with p only if they satisfy the Cirel'son inequalities.

The above serves only as a demonstration of how one could characterize the constraints imposed on outputs (by a theory or empirical generalizations) through the connections compatible with them, in the sense of being embeddable in the same coupling. It should be noted, however, that connections generally do not characterize couplings uniquely. This opens ways for constructing qualified Kolmogorovian models more general than the one presented in this paper.

5   Conclusion

We have shown that the classical, if qualified, Kolmogorovian probability theory is not synonymous with the classical explanation of the input-output relations (especially, in the entanglement paradigm of quantum physics). The latter, since Bell's (1964) pioneering work, has been understood as the existence of a single sample space for all outputs when each output is indexed (identified) only by the inputs that directly influence it. In the qualified Kolmogorovian approach, however, this is only one of a potential infinity of possibilities. Different treatments (combinations of values of all inputs) correspond to stochastically unrelated random variables, and these can be coupled in many different ways. Only one of these ways, with identity connections, corresponds to John Bell's single sample space.

References

1. Aerts, D., Gabora, L., & Sozzo, S.: Concepts and their dynamics: A quantum-theoretic modeling of human thought. Topics in Cognitive Science (in press)

2. Basoalto, R.M., Percival, I.C.: BellTest and CHSH experiments with more than two settings. Journal of Physics A: Mathematical & General, 36, 7411–7423 (2003)
3. Bell, J.: On the Einstein-Podolsky-Rosen paradox. Physics 1, 195-200 (1964)
4. Bruza, P.D. , Kitto, K., Ramm, B.J., Sitbon. L.: A probabilistic framework for analysing the compositionality of conceptual combinations. arXiv:1305.5753 (2013)
5. Cirel'son, B.S.: Quantum generalizations of Bell's inequality. Letters in Mathematical Physics 4, 93–100 (1980)
6. Clauser, J.F., Horne, M.A.: Experimental consequences of objective local theories. Physical Review D 10, 526-535 (1974)
7. Dzhafarov, E.N.: Selective influence through conditional independence. Psychometrika 68, 7-26 (2003)
8. Dzhafarov, E.N., Kujala, J.V.: The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. Frontiers in Quantitative Psychology and Measurement 1:151 doi: 10.3389/fpsyg.2010.0015 (2010)
9. Dzhafarov, E.N., Kujala, J.V.: Selectivity in probabilistic causality: Where psychology runs into quantum physics. Journal of Mathematical Psychology 56, 54-63 (2012a)
10. Dzhafarov, E.N., Kujala, J.V.: Quantum entanglement and the issue of selective influences in psychology: An Overview. Lecture Notes in Computer Science 7620, 184-195 (2012b)
11. Dzhafarov, E.N., & Kujala, J.V.: All-possible-couplings approach to measuring probabilistic context. PLoS ONE 8(5): e61712. doi:10.1371/journal.pone.0061712 (2013a)
12. Dzhafarov, E.N., Kujala, J.V.: Order-distance and other metric-like functions on jointly distributed random variables. Proceedings of the American Mathematical Society, 141, 3291-3301. (2013b)
13. Dzhafarov, E.N., & Kujala, J.V.: On selective influences, marginal selectivity, and Bell/CHSH inequalities. Topics in Cognitive Science (in press)
14. Fine, A.: Hidden variables, joint probability, and the Bell inequalities. Physical Review Letters 48, 291-295 (1982)
15. Khrennikov, A.Yu.: Bell-Boole inequality: Nonlocality or probabilistic incompatibility of random variables? Entropy 10, 19-32 (2008a)
16. Khrennikov, A.Yu.: EPR–Bohm experiment and Bell's inequality: Quantum physics meets probability theory. Theoretical and Mathematical Physics 157, 1448–1460 (2008b)
17. Khrennikov, A.Yu.: Contextual Approach to Quantum Formalism. Berlin: Springer (2009)
18. Kujala, J.V., Dzhafarov, E.N.: Testing for selectivity in the dependence of random variables on external factors. Journal of Mathematical Psychology 52, 128–144 (2008)
19. Landau, L. J.: On the violation of Bell's inequality in quantum theory. Physical Letters A 120, 54–56 (1987)
20. Laudisa, F.: Contextualism and nonlocality in the algebra of EPR observables. Philosophy of Science 64, 478-496 (1997)
21. Schweickert, R., Fisher, D.L., & Sung, K.: Discovering Cognitive Architecture by Selectively Influencing Mental Processes. New Jersey: World Scientific (2012)
22. Thorisson, H.: Coupling, Stationarity, and Regeneration. New York: Springer (2000)
23. Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. (in press). The potential of using quantum theory to build models of cognition. Topics in Cognitive Science.