



Selectivity in probabilistic causality: Where psychology runs into quantum physics

Ehtibar N. Dzhafarov^{a,*}, Janne V. Kujala^b

^a Purdue University, USA

^b University of Jyväskylä, Finland

ARTICLE INFO

Article history:

Received 13 April 2011

Received in revised form

15 December 2011

Available online 23 January 2012

Keywords:

Bell-type inequalities

EPR paradigm

Factorial design

Fine's inequalities

Joint distribution criterion

Probabilistic causality

Mental architectures

Random outputs

Selective influences

Quantum entanglement

Thurstonian scaling

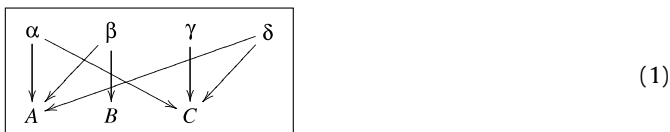
ABSTRACT

Given a set of several inputs into a system (e.g., independent variables characterizing stimuli) and a set of several stochastically non-independent outputs (e.g., random variables describing different aspects of responses), how can one determine, for each of the outputs, which of the inputs it is influenced by? The problem has applications ranging from modeling pairwise comparisons to reconstructing mental processing architectures to conjoint testing. A necessary and sufficient condition for a given pattern of selective influences is provided by the Joint Distribution Criterion, according to which the problem of “what influences what” is equivalent to that of the existence of a joint distribution for a certain set of random variables. For inputs and outputs with finite sets of values this criterion translates into a test of consistency of a certain system of linear equations and inequalities (Linear Feasibility Test) which can be performed by means of linear programming. While new in the behavioral context, both this test and the Joint Distribution Criterion on which it is based have been previously proposed in quantum physics, in dealing with generalizations of Bell inequalities for the quantum entanglement problem. The parallels between this problem and that of selective influences in behavioral sciences are established by observing that noncommuting measurements in quantum physics are mutually exclusive and can therefore be treated as different levels of one and the same factor.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

This paper deals with *diagrams of selective influences*, like this one:



The Greek letters in this diagram represent *inputs*, or *external factors*, e.g., parameters of stimuli whose values can be chosen at will or observed and recorded. The capital Roman letters stand for random outputs characterizing reactions of the system (an observer, a group of observers, stock market, a set of photons, etc.). The arrows show which factor influences which random output. The factors are treated as *deterministic* entities: even if α , β , γ , δ in reality vary randomly (e.g., being randomly generated

by a computer program, or being concomitant parameters of observations, such as age of respondents), for the purposes of analyzing selective influences the random outputs A , B , C are always viewed as *conditioned* upon various combinations of specific values of α , β , γ , δ . The first question to ask is: what is the meaning of the above diagram if the random outputs A , B , C in it are not necessarily stochastically independent? (If they are, the answer is of course trivial.) And once the meaning of the diagram of selective influences is established, how can one determine that this diagram correctly characterizes the dependence of the joint distributions of the random outputs A , B , C on the external factors α , β , γ , δ ?

These questions are important, because the assumption of stochastic independence of the outputs more often than not is either demonstrably false or adopted for expediency alone, with no other justification. At the same time the assumption of selectivity in causal relations between inputs and stochastic outputs is ubiquitous in theoretical modeling, often being built in the very language of the models. For instance, in Thurstone's most general model of pairwise comparisons (Thurstone, 1927) it is assumed that each of the two stimuli is mapped into “its” internal representation, while the two representations are stochastically interdependent random entities. In Dzhafarov (2003a), Dzhafarov and Gluhovsky (2006), and Kujala and Dzhafarov (2008) the

* Correspondence to: Purdue University, Department of Psychological Sciences, 703 Third Street West Lafayette, IN 47907, USA.

E-mail address: ehtibar@purdue.edu (E.N. Dzhafarov).

reader may find other motivating applications for the notion of selective influences: same–different comparisons, conjoint testing, parallel–serial networks of mental operations, response time decompositions, and all conceivable combinations of regression analysis and factor analysis. In this paper we add another motivating example, the quantum entanglement problem in quantum physics.

This paper continues and expands the analysis of selective influences presented in Dzhafarov and Kujala (2010). The familiarity with it can be helpful, but the main concepts, terminology, and notation are recapitulated in Section 2. Unlike in Dzhafarov and Kujala (2010), however, here we do not pursue the goal of maximal generality of formulations, focusing instead on the conceptual set-up that applies to commonly encountered experimental designs. This means a finite number of factors, each having a finite number of values. It also means that the random outcomes influenced by these factors are *random variables* in the narrow sense of the word: their values are vectors of real numbers or elements of countable sets, rather than more complex structures, such as functions or sets. This is done primarily to simplify and shorten exposition, and also because the *Linear Feasibility Test*, a new (for behavioral sciences) application of the Joint Distribution Criterion on which we focus in this paper (Section 3), is confined to finite sets of finite-valued factors and finite-valued random variables. This also allows us to emphasize a simple but important and previously overlooked proposition, *Theorem 2.3*, which essentially says that, when dealing with observable random variables, the unobservable random entities of the theory can also be assumed to be random variables (in the narrow sense). In another respect, however, the present treatment is more general than that in Dzhafarov and Kujala (2010): we allow for *incomplete designs*, those in which some but not necessarily all combinations of the values of the factors serve as allowable treatments. This modification is critical for the possibility of representing any diagram of selective influences, such as (1), in a *canonical form*, with every random output being selectively influenced by one and only one factor.

As it turns out, both the Linear Feasibility Test and the Joint Distribution Criterion on which it is based have their analogues in quantum physics.¹ To appreciate the analogy, however, one has to adopt the interpretation of noncommuting quantum measurements performed on a given component of a quantum-entangled system as mutually exclusive factor levels of the same factor. In Sections 2.6 and 3 we discuss the parallels between the existence of a classical explanation for an entanglement situation in quantum mechanics and the adherence of a behavioral experiment to a diagram of selective influences.

The term “test” in this paper is used in the meaning of necessary (sometimes necessary and sufficient) conditions for diagrams of selective influences. The usage is the same as when we speak of the tests for convergence in calculus or for divisibility in arithmetic. That is, the meaning of the term is non-statistical. We assume that random outputs are known on the population level. General considerations related to statistical tests based on our population level tests are discussed in Section 3.6, but specific statistical issues are outside the scope of this paper.

2. Basic notions

In this section, we establish the terminology, notation, and recapitulate basic facts related to factors, random variables, and the

dependence of the latter on the former. We follow Dzhafarov and Kujala (2010), adding observations related to the factorial designs being incomplete and random outputs being random variables in the narrow sense of the term. At the end of the section we discuss the parallels between the issue of selective influence in behavioral sciences and the quantum entanglement problem.

2.1. Factors, factor points, treatments

A *factor* α is treated as a set of *factor points*, each of which has the format “value (or level) x of factor α ”. In symbols, this can be presented as (x, α) , where ‘ α ’ is the unique name of the set α rather than the set itself. It is convenient to write x^α in place of (x, α) . Thus, if a factor with the name ‘*intensity*’ has three levels, ‘*low*’, ‘*medium*’, and ‘*high*’, then this factor is taken to be the set

$$\text{intensity} = \{ \text{low}^{\text{intensity}}, \text{medium}^{\text{intensity}}, \text{high}^{\text{intensity}} \}.$$

There is no circularity here, for, say, the factor point $\text{low}^{\text{intensity}}$ stands for (*value = low, name = ‘intensity’*) rather than (*value = low, set = intensity*).

We will deal with finite sets of factors $\Phi = \{\alpha_1, \dots, \alpha_m\}$, with each factor $\alpha \in \Phi$ consisting of a finite number of factor points,

$$\alpha = \{v_1^\alpha, \dots, v_{k_\alpha}^\alpha\}.$$

Clearly, $\alpha \cap \beta = \emptyset$ for any distinct $\alpha, \beta \in \Phi$.

A *treatment*, as usual, is defined as the set of factor points containing one factor point from each factor,

$$\phi = \{x_1^{\alpha_1}, \dots, x_m^{\alpha_m}\} \in \alpha_1 \times \dots \times \alpha_m.$$

The *set of treatments* (used in an experiment or considered in a theory) is denoted by $T \subset \alpha_1 \times \dots \times \alpha_m$ and assumed to be nonempty. Note that T need not include all possible combinations of factor points. This is an important consideration in view of the “canonical rearrangement” described below. Also, incompletely crossed designs occur broadly—in an experiment because the entire set $\alpha_1 \times \dots \times \alpha_m$ may be too large, or in a theory because certain combinations of factor points may be physically or logically impossible (e.g., contrast and shape cannot be completely crossed if zero is one of the values for contrast).

2.2. Random variables

We assume the reader is familiar with the notion of a random entity (random variable in the general sense of the term) A associated with an observation space (\mathcal{A}, Σ) , where \mathcal{A} is the set of possible values for A , and Σ a sigma-algebra (set of events) on \mathcal{A} . A *random variable* (in the narrow sense) is a special case of a random entity, defined as follows:

(i) if \mathcal{A} is countable, Σ is the power set of \mathcal{A} , then A is a random variable;

(ii) if \mathcal{A} is an interval of reals, Σ is the Lebesgue sigma-algebra on \mathcal{A} , then A is a random variable;

(iii) if A_1, \dots, A_n are random variables, then any jointly distributed vector (A_1, \dots, A_n) whose observation space is the conventionally understood product of the observations spaces for A_1, \dots, A_n is a random variable.

We use the relational symbol \sim in the meaning of “is distributed as”. $A \sim B$ is well defined irrespective of whether A and B are jointly distributed.

Let, for each treatment $\phi \in T$, there be a vector of jointly distributed random variables $A = (A_1, \dots, A_n)$ with a fixed (product) observation space and the probability measure μ_ϕ that depends

¹ We are grateful to Jerome Busemeyer of Indiana University who pointed out to us that the formulation of the Joint Distribution Criterion in our earlier work has the same formal structure as the identically titled criterion in Fine (1982a,b), in his analysis of quantum entanglement.

on ϕ .² Then we say that we have a *vector of jointly distributed random variables that depends on treatment ϕ* , and write

$$A(\phi) = (A_1, \dots, A_n)(\phi), \quad \phi \in T.$$

A correct way of thinking of $A(\phi)$ is that it represents a *set of vectors of jointly distributed random variables*, each of these vectors being labeled (indexed) by a particular treatment. Any sub-vector of $A(\phi)$ should also be written with the argument ϕ , say, $(A_1, A_2, A_3)(\phi)$. If ϕ is explicated as $\phi = \{x_1^{\alpha_1}, \dots, x_m^{\alpha_m}\}$, we write $A(\phi) = A(x_1^{\alpha_1}, \dots, x_m^{\alpha_m})$.

It is important to note that for distinct treatments ϕ_1 and ϕ_2 the corresponding $A(\phi_1)$ and $A(\phi_2)$ do not possess a joint distribution, they are *stochastically unrelated*. This is easy to understand: since ϕ_1 and ϕ_2 are mutually exclusive conditions for observing values of A , there is no non-arbitrary way of choosing which value $a = (a_1, \dots, a_n)$ observed at ϕ_1 should be paired with which value $a' = (a'_1, \dots, a'_n)$ observed at ϕ_2 . To consider $A(\phi_1)$ and $A(\phi_2)$ stochastically independent and to pair every possible value of $A(\phi_1)$ with every possible value $A(\phi_2)$ is as arbitrary as, say, to consider them positively correlated and to pair every quantile of $A(\phi_1)$ with the corresponding quantile of $A(\phi_2)$.

2.3. Arrow diagrams, canonically (re)arranged

Given a set of factors $\Phi = \{\alpha_1, \dots, \alpha_m\}$ and a vector $A(\phi) = (A_1, \dots, A_n)(\phi)$ of random variables depending on treatment, an *arrow diagram* is a mapping

$$M: \{1, \dots, n\} \rightarrow 2^\Phi \tag{2}$$

(2^Φ being the set of subsets of Φ). Later, in Definition 2.1, the arrows will be interpreted as indicating selective influences, but for now this is unimportant. The set

$$\Phi_i = M(i), \quad (i = 1, \dots, n),$$

is referred to as the subset of factors *corresponding to A_i* . It determines, for any treatment $\phi \in T$, the subtreatments ϕ_{Φ_i} defined as

$$\phi_{\Phi_i} = \{x^\alpha \in \phi : \alpha \in \Phi_i\}, \quad i = 1, \dots, n.$$

Subtreatments ϕ_{Φ_i} across all $\phi \in T$ can be viewed as *admissible values* of the subset of factors Φ_i ($i = 1, \dots, n$). Note that ϕ_{Φ_i} is empty whenever Φ_i is empty.

The simplest arrow diagram is *bijective*, with correspondences



We can simplify the subsequent discussion without sacrificing generality by agreeing to reduce each arrow diagram (in the context of selective influences) to a bijective form, by appropriately redefining factors and treatments. It is obvious how this should be done. Given the subsets of factors Φ_1, \dots, Φ_n determined by an arrow diagram (2), each Φ_i can be viewed as a factor identified with the set of factor points

$$\alpha_i^* = \{(\phi_{\Phi_i})^{\alpha_i^*} : \phi \in T\},$$

in accordance with the notation we have adopted for factor points: $(\phi_{\Phi_i})^{\alpha_i^*} = (\phi_{\Phi_i}, \alpha_i^*)$. If Φ_i is empty, then ϕ_{Φ_i} is empty too, and

² The convenient assumption of the invariance of the observation space for A with respect to ϕ is innocuous: one can always redefine the observation spaces for different treatments ϕ to make them coincide.

the factor α_i^* consists of only the dummy factor point \emptyset^{α_i} (where \emptyset denotes the empty set). The set of treatments T for the original factors $\{\alpha_1, \dots, \alpha_m\}$ should then be redefined for the vector of new factors $(\alpha_1^*, \dots, \alpha_n^*)$ as

$$T^* = \left\{ \left\{ (\phi_{\Phi_1})^{\alpha_1^*}, \dots, (\phi_{\Phi_n})^{\alpha_n^*} \right\} : \phi \in T \right\} \subset \alpha_1^* \times \dots \times \alpha_n^*.$$

We call this (re)definition of factor points, factors, and treatments the *canonical (re)arrangement*. We can say that the random variables following canonical (re)arrangement can be indexed by the corresponding factors. Thus, when convenient, we can write in (3) $A_{\{\alpha_1\}}$ in place of A_1 , $A_{\{\alpha_2\}}$ in place of A_2 , etc. The notation $\phi_{\Phi_i} = \phi_{\{\alpha_i\}}$ then indicates the singleton set $\{x^{\alpha_i}\} \subset \phi$. As usual, we write x^{α_i} in place of $\{x^{\alpha_i}\}$:

$$\phi_{\{\alpha_i\}} = \{x_1^{\alpha_1}, \dots, x_n^{\alpha_n}\}_{\{\alpha_i\}} = x_i^{\alpha_i}.$$

2.4. The criterion

Definition 2.1 (*Selective Influences, Bijective Form*). An arrow diagram (3) is said to be the diagram of selective influences for $(A_1, \dots, A_n)(\phi)$ and $(\alpha_1, \dots, \alpha_n)$, and we write

$$(A_1, \dots, A_n) \leftrightarrow (\alpha_1, \dots, \alpha_n),$$

if, for some random entity R and for any treatment $\phi = \{x_1^{\alpha_1}, \dots, x_n^{\alpha_n}\} \in T$,

$$\begin{aligned} (A_1, \dots, A_n)(\phi) &\sim (f_1(\phi_{\{\alpha_1\}}, R), \dots, f_n(\phi_{\{\alpha_n\}}, R)) \\ &= (f_1(x_1^{\alpha_1}, R), \dots, f_n(x_n^{\alpha_n}, R)), \end{aligned} \tag{4}$$

where $f_i: \alpha_i \times \mathcal{R} \rightarrow \mathcal{A}_i$ ($i = 1, \dots, n$) are some functions, with \mathcal{R} denoting the set of possible values of R .³

This definition is difficult to put to work, as it refers to an existence of a random entity (variable) R without showing how one can find it or prove that it cannot be found. The following criterion (necessary and sufficient condition) for $(A_1, \dots, A_n) \leftrightarrow (\alpha_1, \dots, \alpha_n)$ circumvents this problem.

Criterion 2.2 (*Joint Distribution Criterion, JDC*). A vector of random variables $A(\phi) = (A_1, \dots, A_n)(\phi)$ satisfies a diagram of selective influences (3) if and only if there is a vector of jointly distributed random variables

$$H = \left(\overbrace{H_{x_1^{\alpha_1}}^{\alpha_1}, \dots, H_{x_{k_1}^{\alpha_1}}^{\alpha_1}}^{\text{for } \alpha_1}, \dots, \overbrace{H_{x_1^{\alpha_n}}^{\alpha_n}, \dots, H_{x_{k_n}^{\alpha_n}}^{\alpha_n}}^{\text{for } \alpha_n} \right),$$

one random variable for each factor point of each factor, such that

$$(H_{\phi_{\{\alpha_1\}}}, \dots, H_{\phi_{\{\alpha_n\}}}) \sim A(\phi) \tag{5}$$

for every treatment $\phi \in T$.

See Dzhaferov and Kujala (2010) for a proof. The vector H in the formulation of the JDC is referred to as the *JDC-vector for $A(\phi)$* , or the *hypothetical JDC-vector for $A(\phi)$* , if the existence of such a vector of jointly distributed variables is in question.

The JDC prompts a simple justification for our definition of selective influences. Let, for example, $(A, B, C) \leftrightarrow (\alpha, \beta, \gamma)$, with $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta, 3^\beta\}$, $\gamma = \{1^\gamma, 2^\gamma, 3^\gamma, 4^\gamma\}$. Consider

³ It will be shown below, Theorem 2.3, that random entity R can always be chosen to be a random variable (in the narrow sense).

all treatments ϕ in which the factor point of α is fixed, say, at 1^α . If $(A, B, C) \leftrightarrow (\alpha, \beta, \gamma)$, then in the vectors of random variables

$$(A, B, C) (1^\alpha, 2^\beta, 1^\gamma), (A, B, C) (1^\alpha, 2^\beta, 3^\gamma), (A, B, C) (1^\alpha, 3^\beta, 1^\gamma)$$

the marginal distribution of the variable A is one and the same,

$$A (1^\alpha, 2^\beta, 1^\gamma) \sim A (1^\alpha, 2^\beta, 3^\gamma) \sim A (1^\alpha, 3^\beta, 1^\gamma).$$

But the intuition of selective influences requires more: that we can denote this variable $A(1^\alpha)$ because it *preserves its identity* (and not just its distribution) no matter what other variables it is paired with, $(B, C) (2^\beta, 1^\gamma)$, $(B, C) (2^\beta, 3^\gamma)$, or $(B, C) (3^\beta, 1^\gamma)$. Analogous statements hold for $A(2^\alpha)$, $B(2^\beta)$, $B(3^\beta)$, $C(1^\gamma)$, etc. The JDC formalizes the intuitive notion of variables “preserving their identity” when entering in various combinations with each other: there are jointly distributed random variables

$$H_{1\alpha}, H_{2\alpha}, H_{1\beta}, H_{2\beta}, H_{3\beta}, H_{1\gamma}, H_{2\gamma}, H_{3\gamma}, H_{4\gamma}$$

whose identity is defined by this joint distribution; when $H_{1\alpha}$ is combined with random variables $H_{2\beta}$ and $H_{1\gamma}$, it forms the triad $(H_{1\alpha}, H_{2\beta}, H_{1\gamma})$ whose distribution is the same as that of $(A, B, C) (1^\alpha, 2^\beta, 1^\gamma)$; when the same random variable $H_{1\alpha}$ is combined with random variables $H_{2\beta}$ and $H_{3\gamma}$, the triad $(H_{1\alpha}, H_{2\beta}, H_{3\gamma})$ is distributed as $(A, B, C) (1^\alpha, 2^\beta, 3^\gamma)$; and so on—the key concept being that it is *one and the same* $H_{1\alpha}$ which is being paired with other variables, as opposed to different random variables $A(1^\alpha, 2^\beta, 1^\gamma)$, $A(1^\alpha, 2^\beta, 3^\gamma)$, $A(1^\alpha, 3^\beta, 1^\gamma)$ which are identically distributed. See Dzhafarov and Kujala (2010) for a demonstration that the identity is not generally preserved if all we know is marginal selectivity (as defined in Section 2.5).

The following is an important consequence of JDC.

Theorem 2.3. In Definition 2.1, the random entity R can always be chosen to be a random variable. Moreover, R can be chosen arbitrarily, as any continuously (atomlessly) distributed random variable, e.g., uniformly distributed between 0 and 1.

Proof. The first statement follows from the fact that R can be chosen to coincide with the JDC-vector H , so that

$$f_i(x^{\alpha_i}, H) = H_{x^{\alpha_i}},$$

for $i = 1, \dots, n$, and $x^{\alpha_i} \in \alpha_i$. The JDC-vector H is a random variable. The second statement follows from Theorem 1 in Dzhafarov and Gluhovsky (2006), based on a general result for standard Borel spaces (e.g., in Kechris (1995, p. 116)). \square

2.5. Three basic properties of selective influences

For completeness, we list three other fundamental consequences of JDC (Dzhafarov & Kujala, 2010).

2.5.1. Nestedness

For any subset $\{i_1, \dots, i_k\}$ of $\{1, \dots, n\}$, if $(A_1, \dots, A_n) \leftrightarrow (\alpha_1, \dots, \alpha_n)$ then $(A_{i_1}, \dots, A_{i_k}) \leftrightarrow (\alpha_{i_1}, \dots, \alpha_{i_k})$.

2.5.2. Complete marginal selectivity

For any subset $\{i_1, \dots, i_k\}$ of $\{1, \dots, n\}$, if $(A_1, \dots, A_n) \leftrightarrow (\alpha_1, \dots, \alpha_n)$ then the k -marginal distribution⁴ of $(A_{i_1}, \dots, A_{i_k})(\phi)$ does not depend on points of the factors outside $(\alpha_{i_1}, \dots, \alpha_{i_k})$.

⁴ k -marginal distribution is the distribution of a subset of k random variables ($k \geq 1$) in a set of $n \geq k$ variables. In Townsend and Schweickert (1989) the property was formulated for 1-marginals of a pair of random variables. The adjective “complete” we use with “marginal selectivity” is to emphasize that we deal with all possible marginals rather than with just 1-marginals.

In particular, the distribution of A_i only depends on points of α_i , $i = 1, \dots, n$.

This is, of course, a trivial consequence of the nestedness property, but its importance lies in that it provides the easiest to check necessary condition for selective influences.

2.5.3. Invariance under factor-point-specific transformations

Let $(A_1, \dots, A_n) \leftrightarrow (\alpha_1, \dots, \alpha_n)$ and

$$H = \left(H_{x_1^{\alpha_1}}, \dots, H_{x_{k_1}^{\alpha_1}}, \dots, H_{x_1^{\alpha_n}}, \dots, H_{x_{k_n}^{\alpha_n}} \right)$$

be the JDC-vector for $(A_1, \dots, A_n)(\phi)$. Let F be any function that applies to H componentwise and produces a corresponding vector of random variables

$$F(H) = \left(F(x_1^{\alpha_1}, H_{x_1^{\alpha_1}}), \dots, F(x_{k_1}^{\alpha_1}, H_{x_{k_1}^{\alpha_1}}), \dots, F(x_1^{\alpha_n}, H_{x_1^{\alpha_n}}), \dots, F(x_{k_n}^{\alpha_n}, H_{x_{k_n}^{\alpha_n}}) \right),$$

where we denote by $F(x^\alpha, \cdot)$ the application of F to the component labeled by x^α . Clearly, $F(H)$ possesses a joint distribution and contains one component for each factor point. If we now define a vector of random variables $B(\phi)$ for every treatment $\phi \in T$ as

$$(B_1, \dots, B_n)(\phi) = (F(\phi_{\alpha_1}, A_1), \dots, F(\phi_{\alpha_n}, A_n))(\phi),$$

then it follows from JDC that $(B_1, \dots, B_n) \leftrightarrow (\alpha_1, \dots, \alpha_n)$.⁵ A function $F(x^{\alpha_i}, \cdot)$ can be referred to as a *factor-point-specific transformation* of the random variable A_i , because the random variable is transformed differently for different points of the factor assumed to selectively influence it. We can formulate the property in question by saying that a diagram of selective influences is invariant under all factor-point-specific transformations of the random variables. Note that this includes as a special case transformations which are not factor-point-specific, with

$$F(x_1^{\alpha_i}, \cdot) \equiv \dots \equiv F(x_{k_i}^{\alpha_i}, \cdot) \equiv F(\alpha_i, \cdot).$$

This property is important for construction and use of tests for selective influences (Dzhafarov & Kujala, 2010; Kujala & Dzhafarov, 2008).

2.6. Quantum entanglement and selective influences

In psychology, the notion of selective influences was introduced by Sternberg (1969), in the context of studying “stages” of information processing. Sternberg acknowledged that selective influences can hold even if the durations of the stages being selectively affected are not stochastically independent, but he lacked the mathematical apparatus for dealing with this possibility. Townsend (1984) was the first to study the notion of selectiveness under stochastic interdependence systematically. He proposed to formalize the notion of selectively influenced and stochastically interdependent random variables by the concept of “indirect nonselectiveness”: the conditional distribution of the variable A_1 given any value a_2 of the variable A_2 , depends on α_1 only, and, by symmetry, the conditional distribution of A_2 at any $A_1 = a_1$ depends on α_2 only. Under the name of “conditionally selective influence” this notion was mathematically characterized and generalized in Dzhafarov (1999). It turned out, however, that this notion could not serve as a general definition of selective influences,

⁵ Since it is possible that $F(x^\alpha, H_{x^\alpha})$ and $F(y^\alpha, H_{y^\alpha})$, with $x^\alpha \neq y^\alpha$, have different sets of possible values, strictly speaking, one may need to redefine the functions to ensure that the sets of possible values for $B(\phi)$ is the same for different ϕ . This is, however, not essential (see footnote 2).

because it did not satisfy some intuitive desiderata for such a definition, e.g., the nestedness and marginal selectivity properties formulated in Section 2.5. Variants of Definition 2.1 of the present paper were proposed in Dzhafarov (2003a) and both elaborated and generalized in Dzhafarov and Gluhovsky (2006), Kujala and Dzhafarov (2008); JDC was explicitly formulated in Dzhafarov and Kujala (2010), although clearly implied in the earlier work.

Until very recently (see footnote 1) we were blissfully unaware of the analogous developments in quantum physics. The most conspicuous parallels can be found in Fine (1982a,b), but that work in turn builds on a venerable line of research and thinking: going back first to Bell (1964), and ultimately to Einstein, Podolsky, and Rosen's (1935) paper. The issue in question regards two "noncommuting" measurements, such as those of the momentum and of the location of a particle, or spin measurements along two different axes. For our purposes it is sufficient to state that when one of two noncommuting measurements is performed (without uncertainty about the result), the second one cannot be performed on the same system. The key insight needed to understand the analogy with the problem of selective influences is this: *noncommuting measurements on the same system, being mutually exclusive, can be viewed as levels (mutually exclusive values) of one and the same external factor.*

This is not entirely intuitive. Consider two particles for each of which one can measure its momentum or its location. The analogy requires that one view the measurement on particle 1 as a factor α_1 with two mutually exclusive levels, 1^{α_1} (location measurement) and 2^{α_1} (momentum measurement); and the measurement on particle 2 is a factor α_2 with two mutually exclusive levels, 1^{α_2} and 2^{α_2} , interpreted analogously. The two measurements can be combined in treatments, $(1^{\alpha_1}, 1^{\alpha_2})$, $(1^{\alpha_1}, 2^{\alpha_2})$, etc., but not within a factor, $(1^{\alpha_1}, 2^{\alpha_1})$ or $(1^{\alpha_2}, 2^{\alpha_2})$. The results of each of the measurements is a random variable, A_1 for particle 1 and A_2 for particle 2. The possible values \mathcal{A}_1 for A_1 are possible locations of particle 1 if α_1 is at level 1^{α_1} , but they are possible momentum values for particle 1 if α_1 is at level 2^{α_1} (which makes it awkward but still possible to maintain the convention mentioned in footnote 2). It is easier with spins (Bohm & Aharonov, 1957): for instance, for spin-1/2 particles (such as electrons), \mathcal{A}_1 consists of two possible values of spin in one direction if α_1 is at level 1^{α_1} and of two possible values of spin in another direction if the level is 2^{α_1} . These two two-element sets are more natural to consider "the same".

With all this in mind, the question now can be posed in the familiar to us form: can we say that $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$, or can the measurement (factor) α_1 influence the result (random variable) A_2 and/or α_2 influence A_1 ? In the Einstein–Podolsky–Rosen (EPR) paradigm involving *entangled* particles, the two random outcomes A_1, A_2 are stochastically interdependent, and their joint distribution at every treatment is (correctly) predicted by the quantum theory. The question therefore becomes: are the predicted (and observed) joint distributions of (A_1, A_2) compatible with the hypothesis $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$? Einstein et al. (1935) took $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$ for granted if the two particles are separated in space and measured simultaneously (in some inertial frame of reference).

Bell's (1964) celebrated theorem shows that $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$ is not the case for entangled spin-1/2 particles obeying the laws of quantum mechanics. The reason this result is considered to be of foundational importance ("the most profound discovery in science", repeating the oft-quoted characterization by Stapp (1975)) is that Bell essentially adopted Definition 2.1 for $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$ and identified the random entity R with the set of all hidden variables of a conceivable theory "explaining" the dependence of (A_1, A_2) on (α_1, α_2) : knowing a value of R one would be able to predict, through the functions f_1 and f_2 of Definition 2.1, the values of (A_1, A_2) . In addition to being called "hidden" the variables entailed in R are referred to as "context-independent"

(meaning that the distribution of R and the functions f_1, f_2 do not depend on treatments) and "local" (meaning, essentially, that in the theory involving R and f_1, f_2 the measurement α_1 does not influence A_2 , nor α_2 influences A_1). Bell's (1964) theorem therefore is interpreted as stating that quantum predictions regarding two entangled spin-1/2 particles cannot be explained by any theory involving context-independent and local variables.

A rejection of $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$ in quantum physics can be handled by dispensing with locality (Bohm's approach), but most physicists find this untenable (measurement α_1 cannot influence A_2 if they are separated by a space-like interval). The quantum probability theory can be viewed as a way of allowing for context-dependence while retaining locality. In behavioral applications both locality and context-independence can be targeted when $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$ is rejected, and distinguishing the two is a challenge.

Following the logic of Bell's work, Clauser, Horne, Shimony, and Holt (1969) derived a system of inequalities that are necessary conditions for $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$ in the EPR paradigm with two particles and two measurements (factors) with binary outcomes. These inequalities are subsumed in Fine's (1982a,b) ones (discussed in Section 3.5), which present both necessary and sufficient conditions for $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$, based on JDC. The latter was introduced in Fine's papers for the first time (and called by this name too), although the earlier Suppes and Zanotti (1981) Theorem on Common Causes can also be viewed as a special form of JDC.

Fine's inequalities form a special case of the Linear Feasibility Test considered in the next section. We therefore defer further discussion of the EPR paradigm to Section 3.5, and conclude the present section by the following table of correspondences:

Selective probabilistic causality	Quantum entanglement problem (for spins)
Observed random output	Detected spin value of a given particle
Factor/input	Spin measurement in a given particle
Factor level	Setting (axis) of the spin measurement
Joint distribution criterion	Joint distribution criterion
Canonical diagram of selective influences	"Classical" explanation (by context-independent local variables)

3. Linear feasibility test

In this section we assume that for each random variable $A_i(\phi)$ in $(A_1, \dots, A_n)(\phi)$ the set \mathcal{A}_i of its possible values has m_i elements, $a_1^i, \dots, a_{m_i}^i$. It is arguably the most important special case both because it is ubiquitous and because in all other cases random variables can be discretized into finite number of categories. We are interested in establishing the truth or falsity of the diagram of selective influences (3), where each factor α_j in $(\alpha_1, \dots, \alpha_n)$ contains k_j factor points $x_1^j, \dots, x_{k_j}^j$ (written so instead of more formal $x_1^{\alpha_j}, \dots, x_{k_j}^{\alpha_j}$). The *Linear Feasibility Test* (LFT) to be described is a direct application of JDC to this situation, furnishing a necessary and sufficient condition for the diagram of selective influences $(A_1, \dots, A_n) \leftrightarrow (\alpha_1, \dots, \alpha_n)$.

3.1. The test

In the hypothetical JDC-vector

$$H = (H_{x_1^1}, \dots, H_{x_{k_1}^1}, \dots, H_{x_1^n}, \dots, H_{x_{k_n}^n}),$$

since we assume that, for any point x_j^i of factor α^i and any treatment ϕ containing x_j^i ,

$$H_{x_j^i} \sim A_i(\phi),$$

we know that the set of possible values for the random variable $H_{x_j^i}$ is $\{a_1^i, \dots, a_{m_i}^i\}$, irrespective of j . Denote

$$\Pr[(A_1 = a_{l_1}^1, \dots, A_n = a_{l_n}^n)(x_{j_1}^1, \dots, x_{j_n}^n)] = P \left(\begin{matrix} \text{for r.v.s} & \text{for factor points} \\ l_1, \dots, l_n; & j_1, \dots, j_n \end{matrix} \right), \tag{6}$$

where $l_i \in \{1, \dots, m_i\}$ and $j_i \in \{1, \dots, k_i\}$ for $i = 1, \dots, n$ (“r.v.s” abbreviates “random variables”). Denote

$$\Pr \left[\begin{matrix} H_{x_1^1} = a_{l_{11}}^1, \dots, H_{x_{k_1}^1} = a_{l_{1k_1}}^1 \\ \vdots \\ H_{x_1^n} = a_{l_{n1}}^n, \dots, H_{x_{k_n}^n} = a_{l_{nk_n}}^n \end{matrix} \right] = Q \left(\begin{matrix} \text{for } A_1 & \text{for } A_n \\ l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n} \end{matrix} \right), \tag{7}$$

where $l_{ij} \in \{1, \dots, m_i\}$ for $i = 1, \dots, n$. This gives us $m_1^{k_1} \times \dots \times m_n^{k_n}$ Q -probabilities. A required joint distribution for the JDC-vector H exists if and only if these probabilities can be found subject to $m_1^{k_1} \times \dots \times m_n^{k_n}$ nonnegativity constraints

$$Q(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) \geq 0, \tag{8}$$

and (denoting by n_T the number of treatments in T) $n_T \times m_1 \times \dots \times m_n$ linear equations

$$\sum Q(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) = P(l_1, \dots, l_n; j_1, \dots, j_n), \tag{9}$$

where the summation is across all possible values of the $(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n})$ subject to

$$l_{1j_1} = l_1, \dots, l_{nj_n} = l_n. \tag{6}$$

This can be more compactly formulated in a matrix form. Let the observable probabilities $P(l_1, \dots, l_n; j_1, \dots, j_n)$ constitute components of a $n_T \times m_1 \times \dots \times m_n$ -dimensional column vector \mathbf{P} , with its cells lexicographically enumerated by $(l_1, \dots, l_n; j_1, \dots, j_n)$. Let the hypothetical probabilities $Q(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n})$ constitute components of a $m_1^{k_1} \times \dots \times m_n^{k_n}$ -dimensional column vector \mathbf{Q} , with its cells lexicographically enumerated by $(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n})$. Let \mathbf{M} be a Boolean matrix with $n_T \times m_1 \times \dots \times m_n$ rows and $m_1^{k_1} \times \dots \times m_n^{k_n}$ columns lexicographically enumerated in the same way as, respectively, \mathbf{P} and \mathbf{Q} , such that the entry in the cell in the $(l_1, \dots, l_n; j_1, \dots, j_n)$ th row and $(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n})$ th column is 1 if $l_{1j_1} = l_1, \dots, l_{nj_n} = l_n$; otherwise the entry is 0. Clearly, the vector \mathbf{Q} exists if and only if the system

$$\mathbf{MQ} = \mathbf{P}, \quad \mathbf{Q} \geq 0 \tag{10}$$

⁶ The sum of all Q 's is 1 because it equals the sum of all P 's (across all l_1, \dots, l_n) for any given treatment j_1, \dots, j_n .

(with the inequality understood componentwise) has a solution. This is a typical *linear programming* (LP) problem. More precisely, this is an LP task in the standard form and with a dummy objective function (e.g., a linear combination with zero coefficients). It is known (Karmarkar, 1984; Khachiyan, 1979) that it is always possible, in polynomial time, to either find a solution for such a system or to determine that it does not exist. Many standard software packages can handle this problem (e.g., GNU Linear Programming Kit at <http://www.gnu.org/software/glpk/>).

3.2. Properties of the LP problem

The rank of matrix \mathbf{M} is always strictly smaller than the number of components in \mathbf{P} . This follows from the fact that for any two allowable treatments (j_1, \dots, j_n) and (j'_1, \dots, j'_n) that share a subvector

$$(j_{1'}', \dots, j_{s'}') = (j'_{1'}, \dots, j'_{s'})$$

(where we use $\{1', \dots, s'\}$ to designate s distinct elements chosen from $\{1, \dots, n\}$), and for any fixed (v_1, \dots, v_s) , the sum of all rows of \mathbf{M} corresponding to $(l_1, \dots, l_n; j_1, \dots, j_n)$ th components of \mathbf{P} with $(l_{1'}', \dots, l_{s'}') = (v_1, \dots, v_s)$ is the same Boolean vector as the sum of all rows of \mathbf{M} corresponding to $(l_1, \dots, l_n; j'_1, \dots, j'_n)$ th components of \mathbf{P} with the same property. The upper limit for the rank of matrix \mathbf{M} is given in the following theorem.

Theorem 3.1. The rank of \mathbf{M} for a maximal set of treatments $T = \alpha_1 \times \dots \times \alpha_n$ is

$$(k_1(m_1 - 1) + 1) \dots (k_n(m_n - 1) + 1).$$

Proof. Given any

$$\begin{aligned} \{1', \dots, s'\} &\subset \{1, \dots, n\}, \\ (j_{1'}', \dots, j_{s'}') &\in \{1, \dots, k_{1'}'\} \times \dots \times \{1, \dots, k_{s'}'\}, \\ (l_{1'}', \dots, l_{s'}') &\in \{1, \dots, m_{1'}'\} \times \dots \times \{1, \dots, m_{s'}'\}, \end{aligned}$$

let $\mathbf{V}(1', \dots, s'; j_{1'}', \dots, j_{s'}'; l_{1'}', \dots, l_{s'}')$ denote an $(m_1)^{k_1} \dots (m_n)^{k_n}$ -component Boolean row vector whose components are lexicographically enumerated in the same way as \mathbf{Q} , and such that its $(l_{11}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n})$ th component is 1 if and only if

$$l_{1j_{1'}'} = l_{1'}', \dots, l_{s'j_{s'}'} = l_{s'}'.$$

The rows of matrix \mathbf{M} are $\mathbf{V}(1, \dots, n; j_1, \dots, j_n; l_1, \dots, l_n)$ -vectors. It is easy to check that for any fixed $(1', \dots, s'; j_{1'}', \dots, j_{s'}')$, the sum of the rows of \mathbf{M} corresponding to fixed values $(l_{1'}', \dots, l_{s'}')$ is $\mathbf{V}(1', \dots, s'; j_{1'}', \dots, j_{s'}'; l_{1'}', \dots, l_{s'}')$. It follows that for $s = n, n - 1, \dots, 1$, a vector $\mathbf{V}(1', \dots, s'; j_{1'}', \dots, j_{s'}'; l_{1'}', \dots, l_{s'}')$ in which all $l_{i'}' = 1$ except for $i' \in \{1'', \dots, v''\} \subset \{1', \dots, s'\}$ (a subset of $v < s$ distinct elements), is a linear combination of the vector

$$\mathbf{V}(1'', \dots, v''; j_{1''}', \dots, j_{v''}'; l_{1''}', \dots, l_{v''}')$$

and all the vectors

$$\mathbf{V}(1', \dots, s'; j_{1'}', \dots, j_{s'}'; l_{1'}', \dots, l_{s'}')$$

for which all $l_{i'}' > 1$ and

$$\{j_{1''}', \dots, j_{v''}'; l_{1''}', \dots, l_{v''}'\} \subset \{j_{1'}', \dots, j_{s'}'; l_{1'}', \dots, l_{s'}'\}.$$

As a result the rows of \mathbf{M} are linear combinations of the rows of \mathbf{M}^* consisting of vectors

$$\mathbf{V}(1', \dots, s'; j_{1'}', \dots, j_{s'}'; l_{1'}', \dots, l_{s'}')$$

for all possible

$$\begin{aligned} \{1', \dots, s'\} &\subset \{1, \dots, n\}, \\ (j_{1'}', \dots, j_{s'}') &\in \{1, \dots, k_{1'}'\} \times \dots \times \{1, \dots, k_{s'}'\}, \\ (l_{1'}', \dots, l_{s'}') &\in \{2, \dots, m_{1'}'\} \times \dots \times \{2, \dots, m_{s'}'\}. \end{aligned}$$

By straightforward combinatorics the number of such vectors is $(k_1(m_1 - 1) + 1) \cdots (k_n(m_n - 1) + 1)$.

The rows of \mathbf{M}^* are linearly independent because the column corresponding to the $(l_{11} = 1, \dots, l_{1k_1} = 1, \dots, l_{n1} = 1, \dots, l_{nk_n} = 1)$ th component of \mathbf{Q} contains a single 1, in the row of \mathbf{M}^* corresponding to $s = 0$ (which row contains 1's only). \square

Note that

$$k_i(m_i - 1) + 1 < m_i^{k_i}$$

for all $k_i \geq 2$ and $m_i \geq 1$. This means that

$$(k_1(m_1 - 1) + 1) \cdots (k_n(m_n - 1) + 1) < (m_1)^{k_1} \cdots (m_n)^{k_n},$$

and the system $\mathbf{M}\mathbf{Q} = \mathbf{P}$ is always underdetermined.

Corollary 3.2. *If \mathbf{P} satisfies marginal selectivity, then system (10) is equivalent to*

$$\mathbf{M}^*\mathbf{Q} = \mathbf{P}^*, \quad \mathbf{Q} \geq 0, \tag{11}$$

where \mathbf{M}^* is as defined in the proof above, and \mathbf{P}^* is the “reduced hierarchical” vector with components

$$\Pr \left[\left(A_{l_{i'}} = a_{l_{i'}}^{1'}, \dots, A_{l_{i'}} = a_{l_{i'}}^{s'} \right) \left(x_{j_{i'}}^{1'}, \dots, x_{j_{i'}}^{s'} \right) \right] = P_{l_{i'}, \dots, l_{i'}}^* (l_{i'}, \dots, l_{i'}; j_{i'}, \dots, j_{i'}), \tag{12}$$

where $s = 0, \dots, n, \{1', \dots, s'\} \subset \{1, \dots, n\}$, and $l_{i'} \in \{2, \dots, m_i\}$ for each $i' \in \{1', \dots, s'\}$. \mathbf{M}^* is of full row rank.

To comment on this corollary, it follows from the proof of Theorem 3.1 that $\mathbf{M}\mathbf{Q} = \mathbf{P}$ never has a solution if vector \mathbf{P} violates the equality

$$\sum \Pr \left[\left(A_1 = a_{l_1}^1, \dots, A_n = a_{l_n}^n \right) \left(x_{j_1}^1, \dots, x_{j_n}^n \right) \right] = \sum \Pr \left[\left(A_1 = a_{l_1}^1, \dots, A_n = a_{l_n}^n \right) \left(x_{j_1}^1, \dots, x_{j_n}^n \right) \right],$$

where the summation is across all values of (l_1, \dots, l_n) with a fixed $(l_{1'}, \dots, l_{s'})$. Clearly, this necessary condition is just another way of stating marginal selectivity. Assuming that \mathbf{P} does satisfy marginal selectivity, it can be represented by the “reduced hierarchical” vector \mathbf{P}^* whose components are marginal probabilities of all orders, with $s = 0$ corresponding to the probability 1.

3.3. Examples

Example 3.3. Let $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta\}$, and the set of allowable treatments T consist of all four possible combinations of the factor points. Let A and B be Bernoulli variables, $a_1 = b_1 = 1$, $a_2 = b_2 = 2$, distributed as shown:

$\alpha \beta$	$A B$	Pr	$\alpha \beta$	$A B$	Pr
1 1	1 1	.140	1 2	1 1	.198
	1 2	.360		1 2	.302
	2 1	.360		2 1	.302
	2 2	.140		2 2	.198
2 1	1 1	.189	2 2	1 1	.460
	1 2	.311		1 2	.040
	2 1	.311		2 1	.040
	2 2	.189		2 2	.460

Marginal selectivity here is satisfied trivially: all marginal probabilities are equal 0.5, for all treatments. In the matrix form of the LFT, the column-vector of the above 16 probabilities, $(.140, .360, .360, \dots, .040, .040, .460)^\top$,

using \top for transposition, is denoted by \mathbf{P} . The LFT problem is defined by the system $\mathbf{M}\mathbf{Q} = \mathbf{P}$, $\mathbf{Q} \geq 0$, where the 16×16 Boolean matrix \mathbf{M} is shown below: each column of the matrix corresponds to a combination of values for the hypothetical H -variables (shown above the matrix), while each row corresponds to a combination of a treatment with values of the outputs A, B (shown on the left).

$H_{1\alpha}$	1 1 1 1 1 1 1 2 2 2 2 2 2 2 2	
$H_{2\alpha}$	1 1 1 1 2 2 2 2 1 1 1 1 2 2 2	
$H_{1\beta}$	1 1 2 2 1 1 2 2 1 1 2 2 1 1 2 2	
$H_{2\beta}$	1 2 1 2 1 2 1 2 1 2 1 2 1 2 1 2	
$\alpha \beta$	$A B$	
1 1	1 1	1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0
	1 2	0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 0
	2 1	0 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0
	2 2	0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 1
1 2	1 1	1 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0
	1 2	0 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0
	2 1	0 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0
	2 2	0 0 0 0 0 0 0 0 0 1 0 1 0 1 0 1
2 1	1 1	1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0
	1 2	0 0 1 1 0 0 0 0 0 0 1 1 0 0 0 0
	2 1	0 0 0 0 1 1 0 0 0 0 0 0 1 1 0 0
	2 2	0 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1
2 2	1 1	1 0 1 0 0 0 0 0 1 0 1 0 0 0 0 0
	1 2	0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 0
	2 1	0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0
	2 2	0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1

The linear programming routine of Mathematica™ (using the interior point algorithm) shows that the linear equations (9) have nonnegative solutions corresponding to the JDC-vector

$H_{1\alpha}$	$H_{2\alpha}$	$H_{1\beta}$	$H_{2\beta}$	Pr	$H_{1\alpha}$	$H_{2\alpha}$	$H_{1\beta}$	$H_{2\beta}$	Pr
1	1	1	1	.02708610	2	1	1	1	.15748000
1	1	1	2	.00239295	2	1	1	2	.00204128
1	1	2	1	.16689300	2	1	2	1	.10854100
1	1	2	2	.03358610	2	1	2	2	.00197965
1	2	1	1	.00197965	2	2	1	1	.03358610
1	2	1	2	.10854100	2	2	1	2	.16689300
1	2	2	1	.00204128	2	2	2	1	.00239295
1	2	2	2	.15748000	2	2	2	2	.02708610

The column-vector of these probabilities constitutes $\mathbf{Q} > 0$. This proves that in this case we do have $(A, B) \leftarrow_P (\alpha, \beta)$. \square

Example 3.4. In the previous example, let us change the distributions of (A, B) to the following:

$\alpha \beta$	$A B$	Pr	$\alpha \beta$	$A B$	Pr
1 1	1 1	.450	1 2	1 1	.105
	1 2	.050		1 2	.395
	2 1	.050		2 1	.395
	2 2	.450		2 2	.105
2 1	1 1	.170	2 2	1 1	.110
	1 2	.330		1 2	.390
	2 1	.330		2 1	.390
	2 2	.170		2 2	.110

Once again, marginal selectivity is satisfied trivially, as all marginal probabilities are 0.5, for all treatments. The linear

programming routine of Mathematica™, however, shows that the linear equations (9) have no nonnegative solutions. This excludes the existence of a JDC-vector for this situations, ruling out thereby the possibility of $(A, B) \leftrightarrow (\alpha, \beta)$. □

3.4. Renaming and grouping

Since LFT is both a necessary and sufficient condition for selective influences, if it is passed for $(A_1, \dots, A_n)(\phi)$, it is guaranteed to be passed following any factor-point-specific transformations of these random outputs. All such transformations in the case of discrete random variables can be described as combinations of renaming (factor-point specific one) and coarsening (grouping of some values together). In fact, the outcome of LFT simply does not depend on the values of the random variables involved, only their probabilities matter. Therefore a renaming will not change anything in the system of linear equations and inequalities (8)–(9). An example of coarsening will be redefining A and B , each having possible values 1, 2, 3, 4, into binary variables

$$A^*(\phi) = \begin{cases} 1 & \text{if } A(\phi) = 1, 2, \\ 2 & \text{if } A(\phi) = 3, 4, \end{cases}$$

$$B^*(\phi) = \begin{cases} 1 & \text{if } B(\phi) = 1, 2, 3, \\ 2 & \text{if } B(\phi) = 4. \end{cases}$$

It is clear that any such a redefinition amounts to replacing some of the equations in (9) with their sums. Therefore, if the original system has a solution, so will also the system after such replacements. Of course, the reverse is not generally true: the coarser system can have solutions when the original system does not.

The same is true for coarsening the system by grouping together some of the factor points within factors. Suppose we want to group together points x_1^1 and x_2^1 of factor α_1 containing more than two points. This means that the probabilities $P(l_1, l_2, \dots, l_n; j_1, j_2, \dots, j_n)$ are redefined as⁷

$$P'(l_1, l_2, \dots, l_n; j_1, j_2, \dots, j_n) = \begin{cases} \frac{1}{2}P(l_1, l_2, \dots, l_n; 1, j_2, \dots, j_n) \\ + \frac{1}{2}P(l_1, l_2, \dots, l_n; 2, j_2, \dots, j_n) & \text{if } j_1 = 1, \\ P(l_1, l_2, \dots, l_n; j_1 + 1, j_2, \dots, j_n) & \text{if } j_1 > 1. \end{cases}$$

When we average the original equations for $P(l_1, l_2, \dots, l_n; 1, j_2, \dots, j_n)$ and $P(l_1, l_2, \dots, l_n; 2, j_2, \dots, j_n)$, we get

$$\sum \left\{ \frac{1}{2} \sum_{l_{12}} Q(l_{11} = l_1, l_{12}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) + \frac{1}{2} \sum_{l_{11}} Q(l_{11}, l_{12} = l_1, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) \right\} = P'(l_1, l_2, \dots, l_n; 1, j_2, \dots, j_n),$$

where $l_{2j_2} = l_2, \dots, l_{nj_n} = l_n$ and the outer summation is across all l_{ij} except for the following values for (i, j) : (1, 1), (1, 2), and (i, j_i) ,

⁷ More general mixtures, $\pi P(l_1, l_2, \dots, l_n; 1, j_2, \dots, j_n) + (1 - \pi)P(l_1, l_2, \dots, l_n; 2, j_2, \dots, j_n)$ for $0 < \pi \leq 1$, are dealt with as easily; moreover, $\pi = 1$ formally corresponds to dropping the factor point x_2^1 , considered below. The values of π other than $\frac{1}{2}$ and 1 can be useful if the grouping is done on a sample level, to reflect the differences in sample sizes corresponding to treatments containing x_1^1 and x_2^1 .

$i = 2, \dots, n$. We define a new vector Q' whose dimensionality is less than that of Q by one, putting

$$Q'(l_{11} = l, l_{13}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) = \frac{1}{2} \sum_{l_{12}} Q(l_{11} = l, l_{12}, l_{13}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) + \frac{1}{2} \sum_{l_{11}} Q(l_{11}, l_{12} = l, l_{13}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}),$$

where l has the same range as any of l_{ij} . (For notational simplicity, in Q' we do not re-enumerate (1, 3) as (1, 2), (1, 4) as (1, 3), etc., leaving thereby l_{12} undefined)

For any point of factor α_1 other than x_1^1 and x_2^1 , say, x_3^1 , we have then

$$\sum_{l_{11}, l_{12}} Q(l_{11}, l_{12}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) = P(l_1, l_2, \dots, l_n; 3, j_2, \dots, j_n),$$

which can be presented as

$$\sum_l \sum_t \left\{ \frac{1}{2} \sum_{l_{12}} Q(l_{11} = l, l_{12}, l_{13} = l_1, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) + \frac{1}{2} \sum_{l_{11}} Q(l_{11}, l_{12} = l, l_{13} = l_1, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) \right\} = P(l_1, l_2, \dots, l_n; 3, j_2, \dots, j_n).$$

This is equivalent to

$$\sum Q'(l_{11}, l_{13} = l_1, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) = P'(l_1, l_2, \dots, l_n; j_1 = 2, j_2, \dots, j_n),$$

where $l_{2j_2} = l_2, \dots, l_{nj_n} = l_n$, and the summation is across all l_{ij} except for $(i, j) = (1, 3)$ and $(i, j) = (i, j_i), i = 2, \dots, n$. So we have obtained a solution for the factor-coarsened system from a solution for the original system.

Dropping a point, say, x_2^1 is even simpler: we delete all rows with $j_1 = 2$, and then redefine the Q vector as

$$Q'(l_{11}, l_{13}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}) = \sum_{l_{12}} Q(l_{11}, l_{12}, l_{13}, \dots, l_{1k_1}, \dots, l_{n1}, \dots, l_{nk_n}).$$

If the random variables involved have more than finite number of values and/or the factors consist of more than finite number of factor points, or if these numbers, though finite, are too large to handle the ensuing linear programming problem, then LFT can still be used after the values of the random variables and/or factors have been appropriately grouped. LFT then becomes only a necessary condition for selective influences (with respect to the original system of factors and random variables), and its results will generally be different for different (non-nested) groupings.

Example 3.5. Consider the hypothesis $(A, B) \leftrightarrow (\alpha, \beta)$ with the factors having a finite number of factor points each, and A and B being response times. To use LFT, one can transform the random variable A as, say,

$$A^*(\phi) = \begin{cases} 1 & \text{if } A(\phi) \leq a_{1/4}(\phi), \\ 2 & \text{if } a_{1/4}(\phi) < A(\phi) \leq a_{1/2}(\phi), \\ 3 & \text{if } a_{1/2}(\phi) < A(\phi) \leq a_{3/4}(\phi), \\ 4 & \text{if } A(\phi) > a_{3/4}(\phi), \end{cases}$$

and transform B as

$$B^*(\phi) = \begin{cases} 1 & \text{if } B(\phi) \leq b_{1/2}(\phi), \\ 2 & \text{if } B(\phi) > b_{1/2}(\phi), \end{cases}$$

where $a_p(\phi)$ and $b_p(\phi)$ designate the p th quantiles of, respectively $A(\phi)$ and $B(\phi)$. The initial hypothesis now is reformulated as $(A^*, B^*) \leftrightarrow (\alpha, \beta)$, with the understanding that if it is rejected then the initial hypothesis will be rejected too (a necessary condition only). LFT will now be applied to distributions of the form

α	β	A	B	Pr
x	y	1	1	p_{11}
		1	2	p_{12}
		\vdots	\vdots	\vdots
		4	1	p_{41}
		4	2	p_{42}

where the marginals for A are constrained to 0.25 and the marginals for B to 0.5, for all treatments $\{x^\alpha, y^\beta\}$, yielding a trivial compliance with marginal selectivity. Note that the test may very well uphold $(A^*, B^*) \leftrightarrow (\alpha, \beta)$ even if marginal selectivity is violated for $(A, B)(\phi)$ (e.g., if the quantiles $a_p(x^\alpha, y^\beta)$ change as a function of y^β). □

3.5. Quantum entanglement

Fine's (1982a,b) inequalities relate to the simplest EPR paradigm, with the number of particles $n = 2$, number of spin axes per particle $k_1 = k_2 = 2$, and the number of possible spin values per particle $m_1 = m_2 = 2$ (this value being the same for all spin axes chosen for a given particle). They can be written, with reference to (6) and (12), as

$$\begin{aligned}
 & -1 \leq P(2, 2; j_1, j_2) + P(2, 2; j'_1, j_2) \\
 & + P(2, 2; j_1, j'_2) - P(2, 2; j_1, j_2) \\
 & - P_1^*(2; j'_1) - P_2^*(2; j_2) \leq 0,
 \end{aligned}$$

where $j_1, j'_1 \in \{1, 2\}$, $j_2, j'_2 \in \{1, 2\}$, $j_1 \neq j'_1$, $j_2 \neq j'_2$. These inequalities constitute the necessary and sufficient conditions for $(A_1, A_2) \leftrightarrow (\alpha_1, \alpha_2)$, with marginal selectivity assumed implicitly. Although Fine's derivation of these inequalities is different, they can be derived as solutions of system (11), with \mathbf{P}^* the 9-component vector (using \top for transposition)

$$(1, P_1^*(2; 1), \dots, P_2^*(2; 2), P(2, 2; 1, 1), \dots, P(2, 2; 2, 2))^\top,$$

\mathbf{Q} the 16-component vector

$$(Q(1, 1, 1, 1), \dots, Q(2, 2, 2, 2))^\top,$$

and \mathbf{M}^* the corresponding 9×16 Boolean matrix,

$H_{1\alpha}$	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
$H_{2\alpha}$	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
$H_{1\beta}$	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2
$H_{2\beta}$	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1
α	β	A	B												
.	.	.	.	1	1	1	1	1	1	1	1	1	1	1	1
1	.	2	.	0	0	0	0	0	0	0	1	1	1	1	1
2	.	2	.	0	0	0	0	1	1	1	1	0	0	0	1
.	1	.	2	0	0	1	1	0	0	1	1	0	0	1	1
.	2	.	2	0	1	0	1	0	1	0	1	0	1	0	1
1	1	2	2	0	0	0	0	0	0	0	0	1	1	0	0
1	2	2	2	0	0	0	0	0	0	0	1	0	1	0	1
2	1	2	2	0	0	0	0	0	1	1	0	0	0	0	0
2	2	2	2	0	0	0	0	1	0	1	0	0	0	0	1

In fact, using a standard facet enumeration program (e.g., lrs program at <http://cgm.cs.mcgill.ca/~avis/C/lrs.html>) these inequalities (together with the equalities representing marginal

selectivity) can be derived "mechanically". The essence of the computation is in the fact that a linear system (10) or (11) is feasible if and only if the point \mathbf{P} (respectively, \mathbf{P}^*) belongs to the convex hull of the points corresponding to the columns of \mathbf{M} (respectively, \mathbf{M}^*), which form a subset of the vertices of a unit hypercube. The facet enumeration programs derive inequalities describing this convex hull.

Given a set of numerical (experimentally estimated or theoretical) probabilities, computing the LP problem (10) or (11) is always preferable to dealing with explicit inequalities as their number becomes very large even for moderate-size vectors P . While Fine's inequalities for $n = 2, k_1 = k_2 = 2, m_1 = m_2 = 2$ (assuming marginal selectivity) number just 8, already for $n = 2, k_1 = k_2 = 2$ with $m_1 = m_2 = 3$ (describing, e.g., an EPR experiment with two spin-1 particles, or two spin-1/2 ones and inefficient detectors), our computations yield 1080 inequalities, and for $n = 3, k_1 = k_2 = k_3 = 2$ and $m_1 = m_2 = m_3 = 2$, corresponding to the Greenberger, Horne, and Zeilinger (1989) paradigm with three spin-1/2 particles, this number is 53,792.

The potential of JDC to lead to LFT and provide an ultimate criterion for the entanglement problem has not been utilized in quantum physics until relatively recently, when LFT was proposed in Werner and Wolf (2001a,b) and Basoalto and Percival (2003). Prior to this, criteria (as opposed to just necessary conditions) for the possibility of a classical explanation for an EPR paradigm involving multiple particles, multiple measurement settings, and multiple outcomes per measurements were only known under strong symmetry constraints (de Barros & Suppes, 2001; Garg, 1983; Mermin, 1990; Peres, 1999).

3.6. Sample-level tests

Although this paper is not concerned with statistical questions, it may be useful to mention some of the approaches to constructing sample-level tests based on LFT. As mentioned in Section 3.5, the set of vectors \mathbf{P} for which the system $\mathbf{M}\mathbf{Q} = \mathbf{P}, \mathbf{Q} \geq 0$ has a solution forms a convex polytope. In particular, if the set T of allowable treatments contains all combinations of factors points, the polytope is the $((k_1(m_1 - 1) + 1) \dots (k_n(m_n - 1) + 1) - 1)$ -dimensional convex hull of the points corresponding to the columns of the Boolean matrix \mathbf{M} , which form a subset of the vertices of the $(m_1)^{k_1} \dots (m_n)^{k_n}$ -dimensional unit hypercube. Recently Davis-Stober (2009) developed a statistical theory for testing the hypothesis that a vector of probabilities \mathbf{P} (not necessarily of the same structure as in LFT) belongs to a convex polytope \mathcal{P} against the hypothesis that it does not. Under certain regularity constraints he derived the asymptotic distribution (a convex mixture of chi-square distributions) for the log maximum likelihood ratio statistic

$$-2 \log \frac{\max_{\mathbf{P} \in \mathcal{P}} L(\mathbf{P}|\mathbf{N})}{\max_{\mathbf{P}} L(\mathbf{P}|\mathbf{N})},$$

where \mathbf{N} is the vector of observed absolute frequencies, comprised of the numbers of occurrences of $(l_1, \dots, l_n; j_1, \dots, j_n)$ in the case of LFT. The likelihoods $L(\mathbf{P}|\mathbf{N})$ are computed using the standard theory of multinomial distributions. This theory has been "test-driven" on the polytopes related to the transitivity of preferences problem (Regenwetter, Dana, & Davis-Stober, 2010, 2011). A Bayesian approach to the same problem is presented in Myung, Karabatsos, and Iverson (2005).

Other approaches readily suggest themselves. One of them is to use the known theory of $L(\mathbf{P}|\mathbf{N}) / \max_{\mathbf{P}} L(\mathbf{P}|\mathbf{N})$ to compute a confidence region of possible probability vectors \mathbf{P} for a given empirical vector \mathbf{N} . The hypothesis of selective influences is retained or rejected according as this confidence region contains or

does not contain a point \mathbf{P} that passes LFT. Resampling techniques is another obvious approach, e.g., the permutation test in which the assignment of empirical distributions to different treatments is randomly “reshuffled” so that each distribution generally ends up assigned to a “wrong” treatment. If the proportion of the permuted assignments whose deviation from the LFT polytope does not exceed that of the observed estimate of \mathbf{P} is sufficiently small, the hypothesis of selective influences can be considered supported.

Little is known at present about the computational feasibility and statistical properties of these and similar procedures. In particular (this also applies to Davis-Stober’s test), we do not know their statistical power for different locations of the true vector of probabilities outside the convex polytope described by $\mathbf{MQ} = \mathbf{P}$, $\mathbf{Q} \geq 0$. Nor do we know how the effect size, a measure of deviation of \mathbf{P} from the polytope, should be computed optimally. All of this will have to be investigated separately.

4. Conclusion

Selectiveness in the influences exerted by a set of inputs upon a set of random and stochastically interdependent outputs is a critical feature of many psychological models, often built into the very language of these models. We speak of an internal representation of a given stimulus, as separate from an internal representation of another stimulus, even if these representations are considered random entities and they are not independent. We speak of decompositions of response time into signal-dependent and signal-independent components, or into a perceptual stage (influenced by stimuli) and a memory-search stage (influenced by the number of memorized items), without necessarily assuming that the two components or stages are stochastically independent.

In this paper, we have described the Linear Feasibility Test, an application of the fundamental Joint Distribution Criterion for selective influences to random variables with finite numbers of values. This test can be performed by means of standard linear programming. Due to the fact that any random output can be discretized, the Linear Feasibility Test is universally applicable, although one should keep in mind that if a diagram of selective influences is upheld by the test at some discretization, it may be rejected at a finer or non-nested discretization (but not at a coarser one). Both the Joint Distribution Criterion and the Linear Feasibility Test, although new in the behavioral context, have their direct analogues in quantum physics, in dealing with the problem of the existence of a classical explanation (one with non-contextual, local hidden variables) for outcomes of noncommuting measurements performed on entangled particles. The discovery of these parallels promises to enrich and facilitate our understanding of selective influences.

Acknowledgments

This research has been supported by AFOSR grant FA9550-09-1-0252 to Purdue University and by the Academy of Finland grant 121855 to University of Jyväskylä. We are deeply grateful to Joseph Hout and Jerome Busemeyer whose comments helped us to significantly improve the paper.

References

- Basoalto, R. M., & Percival, I. C. (2003). BellTest and CHSH experiments with more than two settings. *Journal of Physics A: Mathematical & General*, 36, 7411–7423.
- Bell, J. (1964). On the Einstein–Podolsky–Rosen paradox. *Physics*, 1, 195–200.
- Bohm, D., & Aharonov, Y. (1957). Discussion of experimental proof for the paradox of Einstein, Rosen and Podolski. *Physical Review*, 108, 1070–1076.
- Clauser, J. F., Horne, M. A., Shimony, A., & Holt, R. A. (1969). Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23, 880–884.
- Davis-Stober, C. P. (2009). Analysis of multinomial models under inequality constraints: applications to measurement theory. *Journal of Mathematical Psychology*, 53, 1–13.
- de Barros, J. A., & Suppes, P. (2001). Results for six detectors in a three-particle GHZ experiment. In J. Bricmont, D. Durr, M. C. Galavotti, G. Ghirardi, F. Petruccione, & N. Zanghi (Eds.), *Chance in physics: Foundations and perspectives* (pp. 213–223). Berlin: Springer.
- Dzhafarov, E. N. (1999). Conditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, 43, 123–157.
- Dzhafarov, E. N. (2003a). Selective influence through conditional independence. *Psychometrika*, 68, 7–26.
- Dzhafarov, E. N., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, 50, 390–401.
- Dzhafarov, E. N., & Kujala, J. V. (2010). The joint distribution criterion and the distance tests for selective probabilistic causality. *Frontiers in Quantitative Psychology and Measurement*, 1, 151. doi:10.3389/fpsyg.2010.00151.
- Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47, 777–780.
- Fine, A. (1982a). Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23, 1306–1310.
- Fine, A. (1982b). Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48, 291–295.
- Garg, A. (1983). Detector error and Einstein–Podolsky–Rosen correlations. *Physical Review D*, 28, 785–790.
- Greenberger, D. M., Horne, M. A., & Zeilinger, A. (1989). Going beyond Bell’s theorem. In M. Kafatos (Ed.), *Bell’s theorem, quantum theory and conceptions of the universe* (pp. 69–72). Kluwer: Dordrecht.
- Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorica*, 4, 373–395.
- Khachiyan, L. (1979). A polynomial algorithm in linear programming. *Doklady Akademii Nauk SSSR*, 244, 1093–1097.
- Kechris, A. S. (1995). *Classical descriptive set theory*. New York: Springer.
- Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology*, 52, 128–144.
- Mermin, N. D. (1990). Extreme quantum entanglement in a superposition of macroscopically distinct states. *Physical Review Letters*, 65, 1838–1840.
- Myung, J. I., Karabatsos, G., & Iverson, G. J. (2005). A Bayesian approach to testing decision making axioms. *Journal of Mathematical Psychology*, 49, 205–225.
- Peres, A. (1999). All the Bell inequalities. *Foundations of Physics*, 29, 589–614.
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2010). Testing transitivity of preferences on two-alternative forced choice data. *Frontiers in Quantitative Psychology and Measurement*. doi:10.3389/fpsyg.2010.00148.
- Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, 118, 42–56.
- Sternberg, S. (1969). The discovery of processing stages: extensions of Donders method. In: W.G. Koster (Ed.), *Attention and Performance II*. Acta Psychologica. vol. 30. (pp. 276–315).
- Stapp, H. P. (1975). Bell’s theorem and world process. *Nuovo Cimento B*, 29, 270–276.
- Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, 48, 191–199.
- Thurstone, L. L. (1927). A law of comparative judgments. *Psychological Review*, 34, 273–286.
- Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, 28, 363–400.
- Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, 33, 309–327.
- Werner, R.F., & Wolf, M.M. (2001a). All multipartite Bell correlation inequalities for two dichotomic observables per site. arXiv:quant-ph/0102024v1.
- Werner, R.F., & Wolf, M.M. (2001b). Bell inequalities and entanglement. arXiv:quant-ph/0107093v2.