The joint distribution criterion and the distance tests for selective probabilistic causality

Ehtibar N. Dzhafarov[1,2]* and Janne V. Kujala[3]

[1] Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA
[2] Swedish Collegium for Advanced Study, Uppsala, Sweden
[3] University of Jyväskylä, Jyväskylä, Finland

***Correspondence:**
Ehtibar N. Dzhafarov, Department of
Psychological Sciences, Purdue
University, 703 Third Street,
West Lafayette, IN 47907, USA.
e-mail: ehtibar@purdue.edu

A general definition and a criterion (a necessary and sufficient condition) are formulated for an arbitrary set of external factors to selectively influence a corresponding set of random entities (generalized random variables, with values in arbitrary observation spaces), jointly distributed at every treatment (a set of factor values containing precisely one value of each factor). The random entities are selectively influenced by the corresponding factors if and only if the following condition, called the joint distribution criterion, is satisfied: there is a jointly distributed set of random entities, one entity for every value of every factor, such that every subset of this set that corresponds to a treatment is distributed as the original variables at this treatment. The distance tests (necessary conditions) for selective influence previously formulated for two random variables in a two-by-two factorial design (Kujala and Dzhafarov, 2008, *J. Math. Psychol.* 52, 128–144) are extended to arbitrary sets of factors and random variables. The generalization turns out to be the simplest possible one: the distance tests should be applied to all two-by-two designs extractable from a given set of factors.

Keywords: external factors, joint distribution, probabilistic causality, selective influence, systems of random variables, stochastic dependence, stochastically unrelated

A system's behavior, be the system biological, social, or technological, can be thought of as a network of stochastically interdependent random entities. The external world provides inputs (influences, interventions, conditions) presumably affecting some of the components of the network and not affecting the others. The question arises therefore as to how, based on the joint distributions of all these random entities, to distinguish the components affected and not affected by each of these external inputs.

The notion of selective influence under stochastic interdependence was introduced and systematically analyzed in the behavioral context by Townsend (1984), although implicitly it had been used before (Lazarsfeld, 1965; Bloxom, 1972; Schweickert, 1982). Townsend's approach to selective influence (further developed in Townsend and Thomas, 1994, and mathematically characterized in Dzhafarov, 1999) is, however, very different from the present one. In fact, in all non-trivial cases they are incompatible. Our approach gradually developed starting with Dzhafarov (2001), based on Dzhafarov's earlier work on response time analysis (see Dzhafarov, 1997, for an overview). In Dzhafarov (2003) the definition of selective influence adopted in the present paper was given for finite systems of random entities. This notion was put on a more solid probabilistic foundation in Dzhafarov and Gluhovsky (2006), and further developed in Kujala and Dzhafarov (2008). In the latter work, for the first time, workable tests for selective influence were formulated.

The present paper continues this line of research on a higher level of mathematical rigor and arguably the highest possible level of generality. The abstract nature of the mathematical theory makes it rather difficult reading, with notation which, though carefully chosen, may appear complicated. As a partial remedy, we precede the formal development in Sections 2–5 by Section 1 in which we provide a more intuitive account of some of the basic notions and results. We do this on simple examples involving just two random variables influenced by two factors.

1. INTRODUCTION

1.1. WHAT IS SELECTIVE INFLUENCE?

Consider a simple double-detection experiment: there are two stimuli each of which may possess or lack a certain feature (signal property), and an observer has to respond *Yes* (signal present) or *No* (signal absent) to each of the two stimuli. For instance, the stimuli may be two spatially separated line segments in a frontal plane each of which may be either vertical (signal absent) or tilted by a fixed small angle (signal present); the observer says *Yes − Yes* if both lines appear to be tilted, *Yes − No* if the left line appears tilted and the right one not, etc. These responses are *random variables*: A (response to the left stimulus) and B (response to the right one), each with two possible values {*Yes*, *No*} occurring with some probabilities. They are *jointly distributed*, in the sense that by the virtue of co-occurring in the same trial the values of A and B are naturally paired, enabling one to meaningfully pose questions like "What is the joint probability of $A = Yes$ and $B = No$?". The joint distribution of A and B may change depending on the values of the following *external factors*: $\alpha =$ Tilt of the left line, with two possible values, {*absent*, *present*}, and $\beta =$ Tilt of the right line, with the same two values. The combination of factor values chosen, one for each of the factors, is traditionally referred to as a *treatment*. With

this terminology and notation, the population-level (idealized) results of the experiment in question can be presented in the form of four matrices:

Treatment $\begin{bmatrix} \alpha = absent \\ \beta = absent \end{bmatrix}$	B = yes	B = no	
A = yes	p_{11}	p_{12}	$p_{1\cdot}$
A = no	p_{21}	p_{22}	$p_{2\cdot}$
	$p_{\cdot1}$	$p_{\cdot2}$	

Treatment $\begin{bmatrix} \alpha = absent \\ \beta = present \end{bmatrix}$	B = yes	B = no	
A = yes	q_{11}	q_{12}	$q_{1\cdot}$
A = no	q_{21}	q_{22}	$q_{2\cdot}$
	$q_{\cdot1}$	$q_{\cdot2}$	

Treatment $\begin{bmatrix} \alpha = present \\ \beta = absent \end{bmatrix}$	B = yes	B = no	
A = yes	r_{11}	r_{12}	$r_{1\cdot}$
A = no	r_{21}	r_{22}	$r_{2\cdot}$
	$r_{\cdot1}$	$r_{\cdot2}$	

Treatment $\begin{bmatrix} \alpha = present \\ \beta = present \end{bmatrix}$	B = yes	B = no	
A = yes	s_{11}	s_{12}	$s_{1\cdot}$
A = no	s_{21}	s_{22}	$s_{2\cdot}$
	$s_{\cdot1}$	$s_{\cdot2}$	

The letters p, q, r, s here represent theoretical probabilities, with the usual meaning of the subscripts: $p_{1\cdot} = p_{11} + p_{12}$, $p_{2\cdot} = p_{21} + p_{22}$, $p_{1\cdot} + p_{2\cdot} = 1$, etc. It is natural to surmise that, unless the observer does not look at the stimuli at all, the random variable A should depend on (be *influenced* by) the value of α, and B should be influenced by the value of β. It is not obvious, however, whether factor α only (or *selectively*) influences random variable A, without affecting B, and whether factor β only (*selectively*) influences random variable B, without affecting A:

$$\begin{matrix} \alpha & \beta \\ \downarrow & \downarrow \\ A & B \end{matrix},$$
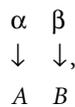
as opposed to the possibilities

$$\begin{matrix} \alpha & \beta & \alpha & \beta & \alpha & \beta \\ \downarrow\searrow\downarrow, & & \downarrow\swarrow\downarrow, & & \downarrow\times\downarrow. \\ A & B & A & B & A & B \end{matrix}$$
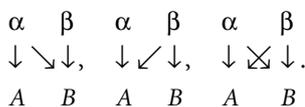
Thus, we will have one of the latter scenarios if the "present" value of α visually masks or enhances the salience of the "present" value of β, or if the values of β somehow affect the level of attention the observer pays to the factor α.

We denote the case when (α, β) *selectively influence* (A, B), *respectively,* by

$$(A, B) \leftarrow\!\!\!\!P \, (\alpha, \beta).$$

What does this mean? The meaning of the relation is only obvious if A and B are *stochastically independent* for all four treatments, i.e., if $p_{ij} = p_{i\cdot}p_{\cdot j}$, $q_{ij} = q_{i\cdot}q_{\cdot j}$, etc., where $i, j \in \{1, 2\}$. In this case all one has to establish to prove $(A, B) \leftarrow\!\!\!\!P \, (\alpha, \beta)$ is that the *marginal distribution* of A is not affected by changes in β and the marginal distribution of B is not affected by changes in α.

To look at this in detail, let the pair of our random variables (A, B) at the four treatments be denoted $(A, B)_{11}$, $(A, B)_{12}$, $(A, B)_{21}$, and $(A, B)_{22}$, where 1, 2 denote "absent" and "present," respectively. If A and B are independent at all four treatments, then the selectiveness $(A, B) \leftarrow\!\!\!\!P \, (\alpha, \beta)$ simply means that the marginal distribution of A does not depend on β (i.e., $p_{1\cdot} = q_{1\cdot}$ and $r_{1\cdot} = s_{1\cdot}$) and the marginal distribution of B does not depend on α (i.e., $p_{\cdot1} = q_{\cdot1}$ and $r_{\cdot1} = s_{\cdot1}$). The problem arises when A and B are not independent for at least one of the treatments: how should one determine then if $(A, B) \leftarrow\!\!\!\!P \, (\alpha, \beta)$? This is the problem addressed in this paper, only we do not confine the consideration to the case of two factors and two random variables. Rather we generalize the problem to an arbitrary set of external factors and an arbitrary (but one-to-one corresponding to the set of factors) set of *random entities*.[1]

For a finite set of random variables the definition of selective influence was given in Dzhafarov (2003) and then refined in Dzhafarov and Gluhovsky (2006) and Kujala and Dzhafarov (2008). Applying it to our example, (A, B) is selectively influenced by (α, β) if and only if one can find functions f and g and a random entity C whose distribution does not depend on α, β, such that

$$(A, B)_{\alpha\beta} \sim \left(f(\alpha, C), g(\beta, C) \right), \tag{1}$$

where $\sim$ stands for "is distributed as".[2] That is, denoting $f(\alpha = 1, C)$ by $f_1(C)$, $g(\beta = 2, C)$ by $g_2(C)$, etc.,

$$(A, B)_{11} \sim (f_1(C), g_1(C)), \quad (A, B)_{12} \sim (f_1(C), g_2(C)),$$
$$(A, B)_{21} \sim (f_2(C), g_1(C)), \quad (A, B)_{22} \sim (f_2(C), g_2(C)).$$

As an example, let C be a random vector (C_0, C_1, C_2) with stochastically independent components having the following interpretation: C_0 is a random entity representing the general level of visual attention, while C_1 and C_2 are stimulus-specific sources of randomness (which, with no loss of generality, can be taken to be uniformly distributed between 0 and 1). Let

---

[1] As explained in Section 2, we distinguish random entities and their special case, random variables. Random entities take on values in arbitrary measurable spaces, while random variables map, or can be redefined to map, into real numbers endowed with the Borel sigma-algebra. Note also that the notion of a random entity (or variable) should always be taken to include deterministic entities (variables) as a special case, the same as the notion of stochastic interdependence, unless otherwise indicated, should be taken to include stochastic independence as a special case.

[2] It is usually the case that the possibility of selectiveness is considered when it is known that the factors are *effective* in their influence upon (A, B), meaning that for at least one value of either of the two factors the change of the other factor from 1 to 2 changes the joint distribution of (A, B). This aspect of the dependence of (A, B) on (α, β) being relatively trivial, we do not include it in the definition of selective influence. In other words, $(A, B) \leftarrow\!\!\!\!P \, (\alpha, \beta)$ is taken to mean that β *does not* influence A and α *does not* influence B, leaving open the question of whether α influences A and/or β influences B (see Dzhafarov, 2003, p. 10).

$$A = \begin{cases} 1 & \text{if } C_1 > h_1(\alpha, C_0) \\ 0 & \text{if } C_1 \leq h_1(\alpha, C_0) \end{cases}, \quad B = \begin{cases} 1 & \text{if } C_2 > h_2(\beta, C_0) \\ 0 & \text{if } C_2 \leq h_2(\beta, C_0) \end{cases},$$

where h_1, h_2 are some measurable functions from the set of possible values of C into interval $[0, 1]$. One can see that A and B are generally stochastically interdependent by virtue of depending on one and the same random entity, C_0, but that A does not depend on β, in the sense that for any given values of the other arguments, $C_0 = c_0$, $C_1 = c_1$, $C_2 = c_2$, and $\alpha = 1$ or 2, the value of A does not change as a function of β; and B does not depend on α in the analogous sense.

The definition of selective influence can also be looked at in a simpler and more fundamental way. The fact that for any given treatment A and B are stochastically related (i.e., paired, whether independent or interdependent) means in Kolmogorov's probability theory that A and B are measurable functions of one and the same random entity. It is always true therefore that

$$(A, B)_{\alpha\beta} \sim \left(f(\alpha, \beta, C_{\alpha\beta}), g(\alpha, \beta, C_{\alpha\beta}) \right).$$

The random entities $C_{11}, C_{12}, C_{21}, C_{22}$, can always be replaced with a single C, e.g., by putting $C = (C_{11}, C_{12}, C_{21}, C_{22})$ and redefining the functions f, g accordingly:

$$(A, B)_{\alpha\beta} \sim \left(f(\alpha, \beta, C), g(\alpha, \beta, C) \right). \tag{2}$$

Comparing this universal representation with (1) we see that the assumption of selective influence is that β in f and α in g are dummy arguments.

1.2. MAIN PROPERTIES OF SELECTIVE INFLUENCE

There are three main properties of the selective influence relation, $(A, B) \hookleftarrow (\alpha, \beta)$.

First, selective influence is invariant with respect to all (measurable) *transformations of the random variables A, B*, even if transformations of A are allowed to depend on values of factor α and transformations of B are allowed to depend on values of factor β. In our example the values of A and B are denoted *yes* and *no*. Clearly, we can encode them 0 and 1, respectively, or by any other two numbers or words. Moreover, we can, if we so choose, denote {*yes, no*} for A by {1, 0} if $\alpha = absent$ but by {−3, 5} if $\alpha = present$; analogously, we can denote {*yes, no*} for B by {*lion, crocodile*} if $\beta = absent$ but by {*zebra, cheetah*} if $\beta = present$. Selective influence, $(A, B) \hookleftarrow (\alpha, \beta)$, if it holds for the original values for A and B must also hold after any such transformations. This follows from the fact that if (1) holds then after any factor-value-specific transformations $F(\alpha, A)$ and $G(\beta, B)$ we have

$$\left(F(\alpha, A), G(\beta, B) \right)_{\alpha\beta} \sim \left(F(\alpha, f(\alpha, C)), G(\beta, g(\beta, C)) \right)$$
$$= \left(f^\star(\alpha, C), g^\star(\beta, C) \right).$$

Second, selective influence implies *marginal selectivity*, the term coined by Townsend and Schweickert (1989) for the situation when the marginal distribution of A does not depend on β and the marginal distribution of B does not depend on α. This is an obvious consequence of (1). The reverse is not true, as illustrated by examples in Dzhafarov (2003) and, more systematically, in Kujala and Dzhafarov (2008). Other examples are given in this paper: in fact, in all our examples where the selective influence relation does not hold marginal selectivity is satisfied.

Third, selective influence relation satisfies the *nestedness property*: if some random variables are selectively influenced by corresponding factors (say, $(A, B, C) \hookleftarrow (\alpha, \beta, \gamma)$ – we need more than two factor-variable pairs for this property to be non-trivial), then any subset of these variables is selectively influenced by the corresponding subset of factors: $(A, B) \hookleftarrow (\alpha, \beta)$, $(A, C) \hookleftarrow (\alpha, \gamma)$, and $(B, C) \hookleftarrow (\beta, z)$. This property is obvious as soon as (1) is generalized to larger sets.

In this paper the three properties of selective influence will be demonstrated on the maximal level of generality, for arbitrary sets of random entities and corresponding sets of factors of arbitrary nature.

1.3. DISTANCE TESTS FOR SELECTIVE INFLUENCE

How can one determine that $(A, B) \hookleftarrow (\alpha, \beta)$? In Kujala and Dzhafarov (2008) two types of necessary conditions for selective influence were formulated, termed cosphericity tests and distance tests. As we only generalize in this paper the latter class of tests, we need not discuss the former. To apply a distance test to our example means to do the following. First, the values of A and B have to be encoded by real numbers. In accordance with what we know about the transformations we can use any functions $f(\alpha, A)$ and $g(\beta, B)$ with numerical values. Second, one chooses a number $\rho \geq 1$. Third, for each of the four treatments $\alpha\beta = 11, 12, 21, 22$ one computes the quantity

$$D\alpha\beta = \sqrt[\rho]{\mathrm{E}\left[\left| A_{\alpha\beta} - B_{\alpha\beta} \right|^\rho \right]},$$

where E denotes expected value and $(A_{\alpha\beta}, B_{\alpha\beta})$ is an alternative (and more convenient) way of designating $(A, B)_{\alpha\beta}$. Note that $D\alpha\beta$ (the same as $\alpha\beta$, 21, etc.) is a string of symbols, with no multiplication involved.

It has been shown in Kujala and Dzhafarov (2008) that if $(A, B) \hookleftarrow (\alpha, \beta)$, then, considering each random variable at each value of the corresponding factor as a point (this yields four points, A_1, A_2, B_1, B_2), these points can be placed in a metric space in which the values $D11, D12, D21, D22$ are, with some caveats, distances between A points and B points ($D11$ between A_1 and B_1, $D12$ between A_1 and B_2, etc.). As these distances, by definition, should satisfy the triangle inequality, we conclude with a bit of algebra (see Section 5) that

$$\max\left\{ D11, D12, D21, D22 \right\} \leq \frac{1}{2}\left(D11 + D12 + D21 + D22 \right).$$

A distance test consists in checking if this inequality is satisfied: if not (at least for one choice of the numerical values and the exponent ρ), then the selective influence relation is ruled out.

In this paper we generalize this test to arbitrary sets of random entities selectively influenced by arbitrary sets of external factors. As it turns out, all one has to do to prove that all random entities, taken one for each value of the corresponding factor, can be embedded in a metric space is to apply the test just described to

all pairs of 2×2 treatments for all pairs of factors. To present this result in an unambiguous form we have to introduce some notation that may appear cumbersome at first: since a value of a factor generally does not itself indicate which factor it is a value of (e.g., *absent* or 1 can be a value of both α and β), we superscript each factor value by the corresponding factor name. In our example it would be $1^\alpha, 2^\alpha, 1^\beta, 2^\beta$. We call these pairs, factor value with factor name, *factor points*. The four distances will now be written $D1^\alpha1^\beta$, $D1^\alpha2^\beta, D2^\alpha1^\beta, D2^\alpha2^\beta$. Note that we could only get away with the previous notation because the identity of the factors in it was encoded by the order of their values within pairs: in $D11$ the first 1 belonged to α and the second one to β. This convention cannot work, of course, for more than two factors. In the new notation the distance test acquires the form

$$\max\left\{D1^\alpha1^\beta, D1^\alpha2^\beta, D2^\alpha1^\beta, D2^\alpha2^\beta\right\}$$
$$\leq \frac{1}{2}\left(D1^\alpha1^\beta + D1^\alpha2^\beta + D2^\alpha1^\beta + D2^\alpha2^\beta\right), \quad (3)$$

where

$$Dx^\alpha y^\beta = Dy^\beta x^\alpha = \sqrt[p]{\mathrm{E}\left[\left|A_{x^\alpha y^\beta} - B_{x^\alpha y^\beta}\right|^p\right]}, \quad x, y \in \{1, 2\}.$$

1.4. THE JOINT DISTRIBUTION CRITERION FOR SELECTIVE INFLUENCE

Compliance with a given set of distance tests is only a necessary condition for selective influence. Is there a way to definitively prove that selective influence $(A, B) \leftarrow\!\!\!/ (\alpha, \beta)$ does hold if it is not ruled out by distance tests? As it turns out, the answer is affirmative, and it is an almost immediate consequence of our definition of selective influence, if presented at a sufficiently high level of mathematical rigor. Stated in intuitive terms and applied to our example, consider four hypothetical random variables, one for each of our factor points: $H_{1^\alpha}, H_{1^\beta}, H_{2^\alpha}, H_{2^\beta}$. Suppose that they are jointly distributed, i.e., we can speak of co-occurring quadruples of values. There are six pairwise combinations of the four factor points but only four of them, those of the form $x^\alpha y^\beta$ ($x, y \in \{1, 2\}$), form treatments, whereas the remaining two, $1^\alpha 2^\alpha$ and $1^\beta 2^\beta$, do not. The four treatments correspond to pairs $(A_{x^\alpha y^\beta}, B_{x^\alpha y^\beta})$ whose joint distributions are well defined. Suppose now that for all those cases when a pair of factor points forms a treatment we have

$$\left(H_{x^\alpha}, H_{y^\beta}\right) \sim \left(A_{x^\alpha y^\beta}, B_{x^\alpha y^\beta}\right), \quad x, y \in \{1, 2\}. \quad (4)$$

Then and only then $(A, B) \leftarrow\!\!\!/ (\alpha, \beta)$, and we can write then A_{x^α} instead of $A_{x^\alpha y^\beta}$ and B_{y^β} instead of $B_{x^\alpha y^\beta}$. We call this the *joint distribution criterion* for selective influence. The joint distributions of $(H_{1^\alpha}, H_{2^\alpha})$ and $(H_{1^\beta}, H_{2^\beta})$ must also be well defined, even though they do not correspond to any pairs of random variables one can choose from the observable $A_{1^\alpha 1^\beta}, B_{1^\alpha 2^\beta}, A_{2^\alpha 1^\beta}, B_{2^\alpha 2^\beta}$.

Let us look at this in detail. The observed joint distributions of $(A_{x^\alpha y^\beta}, B_{x^\alpha y^\beta})$ are represented by four probabilities each, denoted in the four matrices introducing our example by $p_{ij}, q_{ij}, r_{ij}, s_{ij}$. We now switch to a more convenient notation (although again, more cumbersome at first glance):

$$P_{ij}x^\alpha y^\beta = \Pr\left(A_{x^\alpha y^\beta} = i, B_{x^\alpha y^\beta} = j\right), \quad i, j \in \{yes, no\}, \quad x, y \in \{1, 2\},$$

where $P_{ij}x^\alpha y^\beta$ is a string of symbols, with no multiplication implied. To ascertain if $(A, B) \leftarrow\!\!\!/ (\alpha, \beta)$ using (4), we have to see if we can find 16 probabilities

$$Q_{ijkl} = \Pr\left(H_{1^\alpha} = i \wedge H_{1^\beta} = j \wedge H_{2^\alpha} = k \wedge H_{2^\beta} = l\right),$$
$$i, j, k, l \in \{yes, no\}, \quad (5)$$

for four binary variables $\left\{H_{1^\alpha}, H_{2^\beta}, H_{1^\alpha}, H_{2^\beta}\right\}$, subject to the basic constraints

$$Q_{ijkl} \geq 0, \quad \sum_{ijkl} Q_{ijkl} = 1, \quad (6)$$

and such that

$$\sum_{kl} Q_{ijkl} = P_{ij}1^\alpha1^\beta, \quad \sum_{jk} Q_{ijkl} = P_{il}1^\alpha2^\beta,$$
$$\sum_{il} Q_{ijkl} = P_{kj}2^\alpha1^\beta, \quad \sum_{ij} Q_{ijkl} = P_{kl}2^\alpha2^\beta, \quad (7)$$

for all $i, j, k, l \in \{yes, no\}$. Indeed,

$$\Pr\left(H_{1^\alpha} = i \wedge H_{1^\beta} = j\right) \overset{\text{by (5)}}{=} \sum_{kl} Q_{ijkl} = P_{ij}1^\alpha1^\beta$$
$$= \Pr\left(A_{1^\alpha1^\beta} = i \wedge B_{1^\alpha1^\beta} = j\right),$$

which shows that the first of the equations (7) is equivalent to $(H_{1^\alpha}, H_{1^\beta}) \sim (A_{1^\alpha1^\beta}, B_{1^\alpha1^\beta})$, the application of (4) to $x^\alpha y^\beta = 1^\alpha1^\beta$; and analogously for the other three equations.

Note that (7) implies marginal selectivity. For instance, it follows from (7) that

$$\Pr\left(A_{1^\alpha1^\beta} = i\right) = \sum_j P_{ij}1^\alpha1^\beta = \sum_j \sum_{kl} Q_{ijkl} = \sum_{jkl} Q_{ijkl} = \Pr\left(H_{1^\alpha} = i\right),$$

and

$$\Pr\left(A_{1^\alpha2^\beta} = i\right) = \sum_l P_{il}1^\alpha2^\beta = \sum_l \sum_{jk} Q_{ijkl} = \sum_{jkl} Q_{ijkl} = \Pr\left(H_{1^\alpha} = i\right),$$

that is, $\Pr(A_{1^\alpha1^\beta} = i) = \Pr(A_{1^\alpha2^\beta} = i)$.

Example 1.1. Let the dependence of $\{A, B\}$ on $\{\alpha, \beta\}$ be described by the distributions

Treatment = $1^\alpha1^\beta$	B = yes	B = no	
A = yes	0.5	0	0.5
A = no	0	0.5	0.5
	0.5	0.5	

Treatment = $1^\alpha2^\beta$	B = yes	B = no	
A = yes	0	0.5	0.5
A = no	0.5	0	0.5
	0.5	0.5	

Treatment = $2^\alpha1^\beta$	B = yes	B = no	
A = yes	0	0.5	0.5
A = no	0.5	0	0.5
	0.5	0.5	

Treatment = $2^\alpha 2^\beta$	B = yes	B = no	
A = yes	0.5	0	0.5
A = no	0	0.5	0.5
	0.5	0.5	

Note that marginal selectivity here is satisfied trivially, as the marginal distributions remain fixed. Consider the distribution of $\{H_{1^\alpha}, H_{1^\beta}, H_{2^\alpha}, H_{2^\beta}\}$ with

$$Q_{ijkl} = \begin{cases} .5 & \text{for } ijkl = 0011 \\ .5 & \text{for } ijkl = 1100 \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that this distribution satisfies (6) and (7), hence also (4). By the joint distribution criterion, we conclude that $\{A, B\} \leftarrow\!\!\!\!\wp \{\alpha, \beta\}$.

Example 1.2. No such probabilities Q_{ijkl} can be found for the distributions

Treatment = $1^\alpha 1^\beta$	B = yes	B = no	
A = yes	0.5	0	0.5
A = no	0	0.5	0.5
	0.5	0.5	

Treatment = $1^\alpha 2^\beta$	B = yes	B = no	
A = yes	0.5	0	0.5
A = no	0	0.5	0.5
	0.5	0.5	

Treatment = $2^\alpha 1^\beta$	B = yes	B = no	
A = yes	0.5	0	0.5
A = no	0	0.5	0.5
	0.5	0.5	

Treatment = $2^\alpha 2^\beta$	B = yes	B = no	
A = yes	0	0.5	0.5
A = no	0.5	0	0.5
	0.5	0.5	

so $\{A, B\}$ are not selectively influenced by $\{\alpha, \beta\}$ in this case. This can be shown by direct algebra, but there is a simpler method: this dependence of $\{A, B\}$ on $\{\alpha, \beta\}$ fails a distance test. Indeed, let us transform *yes* into 0 and *no* into 1, and let us choose the exponent $\rho = 1$ (although in this example the value of ρ does not matter). Then

$$D1^\alpha 1^\beta = D1^\alpha 2^\beta = D2^\alpha 1^\beta$$
$$= |0-0| \cdot 0.5 + |0-1| \cdot 0 + |1-0| \cdot 0 + |1-1| \cdot 0.5 = 0,$$
$$D2^\alpha 2^\beta = |0-0| \cdot 0 + |0-1| \cdot 0.5 + |1-0| \cdot 0.5 + |1-1| \cdot 0 = 1,$$

and

$$\max\left\{D1^\alpha 1^\beta, D1^\alpha 2^\beta, D2^\alpha 1^\beta, D2^\alpha 2^\beta\right\}$$
$$= 1 > \frac{1}{2} = \left(D1^\alpha 1^\beta + D1^\alpha 2^\beta + D2^\alpha 1^\beta + D2^\alpha 2^\beta\right)/2,$$

which contravenes (3).

In this paper the joint distribution criterion is formulated in complete generality, for arbitrary sets of random entities and corresponding sets of external factors.

1.5. THE NEED FOR GENERALIZATION

In a controlled experiment or systematic survey we usually focus on a small number of random entities, such as which of several responses is given and how long it has taken, and try to selectively target some of them by experimental manipulations, or selectively relate them to concomitant factors. Relatively small networks of random entities and external factors are therefore of paramount practical importance. But a network of random entities and the set of external factors that may be thought to affect them selectively can be quite large, even infinitely large, in theoretical considerations dealing with complex observable behaviors, such as a person's activities within a typical day, or unobservable "mental networks" behind even relatively simple tasks, such as pushing a key in response to a stimulus varying in two binary properties (see Dzhafarov, Schweickert, and Sung, 2004, for an example). Random processes are routinely used in modeling simple forms of decision making (see, e.g., Diederich and Busemeyer, 2003). Any random process can be viewed as a system of stochastically interdependent random entities indexed by "intervention values" (including "no intervention") at every moment of time. An intervention α at moment t_1 can be thought to selectively affect a portion of the random process in some interval $[t_1, t_2]$ (perhaps even with $t_2 = t_1$), and the problem arises as to how to identify such an interval from the observed joint distribution of the random entities constituting the process. It is important therefore to be able to apply the notion of selective influence to arbitrary, finite and infinite, systems of random entities, and external factors.

2. CONVENTIONS AND NOTATION

A *factor* is defined as a non-empty set of *factor points* (a dummy factor can be defined as a set containing a single point). Denoting factors by lowercase Greek letters, α, β, γ, …, the factor points of, say, factor α are formally pairs $(x, '\alpha')$ consisting of a factor value (or level), x, and a unique factor name, 'α' (read: value/level x of factor α). This ensures that no two distinct factors have common points: e.g., level 1 of factor 'size' $(1, 'size')$, is distinct from level 1 of factor 'shape' $(1, 'shape')$. It is convenient to write x^α in place of $(x, '\alpha')$: 1^{shape}, $(50 \text{ db})^{intensity}$, $present^{left\ stilumus}$, etc.

Let Φ be a non-empty set of factors. A set ϕ containing precisely one factor point $(x_\alpha, '\alpha') = x_\alpha^\alpha$ for each factor α in Φ,

$$\phi = \left\{x_\alpha^\alpha\right\}_{\alpha \in \Phi} \in \prod \Phi,$$

is called a *treatment*.[3] When the set of factors Φ is finite, treatments will be presented as strings of factor points, without commas or parentheses: $x^\alpha y^\beta z^\gamma$, $x_1^{\mu_1} x_2^{\mu_2} \ldots x_k^{\mu_k}$, etc.

Example 2.1. If $\Phi = \{\alpha, \beta, \gamma\}$ with $\alpha = \{1^\alpha, 2^\alpha\}$, $\beta = \{1^\beta, 2^\beta\}$, and $\gamma = \{1^\gamma\}$, then the treatments ϕ (written as strings) are $1^\alpha 1^\beta 1^\gamma$, $1^\alpha 2^\beta 1^\gamma$, $2^\alpha 1^\beta 1^\gamma$, and $2^\alpha 2^\beta 1^\gamma$.

A *random entity A* is a triad consisting of a measurable function $f : \mathfrak{A}' \to \mathfrak{A}$, a *sample (probability) space* $(\mathfrak{A}', \Sigma', M')$, and an *observation (measurable) space* $(\mathfrak{A}, \Sigma)$, on which f induces a probability measure M. Traditionally, A is simply identified with f, the sample space and the observation space being assumed implicitly, or A is viewed as the identity function on $\mathfrak{A}$, with $(\mathfrak{A}', \Sigma', M') = (\mathfrak{A}, \Sigma, M)$. The latter view is often the only practical one, as we almost never know anything about a sample space as separate from the observation space.

A *random variable* is a random entity whose observation space is a subset of reals endowed with the Borel sigma-algebra.[4]

Given an arbitrary indexing set Ω, any set of random entities whose measurable functions $\{f_\omega : \mathfrak{A}' \to \mathfrak{A}_\omega\}_{\omega \in \Omega}$ map from one and the same sample space $(\mathfrak{A}', \Sigma', M')$ into respective observation spaces $\{(\mathfrak{A}_\omega, \Sigma_\omega)\}_{\omega \in \Omega}$ possesses a *joint distribution*, i.e., a probability measure M induced by M' on the product space $\otimes_{\omega \in \Omega}(\mathfrak{A}_\omega, \Sigma_\omega)$.[5]

Example 2.2. Let the sample space consist of $\mathfrak{A}' = \{0, 1\} \times \{0, 1\} \times \{0, 1\}$, $\Sigma' = 2^{\mathfrak{A}'}$, and M' derived from elementary probabilities p_{ijk} ($i, j, k \in \{0, 1\}$). Then the random variables A, B, C defined on $(\mathfrak{A}', \Sigma', M')$ by the coordinate projections $a : (i, j, k) \mapsto i$, $b : (i, j, k) \mapsto j$, and $c : (i, j, k) \mapsto k$, respectively ($i, j, k \in \{0, 1\}$), possess the joint distribution $M = M'$ on $(\{0, 1\}, 2^{\{0,1\}}) \otimes (\{0, 1\}, 2^{\{0,1\}}) \otimes (\{0, 1\}, 2^{\{0,1\}}) = (\mathfrak{A}', \Sigma')$.

Two random entities A and B defined on different sample spaces are called (stochastically or probabilistically) *unrelated* (see Dzhafarov and Gluhovsky, 2006). They *do not* possess a joint distribution. Note that two unrelated random variables can be identically distributed – if they map into one and the same observation space on which they induce one and the same probability measure.

Throughout this paper we deal with a set of probabilistically unrelated random entities $\{A_\phi\}_{\phi \in \Pi\Phi}$ indexed by treatments $\phi \in \Pi\Phi$, with measures $\{M_\phi\}_{\phi \in \Pi\Phi}$ induced on one and the same observation space $(\mathfrak{A}, \Sigma)$. For convenience, we refer to A_ϕ as "a random entity A at ϕ", as if A_ϕ and $A_{\phi'}$ for $\phi \neq \phi'$ were "a single" entity A at two

different treatments. Note that A_ϕ and $A_{\phi'}$ are defined on different sample spaces: they do not possess a joint distribution. In particular, they are not mutually independent.[6]

Example 2.3. The random variables A and B in the examples of Section 1 are called A and B by the abuse of language just mentioned. Strictly speaking we deal with four pairwise stochastically unrelated $A_{1^\alpha 1^\beta}, A_{1^\alpha 2^\beta}, A_{2^\alpha 1^\beta}, A_{2^\alpha 2^\beta}$ and four pairwise stochastically unrelated $B_{1^\alpha 1^\beta}, B_{1^\alpha 2^\beta}, B_{2^\alpha 1^\beta}, B_{2^\alpha 2^\beta}$, such that $A_{x^\alpha y^\beta}$ and $B_{z^\alpha u^\beta}$ possess a joint distribution if and only if $x^\alpha y^\beta = z^\alpha u^\beta$.

A *set of random entities* $\{A^\omega\}_{\omega \in \Omega}$ on one and the same sample space is a random entity whose observation space $(\mathfrak{A}, \Sigma)$ is the conventionally understood product of the observation spaces $(\mathfrak{A}_\omega, \Sigma_\omega)$ for $A^\omega, \omega \in \Omega$. If the set of random entities $\{A^\omega\}_{\omega \in \Omega}$ depends on Φ, we present $\{A^\omega\}_{\omega \in \Omega}$ at a treatment ϕ as $\{A_\phi^\omega\}_{\omega \in \Omega}$ instead of the more correct but less convenient $(\{A^\omega\}_{\omega \in \Omega})_\phi$.

Example 2.4. The variables $\{A, B, C\}$ of Example 2.2 depend on the factors $\Phi = \{\alpha, \beta, \gamma\}$ of Example 2.1 if $M = M'$ is viewed as a function of $\phi = 1^\alpha 1^\beta 1^\gamma$, $1^\alpha 2^\beta 1^\gamma$, …. In this example $\mathfrak{A}'_\phi = \mathfrak{A}', \Sigma'_\phi = \Sigma'$, and $\{a_\phi, b_\phi, c_\phi\} = \{a, b, c\}$.

3. SELECTIVE INFLUENCE

In accordance with the previous section, given a set of factors Φ, a corresponding set of random entities is denoted $\{A^\alpha\}_{\alpha \in \Phi}$. For each $\alpha \in \Phi$, the entity A^α may in fact be a shortcut notation for a set of stochastically unrelated random entities indexed by different treatments, $\{A_\phi^\alpha\}_{\phi \in \Pi\Phi}$. In other words, A_ϕ^α is treated as a random entity A corresponding to factor α and taken at treatment ϕ. The complete notation for the set of random entities $\{A^\alpha\}_{\alpha \in \Phi}$ then is

$$\left\{ \left\{ A_\phi^\alpha \right\}_{\alpha \in \Phi} \right\}_{\phi \in \Pi\Phi}, \tag{8}$$

where the elements of $\{A_\phi^\alpha\}_{\alpha \in \Phi}$, for a given ϕ, are stochastically interrelated (possess a joint distribution), while the sets $\{A_\phi^\alpha\}_{\alpha \in \Phi}$ and $\{A_{\phi'}^\alpha\}_{\alpha \in \Phi}$, for distinct ϕ, ϕ', are stochastically unrelated. It is more convenient, however, not to use this explicit notation and to speak instead of $\{A^\alpha\}_{\alpha \in \Phi}$ depending on Φ.

Definition 3.1. Let a set of random entities $\{A^\alpha\}_{\alpha \in \Phi}$ indexed by a set of factors Φ depend on this set of factors (i.e., be presentable as (8)). We say that the dependence of $\{A^\alpha\}_{\alpha \in \Phi}$ on Φ is *marginally selective* (satisfies the property of *marginal selectivity*) if, for any subset $\Phi_1 \subset \Phi$ and any $\phi_1 \in \Pi\Phi_1$, the distribution of $\{A_\phi^\alpha\}_{\alpha \in \Phi_1}$ is the same for all treatments ϕ containing ϕ_1 (that is, it does not depend on $\{x^\beta \in \phi : \beta \in \Phi - \Phi_1\}$).

The notion of marginal selectivity was introduced by Townsend and Schweickert (1989), for two random variables. In Dzhafarov (2003) it was generalized to a finite set of random variables under

---

[3]Strictly speaking, an element of the Cartesian product $\Pi\Phi$ is a choice function, $\{(\alpha, x_\alpha^\alpha)\}_{\alpha \in \Phi}$, whereas a treatment ϕ is the range of a choice function, $\{x_\alpha^\alpha\}_{\alpha \in \Phi}$. We conveniently confuse the two notions. Also for convenience only, in this paper we assume "completely crossed design", i.e., that every member of $\Pi\Phi$ is a possible treatment. With only slight modifications $\Pi\Phi$ can be replaced with any nonempty subset thereof.

[4]A random entity A with $\mathfrak{A}$ a finite or infinite denumerable set and Σ the set of all its subsets can also be (and traditionally is) considered a random variable, because such an $\mathfrak{A}$ can always be injectively mapped into the set of reals, or into a partition of an interval of reals.

[5]Recall that in the product measurable space $\otimes_{\omega \in \Omega}(\mathfrak{A}_\omega, \Sigma_\omega) = (\mathfrak{A}, \Sigma)$ the set $\mathfrak{A}$ is the Cartesian product $\Pi_{\omega \in \Omega} \mathfrak{A}_\omega$, while $\Sigma = \otimes_{\omega \in \Omega} \Sigma_\omega$ is the smallest sigma algebra containing all sets of the form $\mathfrak{a}_{\omega_0} \times \Pi_{\omega \in \Omega - \{\omega_0\}} \mathfrak{A}_\omega$, $\mathfrak{a}_{\omega_0} \in \Sigma_{\omega_0}$.

[6]We could have extended the scope of this definition by allowing A_ϕ to be a function $f_\phi : \mathfrak{A}'_\phi \to \mathfrak{A}_\phi$ relating $(\mathfrak{A}'_\phi, \Sigma'_\phi, M'_\phi)$ to $(\mathfrak{A}_\phi, \Sigma_\phi)$, i.e., by allowing the set and the sigma algebra, not only the measure M_ϕ, to depend on treatment ϕ. This would have, however, made our abuse of language (in treating different A_ϕ's as a single A at different ϕ's) even more abusive. Moreover, this general approach can always be reduced to the set-up with a ϕ-independent $(\mathfrak{A}, \Sigma)$ by putting $\mathfrak{A} = \bigcup_{\phi \in \Pi\Phi}(\mathfrak{A}_\phi \times \{\phi\})$ and Σ the sigma algebra consisting of all countable unions of the sets $\mathfrak{a} \times \{\phi\}$ for all $\mathfrak{a} \in \Sigma_\phi$ and all $\phi \in \Pi\Phi$.

the name of *complete marginal selectivity*. The adjective "complete" (omitted in the present paper for simplicity) distinguishes this notion from a weaker and less useful generalization of Townsend and Schweickert's term: for any factor $\alpha \in \Phi$ and any treatment ϕ, the distribution of A_ϕ^α does not depend on $\{x^\beta \in \phi : \beta \in \Phi - \{\alpha\}\}$.

Note that Definition 3.1 does not mean that for distinct treatments ϕ and ϕ' which include $\phi_1 = \{x^\beta \in \phi : \beta \in \Phi_1\} = \{x^\beta \in \phi' : \beta \in \Phi_1\}$,

$$\left\{A_{\phi_1 \cup (\phi - \phi_1)}^\alpha\right\}_{\alpha \in \Phi_1} = \left\{A_{\phi_1 \cup (\phi' - \phi_1)}^\alpha\right\}_{\alpha \in \Phi_1}.$$

This equality is not legitimate as the two sets of random variables do not possess a joint distribution. One can only say that

$$\left\{A_{\phi_1 \cup (\phi - \phi_1)}^\alpha\right\}_{\alpha \in \Phi_1} \sim \left\{A_{\phi_1 \cup (\phi' - \phi_1)}^\alpha\right\}_{\alpha \in \Phi_1},$$

where, as before, $\sim$ means "is distributed as."

Example 3.2. Let the variables $\{A, B, C\}$ of Example 2.2 be indexed by the factors $\Phi = \{\alpha, \beta, \gamma\}$ of Example 2.1, respectively. That is, $\{A, B, C\}$ stands for $\{A^\alpha, A^\beta, A^\gamma\}$, in accordance with the general notation of Definition 3.1. Then the dependence of $\{A^\alpha, A^\beta, A^\gamma\}$ on $\{\alpha, \beta, \gamma\}$ is marginally selective if and only if

$$\left\{A^\alpha, A^\gamma\right\}_{1^\alpha 1^\beta 1^\gamma} \sim \left\{A^\alpha, A^\gamma\right\}_{1^\alpha 2^\beta 1^\gamma} \& \left\{A^\alpha, A^\gamma\right\}_{2^\alpha 1^\beta 1^\gamma} \sim \left\{A^\alpha, A^\gamma\right\}_{2^\alpha 2^\beta 1^\gamma},$$

$$\left\{A^\beta, A^\gamma\right\}_{1^\alpha 1^\beta 1^\gamma} \sim \left\{A^\beta, A^\gamma\right\}_{2^\alpha 1^\beta 1^\gamma} \& \left\{A^\beta, A^\gamma\right\}_{1^\alpha 2^\beta 1^\gamma} \sim \left\{A^\beta, A^\gamma\right\}_{2^\alpha 2^\beta 1^\gamma},$$

$$A^\alpha_{1^\alpha 1^\beta 1^\gamma} \sim A^\alpha_{1^\alpha 2^\beta 1^\gamma} \& A^\alpha_{2^\alpha 1^\beta 1^\gamma} \sim A^\alpha_{2^\alpha 2^\beta 1^\gamma},$$

$$A^\beta_{1^\alpha 1^\beta 1^\gamma} \sim A^\beta_{2^\alpha 1^\beta 1^\gamma} \& A^\beta_{1^\alpha 2^\beta 1^\gamma} \sim A^\beta_{2^\alpha 2^\beta 1^\gamma},$$

$$A^\gamma_{1^\alpha 1^\beta 1^\gamma} \sim A^\gamma_{1^\alpha 2^\beta 1^\gamma} \sim A^\gamma_{2^\alpha 1^\beta 1^\gamma} \sim A^\gamma_{2^\alpha 2^\beta 1^\gamma}.$$

Under marginal selectivity, it is sometimes admissible, by abuse of notation, to write $\{A_\phi^\alpha\}_{\alpha \in \Phi_1}$ as $\{A_{\phi_1}^\alpha\}_{\alpha \in \Phi_1}$ where $\phi_1 = \{x^\alpha \in \phi : \alpha \in \Phi_1\}$. Thus, in Example 3.2, it may be convenient to present both $\{A^\alpha, A^\gamma\}_{1^\alpha 1^\beta 1^\gamma}$ and $\{A^\alpha, A^\gamma\}_{1^\alpha 2^\beta 1^\gamma}$ as $\{A^\alpha, A^\gamma\}_{1^\alpha 1^\gamma}$, present both $A^\beta_{1^\alpha 1^\beta 1^\gamma}$ and $A^\beta_{2^\alpha 1^\beta 1^\gamma}$ as $A^\beta_{1^\beta}$, etc. This does not lead to complications provided one remembers that, say, $A^\alpha_{1^\alpha 1^\beta 1^\gamma}$ and $A^\alpha_{1^\alpha 2^\beta 1^\gamma}$ are identically distributed rather than identical. One can therefore deal with $A_{1^\alpha}^\alpha$ and $A_{2^\alpha}^\alpha$ in all considerations involving only A^α, or with $A_{1^\beta}^\beta$ and $A_{2^\beta}^\beta$ in all considerations involving only A^β. It would be incorrect, however, to speak of stochastic relationships between $A_{1^\alpha}^\alpha$ or $A_{2^\alpha}^\alpha$ and $A_{1^\beta}^\beta$ or $A_{2^\beta}^\beta$: to depict such relationships one needs both A^α and A^β to be double-indexed by the factor points involved (unless we have selective influence in addition to marginal selectivity, as explained in the next section).

Example 3.3. Continuing Example 3.2, let us return to writing $\{A, B, C\}$ instead of $\{A^\alpha, A^\beta, A^\gamma\}$. We can omit the factor γ when dealing with $\{A, B\}$ only and write $\{A, B\}_{1^\alpha 1^\beta}, \{A, B\}_{1^\alpha 2^\beta}, \{A, B\}_{2^\alpha 1^\beta}, \{A, B\}_{2^\alpha 2^\beta}$. Let the distributions of these random pairs be

Treatment = $1^\alpha 1^\beta$	$B = 0$	$B = 1$	
$A = 0$	0.6	0	0.6
$A = 1$	0	0.4	0.4
	0.6	0.4	

Treatment = $1^\alpha 2^\beta$	$B = 0$	$B = 1$	
$A = 0$	0	0.6	0.6
$A = 1$	0.4	0	0.4
	0.4	0.6	

Treatment = $2^\alpha 1^\beta$	$B = 0$	$B = 1$	
$A = 0$	0.3	0.2	0.5
$A = 1$	0.3	0.2	0.5
	0.6	0.4	

Treatment = $2^\alpha 2^\beta$	$B = 0$	$B = 1$	
$A = 0$	0.25	0.25	0.5
$A = 1$	0.15	0.35	0.5
	0.4	0.6	

We verify that marginal selectivity holds for A and B and introduce the abridged indexing, $A_{1^\alpha}, A_{2^\alpha}$, etc.:

$$A_{1^\alpha 1^\beta} \sim A_{1^\alpha 2^\beta} \sim \begin{array}{|c|c|} \hline A = 0 & A = 1 \\ \hline 0.6 & 0.4 \\ \hline \end{array} = A_{1^\alpha},$$

$$A_{2^\alpha 1^\beta} \sim A_{2^\alpha 2^\beta} \sim \begin{array}{|c|c|} \hline A = 0 & A = 1 \\ \hline 0.5 & 0.5 \\ \hline \end{array} = A_{2^\alpha},$$

$$B_{1^\alpha 1^\beta} \sim B_{2^\alpha 1^\beta} \sim \begin{array}{|c|c|} \hline B = 0 & B = 1 \\ \hline 0.6 & 0.4 \\ \hline \end{array} = B_{1^\beta},$$

$$B_{1^\alpha 2^\beta} \sim B_{2^\alpha 2^\beta} \sim \begin{array}{|c|c|} \hline B = 0 & B = 1 \\ \hline 0.4 & 0.6 \\ \hline \end{array} = B_{2^\beta}.$$

We can use the abridged indexing in all considerations involving A alone and B alone. Consider, however, this: from the joint distribution matrices we have

$$A_{1^\alpha 1^\beta} = B_{1^\alpha 1^\beta}, \qquad A_{1^\alpha 2^\beta} = 1 - B_{1^\alpha 2^\beta},$$

$$A_{2^\alpha 1^\beta} \perp B_{2^\alpha 1^\beta}, \qquad \neg\left(A_{2^\alpha 2^\beta} \perp B_{2^\alpha 2^\beta}\right),$$

where $\perp$ indicates stochastic independence and $\neg$ negation (i.e., $A_{2^\alpha 2^\beta}$ and $B_{2^\alpha 2^\beta}$ are not stochastically independent). If now we attempt to use in these relations the abridged indexing, we will run into a contradiction: from $A_{1^\alpha} = B_{1^\beta}$ and $A_{1^\alpha} = 1 - B_{2^\beta}$ we conclude $B_{1^\beta} = 1 - B_{2^\beta}$, but then it is impossible for B_{1^β} to be independent of A_{2^α} and for B_{2^β} not to be. It is therefore necessary to retain the notation $A_{1^\alpha 1^\beta}, A_{1^\alpha 2^\beta}$, etc. in all considerations involving both A and B even though we know that the dependence of $\{A, B\}$ on $\{\alpha, \beta\}$ is marginally selective.

Definition 3.4. Let a set of random entities $\{A^\alpha\}_{\alpha \in \Phi}$ indexed by a set of factors Φ depend on this set of factors. We say that the set $\{A^\alpha\}_{\alpha \in \Phi}$ is *selectively influenced* by Φ, and write

$$\left\{A^\alpha\right\}_{\alpha \in \Phi} \leftarrow\!\!\!\rho\; \Phi,$$

if, for some random entity C and every $x^\alpha \in \alpha \in \Phi$ there is a measurable function f_{x^α} such that, for every treatment ϕ,

$$\left\{A^{\alpha}_{\phi}\right\}_{\alpha\in\Phi} \sim \left\{f_{x^{\alpha}}(C)\right\}_{x^{\alpha}\in\phi}.$$

Remark 3.5. Alternatively, one could posit, for every treatment ϕ,

$$\left\{A^{\alpha}_{\phi}\right\}_{\alpha\in\Phi} = \left\{f_{x^{\alpha}}(C_{\phi})\right\}_{x^{\alpha}\in\phi},$$

where

$$\left\{C_{\phi}\right\}_{\phi\in\Pi\Phi}$$

is a set of *pairwise unrelated* random entities all distributed as C. This formulation is more cumbersome but it correctly emphasizes the stochastic unrelatedness of $\{A^{\alpha}_{\phi}\}_{\alpha\in\Phi}$ for different treatments ϕ. Definition 3.4, however, is more parsimonious, as the stochastic unrelatedness property is known from the context.

Remark 3.6. If applied to finite sets Φ, Definition 3.4 becomes equivalent to the formulations of selective influence given in Dzhafarov (2003), Dzhafarov and Gluhovsky (2006), and Kujala and Dzhafarov (2008). Even for the finite case, however, the present definition is mathematically more rigorous, and it profits from the precision offered by the notation $x^{\alpha} = (x, \text{`}\alpha\text{'})$ for factor points. More importantly, it can be seen more immediately than the previous definitions to be reformulable into the *joint distribution criterion* for selective influence, as discussed in the next section.

The following statements are obvious.

Lemma 3.7. *If* $\{A^{\alpha}\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$, *then*

(i) $\{A^{\alpha}\}_{\alpha\in\Phi_1} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi_1$ *for any* $\Phi_1 \subset \Phi$;
(ii) *the dependence of* $\{A^{\alpha}\}_{\alpha\in\Phi}$ *on* Φ *is marginally selective.*

That is (refer to Section 1.2), selective influence has the nestedness property and implies marginal selectivity.

The next lemma says that if a set of random entities $\{A^{\alpha}\}_{\alpha\in\Phi}$ is selectively influenced by Φ, then the set of individually transformed versions of these random variables is also selectively influenced by Φ (refer to the first property in Section 1.2). "Individual transformations" of A^{α} can be different for different factor points x^{α}.

Lemma 3.8. *If* $\{A^{\alpha}\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$ *then* $\{B^{\alpha}\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$, *where, for any* $\alpha \in \Phi$, *any* $x^{\alpha} \in \alpha$, *and any treatment* ϕ *containing* x^{α},

$$B^{\alpha}_{\phi} = h_{x^{\alpha}}\left(A^{\alpha}_{\phi}\right),$$

for some measurable function $h_{x^{\alpha}}$.

Proof. By definition,

$$\left\{A^{\alpha}_{\phi}\right\}_{\alpha\in\Phi} \sim \left\{f_{x^{\alpha}}(C)\right\}_{x^{\alpha}\in\phi},$$

which implies

$$\left\{B^{\alpha}_{\phi}\right\}_{\alpha\in\Phi} \sim \left\{h_{x^{\alpha}}\left(f_{x^{\alpha}}(C)\right)\right\}_{x^{\alpha}\in\phi} = \left\{g_{x^{\alpha}}(C)\right\}_{x^{\alpha}\in\phi},$$

where $g_{x^{\alpha}} \equiv h_{x^{\alpha}} \circ f_{x^{\alpha}}$ is a measurable function. $\square$

4. THE JOINT DISTRIBUTION CRITERION

Definition 3.4 suggests a way of looking at the selective influence relation directly in terms of the (product) *observation space* for the system of the random entities involved, making the overt reconstruction of C and the functions $f_{x^{\alpha}}$ unnecessary (or trivial, as in the proof of the theorem below).

Theorem 4.1. *A necessary and sufficient condition for*

$$\left\{A^{\alpha}\right\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$$

is the existence of a jointly distributed system

$$\left\{H_{x^{\alpha}}\right\}_{x^{\alpha}\in\cup\Phi}$$

such that for every subset ϕ *of* $\cup\Phi$ *that forms a treatment (i.e., belongs to* $\Pi\Phi$),

$$\left\{H_{x^{\alpha}}\right\}_{x^{\alpha}\in\phi} \sim \left\{A^{\alpha}_{\phi}\right\}_{\alpha\in\Phi}.$$

Remark 4.2. We call this *the joint distribution criterion* for selective influence.

Proof. The necessity is proved by observing that if $\{A^{\alpha}\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$, then the system

$$\left\{H_{x^{\alpha}}\right\}_{x^{\alpha}\in\cup\Phi} = \left\{f_{x^{\alpha}}(C)\right\}_{x^{\alpha}\in\cup\Phi}$$

is a *jointly distributed* system of random entities. To prove the sufficiency, define

$$C = \left\{H_{x^{\alpha}}\right\}_{x^{\alpha}\in\cup\Phi},$$

and, for every x^{α}, define

$$f_{x^{\alpha}}(C) = \mathrm{Proj}_{x^{\alpha}}(C) = H_{x^{\alpha}},$$

where $\mathrm{Proj}_{x^{\alpha}}$ denotes the x^{α}th coordinate projection. $\square$

As a very simple application of the joint distribution criterion we prove the following (intuitively quite obvious) statement.

Lemma 4.3. *If the dependence of* $\{A^{\alpha}\}_{\alpha\in\Phi}$ *on* Φ *is marginally selective and* $\{A^{\alpha}_{\phi}\}_{\alpha\in\Phi}$ *is a set of mutually independent random entities for every treatment* ϕ, *then* $\{A^{\alpha}\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$.

Proof. For any factor $\alpha \in \Phi$, marginal selectivity implies that the distribution of A^{α}_{ϕ} depends only on the factor point $x^{\alpha} \in \phi$. Form the set $\{H_{x^{\alpha}}\}_{x^{\alpha}\in\cup\Phi}$ consisting of mutually independent random entities such that, for any x^{α}, $H_{x^{\alpha}} \sim A^{\alpha}_{x^{\alpha}}$. Then, for every treatment ϕ, $\{H_{x^{\alpha}}\}_{x^{\alpha}\in\phi} \sim \{A^{\alpha}_{\phi}\}_{\alpha\in\Phi}$, and Theorem 4.1 implies $\{A^{\alpha}\}_{\alpha\in\Phi} \looparrowleft\!\!\!\!\!\!\!\!\!\multimap \Phi$. $\square$

In Section 1.4 we have seen illustrations of the criterion on interdependent random entities. Here is another example.

Example 4.4. Let $\Phi = \{\alpha, \beta\}$, $\alpha = \{1^{\alpha}, 2^{\alpha}\}$, $\beta = \{1^{\beta}, 2^{\beta}\}$, and let $\{A_{\phi}, B_{\phi}\}$ for every treatment ϕ be a pair of Bernoulli variables. Consider the distributions below:

Treatment $= 1^{\alpha}1^{\beta}$	$B = 0$	$B = 1$	
$A = 0$	0.5	0	0.5
$A = 1$	0	0.5	0.5
	0.5	0.5	

Treatment = $1^\alpha 2^\beta$	$B = 0$	$B = 1$	
$A = 0$	0	0.5	0.5
$A = 1$	0.5	0	0.5
	0.5	0.5	

Treatment = $2^\alpha 1^\beta$	$B = 0$	$B = 1$	
$A = 0$	0	0.5	0.5
$A = 1$	0.5	0	0.5
	0.5	0.5	

Treatment = $2^\alpha 2^\beta$	$B = 0$	$B = 1$	
$A = 0$	0	0.5	0.5
$A = 1$	0.5	0	0.5
	0.5	0.5	

The criterion of the joint distribution of $\{H_{1^\alpha}, H_{1^\beta}, H_{2^\alpha}, H_{2^\beta}\}$ rejects the possibility of $\{A, B\} \nleftarrow \{\alpha, \beta\}$, as it can be shown by direct algebra that there are no 16 probabilities Q_{ijkl} generating the distributions in question (cf. Examples 1.1 and 1.2). We do not need to provide such a demonstration as it is obvious in this case from probabilistic considerations that $\{H_{1^\alpha}, H_{1^\beta}, H_{2^\alpha}, H_{2^\beta}\}$ cannot be jointly distributed and satisfy

$$\left\{H_{1^\alpha}, H_{1^\beta}\right\} \sim \left\{A_{1^\alpha 1^\beta}, B_{1^\alpha 1^\beta}\right\}, \quad \left\{H_{1^\alpha}, H_{2^\beta}\right\} \sim \left\{A_{1^\alpha 2^\beta}, B_{1^\alpha 2^\beta}\right\},$$
$$\left\{H_{2^\alpha}, H_{1^\beta}\right\} \sim \left\{A_{2^\alpha 1^\beta}, B_{2^\alpha 1^\beta}\right\}, \quad \left\{H_{2^\alpha}, H_{2^\beta}\right\} \sim \left\{A_{2^\alpha 2^\beta}, B_{2^\alpha 2^\beta}\right\}.$$

Otherwise the four joint distribution matrices shown above would have implied, respectively, $H_{1^\alpha} = H_{1^\beta}$, $H_{1^\alpha} = 1 - H_{2^\beta}$, $H_{2^\alpha} = 1 - H_{1^\beta}$, and $H_{2^\alpha} = 1 - H_{2^\beta}$, which are not mutually compatible equations.

If we take the numerical values of A and B in the last example as they are, then with any exponent $\rho \geq 1$ (see Section 1.3) the distance test is passed:

$$\max\left\{D1^\alpha 1^\beta, D1^\alpha 2^\beta, D2^\alpha 1^\beta, D2^\alpha 2^\beta\right\}$$
$$= 1 \leq \frac{3}{2} = \left(D1^\alpha 1^\beta + D1^\alpha 2^\beta + D2^\alpha 1^\beta + D2^\alpha 2^\beta\right)/2.$$

This is just another demonstration that a distance test is only a necessary condition for selective influence: a dependence of random entities on external factors can pass such a test but still fail the joint distribution criterion. It is instructive to see, however, in reference to Lemma 3.8, that with appropriately chosen transformations of the random variables the distributions in question can be made to fail the respective distance tests. Thus, the possibility of selective influence in Example 4.4 will be rejected if we apply the simple transformation

$$h_{y^\beta}\left(B_{x^\alpha y^\beta}\right) = 1 - B_{x^\alpha y^\beta}, \quad x^\alpha = 1^\alpha, 2^\alpha, \quad y^\beta = 1^\beta, 2^\beta,$$

while leaving $A_{x^\alpha y^\beta}$ untransformed. The distributions then become essentially the same as in Example 1.2. The distance tests therefore can be conjectured to have considerable rejection power if one combines it with adeptly chosen transformations (see the open question we pose at the conclusion of the paper).

In any case, the exceptional simplicity of these tests makes it worthwhile to always consider them before applying the joint distribution criterion.

5. DISTANCE TESTS

In Kujala and Dzhafarov (2008) the distance tests were formulated for two variables influenced by two factors in a two-by-two factorial design. In this section we generalize these tests to arbitrary random variables $\{A^\alpha\}_{\alpha \in \Phi}$ whose dependence on factors Φ is marginally selective. Perhaps surprisingly, we show that this generalization requires nothing more and nothing less than applying the original tests to all possible two-by-two factorial designs one can extract from Φ.

Distance tests can be applied to non-numerical random entities only after they have been numerically transformed (thus, for the distance test applied to Example 1.2 we transformed *yes* into 0 and *no* into 1). In this section therefore we confine our discussion to random variables.

We will need some auxiliary notions and notation conventions. Any finite sequence of factor points $(x_1^{\alpha_1}, \ldots, x_n^{\alpha_n})$ is called a *chain*. Chains will be written as strings, $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$, without commas and parentheses (this generalizes the convention we have already used for chains which are finite treatments). Chains can be denoted by capital Roman letters, $X = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ (from the second half of the alphabet, to distinguish them from random variables and entities for which we use the first half). A chain X may be empty or consist of a single element (factor point), x^α. A subsequence of points belonging to a chain forms its *subchain*.

A *concatenation* of two chains X and Y is written as XY. So, we can have chains $x^\alpha X y^\beta$, $x^\alpha XY y^\beta$, $Xx^\alpha y^\beta Y$, $x^\alpha X y^\beta Z$, etc.

The number of points in a chain X is its *cardinality*, $|X|$, and any chain with the smallest cardinality within a set of chains is referred to as a *minimal* chain (in this set). In particular, one can speak of a *minimal subchain* of a chain among all subchains with a certain property (this notion is used in the proof of Theorem 5.11 below).

Definition 5.1. Let the dependence of a set of random variables $\{A^\alpha\}_{\alpha \in \Phi}$ on factors Φ be marginally selective. Let $\rho \geq 1$ be fixed. For any (x^α, y^β) with $\alpha \neq \beta$, we define

$$Dx^\alpha y^\beta = \left\| A_{x^\alpha y^\beta}^\alpha - A_{x^\alpha y^\beta}^\beta \right\|_\rho,$$

where $\|A - B\|_\rho$ for any jointly distributed A and B is defined as

$$\|A - B\|_\rho = \begin{cases} \sqrt[\rho]{\mathrm{E}\left[|A - B|^\rho\right]} & \text{for } 1 \leq \rho < \infty, \\ \mathrm{ess\,sup}|A - B| & \text{for } \rho = \infty. \end{cases}$$

Remark 5.2. Here ess sup is the essential supremum, the lowest upper bound that holds almost surely; it is the limit of $\|A - B\|_\rho$ as $\rho \to \infty$.

Remark 5.3. Note that $Dx^\alpha y^\beta$ is well-defined only under the assumption that the dependence of a set of random variables $\{A^\alpha\}_{\alpha \in \Phi}$ on factors Φ is marginally selective. Otherwise $\|A_{x^\alpha y^\beta}^\alpha - A_{x^\alpha y^\beta}^\alpha\|_\rho$ would not be determined by $x^\alpha y^\beta$ only, and it would not even be legitimate to index the two variables by $x^\alpha y^\beta$ alone.

We will rely on the following result, whose proof we omit as its only non-trivial part follows from the Minkowski inequality (a somewhat abridged proof can be found in Kujala and Dzhafarov, 2008).

Lemma 5.4. *Given a sample space, let $\mathfrak{R}$ be a set of all random variables A, B, … (jointly distributed) on this space. For any $\rho \geq 1$, $\|A - B\|_\rho$ is an extended metric on $\mathfrak{R}$, provided we do not distinguish A,B identical on a set of measure 1.*

Remark 5.5. The adjective "extended" means that ∞ is included in the set of possible values. The norms $\sqrt[\rho]{E[|A - B|^\rho]}$ and ess sup $|A - B|$, as they only involve non-negative values, always exist, finite or infinite.

Convention 5.6. In the remainder of this section we will tacitly assume that the dependence of $\{A^\alpha\}_{\alpha \in \Phi}$ on Φ is marginally selective. We will also tacitly assume that ρ in the definition of $\|\ldots\|_\rho$ and D is fixed.

For any chain $X = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ such that $\alpha_1 \neq \alpha_n$ and $\alpha_i \neq \alpha_{i+1}$ for $i = 1, \ldots, n-1$, define

$$DX = \sum_{i=1}^{n-1} Dx_i^{\alpha_i} x^{\alpha_{i+1}}$$

(with the understanding that the sum is zero if n is 0 or 1). The operator D always acts upon the entire chain following it, e.g., $Du^\mu Xv^\nu = Dx^\mu x_1^{\alpha_1} + DX + Dx_n^{\alpha_n} x^\nu$.

Definition 5.7. A chain $x^\alpha X y^\beta$ is said to be *compliant with the chain inequality* (or simply, *compliant*) if $Dx^\alpha X y^\beta \geq Dx^\alpha y^\beta$. The chain is said to be *contravening* (the chain inequality) if $Dx^\alpha X y^\beta < Dx^\alpha y^\beta$.

It follows from this definition that if $x^\alpha X y^\beta$ is contravening or compliant, then $\alpha \neq \beta$ (otherwise $Dx^\alpha y^\beta$ is not defined), and no factor in $x^\alpha X y^\beta$ occurs twice in succession. For a chain to be contravening, in addition, X must be non-empty (i.e., $|x^\alpha X y^\beta| \geq 3$; Lemma 5.9 below shows that in fact $|x^\alpha X y^\beta| \geq 4$). A non-contravening chain need not be compliant: it may, e.g., be any chain with fewer than 3 elements, or it can be any chain of the form $x^\alpha X y^\alpha$. Analogously, a non-compliant chain is not necessarily contravening.[7]

Lemma 5.8. *Let $U = Xy^\beta Y z^\gamma Z$ be a contravening chain with a compliant subchain $y^\beta Y z^\gamma$. Then $U^* = Xy^\beta z^\gamma Z$ (i.e., U without Y) is a contravening subchain of U.*

Proof. Let x^α and u^δ be the first and the last elements of U, respectively (then necessarily $\alpha \neq \delta$). Note that x^α may coincide with y^β or u^δ with z^γ (but not both). From

$$Dx^\alpha u^\delta > DXy^\beta + Dy^\beta Y z^\gamma + Dz^\gamma Z$$

and

$$Dy^\beta Y z^\gamma \geq Dy^\beta z^\gamma$$

we get

$$Dx^\alpha u^\delta > DXy^\beta + Dy^\beta z^\gamma + Dz^\gamma Z. \qquad \square$$

Lemma 5.9. *Every triadic chain $x^\alpha y^\beta z^\gamma$ with pairwise distinct α, β, γ is compliant.*

Proof. Denoting the random variables corresponding to the factors α, β, γ, by A, B, C, respectively, marginal selectivity implies

$$\begin{cases} Dx^\alpha y^\beta = \left\| A_{x^\alpha y^\beta} - B_{x^\alpha y^\beta} \right\|_\rho = \left\| A_{x^\alpha y^\beta z^\gamma} - B_{x^\alpha y^\beta z^\gamma} \right\|_\rho, \\ Dy^\beta z^\gamma = \left\| B_{y^\beta z^\gamma} - C_{y^\beta z^\gamma} \right\|_\rho = \left\| B_{x^\alpha y^\beta z^\gamma} - C_{x^\alpha y^\beta z^\gamma} \right\|_\rho, \\ Dx^\alpha z^\gamma = \left\| A_{x^\alpha z^\gamma} - C_{x^\alpha z^\gamma} \right\|_\rho = \left\| A_{x^\alpha y^\beta z^\gamma} - C_{x^\alpha y^\beta z^\gamma} \right\|_\rho. \end{cases}$$

Since $\{A_{x^\alpha y^\beta z^\gamma}, B_{x^\alpha y^\beta z^\gamma}, C_{x^\alpha y^\beta z^\gamma}\}$ are jointly distributed, the statement follows from Lemma 5.4.[8] $\qquad \square$

We are ready now to prove the main theorems regarding the distance tests for selective influence.

Theorem 5.10. *Let $\{A^\alpha\}_{\alpha \in \Phi}$ be selectively influenced by Φ. Then any chain $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ such that $\alpha_i \neq \alpha_{i+1}$ for $i = 1, \ldots, n-1$, and $\alpha_1 \neq \alpha_n$ is compliant,*

$$Dx_1^{\alpha_1} x_n^{\alpha_n} \leq \sum_{i=1}^{n-1} Dx_i^{\alpha_i} x_{i+1}^{\alpha_{i+1}}.$$

Proof. By the joint distribution criterion, there is a jointly distributed system $\{H_{x^\alpha}\}_{x^\alpha \in \cup \Phi}$ such that for any $\{x^\alpha, y^\beta\}$ within a treatment (i.e. with $\alpha \neq \beta$),

$$\left\{ A_{x^\alpha y^\beta}^\alpha, A_{x^\alpha y^\beta}^\beta \right\} \sim \left\{ H_{x^\alpha}, H_{y^\beta} \right\}.$$

But then

$$Dx^\alpha y^\beta = \left\| A_{x^\alpha y^\beta} - B_{x^\alpha y^\beta} \right\|_\rho = \left\| H_{x^\alpha} - H_{y^\beta} \right\|_\rho,$$

and the statement of the lemma follows from Lemma 5.4. $\qquad \square$

For the next theorem, recall that we are following Convention 5.6.

Theorem 5.11. *Every contravening chain X contains a contravening tetradic subchain X' of the form $x^\alpha y^\beta v^\alpha u^\beta$.*

Proof. Let $X' = x^\alpha P u^\beta$ be a minimal contravening subchain of X. Then $\alpha \neq \beta$, and by Lemma 5.9, $|X'| \geq 4$. If for some z^γ in X' we had $\alpha \neq \gamma \neq \beta$, then the subchains $x^\alpha Q z^\gamma$ and $z^\gamma R u^\beta$ with $Qz^\gamma R = P$ would have to be compliant (otherwise X' would not be minimal). Then, by Lemma 5.8, we would have a contravening triadic chain $x^\alpha z^\gamma u^\beta$, which is impossible by Lemma 5.9. For every z^γ in X' therefore, either $\gamma = \alpha$ or $\gamma = \beta$. Since a contravening chain cannot contain repeating superscripts, X' is of the form $x^\alpha y^\beta v^\alpha S u^\beta$.

---

[7]We cannot resist mentioning at this point a surprising mathematical similarity between the conceptual apparatus of (hence also the notation adopted in) the present theory, especially in this section, and that of the completely unrelated theory of "regular well-matched spaces" developed in Dzhafarov and Dzhafarov (2010) for comparative judgments. In particular, factors and factor points seem to be formally homologous to "stimulus areas" and "stimuli," respectively, and the contravening chains of the present theory essentially mirror the "soritical" sequences for comparative judgments, so that the proof of Theorem 5.11 below is almost identical to that of Lemma 3.3 of Dzhafarov and Dzhafarov (2010).

[8]One can easily generalize this reasoning to show that every chain $x_1^{\alpha_1} \ldots x_n^{\alpha_n}$ with pairwise distinct $\{\alpha_1, \ldots \alpha_n\}$ is compliant. As will be apparent from the proof of Theorem 5.11, however, in the present development we should only be concerned with $n = 3$.

But then $v^\alpha S u^\beta$ must be compliant (otherwise X' would not be minimal), and by Lemma 5.8 $x^\alpha y^\beta v^\alpha u^\beta$ is contravening. Since X' is minimal, we conclude that S is empty and $X' = x^\alpha y^\beta v^\alpha u^\beta$. $\quad\square$

It follows that the task of testing the compliance of D with all possible chain inequalities, as stated in Theorem 5.10, is reduced to testing the compliance with only the inequalities involving tetradic chains: if $\{A^\alpha\}_{\alpha\in\Phi} \leftarrow\hspace{-0.5em}\rho\ \Phi$, then, for any chain $x^\mu y^\nu u^\mu v^\nu$ with distinct μ and ν,

$$D x^\mu v^\nu \le D x^\mu y^\nu + D y^\nu u^\mu + D u^\mu v^\nu,$$

and if all such inequalities are satisfied, then there can be no other contravening chains. Given any four factor points $x^\mu, y^\nu, u^\mu, v^\nu$, one can form four different chains with alternating factors and four corresponding inequalities,

$$\begin{cases} D x^\mu v^\nu \le D x^\mu y^\nu + D y^\nu u^\mu + D u^\mu v^\nu, \\ D x^\mu y^\nu \le D x^\mu v^\nu + D v^\nu u^\mu + D u^\mu y^\nu, \\ D u^\mu y^\nu \le D u^\mu v^\nu + D v^\nu x^\mu + D x^\mu y^\nu, \\ D u^\mu v^\nu \le D u^\mu y^\nu + D y^\nu x^\mu + D x^\mu v^\nu. \end{cases}$$

Following Kujala and Dzhafarov (2008), these are easy to see (by adding the left-hand sides to themselves and to the right-hand sides) to be equivalent to the single inequality

$$\max\left\{ D x^\mu y^\nu, D x^\mu v^\nu, D u^\mu y^\nu, D u^\mu v^\nu \right\}$$
$$\le \left(D x^\mu y^\nu + D x^\mu v^\nu + D u^\mu y^\nu + D u^\mu v^\nu \right)/2.$$

We call this a *tetradic inequality*. Note that it is always satisfied if $x^\mu = u^\mu$ or $y^\nu = v^\nu$, so we only have to look at $x^\mu, y^\nu, u^\mu, v^\nu$ with two distinct points of each factor.

The theorem below shows that we have to check all such tetradic inequalities (for any given ρ).

Theorem 5.12. *The tetradic inequalities are mutually independent, in the sense that any one of them can be violated while the rest of them hold.*

Proof. Let μ and ν be distinct factors in Φ, and let, for all points x^μ and y^ν,

$$A_{x^\mu}^\mu \sim A_{y^\nu}^\nu \sim E,$$

where E is some non-singular random variable (i.e., no constant equals E with probability 1). Let $1^\mu, 2^\mu, 1^\nu, 2^\nu$ be distinct fixed points of μ and ν, and let

$$\left(A_{1^\mu 2^\nu}^\mu, A_{1^\mu 2^\nu}^\nu \right) \sim \left(A_{2^\mu 1^\nu}^\mu, A_{2^\mu 1^\nu}^\nu \right) \sim \left(A_{2^\mu 2^\nu}^\mu, A_{2^\mu 2^\nu}^\nu \right) \sim (E, E). \quad (9)$$

Let $A_{x^\alpha}^\alpha$ for any point of any factor $\alpha \notin \{\mu, \nu\}$ be distributed arbitrarily, and let the random variables $\{A_\phi^\alpha\}_{\alpha\in\Phi}$ be mutually independent for any treatment ϕ, except if the latter includes one of the pairs $\{1^\mu, 2^\nu\}$, $\{2^\mu, 1^\nu\}$, or $\{2^\mu, 2^\nu\}$: in those cases the joint distribution of $(A_{u^\mu v^\nu}^\mu, A_{u^\mu v^\nu}^\nu)$ is given by (9), while $\{A_\phi^\alpha\}_{\alpha\in\Phi-\{\mu\}}$ and $\{A_\phi^\alpha\}_{\alpha\in\Phi-\{\nu\}}$ remain the sets of mutually independent variables. It is easy to see that $\{A_\phi^\alpha\}_{\alpha\in\Phi}$ thus defined satisfy the marginal selectivity property.

Now,

$$D 1^\mu 2^\nu = D 2^\mu 1^\nu = D 2^\mu 2^\nu = 0, \quad D 1^\mu 1^\nu = \|E - E'\| > 0,$$

where E and E' are identically distributed and independent. The tetradic inequality on $\{1^\mu, 1^\nu, 2^\mu, 2^\nu\}$ is therefore violated:

$$\max\left\{ D 1^\mu 1^\nu, D 1^\mu 2^\nu, D 2^\mu 1^\nu, D 2^\mu 2^\nu \right\}$$
$$= \|E - E'\| > \frac{\|E - E'\|}{2} = \frac{D 1^\mu 1^\nu + D 1^\mu 2^\nu + D 2^\mu 1^\nu + D 2^\mu 2^\nu}{2}.$$

Clearly, the tetradic inequality holds on any set $\{x^\alpha, y^\beta, s^\alpha, t^\beta\}$ that does not include $\{1^\mu, 2^\nu\}$, $\{2^\mu, 1^\nu\}$, or $\{2^\mu, 2^\nu\}$, as this inequality then only involves mutually independent random variables (Lemma 4.3 and Theorem 5.10). Denoting by 3^μ any point other than 1^μ and 2^μ (if such a point exists), and analogously for 3^ν, it remains to consider the cases $\{1^\mu, 2^\nu, 3^\mu, 3^\nu\}$, $\{1^\mu, 2^\nu, 2^\mu, 3^\nu\}$, $\{1^\mu, 2^\nu, 3^\mu, 1^\nu\}$, $\{2^\mu, 1^\nu, 3^\mu, 3^\nu\}$, $\{2^\mu, 1^\nu, 1^\mu, 3^\nu\}$, $\{2^\mu, 1^\nu, 3^\mu, 2^\nu\}$, and $\{2^\mu, 2^\nu, 3^\mu, 3^\nu\}$ (note that the order of the points is immaterial here). It is easy to check that the four distances in each of these quadruples equal either 0 or $\|E - E'\| > 0$ (with E, E' independent identically distributed), and that the number of zero distances in these quadruples is never greater than two. The tetradic inequality, therefore, always holds: either

$$\max\{\ldots\} = \|E - E'\| < \frac{3\|E - E'\|}{2} = \frac{\text{sum}\{\ldots\}}{2},$$

or

$$\max\{\ldots\} = \|E - E'\| = \frac{2\|E - E'\|}{2} = \frac{\text{sum}\{\ldots\}}{2}.$$

This completes the proof. $\quad\square$

Theorem 5.12 is proved for a fixed $\rho \ge 1$, and for "untransformed" $\{A^\alpha\}_{\alpha\in\Phi}$. The application scope of the distance tests can be significantly broadened by using various values of ρ and by applying to $\{A^\alpha\}_{\alpha\in\Phi}$ various transformations as specified in Lemma 3.8. It is clear that the tetradic inequalities cannot be independent across all ρ and/or all transformations. Since a violation of a tetradic inequality means a strict inequality, the inequality involving the same quadruple of factor points will have to hold also for sufficiently close values of ρ and sufficiently "slight" transformations. This also applies to $\rho = \infty$: every violated inequality for $\rho = \infty$ will have to remain violated for all sufficiently large values of ρ, since the difference between ess sup $|A - B|$ and $\sqrt[\rho]{E[|A - B|^\rho]}$ can be made arbitrarily small (or, if the former is infinite, the latter can be made arbitrarily large).

6. CONCLUSION

We have advanced the theory of selective influence in three ways.

1. The notion of selective influence (together with the related but weaker notion of marginal selectivity) has been generalized to arbitrary sets of random entities whose joint distributions depend on arbitrary sets of external factors by which the random entities are indexed (Definition 3.4).
2. The joint distribution criterion has been formulated for random entities to be selectively influenced by their indexing external factors: this happens if and only if there is a jointly

distributed set of random entities, one for every value of every factor, such that every subset of this set that corresponds to a treatment is distributed as the original entities at this treatment (Theorem 4.1).

3. The distance tests previously formulated for pairs of random variables in two-by-two factorial designs have been generalized to arbitrary sets of random variables. For any quadruple of distinct factor points $x^\mu, y^\nu, u^\mu, v^\nu$, we check whether

$$\max\left\{Dx^\mu y^\nu, Dx^\mu v^\nu, Du^\mu y^\nu, Du^\mu v^\nu\right\}$$
$$\leq \left(Dx^\mu y^\nu + Dx^\mu v^\nu + Du^\mu y^\nu + Du^\mu v^\nu\right)/2,$$

where the function D is as in Definition 5.1, for some choice of $\rho \geq 1$ and of transformations $A^\alpha_\phi \mapsto h_{x^\alpha}(A^\alpha_\phi)$ as specified in Lemma 3.8. If this tetradic inequality is violated, the variables are not selectively influenced by the factors indexing them. It is shown that we do not need to check for compliance with any other chain inequalities (Definition 5.7, Theorems 5.10 and

5.11), and that the tetradic inequalities for different quadruples of factor points (for a given ρ and a given set of transformations) are logically independent (Theorem 5.12).

We conclude by posing an open question. Example 4.4 in Section 4 shows that the distance tests can be passed for all values of ρ while the random variables in question do not selectively depend on the respective factors. At the same time, in this example a distance test can be found to fail after the random variables have been transformed in accordance with Lemma 3.8. The open question is: for random variables which are not selectively influenced (but whose dependence on the corresponding factors is marginally selective), can the distance test be passed under all possible measurable transformations of the variables?

REFERENCES

Bloxom, B. (1972). The simplex in pair comparisons. *Psychometrika* 37, 119–136.

Diederich, A., and Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time and simple response time. *J. Math. Psychol.* 47, 304–322.

Dzhafarov, E. N. (1997). "Process representations and decompositions of response times," in *Choice, Decision and Measurement: Essays in Honor of R. Duncan Luce,* ed. A. A. J. Marley (New York: Erlbaum), 255–278.

Dzhafarov, E. N. (1999). Conditionally selective dependence of random variables on external factors. *J. Math. Psychol.* 43, 123–157.

Dzhafarov, E. N. (2001). Unconditionally selective dependence of random variables on external factors. *J. Math. Psychol.* 45, 421–451.

Dzhafarov, E. N. (2003). Selective influence through conditional independence. *Psychometrika* 68, 7–26.

Dzhafarov, E. N., and Dzhafarov, D. D. (2010). Sorites without vagueness II: Comparative sorites. *Theoria* 76, 25–53.

Dzhafarov, E. N., and Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *J. Math. Psychol.* 50, 390–401.

Dzhafarov, E. N., Schweickert, R., and Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *J. Math. Psychol.* 48, 51–64.

Kujala, J. V., and Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *J. Math. Psychol.* 52, 128–144.

Lazarsfeld, P. F. (1965). "Latent structure analysis," in *Mathematics and Social Sciences,* Vol. 1, eds S. Sternberg, V. Capecchi, T. Kloek, and C. T. Leender (Paris: Mouton), 37–54.

Schweickert, R. (1982). The bias of an estimate of coupled slack in stochastic PERT networks. *J. Math. Psychol.* 26, 1–12.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *J. Math. Psychol.* 28, 363–400.

Townsend, J. T., and Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *J. Math. Psychol.* 33, 309–327.

Townsend, J. T., and Thomas, R. D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *J. Math. Psychol.* 38, 1–34.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any com-mercial or financial relationships that could be construed as a potential conflict of interest.