

Testing for Selectivity

in the Dependence of Random Variables on External Factors

Janne V. Kujala Ehtibar N. Dzhafarov*
University of Jyväskylä Purdue University

Abstract

Random variables A and B , whose joint distribution depends on factors (x, y) , are selectively influenced by x and y , respectively, if A and B can be represented as functions of, respectively, (x, S_A, C) and (y, S_B, C) , where S_A, S_B, C are stochastically independent and do not depend on (x, y) . Selective influence implies selective dependence of marginal distributions on the respective factors: thus no parameter of A may depend on y . But parameters characterizing stochastic interdependence of A and B , such as their mixed moments, are generally functions of both x and y . We derive two simple necessary conditions for selective dependence of (A, B) on (x, y) , which can be used to conduct a potential infinity of selectiveness tests. One condition is that, for any factor values x, x' and y, y' ,

$$s_{xy} \leq s_{xy'} + s_{x'y'} + s_{x'y},$$

where $s_{xy} = E[|f(A_{xy}, x) - g(B_{xy}, y)|^p]^{1/p}$ with arbitrary f, g , and $p \geq 1$, and (A_{xy}, B_{xy}) denoting (A, B) at specific values of x, y . For $p = 2$ this condition is superseded by a more restrictive one:

$$|\rho_{xy}\rho_{xy'} - \rho_{x'y}\rho_{x'y'}| \leq \sqrt{1 - \rho_{xy}^2}\sqrt{1 - \rho_{xy'}^2} + \sqrt{1 - \rho_{x'y}^2}\sqrt{1 - \rho_{x'y'}^2},$$

where ρ_{xy} is the correlation between $f(A_{xy}, x)$ and $g(B_{xy}, y)$. For bivariate normally distributed $(f(A_{xy}, x), g(B_{xy}, y))$ this condition, if satisfied on a 2×2 subset $\{x, x'\} \times \{y, y'\}$, is also sufficient for a selective dependence of (A, B) on (x, y) confined to this subset.

KEYWORDS: multivariate normal distribution, processing architectures, random variables, selective influence, stochastic dependence, stochastic unrelatedness, test scores, Thurstone's general law.

1. Introduction

Selective influence (dependence, attribution) is a relation implied in statements of the following kind:

*Corresponding author. Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2081, USA. E-mail address: ehtibar@purdue.edu (E.N. Dzhafarov).

1. in two medical tests with scores (A, B) , random variables that jointly depend on patient's age (x) and geographic area (y) , A is only affected by x while B is only affected by y (but generally A and B are stochastically interdependent for any given value of x, y);
2. in the hypothetical mental architecture involved in response production, the duration A of one subprocess is a random variable only affected by target's legibility (x) while the duration B of another subprocess is a random variable only affected by the number of alternatives (y) to choose among (but A and B may be stochastically interdependent for given values of x, y);
3. in Thurstone's (1927) general law of comparative judgments, the image A of stimulus x and the image B of stimulus y are random variables which are normally distributed and correlated (but A is the image of x , not of the pair (x, y) , and analogously for B).

For a historical account of the problem of selective influence under stochastic interdependence see Dzhafarov (2003a). Here we only mention that the notion was first considered by Townsend (1984), but its implicit uses can be found in Lazarsfeld (1965), Bloxom (1972), and Schweickert (1982). Since then the assumption of selective influence (without the accompanying assumption of stochastic independence) was prominently used in various contexts (e.g., Dzhafarov, 1992, 1997, 2003b; Dzhafarov & Schweickert, 1995; Dzhafarov, Schweickert, & Sung, 2004; Townsend & Schweickert, 1989; Townsend & Thomas, 1994). The notion of selective influence has undergone significant transformations since Townsend (1984). In this paper we follow the version proposed in Dzhafarov (2003a) and elaborated in Dzhafarov and Gluhovsky (2006).

What both the theory and uses of selective influence have been conspicuously lacking is any possibility of testing for selective influence, except through *marginal selectivity* (Townsend, & Schweicker, 1989). To explain, consider the following four joint distributions of (A, B) -values in a 2×2 factorial design, where the external factor x assumes two values combined with two values of the external factor y :

$$\begin{array}{cc}
 \begin{array}{c} B \\ (x_1, y_1) \\ \begin{array}{|c|c|c|c|} \hline & 0 & 1 & 5 \\ \hline 0 & .24 & .07 & 0 \\ \hline 1 & .07 & .24 & .07 \\ \hline 5 & 0 & .07 & .24 \\ \hline \end{array} \\ A
 \end{array} &
 \begin{array}{c} B \\ (x_1, y_2) \\ \begin{array}{|c|c|c|c|} \hline & 0 & 1 & 5 \\ \hline 0 & .24 & .07 & 0 \\ \hline 1 & .07 & .24 & .07 \\ \hline 5 & 0 & .07 & .24 \\ \hline \end{array} \\ A
 \end{array} \\
 \begin{array}{c} B \\ (x_2, y_1) \\ \begin{array}{|c|c|c|c|} \hline & 0 & 1 & 5 \\ \hline 0 & .24 & .07 & 0 \\ \hline 1 & .07 & .24 & .07 \\ \hline 5 & 0 & .07 & .24 \\ \hline \end{array} \\ A
 \end{array} &
 \begin{array}{c} B \\ (x_2, y_2) \\ \begin{array}{|c|c|c|c|} \hline & 0 & 1 & 5 \\ \hline 0 & 0 & .07 & .24 \\ \hline 1 & .07 & .24 & .07 \\ \hline 5 & .24 & .07 & 0 \\ \hline \end{array} \\ A
 \end{array}
 \end{array} \tag{1}$$

The random variables A and B assume three values each (0, 1, or 5), and the matrices in (1) show the probabilities with which each value of A co-occurs with each value of B . These probabilities, as we see, do not

change as the factor values change from (x_1, y_1) to (x_1, y_2) to (x_2, y_1) , but they change at $(x, y) = (x_2, y_2)$. So we know that the joint distribution of (A, B) depends on the factors (x, y) varying on the set of values $\{x_1, x_2\} \times \{y_1, y_2\}$. We reflect this fact by presenting (A, B) as (A_{11}, B_{11}) when $(x, y) = (x_1, y_1)$, as (A_{12}, B_{12}) when $(x, y) = (x_1, y_2)$, and so on.

The first thing one should do if one suspects that A and B are selectively influenced by x and y ,¹ is to compare the marginal distributions of A_{11} and A_{12} : under the selective influence assumption these distributions, since they correspond to the same value x_1 of x , should be identical. The same statement holds for the marginal distributions of A_{21} and A_{22} , those of B_{11} and B_{21} , and those of B_{12} and B_{22} . If the distributions are different within at least one of these pairs the selective dependence of (A, B) on (x, y) is ruled out. In our example, the marginal selectivity is satisfied trivially: the marginal distributions of A and B are the same in all four matrices:

A			B		
0	1	5	0	1	5
.31	.38	.31	.31	.38	.31

But this does not, of course, make the joint distributions of (A, B) independent of x and y , and leaves the question open as to what properties of these distributions indicate or counterindicate selectiveness in their dependence on x and y . Thus, all mixed moments of (A, B) satisfying marginal selectivity, such as correlation coefficients between A and B , still generally depend on both x and y . In our example the correlation coefficients $\rho(A_{ij}, B_{ij})$ for factor values (x_i, y_j) , $i, j \in \{1, 2\}$, are

	y_1	y_2
x_1	.7299	.7299
x_2	.7299	-.6322

Is this matrix consistent or inconsistent with the hypothesis that A and B are selectively influenced by x and y ? This is precisely the type of question to which we have been lacking answers, and the present paper is aimed at remedying this state of affairs.

In this paper we propose two schemes for generating a potential infinity of simple tests for mixed moments of (A, B) , such that a failure of at least one of them rules out the possibility of selective influence. Thus, having verified that marginal selectivity is satisfied, one can look at the four correlation coefficients like the ones shown above and find out whether they are compatible or incompatible with the selective influence hypothesis. If the correlations pass the test, one can repeat it with nonlinearly transformed random variables. If many such correlation-based tests conducted turn out to be compatible with selective influence, one can proceed to the second block of tests and look at moments of the differences $|A - B|$ (or differences of the transformed A and B) to see if any of these would not rule out selective influence. We will use our opening example throughout to illustrate the computations involved in these tests.

¹We will omit the obvious qualification “respectively” when referring to the correspondence between (A, B) and (x, y) .

Our example also serves to emphasize that we use the word “testing” in a *non-statistical* meaning: aside from occasional remarks dealing with empirical data we are interested in the *population-level* properties of joint distributions, those based on which one can decide whether the random variables in question are or are not selectively influenced. This may disappoint a practically minded reader, but one should realize that no statistical theory can be constructed unless we have a population-level theory first. If the probability values in our example are estimated from relatively small data samples rather than known precisely, then, in essence, a statistical test of selective influence should be constructed from population-level tests applied to all possible joint probability distributions appropriately weighted based on the likelihoods of generating the observed data.

Most of our tests are formulated for 2×2 factorial designs, prominently used in such research areas as reconstruction of information processing architectures (Schweickert, Giorgini, & Dzhafarov, 2000; Roberts & Sternberg, 1992; Sternberg, 1969; Townsend, 1984), but usually with the accompanying assumption of stochastic independence. If the sets of values for x and y are larger than 2×2 , one can apply our tests to some (or all, if possible) 2×2 subsets. With one exception, none of our tests allows one to definitively *establish* selective influence: the tests are based on necessary but not sufficient conditions for selectiveness. The exception mentioned is the case when (A, B) , or some transformations thereof, are bivariate normally distributed. In this case the passage of the correlation-based test on a 2×2 factorial subset is also a sufficient condition for selectiveness on this factorial subset.

1.1. Plan of the paper

In the remainder of this introduction we state (following Dzhafarov, 2003a, and Dzhafarov & Gluhovsky, 2006) a formal definition of selective influence. We then discuss the issue, critical for the present development, of random variables having joint distributions versus being stochastically unrelated (defined on different sample spaces), and how in the latter case they can be presented as if they had joint distributions.

In Section 2 we discuss bivariate normally distributed (A, B) with parameters varying as functions of (x, y) taking their values on arbitrary sets. We establish a necessary and sufficient condition for (A, B) to be selectively influenced by (x, y) in a certain special way. We then restrict (x, y) to a 2×2 factorial set and formulate our first test of selective influence, involving the correlation coefficients corresponding to the four combinations of factor values. For 2×2 factorial sets (but not for larger sets), the restriction of being selectively influenced “in a special way” is subsequently (in Section 3) removed, making the passage of the test both necessary and sufficient for a selective dependence of bivariate normal (A, B) on (x, y) within a 2×2 factorial set.

In Section 3 we generalize this test, but as a necessary condition only, to arbitrarily distributed (A, B) . Outside the context of normal distributions what we get is in fact an infinite set of generally independent correlation-based tests, each applied to a different transformation of (A, B) . We then formulate our second test of selectiveness on a 2×2 factorial set, involving expected values of $|A - B|^p$, with p an arbitrary real

≥ 1 . Again, we get an infinity of generally independent tests, each corresponding to a certain value of p and certain transformations of (A, B) .

The results are summarized in the concluding section in the form of flowcharts presented in Figs. 6, 7, 8, and 9. The reader who wishes to get a glimpse of our tests before proceeding is referred to these flowcharts.

1.2. Preliminaries

We deal with pairs of (real-valued) random variables (A, B) whose joint distribution depends on two distinct variables $(x, y) \in X \times Y$. We indicate the dependence of (A, B) on (x, y) by writing

$$(A, B) = (A_{xy}, B_{xy}), \quad x \in X, y \in Y.$$

The meaning of such dependence is that, for some random entity C with a probability measure which does not depend on (x, y) , the random variables (A_{xy}, B_{xy}) can be represented as

$$\begin{aligned} A_{xy} &= A(x, y, C), \\ B_{xy} &= B(x, y, C), \end{aligned} \tag{2}$$

where $c \mapsto A(x, y, c)$ and $c \mapsto B(x, y, c)$ are measurable functions.² The use of the same symbol for a random variable and for the function used in one specific representation of this random variable should cause no confusion, provided one keeps in mind that a representation of the form (2) is not unique.

We say that (A_{xy}, B_{xy}) are selectively influenced by (x, y) , and symbolically present this as

$$(A, B) \leftarrow_P (x, y) \quad (\text{on } X \times Y), \tag{3}$$

if the functions A, B and the random entity C in (2) can be chosen so that

$$\begin{aligned} A_{xy} &= A_x = A(x, C), \\ B_{xy} &= B_y = B(y, C). \end{aligned} \tag{4}$$

The domain $X \times Y$ in references to selective influence may sometimes be omitted. One has to keep in mind, however, that $(A, B) \leftarrow_P (x, y)$ may very well be true on some $X' \times Y' \subset X \times Y$ but false on $X \times Y$. In the present context it is especially appropriate to note that a representation (4) may hold for all 2×2 subsets of $X \times Y$ without holding on the entire $X \times Y$. Indeed, even if two such subsets differ in one value of one factor only,

$$\{x_1, x_2\} \times \{y_1, y_2\} \quad \text{and} \quad \{x_1, x_2\} \times \{y_1, y'_2\},$$

²Following Dzhamfarov and Gluhovsky (2006), we use the term *random entity* for C to indicate that C is a measurable function from one arbitrary probability space to another. We reserve the term *random variable* for a measurable function into the set of reals endowed with the Borel sigma algebra. Although in (2) C can always be chosen to be a random variable (Theorem 1 in Dzhamfarov & Gluhovsky, 2006), this may not be true when functions A, B are constrained to exhibit selective dependence on (x, y) , as in the representation (4) below.

it is possible that the representations in the form of (4) for these subsets employ differently distributed C ; and even if the distribution is the same, the functions $c \mapsto A(x_1, c)$, $c \mapsto A(x_2, c)$, and $c \mapsto B(y_1, c)$ in the representation (4) for $\{x_1, x_2\} \times \{y_1, y_2\}$ need not be the same as the corresponding functions in the representation (4) for $\{x_1, x_2\} \times \{y_1, y'_2\}$.

It is sometimes more convenient to define (3) as stating the existence of stochastically independent (C, S_A, S_B) whose distributions do not depend on (x, y) , such that the variables (A_{xy}, B_{xy}) can be represented as

$$\begin{aligned} A_{xy} &= A_x = A(x, C, S_A), \\ B_{xy} &= B_y = B(y, C, S_B), \end{aligned} \tag{5}$$

where $(c, s_A) \mapsto A(x, c, s_A)$ and $(c, s_B) \mapsto B(y, c, s_B)$ are measurable functions. The representations (4) and (5) are equivalent (see Dzhafarov & Gluhovsky, 2006), and we make use of both of them in this paper. The representation (5) is preferable in the context of speaking of conditional distributions of (A, B) for a given value c of C , that is, the distributions of the stochastically independent random variables

$$\begin{aligned} A_{x,c} &= A(x, c, S_A), \\ B_{y,c} &= B(y, c, S_B). \end{aligned}$$

1.3. On stochastic unrelatedness and joint distributions

We need to mention a subtlety here which is important for the subsequent development. We understand the representation (2) as equivalent to saying that

$$(A_{xy}, B_{xy}) \sim (A(x, y, C), B(x, y, C)), \tag{6}$$

with \sim standing for “is distributed as.” This expression is more cautious than (2), in the sense that it does not allow one to speak of a joint distribution of (A_{xy}, B_{xy}) and $(A_{x'y'}, B_{x'y'})$ for $(x, y) \neq (x', y')$ – which is clearly correct, for (A_{xy}, B_{xy}) and $(A_{x'y'}, B_{x'y'})$ are *stochastically unrelated*, with no “co-occurrence” scheme defined for the two pairs of their values.

At the same time, the representation (2) leads to no complications (and, as we will see, offers considerable benefits) if one remembers not to confuse “is representable as” with “is.” In (2), the pairs (A_{xy}, B_{xy}) and $(A(x, y, C), B(x, y, C))$ are treated as identical, which is allowable precisely because the two pairs have no joint distribution which would allow one to say that they are different while identically distributed. This representation introduces, of course, a “fictitious” stochastic relationship between (A_{xy}, B_{xy}) and $(A_{x'y'}, B_{x'y'})$, allowing one to consider, say,

$$(A_{xy}, A_{x'y'}) = (A(x, y, C), A(x', y', C))$$

as a jointly distributed pair, with the co-occurrence scheme defined by common values of C . This possibility, however, does not lead to any contradictions: one simply has to realize that if the values of A and B are observable, then the co-occurrence of (A_{xy}, B_{xy}) is also observable, while the co-occurrence of $(A_{xy}, A_{x'y'})$

is not. Put another way, two different representations

$$\begin{aligned} A_{xy} &= A^*(x, y, C^*) & \text{and} & & A_{xy} &= A^{**}(x, y, C^{**}) \\ B_{xy} &= B^*(x, y, C^*) & & & B_{xy} &= B^{**}(x, y, C^{**}) \end{aligned}$$

are equivalent if and only if

$$(A^*(x, y, C^*), B^*(x, y, C^*)) \sim (A^{**}(x, y, C^{**}), B^{**}(x, y, C^{**})),$$

irrespective of whether

$$(A^*(x, y, C^*), A^*(x', y', C^*)) \sim (A^{**}(x, y, C^{**}), A^{**}(x', y', C^{**}))$$

or

$$(B^*(x, y, C^*), B^*(x', y', C^*)) \sim (B^{**}(x, y, C^{**}), B^{**}(x', y', C^{**}))$$

or

$$(A^*(x, y, C^*), B^*(x', y', C^*)) \sim (A^{**}(x, y, C^{**}), B^{**}(x', y', C^{**})).$$

One is free, therefore, to choose a representation of the form (2) for (A_{xy}, B_{xy}) which is most suitable for a specific purpose.

All of this, of course, also applies to the selective influence representations (4) and (5).

For a general discussion of the notions of stochastic unrelatedness and multiple, freely introducible probability spaces (as opposed to a fixed sample space viewpoint, in which any two random variables have joint distribution), see Dzhafarov and Gluhovsky (2006).

2. Normal Distributions

Here, we demonstrate the notion of selective influence on bivariate normally distributed (A_{xy}, B_{xy}) . We formulate a test for a special form of selective influence for such (A_{xy}, B_{xy}) (with x and y combined in a 2×2 factorial design), and we motivate more general tests of selective influence to be introduced in Section 3.

2.1. A special case of selective influence

Let

$$(A_{xy}, B_{xy}) \sim N_2 \left(\mathbf{m} = \begin{bmatrix} \mu_A \\ \mu_B \end{bmatrix}, \mathbf{V} = \begin{bmatrix} \sigma_A^2 & \sigma_A \sigma_B \rho \\ \sigma_A \sigma_B \rho & \sigma_B^2 \end{bmatrix} \right), \quad (7)$$

with N_k indicating a k -variate normal distribution with parameters μ_A , σ_A , μ_B , σ_B , and ρ , each of which generally depends on both x and y . If $(A, B) \leftarrow^P (x, y)$, the marginal selectivity is the first and most obvious test,

$$\begin{aligned} \mu_A &= \mu_A(x), & \sigma_A &= \sigma_A(x), \\ \mu_B &= \mu_B(y), & \sigma_B &= \sigma_B(y). \end{aligned} \quad (8)$$

We will assume henceforth that it is satisfied. We still have of course

$$\rho = \rho(x, y). \quad (9)$$

Let us standardize the distribution in the mean and in the variance, that is, switch from (A_{xy}, B_{xy}) to $\left(\frac{A-\mu_A}{\sigma_A}, \frac{B-\mu_B}{\sigma_B}\right)_{xy}$. Clearly, they are selectively influenced by (x, y) if and only if so are the original variables. For simplicity, we rename the standardized variates back into A, B , to get

$$(A_{xy}, B_{xy}) \sim N_2 \left(\mathbf{m} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} 1 & \rho(x, y) \\ \rho(x, y) & 1 \end{bmatrix} \right). \quad (10)$$

We begin by introducing a class of correlation functions $\rho(x, y)$ for which a representation of the form (5) can be shown to exist. The functions in this class are computed as

$$\rho(x, y) = \sum_{k=1}^n a_k(x) b_k(y), \quad (11)$$

where $n \geq 1$ and the functions $\{a_k(x)\}_{k=1, \dots, n}$, $\{b_k(y)\}_{k=1, \dots, n}$ are subject to the constraints

$$\begin{aligned} \sum_{k=1}^n a_k^2(x) &\leq 1, \\ \sum_{k=1}^n b_k^2(y) &\leq 1. \end{aligned} \quad (12)$$

We achieve a representation of the form (5) by introducing standard normally distributed, stochastically independent

$$C_1 \sim \dots \sim C_n \sim S_A \sim S_B \sim N_1(0, 1),$$

and putting

$$\begin{aligned} A_x &= \sqrt{1 - \sum_{k=1}^n a_k^2(x)} S_A + \sum_{k=1}^n a_k C_k, \\ B_y &= \sqrt{1 - \sum_{k=1}^n b_k^2(y)} S_B + \sum_{k=1}^n b_k C_k. \end{aligned} \quad (13)$$

It is easy to verify that (A_{xy}, B_{xy}) thus defined are distributed as (10), and according to Section 1.3 this is all we need to justify this representation and the conclusion $(A, B) \leftrightarrow \rho(x, y)$. Denoting the vector (C_1, \dots, C_n) by C , we state our

PROPOSITION 1 *If (A, B) is distributed as (10) and ρ is representable as (11), subject to (12), then $(A, B) \leftrightarrow \rho(x, y)$ in a special way: (A_{xy}, B_{xy}) can be presented in the form (5) with multivariate normally distributed (A_x, B_y, C) .*

It turns out that the following converse of this statement is also true:

PROPOSITION 2 *If (A_{xy}, B_{xy}) distributed as (10) can be presented in the form (5) with multivariate normally distributed (A_x, B_y, C) , then ρ is representable as (11), subject to (12).*

To prove this for $C = (C_1, \dots, C_n)$, we can assume with no loss of generality that

$$C_1 \sim \dots \sim C_n \sim N_1(0, 1).$$

Then

$$(A_{xy}, B_{xy}, C) \sim N_{n+2} \left(\mathbf{m}^* = \begin{bmatrix} 0 \\ 0 \\ \mathbf{0} \end{bmatrix}, \mathbf{V}^* = \begin{bmatrix} 1 & \rho & \mathbf{r}_A^T \\ \rho & 1 & \mathbf{r}_B^T \\ \mathbf{r}_A & \mathbf{r}_B & \mathbf{I} \end{bmatrix} \right) \quad (14)$$

where $\mathbf{r}_A, \mathbf{r}_B$ indicate vectors of correlations in column form, T stands for transposition, \mathbf{I} is the $n \times n$ identity matrix, $\mathbf{0}$ is the $n \times 1$ column of zeros. It is clear from (5) that

$$\begin{aligned} \mathbf{r}_A &= \mathbf{r}_A(x), \\ \mathbf{r}_B &= \mathbf{r}_B(y). \end{aligned}$$

Denoting an arbitrary value of C by \mathbf{c} (an $n \times 1$ column), we know how to find the mean and variance matrices $\mathbf{m}_\mathbf{c}, \mathbf{V}_\mathbf{c}$ for (A_{xy}, B_{xy}) given \mathbf{c} (e.g., Tong, 1990):

$$\begin{aligned} \mathbf{m}_\mathbf{c} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{r}_A^T \\ \mathbf{r}_B^T \end{bmatrix} \times \mathbf{c} = \begin{bmatrix} \mathbf{r}_A^T \mathbf{c} \\ \mathbf{r}_B^T \mathbf{c} \end{bmatrix}, \\ \mathbf{V}_\mathbf{c} &= \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} - \begin{bmatrix} \mathbf{r}_A^T \\ \mathbf{r}_B^T \end{bmatrix} \times \begin{bmatrix} \mathbf{r}_A & \mathbf{r}_B \end{bmatrix} = \begin{bmatrix} 1 - \mathbf{r}_A^T \mathbf{r}_A & \rho - \mathbf{r}_A^T \mathbf{r}_B \\ \rho - \mathbf{r}_A^T \mathbf{r}_B & 1 - \mathbf{r}_B^T \mathbf{r}_B \end{bmatrix}. \end{aligned}$$

We verify that the conditional means $\mathbf{r}_A^T \mathbf{c}$ and $\mathbf{r}_B^T \mathbf{c}$ depend on x and y , respectively, and that the same is true for the conditional variances $1 - \mathbf{r}_A^T \mathbf{r}_A$ and $1 - \mathbf{r}_B^T \mathbf{r}_B$. The variances must be nonnegative, whence we get the constraint (12), on putting

$$\mathbf{r}_A = \begin{bmatrix} a_1(x) \\ \vdots \\ a_n(x) \end{bmatrix}, \quad \mathbf{r}_B = \begin{bmatrix} b_1(y) \\ \vdots \\ b_n(y) \end{bmatrix}.$$

It remains to observe that for $A_{x,\mathbf{c}}$ and $B_{y,\mathbf{c}}$ to be independent, as required by (5), we should have

$$\rho = \mathbf{r}_A^T \mathbf{r}_B, \quad (15)$$

which is the same as (11).

We have established that (11)-(12) is a necessary and sufficient condition for the representation (5) to hold with (A_x, B_y, C) multivariate normal.³

2.2. A 2×2 cosphericity test

Let us forget for a while about bivariate normality and look at the geometric meaning of (11)-(12) taken as a statement about correlations between arbitrary random variables. For this purpose it is convenient to

³This result was stated in Dzhafarov (2003a), but with the constraint (12) mistakenly omitted.

introduce $(n + 2)$ -component vectors

$$\mathbf{a}(x) = \begin{bmatrix} a_1(x) \\ \vdots \\ a_n(x) \\ \sqrt{1 - \sum_{k=1}^n a_k^2(x)} \\ 0 \end{bmatrix}, \quad \mathbf{b}(y) = \begin{bmatrix} b_1(y) \\ \vdots \\ b_n(y) \\ 0 \\ \sqrt{1 - \sum_{k=1}^n b_k^2(y)} \end{bmatrix}, \quad (16)$$

so that

$$\begin{aligned} \rho(x, y) &= \mathbf{a}^T \mathbf{b}, \\ \mathbf{a}^T \mathbf{a} &= 1, \\ \mathbf{b}^T \mathbf{b} &= 1. \end{aligned} \quad (17)$$

We see that the correlations $\rho(x, y)$ satisfy (11)-(12) if and only if one can find points $\mathbf{a}(x)$ and $\mathbf{b}(y)$ on a unit (hyper)sphere in \mathbb{R}^{n+2} centered at the origin and satisfying (17). We will denote this (hyper)sphere by \mathbb{S}^{n+1} . This geometric interpretation will help us in constructing a test for whether correlations $\rho(x, y)$ satisfy (11)-(12) in a 2×2 factorial setting.

So let us confine our attention to $X = \{x_1, x_2\}$, $Y = \{y_1, y_2\}$. Denoting

$$\rho_{ij} = \rho(x_i, y_j), \quad i, j \in \{1, 2\},$$

the problem is to characterize the matrices of correlations

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$$

for which one can find four points

$$\mathbf{a}_1 = \mathbf{a}(x_1), \quad \mathbf{a}_2 = \mathbf{a}(x_2), \quad \mathbf{b}_1 = \mathbf{b}(y_1), \quad \mathbf{b}_2 = \mathbf{b}(y_2)$$

in \mathbb{R}^{n+2} , such that

$$\begin{aligned} \rho_{ij} &= \mathbf{a}_i^T \mathbf{b}_j, \\ \mathbf{a}_i^T \mathbf{a}_i &= 1, \quad i, j \in \{1, 2\}, \\ \mathbf{b}_j^T \mathbf{b}_j &= 1, \end{aligned} \quad (18)$$

In other words, we ask the question: for what matrices \mathbf{R} can one find four points on the sphere \mathbb{S}^{n+1} satisfying $\rho_{ij} = \mathbf{a}_i^T \mathbf{b}_j$ ($i, j \in \{1, 2\}$)?

If such points exist, the coordinate system can always be chosen so that

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} a \\ \sqrt{1 - a^2} \\ \mathbf{0} \end{bmatrix}, \quad a^2 \leq 1.$$

where $\mathbf{0}$ is the n -component column of zeros. Let the coordinates of the remaining two points be

$$\mathbf{b}_1 = \begin{bmatrix} u_1 \\ v_1 \\ \vdots \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} u_2 \\ v_2 \\ \vdots \end{bmatrix}.$$

The only constraints imposed by (18) on these coordinates are

$$\begin{aligned} \rho_{11} &= \mathbf{a}_1^T \mathbf{b}_1 = u_1, \\ \rho_{21} &= \mathbf{a}_2^T \mathbf{b}_1 = au_1 + \sqrt{1-a^2}v_1, \\ u_1^2 + v_1^2 &\leq 1, \end{aligned} \tag{19}$$

and

$$\begin{aligned} \rho_{12} &= \mathbf{a}_1^T \mathbf{b}_2 = u_2, \\ \rho_{22} &= \mathbf{a}_2^T \mathbf{b}_2 = au_2 + \sqrt{1-a^2}v_2, \\ u_2^2 + v_2^2 &\leq 1. \end{aligned} \tag{20}$$

Clearly, the values of the remaining coordinates of $\mathbf{b}_1, \mathbf{b}_2$ are irrelevant, and we can always equate the coordinates above the third one to zero. From this we conclude that

PROPOSITION 3 *If any four points satisfying (18) exist, then there are also four points with the same properties lying on the sphere \mathbb{S}^2 :*

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} a \\ \sqrt{1-a^2} \\ 0 \end{bmatrix}, \quad \mathbf{b}_1 = \begin{bmatrix} u_1 \\ v_1 \\ \sqrt{1-u_1^2+v_1^2} \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} u_2 \\ v_2 \\ \sqrt{1-u_2^2+v_2^2} \end{bmatrix}, \tag{21}$$

subject to

$$\begin{aligned} a^2 &\leq 1, \\ u_1^2 + v_1^2 &\leq 1, \\ u_2^2 + v_2^2 &\leq 1. \end{aligned} \tag{22}$$

With the aid of this simplification we can now derive an explicit form of the constraints imposed by (18) on the correlations ρ_{ij} . Assuming first that $a^2 \neq 1$ and using (19)-(20), the inequalities in Proposition 3 are equivalent to

$$\begin{aligned} a^2 &< 1, \\ \rho_{11}^2 + \frac{a^2\rho_{11}^2 + \rho_{21}^2 - 2a\rho_{11}\rho_{21}}{1-a^2} &\leq 1, \\ \rho_{12}^2 + \frac{a^2\rho_{12}^2 + \rho_{22}^2 - 2a\rho_{12}\rho_{22}}{1-a^2} &\leq 1, \end{aligned}$$

which, following some algebra, can be shown to be equivalent to

$$\begin{aligned} a^2 &< 1, \\ \rho_{11}\rho_{21} - \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)} &\leq a \leq \rho_{11}\rho_{21} + \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)}, \\ \rho_{12}\rho_{22} - \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)} &\leq a \leq \rho_{12}\rho_{22} + \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)}. \end{aligned} \tag{23}$$

The excluded case $a = \pm 1$ can now be brought back, for in this case (19)-(20) imply

$$\begin{aligned}\rho_{11} &= \pm \rho_{21} \\ \rho_{12} &= \pm \rho_{22}\end{aligned}\tag{24}$$

(with the same choice of $+$ or $-$ in both equations), and then the last two inequalities in (23) are satisfied. But changing the first inequality in (23) from $a^2 < 1$ to $a^2 \leq 1$ makes it redundant, as it is easy to show that the two intervals

$$\begin{aligned}\left[\rho_{11}\rho_{21} - \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)}, \rho_{11}\rho_{21} + \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)} \right], \\ \left[\rho_{12}\rho_{22} - \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)}, \rho_{12}\rho_{22} + \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)} \right]\end{aligned}\tag{25}$$

are contained in $[-1, 1]$. We conclude that if the four points $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{b}_1, \mathbf{b}_2)$ in (21) exist, then these two intervals have a nonempty intersection, that is,

$$\begin{aligned}\rho_{11}\rho_{21} - \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)} &\leq \rho_{12}\rho_{22} + \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)}, \\ \rho_{12}\rho_{22} - \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)} &\leq \rho_{11}\rho_{21} + \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)}.\end{aligned}\tag{26}$$

Equivalently but more compactly,

$$|\rho_{11}\rho_{21} - \rho_{12}\rho_{22}| \leq \sqrt{(1-\rho_{11}^2)(1-\rho_{21}^2)} + \sqrt{(1-\rho_{12}^2)(1-\rho_{22}^2)}.\tag{27}$$

Conversely, if the intersection of the two intervals in (25) is nonempty, then one can choose an a in accordance with (23) or (24), and then find the coordinates of $\mathbf{b}_1, \mathbf{b}_2$ in accordance with (19)-(20): clearly, the points with thus determined coordinates lie on the sphere \mathbb{S}^2 and satisfy $\rho_{ij} = \mathbf{a}_i^T \mathbf{b}_j$ ($i, j \in \{1, 2\}$).

We have proved the following

PROPOSITION 4 (COSPHERICITY CONDITION) *A matrix of correlations*

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$$

for $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ can be represented as (18) if and only if \mathbf{R} satisfies (27).

We will refer to (27) as the *cosphericity test*, because it amounts (due to Propositions 3 and 4) to finding out whether there are four points $\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2$ lying on the sphere \mathbb{S}^2 such that the cosines $\mathbf{a}_1^T \mathbf{b}_1, \mathbf{a}_1^T \mathbf{b}_2, \mathbf{a}_2^T \mathbf{b}_1, \mathbf{a}_2^T \mathbf{b}_2$ equal $\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}$, respectively (see Fig. 1).

Note that nowhere in the formulation and derivation of Proposition 4 have we used the assumption that the random variables $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ are bivariate normally distributed. In fact, Proposition 4 is based on no distributional assumptions (except for the existence of correlations). This fact will be made use of in Section 3.1, where we broaden the test's applicability and show that when the random variables $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ are selectively influenced by (x, y) on $\{x_1, x_2\} \times \{y_1, y_2\}$, the four points $\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2$ can be viewed as representing, in some well-defined sense, the random variables themselves. For now, we use the cosphericity test in relation to Propositions 1 and 2 to obtain the following

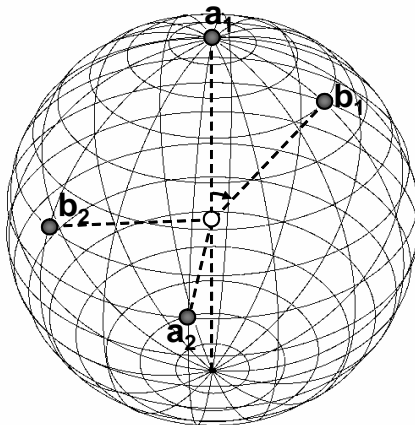


Figure 1. Correlations $\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}$ pass the cospphericity test if there are points $\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2$ on the sphere \mathbb{S}^2 such that (denoting the center of the sphere by \mathbf{o}) $\mathbf{a}_1^T \mathbf{b}_1 = \cos \angle \mathbf{a}_1 \mathbf{o} \mathbf{b}_1 = \rho_{11}$ (the angle shown), and $\mathbf{a}_1^T \mathbf{b}_2 = \cos \angle \mathbf{a}_1 \mathbf{o} \mathbf{b}_2 = \rho_{12}$, $\mathbf{a}_2^T \mathbf{b}_1 = \cos \angle \mathbf{a}_2 \mathbf{o} \mathbf{b}_1 = \rho_{21}$, $\mathbf{a}_2^T \mathbf{b}_2 = \cos \angle \mathbf{a}_2 \mathbf{o} \mathbf{b}_2 = \rho_{22}$.

PROPOSITION 5 *Random variables (A_{ij}, B_{ij}) distributed as*

$$N_2 \left(\mathbf{m} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix} \right), \quad i, j \in \{1, 2\}, \quad (28)$$

with a matrix of correlations

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$$

are representable in the form (5) with multivariate normally distributed (A_i, B_j, C) if and only if \mathbf{R} passes the cospphericity test, (27). The distribution (28) in this formulation can be replaced with any bivariate normal distribution which satisfies marginal selectivity (8).

The last sentence in this proposition is due to the fact that bivariate distributions which satisfy marginal selectivity can always be standardized in the mean and variance without changing their correlation coefficients.

Thus, the correlation matrices

$$\mathbf{R}_1 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{R}_2 = \begin{bmatrix} -.8 & -.6 \\ -.6 & .7 \end{bmatrix}$$

pass the cosphericity test, while

$$\mathbf{R}_3 = \begin{bmatrix} -0.8 & -0.6 \\ -0.6 & .9 \end{bmatrix}$$

does not:

$$\begin{aligned} (\mathbf{R}_1) \quad & |0 \cdot 0 - 0 \cdot 0| < \sqrt{1 - 0^2} \sqrt{1 - 0^2} + \sqrt{1 - 0^2} \sqrt{1 - 0^2} \\ (\mathbf{R}_2) \quad & |(-0.8) \cdot (-0.6) - (-0.6) \cdot (-0.7)| < \sqrt{1 - (-0.8)^2} \sqrt{1 - (-0.6)^2} + \sqrt{1 - (-0.6)^2} \sqrt{1 - (-0.7)^2} \\ (\mathbf{R}_3) \quad & |(-0.8) \cdot (-0.6) - (-0.6) \cdot (.9)| > \sqrt{1 - (-0.8)^2} \sqrt{1 - (-0.6)^2} + \sqrt{1 - (-0.6)^2} \sqrt{1 - .9^2} \end{aligned}$$

We know therefore that if bivariate normally distributed $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ have correlations \mathbf{R}_1 or \mathbf{R}_2 , then $(A, B) \not\propto (x, y)$ on $\{x_1, x_2\} \times \{y_1, y_2\}$. We also know that bivariate normal (A_{ij}, B_{ij}) with correlations \mathbf{R}_3 are not representable by (5) with multivariate normally distributed (A_{ij}, B_{ij}, C) .

The question arises whether the latter statement can be strengthened. Can it be that a matrix of correlations which fails the cosphericity test rule out the representability of the (A_{ij}, B_{ij}) in the form of (5), without specifying the distributions of (A_{ij}, B_{ij}, C) ? Put differently, is a matrix of correlations which fails the cosphericity test incompatible with any form of selective influence relation? As we show in the next section, the answer (for a 2×2 factorial design) is affirmative, and not only for bivariate normal (A_{ij}, B_{ij}) .

3. Selective Influence Tests for Arbitrary Distributions

Here, we first generalize the cosphericity test to (almost) arbitrary bivariate distributions of $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ (as a necessary condition only), and then we construct still more general tests of selective influence involving higher order mixed moments.

3.1. Cosphericity test as a general necessary condition

Let $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ be distributed arbitrarily except for possessing finite means and variances. As here we are only interested in the correlation coefficients

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix},$$

we can assume with no loss of generality that (A_{ij}, B_{ij}) have been standardized in the mean and variance,

$$\begin{aligned} \mathbb{E}[A_{ij}] &= \mathbb{E}[B_{ij}] = 0, \\ \mathbb{E}[A_{ij}^2] &= \mathbb{E}[B_{ij}^2] = 1, \end{aligned} \quad i, j \in \{1, 2\}. \quad (29)$$

Let $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ be representable in the form (4), which will be more convenient to use now than the equivalent (5):

$$A_i(C) = A(x_i, C), \quad B_j(C) = A(y_j, C), \quad i, j \in \{1, 2\}.$$

On the set of values c of C , consider the linear span

$$\text{span}[A_1(c), B_1(c), A_2(c), B_2(c)] = \{\alpha_1 A_1(c) + \beta_1 B_1(c) + \alpha_2 A_2(c) + \beta_2 B_2(c) : (\alpha_1, \beta_1, \alpha_2, \beta_2) \in \mathbb{R}^4\}.$$

The variances of A_1, B_1, A_2, B_2 are finite (in fact, unit),

$$\begin{aligned} \int A_i(c)^2 d\lambda(c) &= 1, \\ \int B_j(c)^2 d\lambda(c) &= 1, \end{aligned} \quad i, j \in \{1, 2\},$$

whence $\text{span}[A_1, B_1, A_2, B_2]$ is an L^2 space (i.e., all functions in it have integrable squares), and then we know that for any $P, Q \in \text{span}[A_1, A_2, B_1, B_2]$, the integral

$$P \cdot Q = \int P(c) Q(c) d\lambda(c)$$

is well-defined and finite. Clearly, $\text{span}[A_1, A_2, B_1, B_2]$ is a vector space with the operation $P \cdot Q$ as its inner product. The dimensionality of $\text{span}[A_1, A_2, B_1, B_2]$ being at most 4, the well-known orthogonalization theorem of functional analysis tells us that one can choose in $\text{span}[A_1, A_2, B_1, B_2]$ a set of at most four functions, O_1, \dots, O_m ($m \leq 4$), which form its orthonormal basis. This means that any function (random variable) P which belongs to $\text{span}[A_1, A_2, B_1, B_2]$ can be represented by a vector $\mathbf{p} \in \mathbb{R}^m$ whose components (p_1, \dots, p_m) are determined by

$$P = \sum_{i=1}^m p_i O_i.$$

It is easy to verify that if P is represented by $\mathbf{p} \in \mathbb{R}^m$ and Q by $\mathbf{q} \in \mathbb{R}^m$ (treated as column-vectors), then

$$P \cdot Q = \sum_{i=1}^m p_i q_i = \mathbf{p}^T \mathbf{q}.$$

We conclude that $\text{span}[A_1, A_2, B_1, B_2]$ endowed with the inner product $P \cdot Q$ is homomorphic to a Euclidean space \mathbb{R}^m ($m \leq 4$) endowed with the scalar product $\mathbf{p}^T \mathbf{q}$.

Now, A_1, B_1, A_2, B_2 are represented in this Euclidean space by some points $\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2$. Since, due to (29),

$$\mathbf{a}_1^T \mathbf{a}_1 = \mathbf{a}_2^T \mathbf{a}_2 = \mathbf{b}_1^T \mathbf{b}_1 = \mathbf{b}_2^T \mathbf{b}_2 = 1$$

and

$$\rho_{ij} = \mathbf{a}_i^T \mathbf{b}_j \quad i, j \in \{1, 2\},$$

the points $\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2$ satisfy (18). Proposition 4 then tells us that the correlations ρ_{ij} have to pass the cosphericity test. This completes the proof of

PROPOSITION 6 *Random variables $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ with a matrix of correlations*

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$$

are selectively influenced (by (x, y) on $\{x_1, x_2\} \times \{y_1, y_2\}$) only if \mathbf{R} passes the cosphericity test, (27).

We do not need to mention in this formulation that (A_{ij}, B_{ij}) are standardized in the mean and variance because \mathbf{R} is invariant with respect to such transformations.

Returning for a moment to bivariate normally distributed (A_{ij}, B_{ij}) , Proposition 6 immediately leads to the following

PROPOSITION 7 (ADDENDUM TO PROPOSITION 5) *Bivariate normally distributed variables $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ satisfy*

$$(A, B) \leftarrow \mathcal{P} (x, y) \quad (\text{on } \{x_1, x_2\} \times \{y_1, y_2\})$$

*if and only if they are representable in the form (5) with multivariate normally distributed (A_i, B_j, C) .*⁴

Indeed, if $(A, B) \leftarrow \mathcal{P} (x, y)$ then, by Proposition 6, the correlations ρ_{ij} pass the cosphericity test; and then, by Proposition 5, (A_{ij}, B_{ij}) are representable in the mentioned special form of (5).

3.2. Multitude and properties of cosphericity tests

If a matrix of correlations \mathbf{R} passes the cosphericity test one cannot claim, unless $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ are bivariate normally distributed, that $(A, B) \leftarrow \mathcal{P} (x, y)$ on $\{x_1, x_2\} \times \{y_1, y_2\}$. It is important in this respect to realize that the power of the cosphericity test to detect the lack of selective influence can be increased by repeatedly applying this test to various transformations of A and B which preserve marginal selectivity.

Indeed, if $(A, B) \leftarrow \mathcal{P} (x, y)$, then also

$$(f(A, x), g(B, y)) \leftarrow \mathcal{P} (x, y), \tag{30}$$

for any functions f, g with measurable $a \mapsto f(a, x)$ and $b \mapsto g(b, y)$. When specialized to a 2×2 factorial setting this becomes

$$(f_i(A_{ij}), g_j(B_{ij})) \leftarrow \mathcal{P} (x, y) \quad (\text{on } \{x_1, x_2\} \times \{y_1, y_2\}), \tag{31}$$

for any measurable $\{f_i\}_{i=1,2}$, $\{g_j\}_{j=1,2}$. The transformed random variables

$$(A_{ij}^*, B_{ij}^*) = (f_i(A_{ij}), g_j(B_{ij}))$$

are characterized by a correlation matrix \mathbf{R}^* which is not generally computable from \mathbf{R} , and its (non)compliance with the cosphericity test is generally independent of that of \mathbf{R} .⁵ This means that by applying a multitude of transformations (f_1, f_2, g_1, g_2) to given $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$, the selective influence hypothesis can be tested by the corresponding multitude of cosphericity tests, the failure of at least one of which would rule out the selectivity $(A, B) \leftarrow \mathcal{P} (x, y)$ on $\{x_1, x_2\} \times \{y_1, y_2\}$.

⁴The statement of this proposition holds true for any $n \times m$ ($n \geq 2, m \geq 2$) factorial design, because the argument used in the proof of Proposition 6 trivially generalizes to $\text{span}[A_1, \dots, A_n, B_1, \dots, B_m]$.

⁵The meaning of ‘‘independence’’ here is simply that \mathbf{R} may pass (fail) the test with \mathbf{R}^* failing (passing) it. One should not confuse this with stochastic independence of two or more statistical tests performed on one and the same sample. We have no knowledge of the sampling distributions of our tests, but their independence in the former, logical meaning of the word will be readily demonstrated on an example below.

Due to Propositions 5 and 7, the situation greatly simplifies and the cosphericity test becomes definitive if the transformations (f_1, f_2, g_1, g_2) can be chosen so that (A_{ij}^*, B_{ij}^*) are bivariate normally distributed. This can only be achieved if the transformations rendering all marginals A_{ij}^* and B_{ij}^* standard normally distributed also make (A_{ij}^*, B_{ij}^*) bivariate normally distributed. If the latter is the case, the correlation matrix \mathbf{R}^* for the transformed variables which passes (fails) the cosphericity test proves (respectively, disproves) the selectiveness hypothesis.⁶ If the bivariate normality cannot be achieved, the selectiveness, if not disproved, can only be *corroborated* – by demonstrating that the cosphericity test (and the distance test, to be discussed later) is passed under multiple transformations.

Thus, the correlation matrix

$$\mathbf{R}_{ex.(1)} = \begin{bmatrix} .7299 & .7299 \\ .7299 & -.6322 \end{bmatrix}$$

computed from our opening example (1) does pass the cosphericity test:

$$|.7299 \cdot .7299 - .7299 \cdot (-.6322)| < \sqrt{1 - .7299^2} \sqrt{1 - .7299^2} + \sqrt{1 - .7299^2} \sqrt{1 - (-.6322)^2}.$$

From this, of course, we cannot conclude that $(A, B) \not\propto (x, y)$, only that $\mathbf{R}_{ex.(1)}$ is not incompatible with this hypothesis. Recall that in this example the random variables A, B vary on the same three-element set $\{0, 1, 5\}$. If we now choose our transformations f_1, f_2, g_1, g_2 to be one and the same nonlinear mapping

$$f_1 \equiv f_2 \equiv g_1 \equiv g_2 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 5 \mapsto 2 \end{cases}, \tag{32}$$

the distributions in our example become

	B^*		B^*		
	(x_1, y_1)	0 1 2	(x_1, y_2)	0 1 2	
A^*	0	.24 .07 0	A^*	0	.24 .07 0
A^*	1	.07 .24 .07	A^*	1	.07 .24 .07
A^*	2	0 .07 .24	A^*	2	0 .07 .24
	B^*		B^*		
	(x_2, y_1)	0 1 2	(x_2, y_2)	0 1 2	
A^*	0	.24 .07 0	A^*	0	0 .07 .24
A^*	1	.07 .24 .07	A^*	1	.07 .24 .07
A^*	2	0 .07 .24	A^*	2	.24 .07 0

(33)

⁶If the distributions of (A_{ij}, B_{ij}) are known on a sample level only, then one can (a) use the empirical distribution function for the combined sample of A_{i1} and A_{i2} (for, due to the marginal selectiveness, the two samples should be identically distributed) and transform it into a normal cumulative curve (separately for $i = 1$ and $i = 2$); (b) do the same with the combined sample of B_{1j} and B_{2j} (separately for $j = 1$ and $j = 2$); and (c) subject the bivariate distributions with the newly obtained marginals to conventional tests of bivariate normality. Statistical uncertainty involved will of course prevent any result of this procedure (and subsequent application of a selectiveness test) from being definitive.

with the correlation matrix

$$\mathbf{R}_{ex.(1)}^* = \begin{bmatrix} .7742 & .7742 \\ .7742 & -.7742 \end{bmatrix}.$$

This correlation matrix does not pass the cosphericity test:

$$|.7742 \cdot .7742 - .7742 \cdot (-.7742)| > \sqrt{1 - .7742^2} \sqrt{1 - .7742^2} + \sqrt{1 - .7742^2} \sqrt{1 - (-.7742)^2}.$$

We conclude that the random variables (A^*, B^*) in (33) and hence also the original random variables (A, B) in (1) are not selectively influenced by (x, y) on $\{x_1, x_2\} \times \{y_1, y_2\}$. Later we will show how the same fact can be established by other means.

The only transformations for which \mathbf{R}^* is a function of \mathbf{R} are linear transformations,

$$\begin{aligned} A_{ij}^* &= \alpha_i A_{ij} + \gamma_i, \\ B_{ij}^* &= \beta_j B_{ij} + \delta_j \end{aligned}, \quad i, j \in \{1, 2\}.$$

To be consistent, the cosphericity test should be passed by \mathbf{R}^* if and only if it is passed by \mathbf{R} . This is indeed the case. As far as the correlations are concerned, the shifts and the absolute values of the scaling coefficients are irrelevant, and we only need to consider the set of 16 reflections

$$\alpha_1 = \pm 1, \alpha_2 = \pm 1, \beta_1 = \pm 1, \beta_2 = \pm 1,$$

where the four signs can be chosen independently. It is easy to verify that as a result of such reflections, \mathbf{R}^* can differ from \mathbf{R} in the signs of 4, 2, or 0 correlations:

$$\begin{bmatrix} -\rho_{11} & \rho_{12} \\ -\rho_{21} & \rho_{22} \end{bmatrix}, \begin{bmatrix} \rho_{11} & -\rho_{12} \\ -\rho_{21} & \rho_{22} \end{bmatrix}, \begin{bmatrix} -\rho_{11} & -\rho_{12} \\ -\rho_{21} & -\rho_{22} \end{bmatrix}, \text{ etc.}$$

(eight different matrices altogether, counting \mathbf{R} itself). Squaring the two sides of the inequality (27) and rearranging terms, we can rewrite the cosphericity test in the form

$$\rho_{11}^2 + \rho_{12}^2 + \rho_{21}^2 + \rho_{22}^2 \leq 2 + 2\rho_{11}\rho_{12}\rho_{21}\rho_{22} + 2\sqrt{(1 - \rho_{11}^2)(1 - \rho_{12}^2)(1 - \rho_{21}^2)(1 - \rho_{22}^2)}, \quad (34)$$

from which it is clear that the test is invariant with respect to any even number of sign changes.

It is also apparent from (34) that the cosphericity test is invariant with respect to all permutations of $(\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22})$ – a nice and rather unexpected form of symmetry. This means that in all computations one may treat the correlation coefficients as an unordered quadruple of numbers.

Figure 2 provides a rough idea of which matrices of correlations \mathbf{R} fail and which pass the cosphericity test. The test is clearly not very restrictive: selective influence is ruled out for only those \mathbf{R} that contain correlations large in absolute value. This observation tells us that repeated applications of the cosphericity test to various transformations of (A_{ij}, B_{ij}) may be critically important. We cannot say anything at this stage about optimal strategies in sequential choices of these transformations, except for the following considerations.

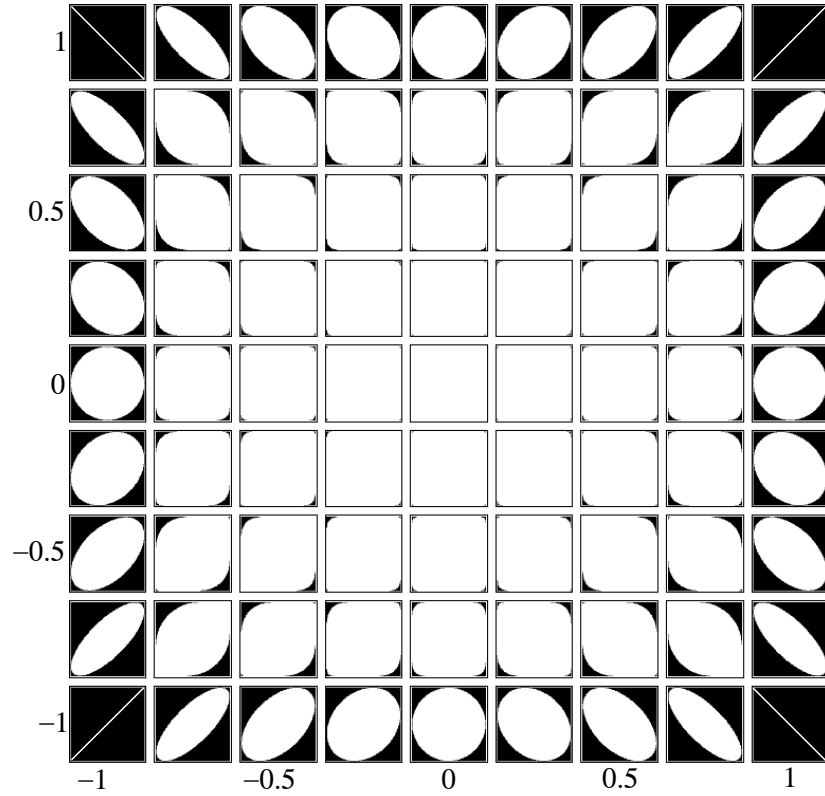


Figure 2. A graphical depiction of the cosphericity test. Let $\{r, s, t, u\}$ be values of the correlation coefficients $\{\rho_{11}, \rho_{12}, \rho_{21}, \rho_{22}\}$, in any order. Each panel corresponds to fixed values of r, s (indicated on the horizontal and vertical margins) and shows for which values of t, u the cosphericity test rules out selective influence (black areas) and for which t, u selective influence remains a possibility (white areas). If the random variables whose correlations one is dealing with are bivariate normally distributed, then the white areas indicate correlations for which selective influence is positively established (on a 2×2 factorial set).

First, we know that the cosphericity test is invariant with respect to linear transformations. So, in conducting the cosphericity test with multiple transformations $(f_1, f_2, g_1, g_2)_1, (f_1, f_2, g_1, g_2)_2, \dots$ (a list which would normally include identity transformations), none of them, to avoid redundancy, should be componentwise linearly dependent on another.

Second, we know that the cosphericity test is most restrictive if more than one of the correlations coefficients are large in absolute value (Fig. 2). This suggests that a search for transformations should be directed at maximizing the absolute values of at least two of the correlations.⁷

3.3. Distance test for correlations

Consider now random variables $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ standardized in the mean and variance,⁸

$$\begin{aligned} \mathbb{E}[A_{ij}] &= \mathbb{E}[B_{ij}] = 0, \\ \mathbb{E}[A_{ij}^2] &= \mathbb{E}[B_{ij}^2] = 1, \end{aligned} \quad i, j \in \{1, 2\}. \quad (35)$$

If $(A, B) \leftrightarrow (x, y)$ on $\{x_1, x_2\} \times \{y_1, y_2\}$, the cosphericity test (due to Propositions 3 and 4) tells us that there are four points $(\mathbf{a}_1, \mathbf{b}_1, \mathbf{a}_2, \mathbf{b}_2)$ on the sphere \mathbb{S}^2 whose scalar products match the correlations ρ_{ij} as shown in Fig. 1. Denoting the Euclidean distance between \mathbf{a}_i and \mathbf{b}_j by σ_{ij} (see Fig. 3) we should have then, by triangle inequality,

$$\begin{aligned} \sigma_{11} &\leq \sigma_{12} + \sigma_{22} + \sigma_{21}, \\ \sigma_{22} &\leq \sigma_{21} + \sigma_{11} + \sigma_{12}, \\ \sigma_{12} &\leq \sigma_{11} + \sigma_{21} + \sigma_{22}, \\ \sigma_{21} &\leq \sigma_{22} + \sigma_{12} + \sigma_{11}. \end{aligned} \quad (36)$$

We will refer to these inequalities as the *distance test for correlations*, because the distances in (36) are easily expressed through correlation coefficients as

$$\sigma_{ij} = \sqrt{2 - 2\rho_{ij}}, \quad i, j \in \{1, 2\}.$$

The four Euclidean distances here have a clear probabilistic meaning,

$$\sigma_{ij} = \sigma[A_i - B_j] = \sqrt{2 - 2\rho_{ij}}, \quad i, j \in \{1, 2\}. \quad (37)$$

where $\sigma[\dots]$ stands for standard deviation. Indeed,

$$\sigma^2[A_i - B_j] = \sigma^2[A_i] + \sigma^2[B_j] - 2\sigma[A_i]\sigma[B_j]\rho_{ij},$$

and since all the standard deviations on the right are unit, we get (37).

⁷If the distributions are known on a sample level only, then one can plot A_{ij} versus B_{ij} for all four combinations of i, j and seek transformations (f_1, f_2, g_1, g_2) which linearize at least two of the plots to a considerable degree (f_i being applied to both A_{i1} and A_{i2} , and g_j to both B_{1j} and B_{2j}).

⁸Unlike in the foregoing we cannot add here “without loss of generality.” The distance tests, as explained in Section 3.4, are not invariant with respect to linear transformations of the random variables involved.

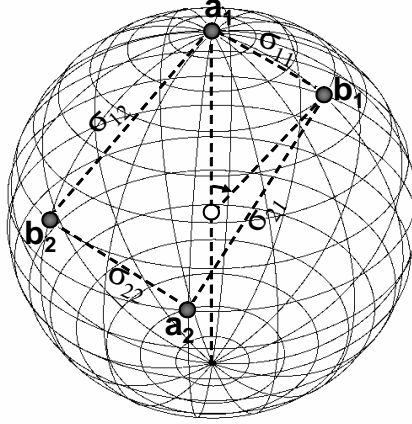


Figure 3. The four Euclidean distances between the \mathbf{a} -points and \mathbf{b} -points of Fig. 1: $\sigma_{ij} = \|\mathbf{a}_i - \mathbf{b}_j\|$. From the triangle $\mathbf{a}_1\mathbf{o}\mathbf{b}_1$ (where \mathbf{o} denotes the center of the sphere), $\|\mathbf{a}_1 - \mathbf{b}_1\|^2 = \|\mathbf{a}_1 - \mathbf{o}\|^2 + \|\mathbf{b}_1 - \mathbf{o}\|^2 - 2\|\mathbf{a}_1 - \mathbf{o}\| \cdot \|\mathbf{b}_1 - \mathbf{o}\| \cdot \cos \angle \mathbf{a}_1\mathbf{o}\mathbf{b}_1 = 1 + 1 - 2\rho_{11}$. The computation of the remaining three distances through correlations is analogous.

An immediate consequence of (36) is that if one of these inequalities is violated (i.e., the distance test for correlations is failed), the cosphericity test should be failed too. Thus, the correlation matrix

$$\mathbf{R}_4 = \begin{bmatrix} .98 & .155 \\ .92 & -.815 \end{bmatrix}$$

fails (36), for it translates into distances

$$\mathbf{D}_4 = \begin{bmatrix} 0.2 & 1.3 \\ 0.4 & 1.91 \end{bmatrix}$$

with

$$1.91 > 0.2 + 1.3 + 0.4.$$

As a result, we know that \mathbf{R}_4 should also fail the cosphericity test, which is easy to check directly:

$$|.98 \cdot .155 - .92 \cdot (-.815)| > \sqrt{1 - .98^2} \sqrt{1 - .155^2} + \sqrt{1 - .92^2} \sqrt{1 - (-.815)^2}.$$

We know from Section 3.2 that the correlation matrix

$$\mathbf{R}_{ex.(1)} = \begin{bmatrix} .7299 & .7299 \\ .7299 & -.6322 \end{bmatrix}$$

of our opening example, (1), passes the cosphericity test. As a result, it should also pass the distance test for correlations: indeed, the matrix of pairwise distances in this case is

$$\mathbf{D}_{ex.(1)} = \begin{bmatrix} 0.7350 & 0.7350 \\ 0.7350 & 1.8068 \end{bmatrix},$$

and the inequalities (36) are clearly satisfied.

The converse, however, is not true: a correlation matrix may pass the distance test for correlations but fail the cosphericity test. Thus, the already introduced correlation matrix

$$\mathbf{R}_3 = \begin{bmatrix} -.8 & -.6 \\ -.6 & .9 \end{bmatrix}$$

fails the cosphericity test, but it translates into distances

$$\mathbf{D}_3 = \begin{bmatrix} 1.90 & 1.79 \\ 1.79 & 0.447 \end{bmatrix}$$

which satisfy (36).

The latter example shows that the distance test for correlations is strictly weaker (less sensitive to violations of selectivity) than the cosphericity test. One can see this by comparing Fig. 2 to Fig. 4, where the quadruples of correlation coefficients failing the distance test are shown by black areas: clearly, these areas are proper subsets of the black areas in Fig. 2 (the quadruples of correlations failing the cosphericity test). Figure 5 provides a detailed illustration of this fact for four subsets of correlation quadruples.

This relationship between the two tests should come as no surprise, as the triangle inequalities used in (36) hold true in arbitrary metric spaces while the cosphericity test pertains only to the Euclidean geometry on a sphere. It turns out, however, that the relative weakness of the distance test is compensated for by its generalizability beyond the scope of second order mixed moments.

3.4. Distance tests in general

The classical Minkowski inequality, when applied to measurable functions F and G on some probability space with a probability measure λ , states that for any $p \geq 1$,

$$\sqrt[p]{\int |F(c) + G(c)|^p d\lambda(c)} \leq \sqrt[p]{\int |F(c)|^p d\lambda(c)} + \sqrt[p]{\int |G(c)|^p d\lambda(c)},$$

with the integration over the entire set of c -values. The inequality holds irrespective of whether the two right-hand integrals are finite (in which case the left-hand integral is finite too) or one of them is infinite (in which case the inequality holds trivially). This inequality can be expressed as

$$\sqrt[p]{\mathbf{E}[|F + G|^p]} \leq \sqrt[p]{\mathbf{E}[|F|^p]} + \sqrt[p]{\mathbf{E}[|G|^p]}, \quad p \geq 1, \quad (38)$$

where the moments must exist as finite or infinite numbers.

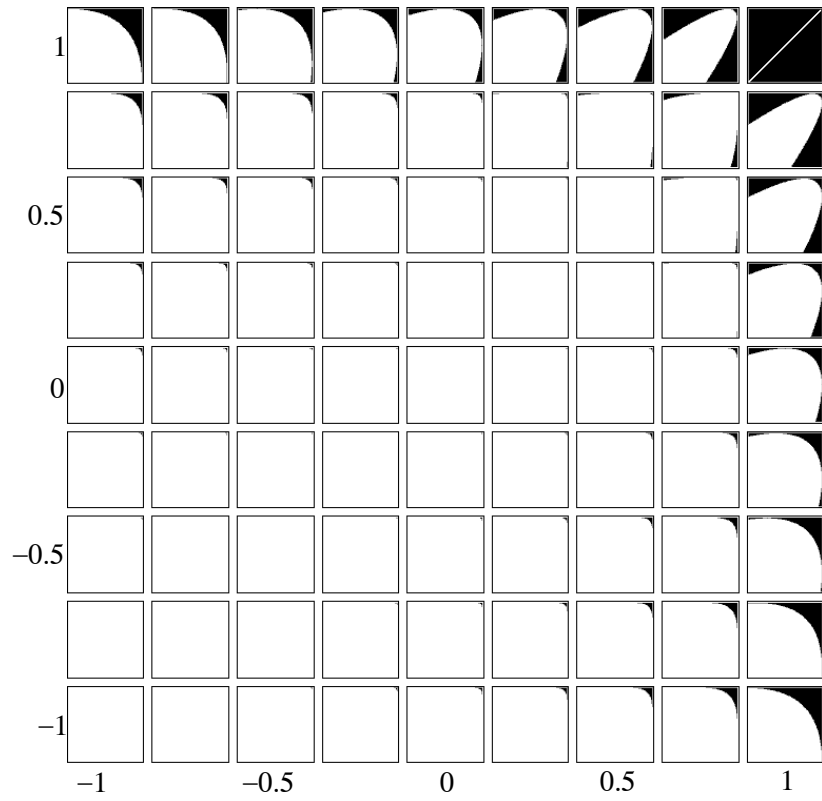


Figure 4. A graphical depiction of the distance test for correlation coefficients, (36). The format is the same as in Fig. 2: the white and black areas indicate correlations satisfying (respectively, violating) the inequalities (36).

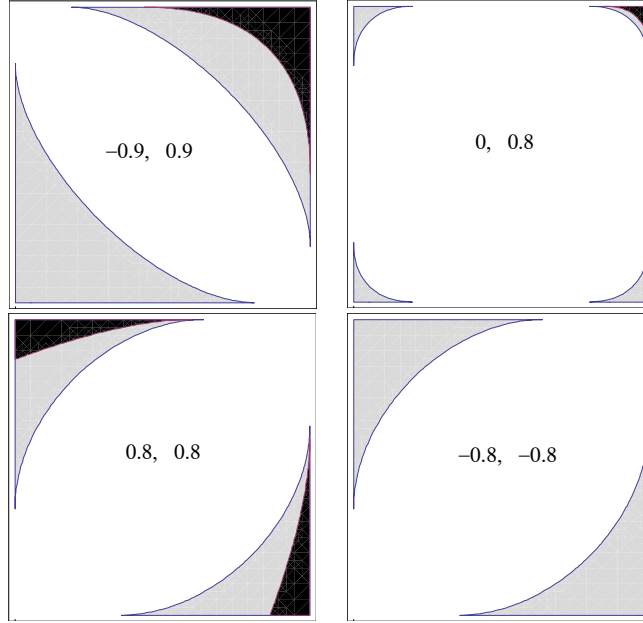


Figure 5. Direct comparison of the cosphericity test and the distance test for correlations. In each panel, two of the correlations (no matter which) are fixed and shown by the numbers in the center, the other two correlations vary on the axes. The dark areas contain correlations violating both tests, in the lighter shaded areas the correlations violate the cosphericity test alone, the correlations in the white areas satisfy both.

It is easy to see now that $\sqrt[p]{\mathbb{E}[|P - Q|^p]}$ is a distance between random variables P and Q : it is zero if and only if $P = Q$,⁹ it is symmetrical, and it satisfies the triangle inequality

$$\sqrt[p]{\mathbb{E}[|P - Q|^p]} \leq \sqrt[p]{\mathbb{E}[|P - R|^p]} + \sqrt[p]{\mathbb{E}[|R - Q|^p]}, \quad (39)$$

which can be seen by applying (38) to $F = P - R$ and $G = R - P$.

If $(A, B) \leftrightarrow (x, y)$, using a representation in the form (4), we have all moments

$$\begin{aligned} \sqrt[p]{\mathbb{E}[|A_x - B_y|^p]} &= \sqrt[p]{\int |A(x, c) - B(y, c)|^p d\lambda(c)}, \\ \sqrt[p]{\mathbb{E}[|A_x - A_{x'}|^p]} &= \sqrt[p]{\int |A(x, c) - A(x', c)|^p d\lambda(c)}, \\ \sqrt[p]{\mathbb{E}[|B_y - B_{y'}|^p]} &= \sqrt[p]{\int |B(y, c) - B(y', c)|^p d\lambda(c)} \end{aligned} \quad (40)$$

well-defined. Note, however, that only the first of three represents a moment for a potentially observable random variable. The variables $A_x - A_{x'}$ and $B_y - B_{y'}$ are “fictitious” random variables because neither $(A_x, A_{x'})$ nor $(B_y, B_{y'})$ has an observable joint distribution (“co-occurrence scheme”). Nevertheless the expressions in (40) are legitimate because, as discussed in Section 1.3, representations (4)-(5) allow us to treat $(A_x, B_y, A_{x'}, B_{y'})$ as if they had a joint distribution, even though it can be chosen at will insofar as it has specified pairwise marginals (A_x, B_y) , $(A_x, B_{y'})$, $(A_{x'}, B_y)$, and $(A_{x'}, B_{y'})$.

⁹The equality here should be taken to designate the equivalence relation $P(c) = Q(c)$ λ -almost everywhere.

Applying (39) to (40) we get

$$\begin{aligned}\sqrt[p]{\mathbb{E}[|A_x - B_y|^p]} &\leq \sqrt[p]{\mathbb{E}[|A_x - A_{x'}|^p]} + \sqrt[p]{\mathbb{E}[|A_{x'} - B_y|^p]}, \\ \sqrt[p]{\mathbb{E}[|A_x - A_{x'}|^p]} &\leq \sqrt[p]{\mathbb{E}[|A_x - B_{y'}|^p]} + \sqrt[p]{\mathbb{E}[|A_{x'} - B_{y'}|^p]},\end{aligned}$$

whence

$$\sqrt[p]{\mathbb{E}[|A_x - B_y|^p]} \leq \sqrt[p]{\mathbb{E}[|A_x - B_{y'}|^p]} + \sqrt[p]{\mathbb{E}[|A_{x'} - B_{y'}|^p]} + \sqrt[p]{\mathbb{E}[|A_{x'} - B_y|^p]}. \quad (41)$$

If we generically denote

$$s_{xy} = \sqrt[p]{\mathbb{E}[|A_{xy} - B_{xy}|^p]}, \quad (42)$$

then, under $(A, B) \leftrightarrow (x, y)$, we have

$$s_{xy} = \sqrt[p]{\mathbb{E}[|A_x - B_y|^p]}, \quad (43)$$

and (41) can be presented as

$$s_{xy} \leq s_{x'y} + s_{xy'} + s_{x'y'}, \quad \text{for all } x, x' \in X \text{ and } y, y' \in Y. \quad (44)$$

It is instructive to see why this need not be true if $(A, B) \not\leftrightarrow (x, y)$. If (4) does not hold, we have to define s_{xy} by the general formula (42) rather than (43). It is easy to see then that we cannot apply the Minkowski inequality to derive (44) for, generally,

$$A_{xy} - B_{xy} \neq [A_{x',y} - B_{x',y}] + [B_{xy} - A_{xy}] + [A_{x'y'} - B_{x'y'}].$$

The inequality (44) is a necessary condition for $(A, B) \leftrightarrow (x, y)$ which holds for any $p \geq 1$. Moreover, as we have observed earlier, if $(A, B) \leftrightarrow (x, y)$, then also

$$(f(A, x), g(B, y)) \leftrightarrow (x, y), \quad (45)$$

for any functions f, g with measurable $a \mapsto f(a, x)$ and $b \mapsto g(b, y)$. This means that under selective influence the condition (44) should hold true for all

$$s_{xy} = \sqrt[p]{\mathbb{E}[|f(A_{xy}, x) - g(B_{xy}, y)|^p]} = \sqrt[p]{\mathbb{E}[|f(A_x, x) - g(B_y, y)|^p]}. \quad (46)$$

Confining our attention again to an $\{x_1, x_2\} \times \{y_1, y_2\}$ factorial design, the choice of transformations f, g amounts to that of four measurable functions f_1, f_2, g_1, g_2 . Denoting¹⁰

$$s_{ij} = s_{x_i y_j} = \sqrt[p]{\mathbb{E}[|f_i(A_{ij}) - g_j(B_{ij})|^p]}, \quad i, j \in \{1, 2\}, \quad (47)$$

we rewrite (44) as

$$\begin{aligned}s_{11} &\leq s_{12} + s_{22} + s_{21}, \\ s_{22} &\leq s_{21} + s_{11} + s_{12}, \\ s_{12} &\leq s_{11} + s_{21} + s_{22}, \\ s_{21} &\leq s_{22} + s_{12} + s_{11}.\end{aligned} \quad (48)$$

¹⁰If s_{ij} in (47) are computed with different choices of (p, f_1, f_2, g_1, g_2) in the same context, it may become necessary to distinguish them notationally, say, $s_{ij}^{(p)}$ or $s_{ij}^{(p, f_1, f_2, g_1, g_2)}$. We avoid doing this in this paper and use s_{xy} as a generic symbol for any choice of (p, f_1, f_2, g_1, g_2) .

PROPOSITION 8 (DISTANCE TEST) *If random variables $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ are selectively influenced by (x, y) on $\{x_1, x_2\} \times \{y_1, y_2\}$, then, given any $p \geq 1$ and any measurable transformations f_1, f_2, g_1, g_2 , the moments s_{ij} defined by (47) satisfy the inequalities (48).*

We refer to (48) as the *distance test* for given p, f_1, f_2, g_1, g_2 , omitting mentioning p and/or the transformations as convenient. To show that $(A, B) \not\#(x, y)$ on $\{x_1, x_2\} \times \{y_1, y_2\}$ (assuming that marginal selectivity is satisfied), it will suffice to find some functions f_1, f_2, g_1, g_2 and some $p \geq 1$ for which the distance test is failed. The previously considered distance test for correlations corresponds to $p = 2$ provided the random variables involved are all standardized in the mean and variance.

Let us illustrate the logical independence of the distance test for different choices of f_1, f_2, g_1, g_2 and p on our opening example, (1). Applying the value $p = 1$ to (1) we get

$$\mathbf{D}_{ex.(1)}^{(p=1)} = \begin{bmatrix} \mathbb{E}[|A_{11} - B_{11}|] = 0.7 & \mathbb{E}[|A_{12} - B_{12}|] = 0.7 \\ \mathbb{E}[|A_{21} - B_{21}|] = 0.7 & \mathbb{E}[|A_{22} - B_{22}|] = 0.31 \end{bmatrix},$$

and we see that the distance test is failed:

$$0.31 > 0.7 + 0.7 + 0.7.$$

If we instead apply to (1) the value $p = 1.6$, however, we get

$$\mathbf{D}_{ex.(1)}^{(p=1.6)} = \begin{bmatrix} \sqrt[1.6]{\mathbb{E}[|A_{11} - B_{11}|^{1.6}]} = 1.249 & \sqrt[1.6]{\mathbb{E}[|A_{12} - B_{12}|^{1.6}]} = 1.249 \\ \sqrt[1.6]{\mathbb{E}[|A_{21} - B_{21}|^{1.6}]} = 1.249 & \sqrt[1.6]{\mathbb{E}[|A_{22} - B_{22}|^{1.6}]} = 3.590 \end{bmatrix},$$

which passes the distance test:

$$1.249 < 1.249 + 1.249 + 3.590,$$

$$3.590 < 1.249 + 1.249 + 1.249.$$

The situation reverses once again if we first transform the random variables by using the f_1, f_2, g_1, g_2 in (32) to obtain the random variables (33), and if we then apply to (33) the same $p = 1.6$ as in the previous computation:

$$\mathbf{D}_{ex.(1)}^{*(p=1.6)} = \begin{bmatrix} \sqrt[1.6]{\mathbb{E}[|A_{11}^* - B_{11}^*|^{1.6}]} = 0.4513 & \sqrt[1.6]{\mathbb{E}[|A_{12}^* - B_{12}^*|^{1.6}]} = 0.4513 \\ \sqrt[1.6]{\mathbb{E}[|A_{21}^* - B_{21}^*|^{1.6}]} = 0.4513 & \sqrt[1.6]{\mathbb{E}[|A_{22}^* - B_{22}^*|^{1.6}]} = 1.411 \end{bmatrix},$$

and the distance test is failed:

$$1.411 > 0.4513 + 0.4513 + 0.4513.$$

The outcome of the distance test therefore can change depending on both one's choice of p and one's choice of f_1, f_2, g_1, g_2 .

3.5. Multitude and properties of distance tests

One evident property of (48) is that, analogous to the cosphericity test, the distance test is invariant with respect to all permutations of $(s_{11}, s_{12}, s_{21}, s_{22})$: in all computations, these distances can be treated as an unordered quadruple of numbers. This property can be made even more apparent if we rewrite (48) as

$$\max\{s_{11}, s_{12}, s_{21}, s_{22}\} \leq \frac{s_{11} + s_{12} + s_{21} + s_{22}}{2}. \quad (49)$$

This inequality follows from the obvious fact that (48) is equivalent to

$$\max\{s_{11}, s_{12}, s_{21}, s_{22}\} \leq (s_{11} + s_{12} + s_{21} + s_{22}) - \max\{s_{11}, s_{12}, s_{21}, s_{22}\}.$$

The distance test can be presented in other compact forms:

$$|s_{11}^2 + s_{22}^2 - s_{12}^2 - s_{21}^2| \leq 2s_{11}s_{22} + 2s_{12}s_{21}. \quad (50)$$

or

$$|s_{11}^2 + s_{12}^2 - s_{21}^2 - s_{22}^2| \leq 2s_{11}s_{12} + 2s_{21}s_{22}. \quad (51)$$

The equivalence of these inequalities to (48) is shown by simple algebra.

It is not possible to provide a graphical illustration for the general distance test analogous to Figs. 2 and 4 because the range of possible values for $s_{11}, s_{12}, s_{21}, s_{22}$ is unbounded.

Like in the case of the cosphericity test, the distance tests using different values of p and/or different transformations f_1, f_2, g_1, g_2 are logically independent of each other. Unlike for the cosphericity test, however, this applies even to linear transformations, including pure shifts

$$\begin{aligned} A_{ij}^* &= A_{ij} + \alpha_i, \\ B_{ij}^* &= B_{ij} + \beta_j, \end{aligned}$$

and simple scaling

$$\begin{aligned} A_{ij}^* &= \alpha_i A_{ij}, \\ B_{ij}^* &= \beta_j B_{ij}. \end{aligned}$$

We illustrate this fact on the random variables in (33), treating them now as our original variables (A_{ij}, B_{ij}) . These variables fail the distance test with $p = 1$:

$$\mathbf{D}_{ex.(33)}^{(p=1)} = \begin{bmatrix} \text{E}[|A_{11} - B_{11}|] = 0.28, & \text{E}[|A_{12} - B_{12}|] = 0.28 \\ \text{E}[|A_{21} - B_{21}|] = 0.28 & \text{E}[|A_{22} - B_{22}|] = 1.24 \end{bmatrix}$$

and (using the form (49) of the test)

$$\max\{0.28, 0.28, 0.28, 1.24\} > \frac{0.28 + 0.28 + 0.28 + 1.24}{2}.$$

If, however, we first apply to (33) the scaling transformations

$$f_1 \equiv f_2 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 2 \\ 2 \mapsto 4 \end{cases}, \quad g_1 \equiv g_2 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases}, \quad (52)$$

to obtain

		B^*					B^*		
	(x_1, y_1)	0	1	2		(x_1, y_2)	0	1	2
A^*	0	.24	.07	0	A^*	0	.24	.07	0
	2	.07	.24	.07		2	.07	.24	.07
	4	0	.07	.24		4	0	.07	.24
		B^*					B^*		
	(x_2, y_1)	0	1	2		(x_2, y_2)	0	1	2
A^*	0	.24	.07	0	A^*	0	0	.07	.24
	2	.07	.24	.07		2	.07	.24	.07
	4	0	.07	.24		4	.24	.07	0

the transformed variables will pass the distance test with $p = 1$:

$$\mathbf{D}_{ex.(33)}^{*(p=1)} = \begin{bmatrix} \mathbb{E}[|A_{11}^* - B_{11}^*|] = 1.14 & \mathbb{E}[|A_{12}^* - B_{12}^*|] = 1.14 \\ \mathbb{E}[|A_{21}^* - B_{21}^*|] = 1.14 & \mathbb{E}[|A_{22}^* - B_{22}^*|] = 2.1 \end{bmatrix}$$

and

$$\max\{1.14, 1.14, 1.14, 2.1\} < \frac{1.14 + 1.14 + 1.14 + 2.1}{2}.$$

The same difference in the test outcomes can be shown by applying to (33) $p = 1$ without and with pure shifts

$$f_1 \equiv f_2 : \begin{cases} 0 \mapsto 10 \\ 1 \mapsto 11 \\ 2 \mapsto 12 \end{cases}, \quad g_1 \equiv g_2 : \begin{cases} 0 \mapsto 0 \\ 1 \mapsto 1 \\ 2 \mapsto 2 \end{cases}.$$

One known to us exception is the case when $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ are standardized in the mean and the distance test is passed with $p = 2$. Then it can be shown that the distance test with $p = 2$ has to be passed by the shifted random variables $A_{ij}^* = A_{ij} + \alpha_i$ and $B_{ij}^* = B_{ij} + \beta_j$.¹¹ But $p = 2$ is precisely the case when the distance test need not be performed at all, as this test is superseded by the more restrictive cosphericity test. We have already seen this for the case when $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ are standardized in the mean and variance (Fig. 5), so the proposition stated next generalizes this observation.

PROPOSITION 9 *If $\{(A_{ij}, B_{ij})\}_{i,j \in \{1,2\}}$ (satisfying marginal selectivity) pass the cosphericity test, then they must also pass the distance test with $p = 2$.*

¹¹To outline the demonstration: if $\sigma_{ij} = \sqrt{\mathbb{E}[(A_{ij} - B_{ij})^2]}$, then $\sqrt{\mathbb{E}[(A_{ij}^* - B_{ij}^*)^2]} = \sqrt{\sigma_{ij}^2 + d_{ij}^2}$, where $d_{ij} = \alpha_i - \beta_j$; if $\sigma_{11} \leq \sigma_{12} + \sigma_{22} + \sigma_{21}$ (distance test with $p = 2$ is passed), then $\sqrt{\sigma_{11}^2 + d_{11}^2} \leq \sqrt{(\sigma_{12} + \sigma_{22} + \sigma_{21})^2 + d_{11}^2} = \sqrt{(\sigma_{12} + \sigma_{22} + \sigma_{21})^2 + (d_{12} + d_{22} + d_{21})^2}$; the latter expression, following some tedious algebra and using the Cauchy–Schwarz inequality, is seen to be less than $\sqrt{\sigma_{12}^2 + d_{12}^2} + \sqrt{\sigma_{22}^2 + d_{22}^2} + \sqrt{\sigma_{21}^2 + d_{21}^2}$.

Indeed, let (A_{ij}, B_{ij}) have means $(\mu_{A,i}, \mu_{B,j})$, variances $(\sigma_{A,i}^2, \sigma_{B,j}^2)$, and correlations ρ_{ij} ($i, j \in \{1, 2\}$). Then one can define random variables (P_{ij}, Q_{ij}) such that

$$(P_{ij}, Q_{ij}) \sim N_2 \left(\mathbf{m}_{ij} = \begin{bmatrix} \mu_{A,i} \\ \mu_{B,j} \end{bmatrix}, \mathbf{V}_{ij} = \begin{bmatrix} \sigma_{A,i}^2 & \sigma_{A,i}\sigma_{B,j}\rho_{ij} \\ \sigma_{A,i}\sigma_{B,j}\rho_{ij} & \sigma_{B,j}^2 \end{bmatrix} \right).$$

If the matrix of correlations

$$\mathbf{R} = \begin{bmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{bmatrix}$$

passes the cosphericity test, then we know from Proposition 5 that $(P, Q) \not\leftrightarrow (x, y)$ on $\{x_1, x_2\} \times \{y_1, y_2\}$. This means that no selectivity test involving only the means $(\mu_{A,i}, \mu_{B,j})$, variances $(\sigma_{A,i}^2, \sigma_{B,j}^2)$, and correlations ρ_{ij} can reject $(A, B) \leftrightarrow (x, y)$, because then it will also have to reject $(P, Q) \leftrightarrow (x, y)$. But the distance test with $p = 2$ is based precisely on these means, variances, and correlations:

$$s_{ij} = \mathbb{E} \left[(A_{ij} - B_{ij})^2 \right] = (\sigma_{A,i}^2 + \mu_{A,i}^2) + (\sigma_{B,j}^2 + \mu_{B,j}^2) - 2(\sigma_{A,i}\sigma_{B,j}\rho_{ij} + \mu_{A,i}\mu_{B,j}).$$

This completes the proof of the proposition.

4. Conclusion

Our main results are summarized in the flowcharts presented in Figs. 6, 7, 8, and 9.

The potential infinity of the selective influence tests raises the question of how to apply them (with different choices of p and the transformations) in an “optimal order,” to maximize the chances of detecting a violation of selective influence with a minimum number of tests. We do not have an answer to this question. An obvious prescription is that one has to look first of all for transformations rendering the random variables bivariate normally distributed. If they exist, the cosphericity test unambiguously decides between the presence and absence of a selective influence relation. If they cannot be found, it is reasonable to seek transformations rendering at least two of the correlation coefficients as large as possible in absolute values. More work is needed to develop more refined recommendations.

One limitation of both the cosphericity and distance tests is that they are formulated for 2×2 factorial designs, even though the latter play a prominent role in experimentation. For larger factor sets $X \times Y$ one can, of course, apply the tests to various 2×2 subsets of $X \times Y$, and a failure of any of our tests on any of such subsets will rule out the possibility of selective influence on $X \times Y$. As mentioned in Section 1.2, however, even if selective influence is positively established on all 2×2 subsets (which our tests can only achieve in the case of bivariate normality), this will not guarantee the selective influence relation for the entire $X \times Y$. More work is needed to formulate appropriate generalizations of our tests for arbitrary $n \times m$ factorial designs, and even more generally, for k random variables (R_1, \dots, R_k) influenced by k factors (v_1, \dots, v_k) varying on an $n_1 \times \dots \times n_k$ set of combined values.

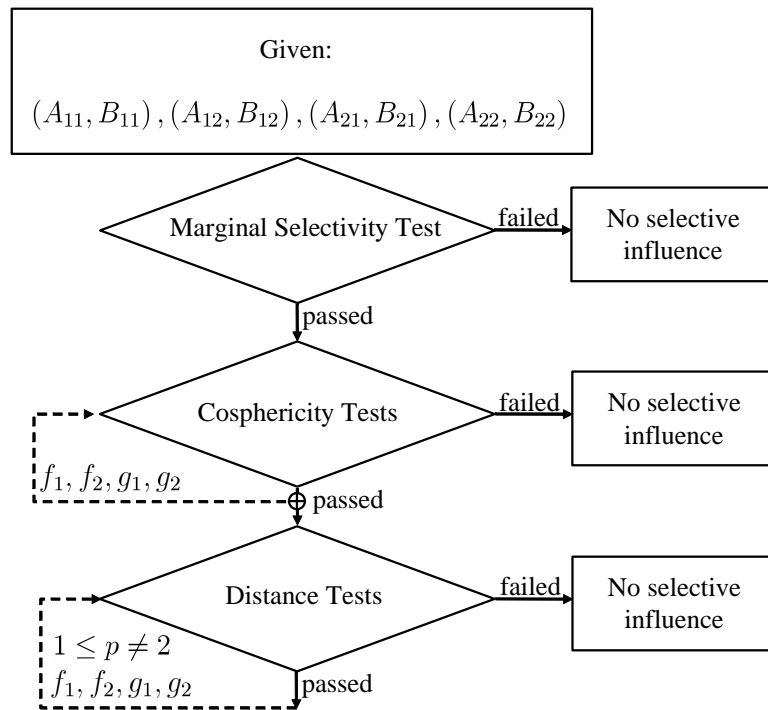


Figure 6. The overall flowchart for testing selectiveness in a 2×2 factorial setting. The symbol \oplus indicates a choice between two paths. The dashed lines show that the cosphericity test and the distance test can be repeated ad infinitum with different transformations f_1, f_2, g_1, g_2 and (in the case of the distance test) exponent p . A more detailed view of the testing blocks is given on Figs. 7, 8, and 9.

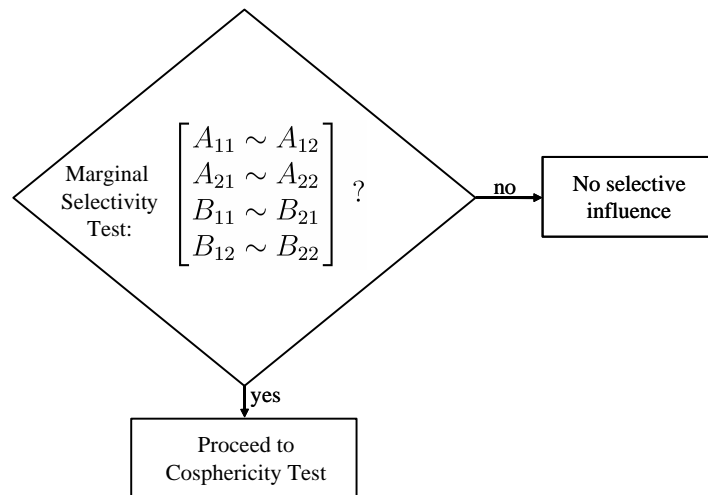


Figure 7. The marginal selectivity test. The symbol \sim stands for “is distributed as.”

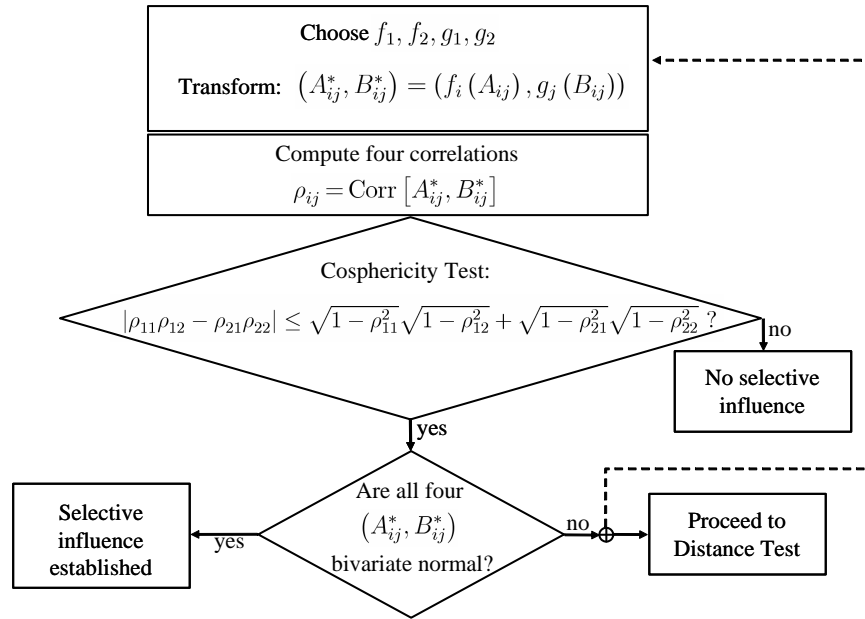


Figure 8. The cosphericity test. If transformations f_1, f_2, g_1, g_2 can be found which render the random variables bivariate normally distributed, the test becomes definitive: if it is passed (failed), the selective influence hypothesis is established (respectively, ruled out). Otherwise the test can be repeated with various transformations f_1, f_2, g_1, g_2 which are componentwise linearly independent (see Sections 3.1 and 3.2 for details).

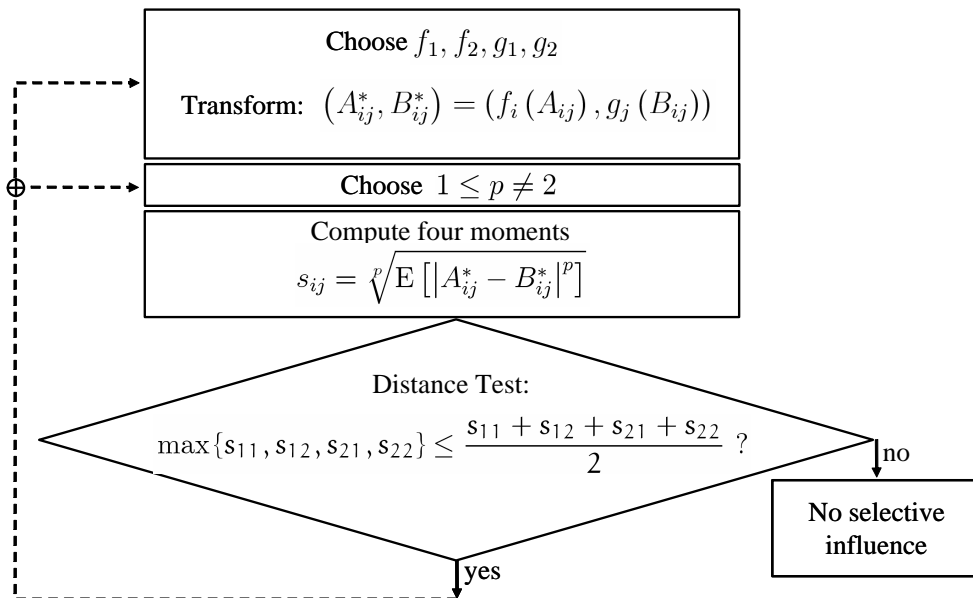


Figure 9. The distance test. It can be repeated with various (even componentwise linearly dependent) transformations f_1, f_2, g_1, g_2 and different values of $p \geq 1$. The value $p = 2$ need not be used as the distance test then is superseded by the cosphericity test (see Sections 3.3 and 3.4 for details).

Acknowledgement. This research has been supported by NSF grant SES 0620446 and AFOSR grant FA9550-06-1-0288 to Purdue University, as well as by the Commission of the European Communities, Marie Curie Excellence Grant MEXT-CT-2004-014203 to University of Jyväskylä. The authors are grateful to two anonymous reviewers for many helpful comments.

References

- Bloxom, B. (1972). The simplex in pair comparisons. *Psychometrika*, *37*, 119–136.
- Dzhafarov, E.N. (1992). The structure of simple reaction time to step-function signals. *Journal of Mathematical Psychology*, *36*, 235–268.
- Dzhafarov, E.N. (1997). Process representations and decompositions of response times. In A.A.J. Marley (Ed.), *Choice, Decision and Measurement: Essays in Honor of R. Duncan Luce* (pp. 255–278). Mahwah, NJ: Erlbaum.
- Dzhafarov, E.N. (2003a). Selective influence through conditional independence. *Psychometrika*, *68*, 7–26.
- Dzhafarov, E.N. (2003b). Thurstonian-type representations for “same-different” discriminations: Probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, *47*, 229–243.
- 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., & Schweickert, R. (1995). Decompositions of response times: An almost general theory. *Journal of Mathematical Psychology*, *39*, 285–314.
- Dzhafarov, E.N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, *48*, 51–64.
- Lazarsfeld, P.F. (1965). Latent structure analysis. In S. Sternberg, V. Capecchi, T. Kloek, & C.T. Leenders (Eds.), *Mathematics and Social Sciences*, v. 1 (pp. 37–54). Paris: Mouton.
- Roberts, S., & Sternberg, S. (1992). The meaning of additive reaction-time effects: Tests of three alternatives. In D.E. Meyer and S. Kornblum (Eds.), *Attention and Performance XIV* (pp. 611–654). Cambridge, MA: MIT Press.
- Schweickert, R. (1982). The bias of an estimate of coupled slack in stochastic PERT networks. *Journal of Mathematical Psychology*, *26*, 1–12.
- Schweickert, R., Giorgini, M., & Dzhafarov, E.N. (2000). Selective influence and response time cumulative distribution functions in serial-parallel networks. *Journal of Mathematical Psychology*, *44*, 504–535.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders's method. *Acta Psychologica*, *30*, 276-315.

Tong, Y.L. (1990). *The multivariate normal distribution*. New York: Springer-Verlag.

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.

Townsend, J.T., & Thomas, R.D. (1994). Stochastic dependencies in parallel and serial models: Effects on systems factorial interactions. *Journal of Mathematical Psychology*, *38*, 1-34.

Thurstone, L.L. (1927). A law of comparative judgments. *Psychological Review*, *34*, 273-286.