PSYCHOMETRIKA—VOL. 68, NO. 1, 7–25
MARCH 2003

SELECTIVE INFLUENCE THROUGH CONDITIONAL INDEPENDENCE

EHTIBAR N. DZHAFAROV

PURDUE UNIVERSITY

Let each of several (generally interdependent) random vectors, taken separately, be influenced by a particular set of external factors. Under what kind of the joint dependence of these vectors on the union of these factor sets can one say that each vector is selectively influenced by "its own" factor set? The answer proposed and elaborated in this paper is: One can say this if and only if one can find a factor-independent random vector given whose value the vectors in question are conditionally independent, with their conditional distributions selectively influenced by the corresponding factor sets. Equivalently, the random vectors should be representable as deterministic functions of "their" factor sets and of some mutually independent and factor-independent random variables, some of which may be shared by several of the functions.

Key words: selective influence, conditional independence, marginal selectivity, processing architectures, Thurstonian modeling.

1. Problem

This paper presents a generalization and improvement for the definition proposed in Dzhafarov (2001a) for selectiveness in the dependence of several random variables upon several (sets of) external factors. This generalization links the notion of selective influence with that of conditional independence, routinely used in psychometric theorizing.

The problem can be stated as follows (this formulation differs from the one adopted in Dzhafarov, 2001a). Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be random vectors (as a special case, random variables) whose marginal distributions depend on, respectively, sets of external factors $\Gamma_1, \ldots, \Gamma_n$, not necessarily disjoint or nonempty. If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are mutually independent, one can say, quite unambiguously, that $\mathbf{X}_i$ is being *selectively influenced* by factors belonging to Γ_i ($i = 1, \ldots, n$). If, however, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are not necessarily mutually independent, one faces a conceptual problem: Considering the joint distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ as depending on the external factors $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_n$, what conditions should be imposed on this dependence to enable one to speak of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ as being *selectively influenced* by, respectively, $\Gamma_1, \ldots, \Gamma_n$? (Since $\Gamma_1, \ldots, \Gamma_n$ are not necessarily disjoint, the selectiveness should be taken to mean that $\mathbf{X}_i$ is not influenced by factors outside Γ_i, $i = 1, \ldots, n$).

In the concluding section of this paper this problem and its solution are generalized to random entities taking on their values in arbitrary measure spaces. Until then, however, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are assumed to be random vectors with real-valued (though not necessarily continuous) components. The dimensionality of these vectors is usually left unspecified (as a special case, a random vector can be a random variable).

The three examples given in the next three sections should provide a motivation for and clarify the meaning of the problem of selective influence.

First, however, the following conventions should be noted. Boldface capital Roman letters $\mathbf{X}, \mathbf{Y}$, and so forth, always denote real-valued random vectors. A finite sequence of random vectors, say, $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, is also viewed as a random vector. Realizations of random vectors are always denoted by lowercase italics, x, y, $(x_1, \ldots, x_n)$, etcetera. The term "external factor"

This research has been supported by the NSF Grant SES-0001925. I am grateful to Hans Colonius and Richard Schweickert for helpful critical comments.

Requests for reprints should be sent to Ehtibar N. Dzhafarov, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2004. E-Mail: ehtibar@purdue.edu

0033-3123/2003-1/2001-0952-A $00.75/0

(or simply "factor") refers to an observable experimental manipulation, condition, or covariate upon which the distribution of a random vector of interest may or is known to depend. Factors may be continuous or discrete, and they are always treated as deterministic quantities. Factors and factor sets are always denoted by Greek letters.

2. Motivating Example: Multivariate Normal Distribution

Let $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ (generally interdependent random variables, say, performance scores in n tests) be n-variate normally distributed, and let the mean-variance pair (μ_i, σ_i^2) of $\mathbf{X}_i$ be known to depend on an external factor $\gamma_i (i = 1, \ldots, n)$. Assume that $\gamma_1, \ldots, \gamma_n$ (say, sex, age, blood pressure, etc.) are all distinct. Considering now the dependence of the n-variate normal distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ upon the set $\{\gamma_1, \ldots, \gamma_n\}$, we see that it is entirely characterized by the dependence on these factors of the correlation coefficients ρ_{ij} $(i, j = 1, \ldots, n, i \neq j)$. Generally speaking, each ρ_{ij} may depend on all factors $\gamma_1, \ldots, \gamma_n$. The question arises: Under what pattern of the dependence of the correlations ρ_{ij} upon $\gamma_1, \ldots, \gamma_n$ can one speak of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ being selectively influenced by $\gamma_1, \ldots, \gamma_n$, respectively? Should, for example, the notion of selectiveness imply that no ρ_{ij} may depend on any of these factors, or even that all ρ_{ij} must be zero?

The applicability of this example is not, of course, confined to performance scores. In particular, the two motivating examples below, dealing with response time decompositions and Thurstonian-type modeling, may very well have the same mathematical structure. The dependence of a multivariate normal distribution on external factors is used in Dzhafarov (1999, 2001a) as a "testing ground" for possible meanings of selective influence. It plays a similar role in the present paper.

Note that the situation where each $\mathbf{X}_i$ in the n-variate normally distributed $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is selectively influenced by a one-element factor set $\Gamma_i = \{\gamma_i\}$ is not the only one of interest. For example, with mutually independent $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ (i.e., with $\rho_{12} = \rho_{23} = \rho_{13} = 0$) it is perfectly reasonable to speak of $\mathbf{X}_1$ being selectively influenced by, say, $\Gamma_1 = \{\gamma_1, \gamma_2\}$, $\mathbf{X}_2$ being selectively influenced by $\Gamma_2 = \{\gamma_2, \gamma_3\}$, and $\mathbf{X}_3$ by $\Gamma_3 = \{\gamma_3\}$. The selectiveness here means that $\mathbf{X}_1$ is not being influenced by $\{\gamma_3\}$, $\mathbf{X}_2$ not being influenced by $\{\gamma_1\}$, and $\mathbf{X}_3$ by $\{\gamma_1, \gamma_2\}$. It is natural, therefore, to ask whether and how the same pattern of selectiveness can be established when $\rho_{12}, \rho_{23}, \rho_{13}$ are non-zero and may depend on $\gamma_1, \gamma_2, \gamma_3$.

3. Motivating Example: Processing Architectures

In the analysis of information processing architectures it is common to think of response times to stimuli as being composed of several time components, each defined by an attribute of stimulus or observation mode that influences this component selectively (Dzhafarov, 1997). Thus, simple reaction time to stimulus varying in intensity γ can be thought of as the sum of a signal-dependent component $\mathbf{X}_1$ (selectively influenced by $\Gamma_1 = \{\gamma\}$) and a signal-independent component $\mathbf{X}_2$ (selectively influenced by $\Gamma_2 = \emptyset$). In addition to stochastically independent $\mathbf{X}_1, \mathbf{X}_2$, with which the meaning of this selectiveness is trivial, Dzhafarov (1992) and Dzhafarov and Rouder (1996) argue in favor of considering positively correlated $\mathbf{X}_1, \mathbf{X}_2$. This leads to a special case of the problem posed in section 1: Under what dependence of $(\mathbf{X}_1, \mathbf{X}_2)$ on $\Gamma_1 \cup \Gamma_2 = \{\gamma\}$ would one say that $(\mathbf{X}_1, \mathbf{X}_2)$ are selectively influenced by, respectively $(\{\gamma\}, \emptyset)$?

Choice response times too are often presented as compositions $H(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ of durations selectively influenced by certain factor sets. In Sternberg's classical paradigm, for example (Roberts & Sternberg, 1993; Sternberg, 1969), $\mathbf{X}_1$ may be defined as the time of a process selectively influenced by memory set, while $\mathbf{X}_2$ may be defined as the time of a process selectively influenced by stimulus contrast. The focus of interest here is to determine the relationship between these two component processes in the "mental architecture": for example, do these processes develop in parallel or does one of them have to terminate before the other begins? Again, the problem of section 1 arises when $\mathbf{X}_1, \mathbf{X}_2$ are not assumed to be stochastically independent (Cortese & Dzhafarov, 1996; Dzhafarov & Cortese, 1996; Dzhafarov & Schweickert, 1995).

In relation to the previous section, note that some transformations $h_1(\mathbf{X}_1), \ldots, h_n(\mathbf{X}_n)$ (e.g., logarithms) of the hypothetical response time components may very well be assumed to be multivariate normally distributed.

Historically, the problem of developing a viable definition of selective influence under stochastic interdependence was first posed in the context of response time decompositions (Townsend, 1984; Townsend & Thomas, 1994). The solution proposed by Townsend ("indirect nonselectivity") is briefly discussed in section 7.

4. Motivating Example: Thurstonian-type Modeling

Pairwise comparisons (in terms of "greater-less" or "same-different") are often modeled by assuming that the two stimuli being compared, α, β, are mapped into random vectors ("images" of these stimuli) $\mathbf{X}, \mathbf{Y}$, respectively, and that the decision in a given trial is uniquely determined by their realizations x, y. If $\mathbf{X}, \mathbf{Y}$ are stochastically independent, with densities $f(x; \alpha)$ and $g(y; \beta)$, then the probability of a specified response ("$\alpha > \beta$" or "$\alpha \neq \beta$") is given by

$$\mathrm{P}(\alpha, \beta) = \int_{(x,y)\in G} f(x; \alpha) g(y; \beta)\, dx\, dy, \tag{1}$$

where G is the class of all (x, y) that are mapped into the answer in question. Thus, in the classical Thurstonian theory (Thurstone, 1927a, 1927b) $\mathrm{P}(\alpha, \beta)$ is the probability of "$\alpha > \beta$" (with respect to some semantically unidimensional attribute, say, "brightness"), $\mathbf{X}, \mathbf{Y}$ take on their values in Re (the axis presumably representing this attribute in one's perception), f, g are normal densities, and G consists of all (x, y) such that $x > y$. In Luce and Galanter's (1963) modification of Thurstone's modeling scheme $\mathrm{P}(\alpha, \beta)$ is the probability of "$\alpha \neq \beta$", the functions f, g are again normal densities in Re, and G consists of all (x, y) such that $x - y > \varepsilon_1$ or $y - x > \varepsilon_2$" (where $\varepsilon_1, \varepsilon_2$ are some positive constants).

When $\mathbf{X}, \mathbf{Y}$ are not assumed to be stochastically independent, denoting their joint density by $h(x, y; \alpha, \beta)$, we have

$$\mathrm{P}(\alpha, \beta) = \int_{(x,y)\in G} h(x, y; \alpha, \beta)\, dx\, dy. \tag{2}$$

At the same time, we still wish to treat $\mathbf{X}$ as an image of α, and $\mathbf{Y}$ as an image of β. This leads one to the conceptual problem posed in section 1: For what kind of densities $h(x, y; \alpha, \beta)$ the selective correspondence between stimuli and "their" images can be established in a nonarbitrary fashion? If $(\mathbf{X}, \mathbf{Y})$ are posited to be bivariate normally distributed, this question becomes a special case of the problem posed in section 2. Thurstone (1927a, 1927b), apparently, considered the selectiveness in the dependence of the marginal distributions of $\mathbf{X}$ and $\mathbf{Y}$ on, respectively, α and β sufficient for the selective attribution of $\mathbf{X}$ to α and $\mathbf{Y}$ to β, irrespective of how the correlation $\rho_{\mathbf{XY}}$ depends on (α, β). In this paper, however, the selectiveness on the level of marginals is only considered necessary but not sufficient for speaking of selective influence.

The problem of selective correspondence between stimuli and images (equivalently, selective influence of stimuli upon their images) becomes especially apparent if $\mathbf{X}, \mathbf{Y}$ are assumed to take on their values in Re^m, with an arbitrary dimensionality m, as it is done in Ashby and Perrin (1988), Dzhafarov (2001b), Ennis (1992), Ennis, Palen, and Mullen (1988), Suppes and Zinnes (1963), Zinnes and MacKey (1983), and Thomas (1996, 1999). When no semantically unidimensional attributes are involved, there is no good reason to assume that images $\mathbf{X}, \mathbf{Y}$ are unidimensional. In this case, unless the problem of selective correspondence is given a satisfactory solution, nothing prevents one from interpreting (2) as a model in which the stimulus pair (α, β) is mapped into a single random vector $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$, with no particular attribution of its components to α and β taken separately.

5. Definition of Selective Influence

Let Γ denote the set of all factors being considered, and $\Lambda \subseteq \Gamma$. The notation

$$\mathbf{X} \leftarrow \Lambda$$

means that the distribution of the random vector $\mathbf{X}$ does not depend on factors belonging to $\Gamma - \Lambda$ (so it may or may nor depend on factors belonging to Λ). The formulation of the selective influence problem in section 1 presupposes that

$$\mathbf{X}_1 \leftarrow \Gamma_1, \quad \ldots, \quad \mathbf{X}_n \leftarrow \Gamma_n, \tag{3}$$

a relation that Townsend and Schweickert (1989) call the *marginal selectivity*. The formulation of the selective influence problem also presupposes that the set of all factors being considered is confined to

$$\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_n.$$

In particular, the joint distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is not related to any factors outside Γ, that is,

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \leftarrow \Gamma. \tag{4}$$

Note that this is not a logical necessity: The joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ may very well depend on factors that do not influence any of the marginal distributions (think, e.g., of a bivariate normal distribution with the means and variances that do not depend on some factor while the correlation coefficient does). We simply agree to consider the dependence of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ on Γ only.

(The predication of selective influence upon (3) and (4) greatly simplifies the discussion, as compared to Dzhafarov, 2001a. In particular, this makes superfluous the notion of factor effectiveness, the distinction between selective influence and selective attributability, and the issue of uniqueness, all prominently entering the theory developed in Dzhafarov, 2001a.) Let the notation

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \looparrowleft (\Gamma_1, \ldots, \Gamma_n) \tag{5}$$

stand for the proposition "$\mathbf{X}_1, \ldots, \mathbf{X}_n$ are selectively influenced by $\Gamma_1, \ldots, \Gamma_n$, respectively".

Definition 1. Given $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $(\Gamma_1, \ldots, \Gamma_n)$ satisfying

$$\mathbf{X}_1 \leftarrow \Gamma_1, \quad \ldots, \quad \mathbf{X}_n \leftarrow \Gamma_n,$$
$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \leftarrow \Gamma_1 \cup \ldots \cup \Gamma_n,$$

the selective influence relation

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \looparrowleft (\Gamma_1, \ldots, \Gamma_n)$$

means that one can find mutually stochastically independent random vectors $\mathbf{C}, \mathbf{S}_1, \ldots, \mathbf{S}_n$ whose distributions do not depend on $\Gamma_1 \cup \ldots \cup \Gamma_n$, and some measurable functions $f_1, \ldots, f_n$, such that

$$\mathbf{X}_1 = f_1(\mathbf{C}, \mathbf{S}_1, \Gamma_1), \quad \ldots, \quad \mathbf{X}_n = f_n(\mathbf{C}, \mathbf{S}_n, \Gamma_n). \tag{6}$$

The random vector $\mathbf{C}$ can be referred to as a *common source of randomness* for $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, whereas the random vectors $\mathbf{S}_1, \ldots, \mathbf{S}_n$ can be called *specific sources of randomness* for $\mathbf{X}_1, \ldots, \mathbf{X}_n$, respectively. That the vector $(\mathbf{C}, \mathbf{S}_1, \ldots, \mathbf{S}_n)$ does not depend on any factors from

$\Gamma_1 \cup \ldots \cup \Gamma_n$ can be written as

$$(\mathbf{C}, \mathbf{S}_1, \ldots, \mathbf{S}_n) \leftarrow \emptyset.$$

This independence property is critical for the logic of the definition.

It is worth mentioning that the definition imposes no restrictions on how the functions $f_1, \ldots, f_n$ depend on $\mathbf{C}$ or on their specific sources of randomness $\mathbf{S}_1, \ldots, \mathbf{S}_n$. In particular, $\mathbf{C}$ is allowed to be a dummy argument for all $f_1, \ldots, f_n$, in which case the random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ will be stochastically independent,

$$\mathbf{X}_1 = f_1(\mathbf{S}_1, \Gamma_1), \quad \ldots, \quad \mathbf{X}_n = f_n(\mathbf{S}_n, \Gamma_n).$$

The stochastic independence also results if $\mathbf{C}$ is decomposable into subvectors $(\mathbf{C}_1, \ldots, \mathbf{C}_n)$, with f_i depending on $\mathbf{C}_i$ $(i = 1, \ldots, n)$ but not on the rest of the subvectors. In this case the pairs $(\mathbf{C}_i, \mathbf{S}_i)$ are de facto specific sources of randomness.

At the opposite extreme, when dealing with (unidimensional) random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$, the specific sources of randomness $\mathbf{S}_1, \ldots, \mathbf{S}_n$ may all be dummy arguments for $f_1, \ldots, f_n$,

$$\mathbf{X}_1 = f_1(\mathbf{C}, \Gamma_1), \quad \ldots, \quad \mathbf{X}_n = f_n(\mathbf{C}, \Gamma_n),$$

and if $\mathbf{C}$ is a random variable upon which all the functions depend monotonically, then $\mathbf{X}_1, \ldots, \mathbf{X}_n$ will all be monotonic functions of each other (for this form of stochastic relationship see Dzhafarov, 1992; Dzhafarov & Rouder, 1996; Dzhafarov & Schweickert, 1995).

If the functions $f_1, \ldots, f_n$ are linear in their second argument, one can recognize in (6), for any fixed value of $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_n$, the familiar ("nonlinear") factor-analytic structure (McDonald, 1967, 1982),

$$\mathbf{X}_i = a_i(\mathbf{C}, \Gamma_i) + b_i(\Gamma_i)\mathbf{S}_i, \qquad i = 1, \ldots, n, \tag{7}$$

where a_i, b_i are some functions. In the factor-analytic context the term "factors" (common and specific) is traditionally used to designate the sources of randomness $\mathbf{C}, \mathbf{S}_1, \ldots, \mathbf{S}_n$. (In the present paper the term "factor" is reserved for the elements of Γ, but the sources of randomness, if one so wishes, can also be referred to as "factors" provided the term is used with some qualifying adjective, like "internal", or "latent".)

It is also worth mentioning that the definition formally applies to $n = 1$, in which case

$$\mathbf{X} \hookleftarrow \Gamma \quad \text{if and only if} \quad \mathbf{X} \leftarrow \Gamma.$$

The "only if" part being obvious, note that if $\mathbf{X} \leftarrow \Gamma$ then $\mathbf{X}$ can always be presented as $\mathbf{X} = f(\mathbf{S}, \Gamma)$, where $\mathbf{S}$ is a vector of stochastically independent random variables uniformly distributed between 0 and 1 (see Comment 1 in the Appendix). In other words, if $\mathbf{X} \leftarrow \Gamma$ then (6) is satisfied, with $\mathbf{C}$ being a dummy argument.

It is easy to see that the definition of selective influence satisfies the following property, called the *nestedness of selective influence*. (The term is introduced in Dzhafarov, 2001a, but its meaning here is somewhat modified.) Let $(1, \ldots, n)$ be partitioned into two subvectors, presentable, after an appropriate permutation, as $(1, \ldots, k)$ and $(k + 1, \ldots, n)$, $0 < k < n$. The nestedness property means that (5) implies

$$\begin{gathered}
(\mathbf{X}_1, \ldots, \mathbf{X}_k) \hookleftarrow (\Gamma_1, \ldots, \Gamma_k), \\
(\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n) \hookleftarrow (\Gamma_{k+1}, \ldots, \Gamma_n), \\
((\mathbf{X}_1, \ldots, \mathbf{X}_k), (\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n)) \hookleftarrow ((\Gamma_1 \cup \ldots \cup \Gamma_k), (\Gamma_{k+1} \cup \ldots \cup \Gamma_n)).
\end{gathered} \tag{8}$$

These relations follow from (6) trivially. Indeed,

$$\mathbf{X}_1 = f_1(\mathbf{C}, \mathbf{S}_1, \Gamma_1), \quad \ldots, \quad \mathbf{X}_k = f_k(\mathbf{C}, \mathbf{S}_k, \Gamma_k),$$
$$\mathbf{X}_{k+1} = f_{k+1}(\mathbf{C}, \mathbf{S}_{k+1}, \Gamma_{k+1}), \quad \ldots, \quad \mathbf{X}_n = f_n(\mathbf{C}, \mathbf{S}_n, \Gamma_n),$$

which is equivalent to the first two lines of (8). Denoting $\mathbf{T}_1 = (\mathbf{S}_1, \ldots, \mathbf{S}_k)$, $\mathbf{T}_2 = (\mathbf{S}_{k+1}, \ldots, \mathbf{S}_n)$, one also observes that there are (vectorial) measurable functions F_1, F_2 such that

$$(\mathbf{X}_1, \ldots, \mathbf{X}_k) = F_1(\mathbf{C}, \mathbf{T}_1, \Gamma_1 \cup \ldots \cup \Gamma_k), (\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n) = F_2(\mathbf{C}, \mathbf{T}_2, \Gamma_{k+1} \cup \ldots \cup \Gamma_n),$$

which is equivalent to the third line of (8).

It immediately follows from (8), of course, that the marginal distribution of any subvector $(\mathbf{X}_1, \ldots, \mathbf{X}_k)$ cannot depend on factors outside $\Gamma_1 \cup \ldots \cup \Gamma_k$, that is,

$$(\mathbf{X}_1, \ldots, \mathbf{X}_k) \looparrowleft \Gamma_1 \cup \ldots \cup \Gamma_k. \tag{9}$$

This property, which is understood as holding for all subvectors of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, is referred as the *complete marginal selectivity* in the dependence of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ on $\Gamma_1 \cup \ldots \cup \Gamma_n$. The marginal selectivity (3) as defined by Townsend and Schweickert (1989) is then a special case that can be called the *simple marginal selectivity*. In the present work, therefore, simple marginal selectivity is a necessary condition for complete marginal selectivity, which is a necessary condition for nestedness, which in turn is a necessary condition for selective influence,

$$\boxed{\text{selective influence, (5)}}$$
$$\Downarrow$$
$$\boxed{\text{nestedness, (8)}}$$
$$\Downarrow \tag{10}$$
$$\boxed{\text{compete marginal selectivity, (9)}}$$
$$\Downarrow$$
$$\boxed{\text{simple marginal selectivity. (3)}}$$

It is easy to see that simple marginal selectivity does not imply complete marginal selectivity (unless $n = 2$, in which case the two concepts coincide). Think, for example, of a trivariate normally distributed $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, with their marginal mean-variance pairs being functions of $\gamma_1, \gamma_2, \gamma_3$, respectively, but with $\rho_{12} = \rho_{12}(\gamma_1, \gamma_2, \gamma_3)$. Clearly, the subvector $(\mathbf{X}_1, \mathbf{X}_2)$ is not influenced by $\{\gamma_1, \gamma_2\}$ only. This is one reason why simple marginal selectivity cannot be taken as a competing definition for the concept of selective influence.

The complete marginal selectivity is a significantly more powerful concept, but, as shown in section 10, it is weaker than the notion of selective influence. To use complete marginal selectivity as a competing definition for selective influence would be intellectually unsatisfactory: A good definition of selective influence should provide some form of “explanation” of why the random vectors under consideration are stochastically interdependent and why they are nevertheless selectively influenced by different factor sets. Definition 1 provides such an explanation: $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are generally interdependent because $f_1, \ldots, f_n$ depend on a common random vector $\mathbf{C}$, and $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are selectively influenced by $\Gamma_1, \ldots, \Gamma_n$, respectively, because each function only depends on “its own” factor set. It is natural to interpret $\mathbf{C}$ as reflecting certain factor-independent processes or characteristics affecting all variables under consideration. One can think, for example, of the general arousal, or attention level in the case of response times (section 3), or of individual aptitudes when $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are measured across a population of subjects, as in the case of interdependent performance scores (section 2).

Note that one can always add to (6) the dummy equality

$$\mathbf{C} = \mathbf{C}$$

to obtain

$$(\mathbf{C}, \mathbf{X}_1, \ldots, \mathbf{X}_n) \looparrowleft (\emptyset, \Gamma_1, \ldots, \Gamma_n),$$

with the nestedness property applicable now to all combinations of $\mathbf{X}_i$'s and $\mathbf{C}$. It follows, in particular, that the joint distribution of $(\mathbf{C}, \mathbf{X}_i)$ may only depend on Γ_i, and, more generally, the joint distribution of $(\mathbf{C}, \mathbf{X}_1, \ldots, \mathbf{X}_k)$ may only depend on $\Gamma_1 \cup \ldots \cup \Gamma_k$.

6. Two Illustrations

Although Definition 1 of selective influence has not been previously proposed in the literature (see the history of the issue in the next section), constructions conforming to this definition have been used. The two illustrations given below relate to the motivating examples given in sections 3 and 4.

The first illustration is taken from Schweickert (1982). Presented in a modified notation, the durations $(\mathbf{X}_1, \mathbf{X}_2)$ of two processes in a certain network of processes are posited in this paper to be influenced by two factors, γ_1, γ_2, both varying on two levels, 0 and 1. The relationship can be written as

$$\mathbf{X}_1 = \mathbf{X}_{10} + \gamma_1 \mathbf{X}_{11}, \quad \mathbf{X}_2 = \mathbf{X}_{20} + \gamma_2 \mathbf{X}_{21},$$

where $\mathbf{X}_{10}$, $\mathbf{X}_{11}$, $\mathbf{X}_{20}$, $\mathbf{X}_{21}$ are nonnegative random variables whose joint distribution does not depend on γ_1 or γ_2. The "baseline durations" $\mathbf{X}_{10}, \mathbf{X}_{20}$ are allowed to be stochastically interdependent, and so are the "prolongations" $\mathbf{X}_{11}, \mathbf{X}_{21}$, the two pairs being stochastically independent.

It is easy to see then that $(\mathbf{X}_1, \mathbf{X}_2)$ are stochastically interdependent while being selectively influenced by $(\{\gamma_1\}, \{\gamma_2\})$, in the sense of Definition 1. The formal compliance is achieved by modeling $\mathbf{X}_{10}$, $\mathbf{X}_{11}$, $\mathbf{X}_{20}$, $\mathbf{X}_{21}$ by functions $g_{10}(\mathbf{C}_0, \mathbf{S}_{10})$, $g_{11}(\mathbf{C}_1, \mathbf{S}_{11})$, $g_{20}(\mathbf{C}_0, \mathbf{S}_{20})$, $g_{21}(\mathbf{C}_1, \mathbf{S}_{21})$, respectively, with mutually independent $\mathbf{C}$'s and $\mathbf{S}$'s, and by putting in Definition 1 $(\mathbf{C}_0, \mathbf{C}_1) = \mathbf{C}$, $(\mathbf{S}_{10}, \mathbf{S}_{11}) = \mathbf{S}_1$, and $(\mathbf{S}_{20}, \mathbf{S}_{21}) = \mathbf{S}_2$:

$$\mathbf{X}_1 = g_{10}(\mathbf{C}_0, \mathbf{S}_{10}) + \gamma_1 g_{11}(\mathbf{C}_1, \mathbf{S}_{11}) = f_1(\mathbf{C}, \mathbf{S}_1, \{\gamma_1\}),$$
$$\mathbf{X}_2 = g_{20}(\mathbf{C}_0, \mathbf{S}_{20}) + \gamma_2 g_{21}(\mathbf{C}_1, \mathbf{S}_{21}) = f_2(\mathbf{C}, \mathbf{S}_2, \{\gamma_2\}).$$

The second illustration is taken from Bloxom (1972), dealing with a version of the classical Thurstonian theory for modeling "greater-less" discrimination probabilities (section 4). Again, the notation is modified to better suit our purposes. Let stimuli α and β take on their values in a finite set $\{1, \ldots, m\}$. Bloxom assumes that following some permutation of these indices (and renumbering them as $1, \ldots, m$ again), there are some (unidimensional) mutually independent random variables $(\mathbf{C}_1, \ldots, \mathbf{C}_m, \mathbf{G}_1 \ldots, \mathbf{G}_m)$ such that, denoting by $\mathbf{X}$, $\mathbf{Y}$ the random images of α and β, respectively,

$$\mathbf{X} = \sum_{i=1}^{\alpha} \mathbf{C}_i + \mathbf{G}_\alpha, \qquad \alpha = 1, \ldots, m,$$

$$\mathbf{Y} = \sum_{i=1}^{\beta} \mathbf{C}_i + \mathbf{G}_\beta, \qquad \beta = 1, \ldots, m.$$

(In Bloxom's paper, which focuses on the Thurstonian Case II, the random variability in $\mathbf{C}_j$ and $\mathbf{G}_j$, $j = 1, \ldots, m$, is taken to reflect individual differences within a population of subjects.)

One way of demonstrating the compliance of this model with Definition 1 is to denote by $G(s, j)$ the quantile of rank s $(0 \leq s \leq 1)$ of the random variable $\mathbf{G}_j$ $(j = 1, \ldots, m)$, and to introduce two random variables $\mathbf{S}_1, \mathbf{S}_2$, uniformly distributed between 0 and 1 and such that $\mathbf{C}_1, \ldots, \mathbf{C}_m, \mathbf{S}_1, \mathbf{S}_2$ are mutually independent. Then $G(\mathbf{S}_1, j)$ and $G(\mathbf{S}_2, j)$ are both distributed as $\mathbf{G}_j$, and the model can be presented as

$$\mathbf{X} = \sum_{i=1}^{m} h(\alpha - i)\mathbf{C}_i + G(\mathbf{S}_1, \alpha) = f_1(\mathbf{C}, \mathbf{S}_1, \{\alpha\}), \qquad \alpha = 1, \ldots, m,$$

$$\mathbf{Y} = \sum_{i=1}^{m} h(\beta - i)\mathbf{C}_i + G(\mathbf{S}_2, \beta) = f_2(\mathbf{C}, \mathbf{S}_2, \{\beta\}), \qquad \beta = 1, \ldots, m,$$

where $h(j - i)$ is the Heaviside function,

$$h(j - i) = \begin{cases} 1 & \text{if } j \geq i \\ 0 & \text{if } j < i, \end{cases}$$

and $\mathbf{C}$ denotes $(\mathbf{C}_1, \ldots, \mathbf{C}_m)$.

7. Comparison With Previous Solutions

The historically first attempt to systematically explain how the selectiveness of influence can coexist with the stochastic interdependence of the influenced variables was made by Townsend and his colleagues (Townsend, 1984; Townsend & Thomas, 1994). Townsend's solution, termed the *indirect nonselectivity*, consists in treating the relation (5) as indicating that the conditional distribution of $\mathbf{X}_i$, given

$$(\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_n) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n),$$

depends on Γ_i only, for any $i = 1, \ldots, n$ and any possible values $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$.

Somewhat generalized and renamed into the *conditional selectivity* (a term more convenient in the context of discussing different approaches to the selectiveness of influence), this property is completely characterized in Dzhafarov (1999). One of the simplest consequences of this characterization is that the conditional selectivity is not generally compatible even with the simple marginal selectivity, (3), in a contrast with the implication chain (10). Moreover, some of the constraints the conditional selectivity imposes on the marginal distributions are outright unacceptable for application purposes. Thus, in relation to section 2, the dependence of a bivariate normally distributed $(\mathbf{X}_1, \mathbf{X}_2)$ (with nonzero ρ_{12}) on two distinct factors γ_1, γ_2 is conditionally selective with respect to $(\{\gamma_1\}, \{\gamma_2\})$ only if, first,

$$\sigma_1^2 = \sigma_1^2(\gamma_1, \gamma_2), \sigma_2^2 = \sigma_2^2(\gamma_1, \gamma_2),$$

and, second, μ_1, μ_2 do not depend on either of the two factors (see Dzhafarov, 1999, 2001a for details). This dependence seems too odd to be interesting, in addition to lacking the marginal selectivity $\mathbf{X}_1 \leftarrow \{\gamma_1\}, \mathbf{X}_2 \leftarrow \{\gamma_2\}$. This example shows that conditional selectivity (indirect nonselectivity), although it may play a role in modeling information processing architectures, cannot be viewed as a viable definition of selectiveness under stochastic interdependence.

A different solution, called the *unconditional selectivity*, is proposed in Dzhafarov (2001a), based on Dzhafarov (1997) and Dzhafarov and Schweickert (1995). In this solution the relation (5) is treated as indicating that $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ can be presented in the form

$$\mathbf{X}_1 = f_1(\mathbf{R}, \Gamma_1), \quad \ldots, \quad \mathbf{X}_n = f_n(\mathbf{R}, \Gamma_n), \tag{11}$$

where $(f_1, \ldots, f_n)$ are some measurable (generally vectorial) functions and $\mathbf{R}$ is a random vector whose distribution does not depend on any of the factors from $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_n$. This vector is interpreted as an "internal source of randomness" for $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. Since it generally consists of several components, and each of the functions $f_1, \ldots, f_n$ may very well be independent of some of these components, (11) can be viewed as being less explicit than but equivalent to (6). Unfortunately, in order to derive necessary and sufficient conditions for the representation (11), Dzhafarov (2001a) imposes on the functions $(f_1, \ldots, f_n)$ and on the joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ constraints that appear to be both excessively strong and somewhat ad hoc. In addition, the theory only applies to factor sets $\Gamma_1, \ldots, \Gamma_n$ that are pairwise disjoint. The necessary and sufficient conditions themselves, given in Dzhafarov (2001a) in terms of the first-order conditional distributions of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, are rather nontransparent and do not seem to be easily applicable. In particular, they are not useful in dealing with the problem described in section 4.

By contrast, Definition 1 is predicated on no restrictions on the joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$, and it imposes no restrictions on the functions $f_1, \ldots, f_n$ (except for measurability) or sets of external factors $\Gamma_1, \ldots, \Gamma_n$. We will see also that this definition has a considerably greater working power.

8. Conditional Independence

The most basic and focal for the present development fact about $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ selectively influenced by $(\Gamma_1, \ldots, \Gamma_n)$ in the sense of Definition 1 is this: For any given (generally vectorial) value of the common source of randomness $\mathbf{C}$, the random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are *conditionally independent*, with their distributions depending on $\Gamma_1, \ldots, \Gamma_n$, respectively. Indeed,

$$\mathbf{X}_{1|\mathbf{C}=c} = f_1(c, \mathbf{S}_1, \Gamma_1), \quad \ldots, \quad \mathbf{X}_{n|\mathbf{C}=c} = f_n(c, \mathbf{S}_n, \Gamma_n), \tag{12}$$

with stochastically independent $\mathbf{S}_1, \ldots, \mathbf{S}_n$. Thus, in the example with factor analysis, (7), fixing a value of $\mathbf{C}$ makes $\mathbf{X}_{1|\mathbf{C}=c}, \ldots, \mathbf{X}_{n|\mathbf{C}=c}$ stochastically independent variables with means (but not variances) depending on this value,

$$\mathbf{X}_{i|\mathbf{C}=c} = a_i(c, \Gamma_i) + b_i(\Gamma_i)\mathbf{S}_i, \qquad i = 1, \ldots, n.$$

Conversely, the representation (6) holds if there is a random vector $\mathbf{C}$ whose distribution does not depend on $\Gamma_1 \cup \ldots \cup \Gamma_n$ and such that for any value c, $\mathbf{X}_{1|\mathbf{C}=c}, \ldots, \mathbf{X}_{n|\mathbf{C}=c}$ are stochastically independent vectors whose distributions depend on $\Gamma_1, \ldots, \Gamma_n$, respectively. To see why, recall that any p-component random vector $\mathbf{Z}$ ($p \geq 1$) can be presented as $\mathbf{Z} = \varphi(\mathbf{S})$, where $\mathbf{S}$ is a p-component vector of stochastically independent variables uniformly distributed between 0 and 1 (see Comment 1 in the Appendix). Putting

$$\mathbf{X}_{1|\mathbf{C}=c} = \varphi_{1,c,\Gamma_1}(\mathbf{S}_1), \quad \ldots, \quad \mathbf{X}_{n|\mathbf{C}=c} = \varphi_{n,c,\Gamma_n}(\mathbf{S}_n),$$

where φ_{i,c,Γ_i} is a function that may be different depending on both c and Γ_i, whereas $\mathbf{S}_i$ is one and the same for all c and Γ_i ($i = 1, \ldots, n$), we can rename $\varphi_{i,c,\Gamma_i}(\mathbf{S}_i)$ into $f_i(c, \mathbf{S}_i, \Gamma_i)$ and obtain thence the representation (6).

This simple argument is summarized in the following formal statement.

Proposition 1. Given $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and $(\Gamma_1, \ldots, \Gamma_n)$ satisfying

$$\mathbf{X}_1 \hookleftarrow \Gamma_1, \quad \ldots, \quad \mathbf{X}_n \hookleftarrow \Gamma_n,$$

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \hookleftarrow \Gamma_1 \cup \ldots \cup \Gamma_n,$$

the selective influence relation

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \looparrowleft (\Gamma_1, \ldots, \Gamma_n)$$

holds if and only if one can find a random vector $\mathbf{C}$ (whose distribution does not depend on $\Gamma_1 \cup \ldots \cup \Gamma_n$) such that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are conditionally independent given any value of $\mathbf{C}$, with their conditional distributions depending on $\Gamma_1, \ldots, \Gamma_n$, respectively.

This proposition is central for the present development, and it can be viewed as an alternative (and equivalent) definition of selective influence.

Conditional independence is, of course, one of the most basic concepts in constructing psychometric and statistical models, being routinely used in item-response theories, analysis of distribution mixtures, and other constructs involving latent variables (see, e.g., Lazarsfeld, 1965; Lindsay, 1995). Suppes and Zanotti (1981) associate this concept with a general philosophical principle, called by them the *common cause criterion*:

> The primary criterion of adequacy of probabilistic causal analysis is that the causal variable should render the simultaneous phenomenological data conditionally independent. (p. 191)

My preference is to distinguish, at least epistemologically, between "external (observable) causes" of variation and "internal (unobservable) causes" of (necessarily random) variation, and further distinguishing "common internal causes" from "specific" ones. What makes the present use of conditional independence different from the previous ones is that the (hypothetical or observable) "effects" $\mathbf{X}_1, \ldots, \mathbf{X}_n$, when rendered conditionally independent by finding their "common internal causes" $\mathbf{C}$, are *selectively* linked to their respective "external causes" $\Gamma_1, \ldots, \Gamma_n$.

To explicate the precise meaning of the conditional independence in Proposition 1, let A_i be a Lebesgue measurable set within the domain of $\mathbf{X}_i$ ($i = 1, \ldots, n$), and let ω denote the probability measure associated with the conditioning vector $\mathbf{C}$. Then Proposition 1 says that the meaning of (5) is this: For any $A_1, \ldots, A_n$,

$$\Pr[\mathbf{X}_1 \in A_1, \ldots, \mathbf{X}_n \in A_n; \Gamma] = \int_{\mathrm{Dom}(\mathbf{C})} \left[\prod_{i=1}^{n} \Pr(\mathbf{X}_i \in A_i | \mathbf{C} = c; \Gamma_i) \right] d\omega(c), \tag{13}$$

where $\Gamma = \Gamma_1 \cup \ldots \cup \Gamma_n$ and the integration (in the Lebesgue sense) is over the domain of $\mathbf{C}$. The nestedness property (8) can be derived from this representation directly, without referring to (6). Thus, choosing $A_i = \mathrm{Dom}(\mathbf{X}_i)$ for all $i > k$, and observing that

$$\Pr(\mathbf{X}_i \in \mathrm{Dom}(\mathbf{X}_i) | \mathbf{C} = c; \Gamma_i) = 1 \quad \text{a.s. } [\omega]$$

(a.s. $[\omega]$ standing for "almost surely with respect to measure ω"), one obtains

$$\Pr[\mathbf{X}_1 \in A_1, \ldots, \mathbf{X}_k \in A_k; \Gamma_1 \cup \ldots \cup \Gamma_k] = \int_{\mathrm{Dom}(\mathbf{C})} \left[\prod_{i=1}^{k} \Pr(\mathbf{X}_i \in A_i | \mathbf{C} = c; \Gamma_i) \right] d\omega(c),$$

which, by (13), is equivalent to

$$(\mathbf{X}_1, \ldots, \mathbf{X}_k) \looparrowleft (\Gamma_1, \ldots, \Gamma_k).$$

The conformity with the third line of (8) follows equally easily, on using basic measure-theoretic constructs (see Comment 2 in the Appendix).

If the conditional distributions of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ (given a value of $\mathbf{C}$) have conventional probability density (or probability mass) functions $\psi_{1|\mathbf{C}}, \ldots, \psi_{n|\mathbf{C}}$, then (13) can also be written as

$$\psi_{1\ldots n}(x_1, \ldots, x_n; \Gamma) = \int_{\mathrm{Dom}(\mathbf{C})} \left[\prod_{i=1}^{n} \psi_{i|\mathbf{C}}(x_i | c; \Gamma_i) \right] d\omega(c), \tag{14}$$

where $\psi_{1\ldots n}$ is the joint probability density (mass) function for $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$. Obviously, if the conditioning vector $\mathbf{C}$ has a density (mass) function $\psi(c)$, then $d\omega(c)$ in this equation can be replaced with $\psi(c)dc$.

9. Applications

For stochastically independent $\mathbf{X}_1, \ldots, \mathbf{X}_n$, (13) and (14) acquire the forms

$$\Pr[\mathbf{X}_1 \in A_1, \ldots, \mathbf{X}_k \in A_k; \Gamma_1 \cup \ldots \cup \Gamma_k] = \prod_{i=1}^{k} \Pr(\mathbf{X}_i \in A_i; \Gamma_i) \tag{15}$$

and

$$\psi_{1\ldots n}(x_1, \ldots, x_n; \Gamma) = \prod_{i=1}^{n} \psi_i(x_i; \Gamma_i). \tag{16}$$

These equations have precisely the same structure as the integrands in (13) and (14) taken at a given $\mathbf{C} = c$, and it is to this fact that the new definition of selective influence owes its working power. One can reasonably expect that any property derived from (15) and (16) for stochastically independent $\mathbf{X}_1, \ldots, \mathbf{X}_n$ will also be shared by the integrands in (13) and (14), for any given $\mathbf{C} = c$. Integration over the domain of $\mathbf{C}$ will often preserve this property of the integrand, yielding thereby an easy and natural generalization from stochastically independent $\mathbf{X}_1, \ldots, \mathbf{X}_n$ to stochastically interdependent $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with the same pattern of selective influence. Even when this is not the case, the multiplicative decomposability of the integrands in (13) and (14) greatly facilitates theoretical analysis of the joint distributions.

To illustrate, consider a problem taken from Schweickert, Giorgini, and Dzhafarov (2000). This problem falls within the scope of the motivating example given in section 3. Let there be two binary (0-1) factors γ_1, γ_2 known to influence a response time $\mathbf{T}$. Denote its distribution function by $G(t; \gamma_1, \gamma_2)$. Let $\mathbf{T}$ be modeled as

$$\mathbf{T} = H(\mathbf{T}_1, \ldots, \mathbf{T}_n; \gamma_1, \gamma_2),$$

where $\mathbf{T}_1, \ldots, \mathbf{T}_n$ ($n \geq 2$) are durations of hypothetical processes constituting a "*parallel-serial network with AND gates*". This means that H can be written entirely in terms of operations $\ldots + \ldots$ and $\max\{\ldots, \ldots\}$, with each argument $\mathbf{T}_1, \ldots, \mathbf{T}_n$ entering in the expression once and only once. (If max in this definition is replaced with min, the parallel-serial network is said to have OR gates.) Different levels of γ_1, γ_2 do not change the function H (i.e., the processing architecture remains fixed), but they are assumed to *selectively influence* two of the durations $\mathbf{T}_1, \ldots, \mathbf{T}_n$ (say, $\mathbf{T}_1, \mathbf{T}_2$, respectively). Suppose $\mathbf{T}_1$ and $\mathbf{T}_2$ are in parallel (this means that in the expression for H no term containing $\mathbf{T}_1$ is connected by $+$ with a term containing $\mathbf{T}_2$). We have

$$(\mathbf{T}_1, \mathbf{T}_2, \overbrace{\mathbf{T}_3, \ldots, \mathbf{T}_n}) \hookleftarrow (\{\gamma_1\}, \{\gamma_2\}, \overbrace{\emptyset, \ldots, \emptyset}). \tag{17}$$

Denoting the marginal distributions of $\mathbf{T}_1, \mathbf{T}_2$ by $F_1(t; \gamma_1)$ and $F_2(t; \gamma_2)$, respectively, assume also that $F_1(t; 0) \geq F_1(t; 1)$ and $F_2(t; 0) \geq F_2(t; 1)$, for all $t \geq 0$ (the *stochastic dominance* assumption, meaning that switching from level 0 to level 1 results in a prolongation of the processes). Under these conditions, as proved in Schweickert, Giorgini, Dzhafarov, (2000), if $\mathbf{T}_1, \ldots, \mathbf{T}_n$ are mutually stochastically independent, then, for all $t \geq 0$,

$$G(t; 1, 1) - G(t; 0, 1) - G(t; 1, 0) + G(t; 0, 0) \geq 0.$$

(The inequality reverses in the case of the "parallel-serial network with OR gates".)

The question is whether this result also holds for stochastically interdependent $(\mathbf{T}_1, \ldots, \mathbf{T}_n)$. Due to (17) one can invoke Proposition 1: There is a random vector $\mathbf{C}$ such that given any its value, $\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}_3, \ldots, \mathbf{T}_n$ are conditionally independent, with their conditional distributions depending on, respectively, $\{\gamma_1\}, \{\gamma_2\}, \emptyset, \ldots, \emptyset$. Assume that the stochastic dominance assumption holds now for every value of $\mathbf{C}$, that is, $F_1(t|\mathbf{C} = c; 0) \geq F_1(t|\mathbf{C} = c; 1)$ and $F_2(t|\mathbf{C} = c; 0) \geq F_2(t|\mathbf{C} = c; 1)$, for all $t \geq 0$ and all c. Then, by the cited result from Schweickert et al. (2000),

$$G(t|\mathbf{C} = c; 1, 1) - G(t|\mathbf{C} = c; 0, 1) - G(t|\mathbf{C} = c; 1, 0) + G(t|\mathbf{C} = c; 0, 0) \geq 0$$

for all $t \geq 0$ and all c. Denoting, as in 13, the distribution function for $\mathbf{C}$ by ω, it is easy to see that

$$\begin{aligned} &G(t; 1, 1) - G(t; 0, 1) - G(t; 1, 0) + G(t; 0, 0) \\ &= \int_{\mathrm{Dom}(\mathbf{C})} [G(t|\mathbf{C} = c; 1, 1) - G(t|\mathbf{C} = c; 0, 1) - G(t|\mathbf{C} = c; 1, 0) + G(t|\mathbf{C} = c; 0, 0)]\, d\omega(c), \end{aligned}$$

and the integral is nonnegative because so is the integrand. The generalizability of the inequality for OR gates is proved analogously. A comprehensive analysis of a broad class of "directed acyclic networks" with stochastically interdependent but selectively influenced constituent durations is given elsewhere (Dzhafarov, Schweickert, Sung, submitted).

To illustrate other applications of Definition 1 and Proposition 1, consider the problem outlined in section 4, in relation to Thurstone's (1927a, 1927b) and Luce and Galanter's (1963) models. Assuming $(\mathbf{X}, \mathbf{Y})$ is bivariate normally distributed and the marginal selectivity with respect to (α, β) holds, what dependence of $\rho_{\mathbf{XY}}$ on α and β would allow one to speak of

$$(\mathbf{X}, \mathbf{Y}) \looparrowleft (\{\alpha\}, \{\beta\}),$$

that is, to consider $\mathbf{X}$ as selectively corresponding to α and $\mathbf{Y}$ as selectively corresponding to β? In accordance with Proposition 1, the question is whether one can find a random vector $\mathbf{C}$ (independent of α, β) such that given $\mathbf{C} = c$ the $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent, with their conditional distributions depending on α and β, respectively.

Let us consider this problem under the additional constraint that the joint distribution of $(\mathbf{X}, \mathbf{Y}, \mathbf{C})$ be multivariate normal. If $\mathbf{C}$ is a vector of random variables $(\mathbf{C}_1, \ldots, \mathbf{C}_p)$, then $(\mathbf{X}, \mathbf{Y}, \mathbf{C}_1, \ldots, \mathbf{C}_p)$ has $p+2$ real-valued components. It is easy to see that in the absence of any restrictions imposed on $(f_1, \ldots, f_n)$ in (6), $\mathbf{C}_1, \ldots, \mathbf{C}_p$ can be taken to be independent standard normally distributed variables. In view of the observation made in the last paragraph of section 5,

$$(\mathbf{X}, \mathbf{Y}, \mathbf{C}_1, \ldots, \mathbf{C}_p) \looparrowleft (\{\alpha\}, \{\beta\}, \emptyset, \ldots, \emptyset),$$

whence the correlation coefficients $\rho_{\mathbf{X}i}$ and $\rho_{\mathbf{Y}i}$ between, respectively, $\mathbf{X}$ and $\mathbf{C}_i$ and $\mathbf{Y}$ and $\mathbf{C}_i$ $(i = 1, \ldots, p)$ may only depend on, respectively, α and β. It follows now by straightforward algebra (see, e.g., Theorems 3.3.4 and 3.4.3 in Tong, 1990) that $\mathbf{X}, \mathbf{Y}$ are conditionally independent given $\mathbf{C}$ if and only if

$$\rho_{\mathbf{XY}}(\alpha, \beta) = \sum_{i=1}^{p} \rho_{\mathbf{X}i}(\alpha)\rho_{\mathbf{Y}i}(\beta). \tag{18}$$

Put differently, a bivariate normally distributed $(\mathbf{X}, \mathbf{Y})$ with the marginal selectivity with respect to (α, β) is selectively influenced by (α, β) if, for some $p = 1, 2, \ldots,$ one can find functions

$$-1 \leq \rho_{\mathbf{X}i}(\alpha), \rho_{\mathbf{Y}i}(\beta) \leq 1, \quad i = 1, \ldots, p,$$

such that $\rho_{\mathbf{XY}}(\alpha, \beta)$ satisfies (18). Note that "if" in this statement cannot be complemented by "only if" because it is not generally necessary that $(\mathbf{X}, \mathbf{Y}, \mathbf{C})$ be multivariate normally distributed.

Not surprisingly, there are functions $\rho_{\mathbf{XY}}(\alpha, \beta)$ that cannot be decomposed in accordance with (18), for any p. An example of one such function is $\rho_{\mathbf{XY}}(\alpha, \beta) = |\alpha - \beta|$, where we assume that α and β continuously vary within [0, 1] (see Comment 3 in the Appendix). By choosing p sufficiently large, however, we know from the classical Weierstrass theorem and related results that any well-behaved (e.g., continuous, or absolutely integrable) $\rho_{\mathbf{XY}}(\alpha, \beta)$ taken on any compact region of (α, β) (generally a subset of $\mathrm{Re}^k \times \mathrm{Re}^k$, $k \geq 1$) can be approximated by (18) to any degree of precision. One consequence of this fact is that if p is not restricted, the hypothesis of selective influence $(\mathbf{X}, \mathbf{Y}) \looparrowleft (\alpha, \beta)$ cannot be rejected on a sample level, with only estimated $\rho_{\mathbf{XY}}$ available. This issue is discussed in the concluding section of this paper.

The analysis just presented can be trivially generalized to incorporate the problem posed in section 2. An n-variate normally distributed $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ possessing the simple marginal selectivity with respect to $(\Gamma_1, \ldots, \Gamma_n)$ is selectively influenced by $(\Gamma_1, \ldots, \Gamma_n)$ if, for some $p = 1, 2, \ldots$, one can find functions

$$-1 \leq \rho_{i\lambda}(\Gamma_i) \leq 1, \quad i = 1, \ldots, n, \lambda = 1, \ldots, p$$

such that

$$\rho_{ij}(\Gamma_i, \Gamma_j) = \sum_{\lambda=1}^{p} \rho_{i\lambda}(\Gamma_i)\rho_{j\lambda}(\Gamma_j), \quad i, j = 1, \ldots, n, i \neq j. \tag{19}$$

If these equalities hold, $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ can be embedded into a $(p+n)$-variate normally distributed $(\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{C}_1, \ldots, \mathbf{C}_p)$, with the common sources of randomness $\mathbf{C}_1, \ldots, \mathbf{C}_p$ being mutually independent standard normally distributed.

10. Selective Influence and Complete Marginal Selectivity

Refer to the chain of implications (10). As explained in the paragraph following it, simple marginal selectivity, (3), does not imply selective influence, (5), because it does not imply complete marginal selectivity, (9). This leaves open the possibility, however, that selective influence could be implied by complete marginal selectivity. If this were the case, then, as a special case, the selective influence relation

$$(\mathbf{X}_1, \mathbf{X}_2) \looparrowleft (\Gamma_1, \Gamma_2), \tag{20}$$

understood in the sense of Definition 1-Proposition 1, would have to follow from the marginal selectivity, in this case both simple and complete,

$$\mathbf{X}_1 \leftarrow \Gamma_1, \mathbf{X}_2 \leftarrow \Gamma_2. \tag{21}$$

It can be proved by counter-example that this is not the case.

Proposition 2. The marginal selectivity (21) may hold while the selective influence (20) does not. Hence selective influence, (5), is not implied by complete marginal selectivity, (9).

Proof. Consider binary (zero-one) random variables $\mathbf{X}_1, \mathbf{X}_2$ whose joint distribution depends on two continuous factors, $\gamma_1 \in [0, 1]$ and $\gamma_2 \in [0, 1]$. Let

$$P(\gamma_1, \gamma_2) = \Pr(\mathbf{X}_1 = 0, \mathbf{X}_2 = 0; \gamma_1, \gamma_2) = \frac{\min(\gamma_1, \gamma_2)}{1 + |\gamma_1 - \gamma_2|},$$
$$P_1(\gamma_1) = \Pr(\mathbf{X}_1 = 0; \gamma_1) = \gamma_1, \quad P_2(\gamma_2) = \Pr(\mathbf{X}_2 = 0; \gamma_2) = \gamma_2. \tag{22}$$

The marginal selectivity $\mathbf{X}_1 \leftarrow \{\gamma_1\}, \mathbf{X}_2 \leftarrow \{\gamma_2\}$ clearly holds. The selective influence relation $(\mathbf{X}_1, \mathbf{X}_2) \hookleftarrow (\{\gamma_1\}, \{\gamma_2\})$ would mean here that for some random vector $\mathbf{C}$ associated with a probability measure ω,

$$\mathrm{P}(\gamma_1, \gamma_2) = \int_{c \in \mathrm{Dom}(\mathbf{C})} Q_1(c, \gamma_1) Q_2(c, \gamma_2)\, d\omega(c),$$

$$P_1(\gamma_1) = \int_{c \in \mathrm{Dom}(\mathbf{C})} Q_1(c, \gamma_1)\, d\omega(c),$$

$$P_2(\gamma_2) = \int_{c \in \mathrm{Dom}(\mathbf{C})} Q_2(c, \gamma_2)\, d\omega(c),$$

where

$$Q_1(c, \gamma_1) = \Pr(\mathbf{X}_1 = 0 | \mathbf{C} = c; \gamma_1), \quad Q_2(c, \gamma_2) = \Pr(\mathbf{X}_2 = 0 | \mathbf{C} = c; \gamma_2).$$

Let such a representation for P, P_1, P_2 exist. For every $\gamma \in [0, 1]$ let $\mathrm{Dom}(\mathbf{C})$ be partitioned into

$$C_{00}(\gamma) = \{c : Q_1(c, \gamma) = 0, Q_2(c, \gamma) = 0\}, C_{++}(\gamma) = \{c : Q_1(c, \gamma) > 0, Q_2(c, \gamma) > 0\},$$
$$C_{0+}(\gamma) = \{c : Q_1(c, \gamma) = 0, Q_2(c, \gamma) > 0\}, C_{+0}(\gamma) = \{c : Q_1(c, \gamma) > 0, Q_2(c, \gamma) = 0\}.$$

From (22),

$$\mathrm{P}(\gamma, \gamma) = \int_{c \in C_{++}(\gamma)} Q_1(c, \gamma) Q_2(c, \gamma)\, d\omega(c) = \gamma. \tag{23}$$

Comparing this with

$$P_1(\gamma) = \int_{c \in C_{+0}(\gamma)} Q_1(c, \gamma)\, d\omega(c) + \int_{c \in C_{++}(\gamma)} Q_1(c, \gamma)\, d\omega(c) = \gamma,$$

$$P_2(\gamma) = \int_{c \in C_{0+}(\gamma)} Q_2(c, \gamma)\, d\omega(c) + \int_{c \in C_{++}(\gamma)} Q_2(c, \gamma)\, d\omega(c) = \gamma,$$

and noting that $Q_i(c, \gamma) \geq Q_1(c, \gamma) Q_2(c, \gamma)$, $i = 1, 2$, we conclude that

$$\omega[C_{+0}(\gamma)] = 0, \omega[C_{0+}(\gamma)] = 0,$$
$$c \in C_{++}(\gamma) \Rightarrow Q_1(c, \gamma) = Q_2(c, \gamma) = 1 \text{ a.s. } [\omega].$$

Then, from (23), $\omega[C_{++}(\gamma)] = \gamma$. In particular, $\omega[C_{++}(1)] = 1$, whence

$$Q_1(c, 1) = Q_2(c, 1) = 1 \text{ a.s. } [\omega].$$

But then, for any $\gamma \in (0, 1)$,

$$\mathrm{P}(\gamma, 1) = \int_{c \in C_{++}(\gamma)} Q_1(c, \gamma) Q_2(c, 1)\, d\omega(c) = \gamma,$$

which contradicts the definition of P in (22). □

There is, however, a degenerate case when even simple marginal selectivity implies selective influence. This is the case $\Gamma_1 = \ldots = \Gamma_n = \Gamma$. Formally, the statements

$$\mathbf{X}_1 \leftarrow \Gamma, \quad \ldots, \quad \mathbf{X}_n \leftarrow \Gamma \tag{24}$$

and

$$(\mathbf{X}_1, \ldots, \mathbf{X}_n) \looparrowleft (\Gamma, \ldots, \Gamma) \tag{25}$$

fall within the scope of Definition 1-Proposition 1, even though one would not normally associate any sense of "selectiveness" with this case.

Proposition 3. The (degenerate) simple marginal selectivity (24) implies the (degenerate) selective influence (25).

Proof. For any fixed Γ, applying Comment 1 from the Appendix to $\mathbf{Z} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, we have

$$\mathbf{X}_1 = \varphi_{1,\Gamma}(\mathbf{V}_1), \mathbf{X}_2 = \varphi_{2,\Gamma}(\mathbf{V}_1, \mathbf{V}_2), \quad \ldots, \quad \mathbf{X}_n = \varphi_{n,\Gamma}(\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_n),$$

where $\mathbf{V}_i$ $(i = 1, \ldots, n)$ is a vector of stochastically independent variables uniformly distributed between 0 and 1, and $\mathbf{V}_1, \ldots, \mathbf{V}_n$ are mutually stochastically independent. Putting $\mathbf{C} = (\mathbf{V}_1, \ldots, \mathbf{V}_n)$, the functions $\varphi_{i,\Gamma}(\mathbf{V}_1, \ldots, \mathbf{V}_i)$ can be written as $f_i(\mathbf{C}, \Gamma)$, whence

$$\mathbf{X}_1 = f_1(\mathbf{C}, \Gamma), \quad \ldots, \quad \mathbf{X}_n = f_n(\mathbf{C}, \Gamma), \tag{26}$$

a special case of (6). □

Note that fixing $\mathbf{C} = c$ in (26) makes $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ a vector of deterministic functions of Γ (see, however, Comment 4 in the Appendix). Formally, deterministic quantities are stochastically independent, in agreement with Proposition 1. With this in mind, for $\Gamma_1 = \ldots = \Gamma_n = \emptyset$ Proposition 3 implies that any set of random vectors (with a fixed joint distribution) can be made conditionally independent. This generalizes the sufficiency part of the *Theorem on Common Causes* by Suppes and Zanotti (1981), proved there by a very different argument for the case when $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are binary variables (but correctly claimed to hold generally).

11. Concluding Remarks

To summarize, if $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ is selectively influenced by $(\Gamma_1, \ldots, \Gamma_n)$ (where the factor sets $\Gamma_1, \ldots, \Gamma_n$ may be overlapping or empty), then $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ exhibits the complete marginal selectivity with respect to $(\Gamma_1, \ldots, \Gamma_n)$: Any subvector of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ depends only on the union of the corresponding factor sets. The complete marginal selectivity, however, is not equivalent to the selective influence relation. The latter holds if and only if one can embed $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ in a vector $(\mathbf{X}_1, \ldots, \mathbf{X}_n, \mathbf{C})$ with the marginal distribution of $\mathbf{C}$ being independent of $\Gamma_1 \cup \ldots \cup \Gamma_n$, such that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are conditionally independent given any value $\mathbf{C} = c$, with the conditional distributions of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ depending on $\Gamma_1, \ldots, \Gamma_n$, respectively. Equivalently, one can find mutually independent $\mathbf{C}, \mathbf{S}_1, \ldots, \mathbf{S}_n$ such that $\mathbf{X}_i = f_i(\mathbf{C}, \mathbf{S}_i, \Gamma_i)$ for $i = 1, \ldots, n$, with $f_1, \ldots, f_n$ being arbitrary measurable functions. This approach to the notion of selective influence generalizes the one proposed in Dzhafarov (2001a), and it is better suited for applied purposes, such as described in sections 2, 3, and 4. It includes as a special, or degenerate case most of the "theory of common causes" proposed by Suppes and Zanotti (1981).

What can be considered the main weakness of the theory presented in this paper is that given the joint distribution of $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ as a function of $\Gamma_1 \cup \ldots \cup \Gamma_n$, and assuming that the complete marginal selectivity is satisfied, neither Definition 1 nor Proposition 1 provide an algorithm for finding the conditioning random vector $\mathbf{C}$ or determining that no such $\mathbf{C}$ can be found. Proposition 2 tells us that situations in which $\mathbf{C}$ cannot be found do exist, but we do not have a characterization (necessary and sufficient conditions) for these situations. Thus, for n-variate normally distributed $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ with

$$\mu_i = \mu_i(\Gamma_i), \sigma_i^2 = \sigma_i^2(\Gamma_i), \rho_{ij} = \rho_{ij}(\Gamma_i, \Gamma_j), \qquad i, j = 1, \dots, n, i \neq j,$$

the complete marginal selectivity with respect to $(\Gamma_1, \dots, \Gamma_n)$ is satisfied, and we know that if all $\rho_{ij}(\Gamma_i, \Gamma_j)$ can be presented in accordance with (19), then one can embed $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ into an $(n+p)$-variate normally distributed $(\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{C})$, with $\mathbf{C}$ playing the role of the conditioning vector of Proposition 1. We do not know, however, whether or how a conditioning vector $\mathbf{C}$ can be found if (19) is not satisfied for some or all of the correlations $\rho_{ij}(\Gamma_i, \Gamma_j)$.

It is worth emphasizing that the distributions of $\mathbf{C}, \mathbf{S}_1, \dots, \mathbf{S}_n$ can be chosen essentially arbitrarily. As follows from the argument presented in Comment 1 (Appendix), they all can be rendered uniformly distributed in standard unit (hyper)cubes. As a result, the problem of finding a conditioning vector $\mathbf{C}$ is never the problem of finding its separate distribution, but rather that of finding the joint distribution of $(\mathbf{X}_1, \dots, \mathbf{X}_n, \mathbf{C})$.

Another notable aspect of the theory is that the hypothesis $(\mathbf{X}_1, \dots, \mathbf{X}_n) \hookleftarrow (\Gamma_1, \dots, \Gamma_n)$ may be difficult if not impossible to reject when $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ is known on a sample level only. Depending on one's purposes, one might consider this a weakness or an advantage. It is mentioned in section 9, for example, that by choosing p sufficiently large, (19) can approximate any well-behaved $\rho_{ij}(\Gamma_i, \Gamma_j)$ on any compact area of the factor values to any degree of precision. As a result, any set of data that can be modeled by n-variate normally distributed $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ possessing the complete marginal selectivity with respect to $(\Gamma_1, \dots, \Gamma_n)$, can also be modelled by n-variate normally distributed $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ selectively influenced by $(\Gamma_1, \dots, \Gamma_n)$. One may ask oneself, however, whether it might be reasonable, in this context as well as generally, in Definition 1, to require that the dimensionality of $\mathbf{C}$ be smaller than the combined dimensionality of $(\mathbf{X}_1, \dots, \mathbf{X}_n)$. This constraint is natural in the context of nonlinear factor analysis and other latent variable constructs, but whether it can be justified in the general definition of selective influence remains unclear.

All random entities throughout this paper have been treated as random vectors with real-valued components. This is sufficient for most but not all applied purposes. Thus, Regenwetter and Marley (2001) consider "random relations" and "random functions" as conceptual alternatives to real-valued "random utilities". In the context of Thurstonian-type modeling described in section 4, it is natural to consider the possibility that the random images into which stimuli map may be more complex than being representable by vectors with real-valued components. One might think of perceptual images as pictorial templates or processes developing in time, in which case they should be represented by functions or relations rather than real-component vectors. In view of these theoretical possibilities, it seems useful to observe that the approach to the notion of selective influence presented in this paper can be easily generalized to random entities taking on their values in arbitrary measure spaces.

It is more convenient to construct this generalization for Proposition 1 rather than for Definition 1. (To follow the remainder the reader should be familiar with basic notions of abstract measure theory.) Let $R_1, \dots, R_n$ be arbitrary sets with sigma-algebras $\Sigma_1, \dots, \Sigma_n$, respectively. Let Σ be the minimal sigma-algebra on $R_1 \times \dots \times R_n$ that contains the product $\Sigma_1 \times \dots \times \Sigma_n$. Given sets of external factors $\Gamma_1, \dots, \Gamma_n$, let $\pi(A; \Gamma_1 \cup \dots \cup \Gamma_n)$ be a sigma-finite measure defined for all sets (events)$A \in \Sigma$.

Consider now a set C with a sigma-algebra Σ_C and a measure ω defined on Σ_C. Suppose that for $[\omega]$-almost all $c \in C$ there exist probability measures $\pi_1(A_1|c; \Gamma_1), \dots, \pi_n(A_n|c; \Gamma_n)$ defined for all $A_1 \in \Sigma_1, \dots, A_n \in \Sigma_n$, respectively. These measures (that depend on c and on the factors) are interpreted as conditional probability measures on these sigma-algebras given c. Suppose that for all $A_1 \in \Sigma_1, \dots, A_n \in \Sigma_n$,

$$\pi(A_1 \times \dots \times A_n; \Gamma_1 \cup \dots \cup \Gamma_n) = \int_{c \in C} [\pi_1(A_1|c; \Gamma_1) \dots \pi_1(A_n|c; \Gamma_n)]\, d\omega(c)$$

(where the integrand may be undefined on a factor-independent set of $[\omega]$-measure zero). By a standard measure-theoretic construction, this implies the more general equality

$$\pi(A; \Gamma_1 \cup \ldots \cup \Gamma_n) = \int_{c \in C} \int_{(r_1,\ldots,r_n)\in A} [d\pi_1(r_1|c; \Gamma_1) \ldots d\pi_1(r_n|c; \Gamma_n)]\, d\omega(c),$$

for any $A \in \Sigma$.

If such a measure space (C, Σ_C, ω) exists, we say that the random entities $(\mathbf{R}_1, \ldots, \mathbf{R}_n)$ are selectively influenced by $(\Gamma_1, \ldots, \Gamma_n)$, respectively, and write

$$(\mathbf{R}_1, \ldots, \mathbf{R}_n) \hookleftarrow (\Gamma_1, \ldots, \Gamma_n).$$

A "random entity" $\mathbf{R}_i$ here is simply a variable taking on its values on the set R_i $(i = 1, \ldots, n)$. This completes the generalization.

Appendix: Technical Comments

1. The explanation is given in Dzhafarov (1999, p. 125) and is recounted here for reader's convenience. If $\mathbf{Z}$ is a random variable with a distribution function $F(z)$, its quantile function $Q(u)$, $0 < u < 1$ is defined as $\min\{z : F(z) > u\}$. Plainly, $\mathbf{Z} = Q(\mathbf{U})$, where $\mathbf{U}$ is standard uniformly distributed. If $\mathbf{Z}$ is a vector with components $(\mathbf{Z}_1, \ldots, \mathbf{Z}_p)$, $p > 1$, let $Q_i(u; z_1, \ldots, z_{i-1})$ be the quantile function corresponding to the conditional distribution

$$F_i(z; z_1, \ldots, z_{i-1}) = \Pr[\mathbf{Z}_i \le z|(\mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1}) = (z_1, \ldots, z_{i-1})].$$

(If this conditional distribution is not defined on a zero-measure set of $(z_1, \ldots, z_{i-1})$, it can be additionally defined on this set arbitrarily.) Then $\mathbf{Z}_i = Q_i(\mathbf{U}; \mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1})$, where $\mathbf{U}$ is standard uniformly distributed and independent of $\mathbf{Z}_1, \ldots, \mathbf{Z}_{i-1}$. Choosing now $\mathbf{U}_1, \ldots, \mathbf{U}_p$ to be independent standard uniformly distributed variables,

$$\mathbf{Z}_1 = Q_1(\mathbf{U}_1), \mathbf{Z}_2 = Q_2(\mathbf{U}_2; \mathbf{Z}_1) = Q_2(\mathbf{U}_2; Q_1(\mathbf{U}_1)), \text{etc.},$$

so $\mathbf{Z}$ is a vectorial function φ of $\mathbf{S} = (\mathbf{U}_1, \ldots, \mathbf{U}_p)$.

2. Denoting $\mathbf{Y}_1 = (\mathbf{X}_1, \ldots, \mathbf{X}_k)$, $\mathbf{Y}_2 = (\mathbf{X}_{k+1}, \ldots, \mathbf{X}_n)$, $\Lambda_1 = \Gamma_1 \cup \ldots \cup \Gamma_k$, $\Lambda_2 = \Gamma_{k+1} \cup \ldots \cup \Gamma_n$, the equation

$$\Pr[\mathbf{Y}_1 \in B_1, \mathbf{Y}_2 \in B_2; \Gamma] = \int_{\mathrm{Dom}(\mathbf{C})} \Pr(\mathbf{Y}_1 \in B_1|\mathbf{C} = c; \Lambda_1) \Pr(\mathbf{Y}_2 \in B_2|\mathbf{C} = c; \Lambda_2)\, d\omega(c)$$

follows from (13) trivially if B_1, B_2 are (hyper)rectangular in shape,

$$B_1 = A_1 \times \ldots \times A_k, B_2 = A_{k+1} \times \ldots \times A_n.$$

Since measures for arbitrary Borel sets are constructed by approximating them by unions of (hyper)rectangular sets, the equation above also holds for arbitrary Lebesque measurable B_1, B_2.

3. The logic of this proof, in a different context, is suggested by A. Eremenko (personal communication, 2000). Let $p \ge 1$ be the smallest number for which one can find some functions $\{a_i(\alpha)\}_{i=1}^{i=p}$, $\{b_i(\beta)\}_{i=1}^{i=p}$ (with no restrictions on the range) such that

$$|\alpha - \beta| = \sum_{i=1}^{p} a_i(\alpha) b_i(\beta).$$

Then the matrix $[b_i(\beta_j)]$ $(i, j = 1, \ldots, p)$ must be nonsingular for some $(\beta_1, \ldots, \beta_p)$ (see Lemma 1.1 in Gauchman & Rubel, 1989). As a result, $\{a_i(\alpha)\}_{i=1}^{i=p}$ can be presented as linear combinations of $\{|\alpha - \beta_i|\}_{i=1}^{i=p}$, whence we conclude that $\{a_i(\alpha)\}_{i=1}^{i=p}$ are differentiable at any $\alpha \notin \{\beta_1, \ldots, \beta_p\}$. But then $|\alpha - \beta|$ must be differentiable at any such α irrespective of β, which is wrong for $\beta = \alpha$. This argument applies to all functions of (α, β) that possess some regularity feature in α (continuity, differentiability, etc.) at all (α, β) except when $\alpha = h(\beta)$, for some function h.

4. We could also define $\mathbf{C} = (\mathbf{V}_1, \ldots, \mathbf{V}_{n-1})$, and rewrite $\varphi_{i,\Gamma}(\mathbf{V}_1, \ldots, \mathbf{V}_i)$ as $f_i(\mathbf{C}; \Gamma)$ when $i < n$ and as $f_n(\mathbf{C}, \mathbf{V}_n; \Gamma)$ for $i = n$. Then fixing $\mathbf{C} = c$ would render all *but one* of $(\mathbf{X}, \ldots, \mathbf{X}_n)$ deterministic quantities.

References

Ashby, F.G., & Perrin, N.A. (1988). Towards a unified theory of similarity and recognition. *Psychological Review*, *95*, 124–130.

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

Cortese, J.M., & Dzhafarov, E.N. (1996). Empirical recovery of response time decomposition rules II: Discriminability of serial and parallel architectures. *Journal of Mathematical Psychology*, *40*, 203–218.

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. (1999). Conditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, *43*, 123–157.

Dzhafarov, E.N. (2001a). Unconditionally selective dependence of random variables on external factors. *Journal of Mathematical Psychology*, *45*, 421–451.

Dzhafarov, E.N. (2001b). Fechnerian scaling and Thurstonian modeling. In E. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001* (pp. 42–47). Lengerich: Pabst Science Publishers.

Dzhafarov, E.N., & Cortese, J.M. (1996). Empirical recovery of response time decomposition rules I: Sample-level Decomposition tests. *Journal of Mathematical Psychology*, *40*, 185–202.

Dzhafarov, E.N., & Rouder, J.N. (1996). Empirical discriminability of two models for stochastic relationship between additive components of response time. *Journal of Mathematical Psychology*, *40*, 48–63.

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. (submitted). *Mental architectures with selectively influenced but stochastically interdependent components.*

Ennis, D.M. (1992). Modeling similarity and identification when there momentary fluctuations in psychological magnitudes. In F.G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 279–298). Hillsdale, NJ: Erlbaum.

Ennis, D.M., Palen, J.J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology*, *32*, 449–465.

Gauchman, H., & Rubel, L. (1989). Sums of products of functions of x times functions of y. *Linear Algebra and its Applications*, *125*, 19–63.

Lazarsfeld, P.F. (1965). Latent structure analysis. In S. Sternberg, V. Capecchi, T. Kloek, & C.T. Leenders (Eds.), *Mathematics and Social Sciences*, (Vol. 1, pp. 37–54). Paris: Mouton.

Lindsay, B.G. (1995). *Mixture models: Theory, geometry, and applications*. Hayward, CA: Institute of Mathematical Statistics Press.

Luce, R.D., & Galanter, E. (1963). Discrimination. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103–189). New York, NY: Wiley.

McDonald, R. P. (1967). Nonlinear factor analysis. *Psychometrika Monographs*, No. 15.

McDonald, R. P. (1982). Linear versus nonlinear models in item response theory. *Applied Psychological Measurement*, *6*, 379–396.

Regenwetter, M., & A.A.J. Marley (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology*, *45*, 864–912.

Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D.E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (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' method. In W.G. Koster (Ed.), *Attention and Performance II. Acta Psychologica*, *30*, 276–315.

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

Suppes, P., & Zinnes, J.L. (1963). Basic measurement theory. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (Vol. 1, pp. 3–76). New York, NY: Wiley.
Thomas, R.D. (1996). Separability and independence of dimensions within the same-different judgment task. *Journal of Mathematical Psychology*, *40*, 318–341.
Thomas, R.D. (1999). Assessing sensitivity in a multidimensional space: Some problems and a definition of a general d'. *Psychonomic Bulletin and Review*, *6*, 224–238.
Thurstone, L.L. (1927a). Psychophysical analysis. *American Journal of Psychology*, *38*, 368–389.
Thurstone, L.L. (1927b). A law of comparative judgments. *Psychological Review*, *34*, 273–286.
Tong, Y.L. (1990). *The multivariate normal distribution*. New York, NY: 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.
Zinnes, J.L., & MacKay, D.B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika*, *48*, 27–48.

Manuscript received 28 NOV 2001
Final version received 29 JUN 2002