Journal of Mathematical Psychology **43**, 123–157 (1999)
Article ID jmps.1998.1231, available online at http://www.idealibrary.com on IDEAL®

Conditionally Selective Dependence of Random Variables on External Factors

Ehtibar N. Dzhafarov

Purdue University and Hanse-Wissenschaftskolleg

Selective influence of experimental factors upon observable or hypothetical random variables is a key concept in the analysis of processing architectures and response time decompositions. This paper deals with the notion of conditionally selective influence, defined as follows. Let $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ be stochastically interdependent random variables (e.g., hypothetical components of response time), and let Φ be a set of external factors affecting the joint distribution of $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$. A subset of factors Λ_i conditionally selectively influences $\mathbf{X}_i$ if at any fixed values of the remaining random variables the conditional distribution of $\mathbf{X}_i$ only depends on factors inside Λ_i. The notion of conditional selectivity generalizes the relationship between factors and random variables described in Townsend (1984) as "indirect nonselectivity." This paper establishes the structure of the joint distribution of $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ that is necessary and sufficient for $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ to be conditionally selectively influenced by (not necessarily disjoint) factor subsets $\{\Lambda_1, ..., \Lambda_n\}$, respectively. The notion of conditional selectivity is compared to that of unconditional selectivity, defined as follows. A subset of factors Γ_i unconditionally selectively influences $\mathbf{X}_i$ if the latter can be presented as a deterministic function of Γ_i and of some random variables (the same for all $\mathbf{X}_i$, $i = 1, ..., n$) whose joint distribution does not depend on any factors from Φ. The two forms of selective influence are generally incompatible.

1. INTRODUCTION

1.1. *Selective influence.* Selective influence of experimental factors upon observable or hypothetical random variables (such as response times or response time components) has been a key concept in the analysis of "mental architectures" since

Correspondence and requests for reprints should be addressed to Ehtibar N. Dzhafarov, Department of Psychological Sciences, Purdue University, 1364 Psychological Sciences Building, West Lafayette, IN 47907-1364. I am grateful to M. Koppen for constructive critical comments. Among other things, his suggestions led to significant simplifications in the notation and some of the derivations.

0022-2496/99 $30.00

the pioneering publication of Sternberg (1969). The concept is very simple when the random variables in question are (assumed to be) stochastically independent. As an example, let $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ be hypothetical stochastically independent components of observable response time, which means that the response time is assumed to be some algebraic combination of $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ (Dzhafarov, 1997; Dzhafarov & Schweickert, 1995). Let Φ be the set of factors manipulated in the experiment in a mutually independent fashion. We say that the component $\mathbf{X}_i$ is selectively influenced by factors from a subset $\Gamma_i \subseteq \Phi$ if the distribution of $\mathbf{X}_i$ remains unchanged whenever all factors in Γ_i are fixed, irrespective of the values of the factors from the complementary subset $\Phi - \Gamma_i$, whereas the distribution of $\mathbf{X}_i$ is different for at least two different values of Γ_i, at some fixed values of $\Phi - \Gamma_i$. (A value of a set is, of course, the set of values of its elements.)

Note that the factor subsets $\{\Gamma_1, ..., \Gamma_n\}$ selectively influencing $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$, respectively, need not be pairwise disjunctive; they may overlap or even coincide. The selectiveness in the above definition only refers to the fact that some factors (namely, those from $\Phi - \Gamma_i$) are excluded from the list of the factors that may influence $\mathbf{X}_i$. If, for example, the set of random variables is $\{\mathbf{X}_1, \mathbf{X}_2\}$ and Φ consists of two independently manipulated factors $\{\gamma_1, \gamma_2\}$, then the above definition of selective influence encompasses all 16 possible situations: $\mathbf{X}_1$ being selectively influenced by the empty set, or $\{\gamma_1\}$, or $\{\gamma_2\}$, or $\{\gamma_1, \gamma_2\}$, combined with the same possibilities for $\mathbf{X}_2$.

1.2. *Stochastic interdependence and selective influence.* The notion of selective influence becomes less straightforward once the random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ are allowed to be stochastically interdependent. In the literature on mental architectures and processing time decompositions one can find two distinctly different approaches to this problem. According to one of them, derived from Dzhafarov (1992, 1997) and Dzhafarov and Schweickert (1995), the random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ are selectively influenced by, respectively, subsets $\{\Gamma_1, ..., \Gamma_n\}$ of a factor space Φ if these random variables can be represented as

$$\begin{aligned}
\mathbf{X}_1 &= X_1(\mathbf{P}_1, ..., \mathbf{P}_n; \Gamma_1) \\
&\cdots \\
\mathbf{X}_i &= X_i(\mathbf{P}_1, ..., \mathbf{P}_n; \Gamma_i) \qquad (1) \\
&\cdots \\
\mathbf{X}_n &= X_n(\mathbf{P}_1, ..., \mathbf{P}_n; \Gamma_n),
\end{aligned}$$

where $\{X_1, ..., X_n\}$ are some (deterministic) functions, and $\{\mathbf{P}_1, ..., \mathbf{P}_n\}$ are stochastically independent random variables ("sources of randomness") with arbitrary distribution functions strictly increasing on their respective supports (e.g., uniformly distributed between 0 and 1).

This definition of selective influence is based on the observation (Dzhafarov, 1997) that any set of random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ whose joint distribution depends on factors from a set Φ can be represented as

$$
\begin{aligned}
\mathbf{X}_1 &= X_1(\mathbf{P}_1, ..., \mathbf{P}_n; \Phi)\\
&\cdots\\
\mathbf{X}_i &= X_i(\mathbf{P}_1, ..., \mathbf{P}_n; \Phi)\\
&\cdots\\
\mathbf{X}_n &= X_n(\mathbf{P}_1, ..., \mathbf{P}_n; \Phi),
\end{aligned}
\tag{2}
$$

with the same meaning of the symbols as in (1).

For instance, choosing stochastically independent $\{\mathbf{P}_1, ..., \mathbf{P}_n\}$ uniformly distributed between 0 and 1, one can always represent $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ as

$$
\begin{aligned}
\mathbf{X}_1 &= Q_1(\mathbf{P}_1; \Phi)\\
&\cdots\\
\mathbf{X}_i &= Q_i(\mathbf{X}_1, ..., \mathbf{X}_{i-1}, \mathbf{P}_i; \Phi)\\
&\cdots\\
\mathbf{X}_n &= Q_n(\mathbf{X}_1, ..., \mathbf{X}_{n-1}, \mathbf{P}_n; \Phi),
\end{aligned}
\tag{3}
$$

where Q_1 is the quantile function[1] for $\mathbf{X}_1$, and, for any $x_1, ..., x_{i-1}$, Q_i is the quantile function for the conditional distribution of $\mathbf{X}_i$ given $\mathbf{X}_1 = x_1, ..., \mathbf{X}_{i-1} = x_{i-1}$. Note that $\mathbf{P}_i$ is stochastically independent of $\{\mathbf{X}_1, ..., \mathbf{X}_{i-1}\}$, $i = 2, ..., n$. This universal representation is used for simulating arbitrarily distributed vectors of random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ by means of standard generators of random numbers, $\{\mathbf{P}_1, ..., \mathbf{P}_n\}$ (Yermalov, 1971). Since

$$
\begin{aligned}
Q_i(\mathbf{X}_1, ..., \mathbf{X}_{i-1}, \mathbf{P}_i; \Phi) &= Q_i[Q_1(\mathbf{P}_1; \Phi), ..., Q_{i-1}(\mathbf{X}_1, ..., \mathbf{X}_{i-2}, \mathbf{P}_{i-1}; \Phi), \mathbf{P}_i; \Phi]\\
&= X_i(\mathbf{P}_1, ..., \mathbf{P}_i; \Phi), \qquad i = 1, ..., n,
\end{aligned}
\tag{4}
$$

representation (2) can be viewed as a symmetrical rendering of (3); the two representations are equivalent because (3) is universally true, and, through (4), it implies (2).

By imposing different restrictions on the functions $\{X_1, ..., X_n\}$ one can obtain various forms of interdependence between $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ as special cases of (1). Thus the case of mutual stochastic independence of these random variables corresponds to the representation

$$\mathbf{X}_i = X_i(\mathbf{P}_i; \Gamma_i), \qquad i = 1, ..., n.$$

At another extreme, if all $\{X_1, ..., X_n\}$ are strictly monotonic functions of one and the same source of randomness, $\mathbf{P}_i = \mathbf{P}$, so that (1) reduces to

$$\mathbf{X}_i = X_i(\mathbf{P}; \Gamma_i), \qquad i = 1, ..., n,$$

[1] That is, the inverse of the distribution function, defined, e.g., as the infimum of the variable's values exceeding a given cumulative probability.

then any random variable $\mathbf{X}_i$ becomes a deterministic function of any other random variable $\mathbf{X}_j$, for any fixed values of the factor subsets Γ_i and Γ_j influencing them selectively:

$$\mathbf{X}_j = X_{ij}(\mathbf{X}_i; \Gamma_i \cup \Gamma_j), \qquad i, j = 1, ..., n.$$

These extreme special cases of stochastic in(ter)dependence under selective influence are analyzed in Dzhafarov and Schweickert (1995), Cortese and Dzhafarov (1996) and Dzhafarov and Cortese (1996).

1.3. *Conditionally selective influence.* The focus of this paper, however, is on another approach to extending the notion of selective influence to incorporate stochastically interdependent random variables, proposed in Townsend (1984) and Townsend and Thomas (1994). This approach consists in allowing the marginal distribution of any of the variables $\mathbf{X}_i$ to generally depend on all factors in the factor space Φ, but distinguishing between a subset $\Lambda_i \subseteq \Phi$ of factors that influence $\mathbf{X}_i$ *directly* and the complementary subset $\Phi - \Lambda_i$ of factors influencing $\mathbf{X}_i$ *indirectly*, through the values of other variables from the list $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$.

The following definition is designed to facilitate the comparison of this notion to (1). The random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ are directly influenced by, respectively, subsets $\{\Lambda_1, ..., \Lambda_n\}$ of a factor space Φ if these random variables can be represented by the following self-referencing system of equations:

$$\begin{aligned} \mathbf{X}_1 &= X_1(\mathbf{P}_1, \mathbf{X}_2, ..., \mathbf{X}_n; \Lambda_1) \\ &\cdots \\ \mathbf{X}_i &= X_i(\mathbf{X}_1, ..., \mathbf{X}_{i-1}, \mathbf{P}_i, \mathbf{X}_{i+1}, ..., \mathbf{X}_n; \Lambda_i) \\ &\cdots \\ \mathbf{X}_n &= X_n(\mathbf{X}_1, ..., \mathbf{X}_{n-1}, \mathbf{P}_n; \Lambda_n), \end{aligned} \tag{5}$$

where all symbols have the same meaning as in (1), and, in addition, $\mathbf{P}_i$ is stochastically independent of $\{\mathbf{X}_1, ..., \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, ..., \mathbf{X}_n\}$, $i = 1, ..., n$.

The rationale for this definition is analogous to that of (1): it is derived from the possibility of representing any set of random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ (whose joint distribution depends on Φ) in the form

$$\begin{aligned} \mathbf{X}_1 &= X_1(\mathbf{P}_1, \mathbf{X}_2, ..., \mathbf{X}_n; \Phi) \\ &\cdots \\ \mathbf{X}_i &= X_i(\mathbf{X}_1, ..., \mathbf{X}_{i-1}, \mathbf{P}_i, \mathbf{X}_{i+1}, ..., \mathbf{X}_n; \Phi). \\ &\cdots \\ \mathbf{X}_n &= X_n(\mathbf{X}_1, ..., \mathbf{X}_{n-1}, \mathbf{P}_n; \Phi), \end{aligned} \tag{6}$$

with the same meaning of the symbols. This follows from the universal representation (3) by renaming the variables $\mathbf{X}_1, ..., \mathbf{X}_n$ so that each of them is successively placed at the end of the variables' list.

Returning to (5), if the value of the arguments $\mathbf{X}_1, ..., \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, ..., \mathbf{X}_n$ in the function X_i are fixed, the remaining function can be written as

$$X_i(\mathbf{P}_i; \Lambda_i \mid x_1, ..., x_{i-1}, x_{i+1}, ..., x_n).$$

Clearly, this is a random variable, and it can be viewed as the *conditional random variable* $\mathbf{X}_i$, given $\mathbf{X}_1 = x_1, ..., \mathbf{X}_{i-1} = x_{i-1}, \mathbf{X}_{i+1} = x_{i+1}, ..., \mathbf{X}_n = x_n$:

$$\begin{aligned} &\mathbf{X}_i \mid_{\mathbf{X}_1 = x_1, ..., \mathbf{X}_{i-1} = x_{i-1}, \mathbf{X}_{i+1} = x_{i+1}, ..., \mathbf{X}_n = x_n} \\ &\quad = X_i(\mathbf{P}_i; \Lambda_i \mid x_1, ..., x_{i-1}, x_{i+1}, ..., x_n), \qquad i = 1, ..., n. \end{aligned} \tag{7}$$

The distribution of this conditional random variable (in more conventional terms, the conditional distribution of $\mathbf{X}_i$ given $\mathbf{X}_1 = x_1, ..., \mathbf{X}_{i-1} = x_{i-1}, \mathbf{X}_{i+1} = x_{i+1}, ..., \mathbf{X}_n = x_n$) only depends on factors from the subset Λ_i. Substantively, this observation generalizes the original definition of direct and indirect influence (Townsend, 1984), making it applicable to n arbitrary random variables (without the context of any specific "mental architecture") and n arbitrary factor subsets $\{\Lambda_1, ..., \Lambda_n\}$, not necessarily disjunctive or distinct.

The term coined by Townsend (1984) for the situation described in (7) is *indirect nonselectivity*. In this paper, however, I use the mathematically more descriptive term *conditionally selective influence*, as opposed to the *unconditionally selective influence* represented by (1). In other words, if (5) holds (equivalently, if (7) holds), we say that $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ are conditionally selectively influenced by $\{\Lambda_1, ..., \Lambda_n\}$, respectively, while if (1) holds, we say that $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ are (unconditionally) selectively influenced by $\{\Gamma_1, ..., \Gamma_n\}$, respectively.

1.4. *Unconditional and conditional selectivity.* Conditional selectivity is generally excluded by unconditional selectivity, that is, generally

$$\{\Lambda_1, ..., \Lambda_n\} \neq \{\Gamma_1, ..., \Gamma_n\}$$

for one and the same vector $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$. Indeed, by fixing the values of, say, $\{\mathbf{X}_2, ..., \mathbf{X}_n\}$ in (1), one imposes $n-1$ constraints on $\{\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_n\}$, and these constraints induce a conditional distribution of

$$\{\mathbf{P}_1, ..., \mathbf{P}_n\} \mid_{\mathbf{X}_2 = x_2, ..., \mathbf{X}_n = x_n}$$

that generally depends on factors from $\Gamma_2 \cup \cdots \cup \Gamma_n$. The remaining conditional random variable, therefore,

$$\mathbf{X}_1 \mid_{\mathbf{X}_2 = x_2, ..., \mathbf{X}_n = x_n} = X_1(\{\mathbf{P}_1, ..., \mathbf{P}_n\} \mid_{\mathbf{X}_2 = x_2, ..., \mathbf{X}_n = x_n}; \Gamma_1),$$

generally depends on $\Gamma_1 \cup \Gamma_2 \cup \cdots \cup \Gamma_n$, rather than just on Γ_1.

1.5. *Problem.* Although representation (5) is conceptually simple and convenient for comparing the two forms of selective influence, conditional and unconditional, its self-referencing structure makes it difficult to study the relationship among the random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ directly. The question arises whether a greater insight in this relationship can be achieved if instead of the random variables per se one investigates their joint distribution functions. To realize this approach one should establish *the structure of the joint distribution of* $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ *that is sufficient and necessary for these random variables to be conditionally selectively influenced by* $\{\Lambda_1, ..., \Lambda_n\}$, *respectively*. This is the target issue of the present paper.

2. CONVENTIONS AND PRELIMINARIES

In this section I introduce notation, terminology, and technical conventions to be used subsequently. It can also serve as a reference section to be consulted recurrently while reading the remainder of the paper.

2.1. *Context system.* It is apparent from (5) and (7) that the answer to the question whether a given random variable, $\mathbf{X}_i$, is conditionally selectively influenced by some factor subset Λ_i depends on what other random variables are in play and what other factors belong to the factor space Φ. In other words, the concept of conditional selectivity is context-dependent, and I accordingly refer to the system $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$ as the *context system*. Given a context system, the subset of factors that conditionally selectively influence $\mathbf{X}_i$ is denoted by $\Lambda(\mathbf{X}_i)$. Thus in (5), $\Lambda(\mathbf{X}_1) = \Lambda_1 \;\&\cdots\&\; \Lambda(\mathbf{X}_n) = \Lambda_n$.

The following simple observation is required in the proof of the main theorem of this paper (Theorem 3), and it is also of interest by itself. Consider a context system $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, and let some of the random variables, say, $\mathbf{X}_1, ..., \mathbf{X}_r$ $(r < n)$, be conditioned on some fixed values of the remaining variables, $\mathbf{X}_{r+1} = x_{r+1} \;\&\cdots\&\; \mathbf{X}_n = x_n$. Then, for the *conditional context system*

$$\{\mathbf{X}_1, ..., \mathbf{X}_r; \Phi\}\,|_{\mathbf{X}_{r+1} = x_{r+1} \,\&\cdots\&\, \mathbf{X}_n = x_n},$$

the following is true:

$$\Lambda(\mathbf{X}_i\,|_{\mathbf{X}_{r+1} = x_{r+1} \,\&\cdots\&\, \mathbf{X}_n = x_n}) \subseteq \Lambda(\mathbf{X}_i), \qquad i = 1, ..., r. \tag{8}$$

In other words, the factor subsets conditionally selectively influencing $\mathbf{X}_1, ..., \mathbf{X}_r$ at some fixed values of the remaining variables are included in the corresponding factor subsets conditionally selectively influencing $\mathbf{X}_1, ..., \mathbf{X}_r$ when the remaining variables vary freely. This immediately follows from the fact that

$$\begin{aligned}\{\mathbf{X}_1\,|_{\mathbf{X}_2 = x_2 \,\&\cdots\&\, \mathbf{X}_r = x_r}\}\,&|_{\mathbf{X}_{r+1} = x_{r+1} \,\&\cdots\&\, \mathbf{X}_n = x_n}\\ &= \mathbf{X}_1\,|_{\mathbf{X}_2 = x_2 \,\&\cdots\&\, \mathbf{X}_r = x_r \,\&\, \mathbf{X}_{r+1} = x_{r+1} \,\&\cdots\&\, \mathbf{X}_n = x_n}.\end{aligned} \tag{9}$$

2.2. *Factors.* The factors belonging to the factor space Φ are assumed to have *completely crossable levels* (which can always be achieved by appropriate parametrization). It is also assumed that all factors are *effective*, in the following sense: every factor possesses at least two different values corresponding to two different joint distributions of $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ at some fixed values of the remaining factors. As shown below (in corollaries to Theorems 2 and 3), this amounts to saying that any factor conditionally selectively influences at least one random variable, that is, $\Lambda(\mathbf{X}_1) \cup \cdots \cup \Lambda(\mathbf{X}_n) = \Phi$.

2.3. *Density.* The joint distribution of $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ is assumed to be representable by a *density function* $\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]$, defined on some unspecified region of $\{x_1, ..., x_n\}$. For all applied purposes the density can be understood either as a conventional piecewise continuous density in n-dimensional Euclidean space, or as a probability mass function on a Cartesian product of n sets of counting indices (in which case all integrals should be understood as simple summation). However, $\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]$ may also be treated in a general way, as a Radon–Nikodym derivative with respect to a product-measure imposed on $\{x_1, ..., x_n\}$. All results established in this paper hold under this general point of view in the "almost everywhere" sense.

2.4. *Conditional density.* The focal object of our discussion is the *conditional density*

$$\begin{aligned}
&\Psi[\mathbf{X}_i = x_i \mid \mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_{i-1} = x_{i-1} \,\&\, \mathbf{X}_{i+1} = x_{i+1} \,\&\cdots\&\, \mathbf{X}_n = x_n] \\
&\quad = \frac{\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]}{\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_{i-1} = x_{i-1} \,\&\, \mathbf{X}_{i+1} = x_{i+1} \,\&\cdots\&\, \mathbf{X}_n = x_n]} \\
&\quad = \frac{\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]}{\int \Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]\, \mathrm{d}x_i},
\end{aligned} \tag{10}$$

for $i = 1, ..., n$. Thus we say that $\mathbf{X}_i$ is conditionally selectively influenced by Λ_i, that is, $\Lambda(\mathbf{X}_i) = \Lambda_i$, if the value of (10) is fixed whenever Λ_i is fixed. To ascertain that $\Lambda(\mathbf{X}_i) = \Lambda_i$ rather than merely $\Lambda(\mathbf{X}_i) \subset \Lambda_i$, one has to add to the previous sentence that every factor from Λ_i possesses at least two different values corresponding to two different values of (10), at some fixed values of the remaining factors from Φ.

2.5. *Convention on* 0/0. Conditional density (10) is either well-defined for all values of x_i, or it is indefinite (0/0) for all values of x_i, depending on whether

$$\int \Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]\, \mathrm{d}x_i$$

is positive or zero (in both cases x_i is considered varying while other x-arguments and factors are fixed). Throughout this paper 0/0 is treated as a non-removable singularity assigned a special value, *indef.* Consequently, if as a result of changing the value of a factor λ (at some fixed values of other factors) the value of (10) changes from *indef* to zero or to a positive number, then the conditional probability in (10) depends on λ, that is, $\lambda \in \Lambda(\mathbf{X}_i)$; if, however, the value of (10) remains *indef* under all changes of λ (at any fixed values of other factors), then the conditional probability does not depend on λ, that is, $\lambda \notin \Lambda(\mathbf{X}_i)$.

2.6. *Positive domain.* Given a function $F(x_1, ..., x_n; \Phi)$, the set of values of its arguments $\{x_1, ..., x_n\}$ at which the function is positive generally depends on the value of its factor set Φ. This area is referred to as the *positive domain* of the function. Note that the factors from Φ in this definition are treated as parameters, rather than arguments of the function. If $F(x_1, ..., x_n; \Phi)$ is a density function, then its positive domain is traditionally referred to as its *support*. Of special interest is the support of the joint density $\Psi[\mathbf{X}_1 = x_1 \,\&\, \cdots \,\&\, \mathbf{X}_n = x_n]$, always denoted in this paper by $\mathfrak{R}_{1 \cdots n} = \mathfrak{R}_{1 \cdots n}\{\Phi\}$. It is assumed throughout this paper that $\mathfrak{R}_{1 \cdots n}$ is a region of non-zero measure at all values of Φ. The projection of this region on the x_i-axis is denoted by $\mathfrak{R}_i = \mathfrak{R}_i\{\Phi\}$ $(i = 1, ..., n)$; its projection on the $x_i x_j$-plane is denoted by $\mathfrak{R}_{ij} = \mathfrak{R}_{ij}\{\Phi\}$ $(i, j = 1, ..., n;\ i \neq j)$; and so on, for all projection hyperplanes corresponding to the subsets of the arguments $\{x_1, ..., x_n\}$. Figure 1 illustrates these concepts for $n = 2$.

2.7. *Absorption principle.* Many formulations found in this paper have the following form: a certain function is representable as a sum of some functions, $\sum F_i(A_i)$, or a product of some nonnegative functions, $\prod F_i(A_i)$, where the A_i's are certain sets of arguments, with no other restrictions imposed on F_i's. The *absorption principle* is a simple observation that if $A_j \subseteq A_i$ for some i and j, then the function $F_j(A_j)$ can be eliminated from such a representation. As an example, to say that a function is representable as

$$F_1(a_1, a_2, a_3) + F_2(a_1, a_4) + F_3(a_2, a_3)$$

is equivalent to saying that it is representable as

$$F_1(a_1, a_2, a_3) + F_2(a_1, a_4).$$

Indeed, $F_1(a_1, a_2, a_3) + F_3(a_2, a_3)$ is some function of $\{a_1, a_2, a_3\}$ that, to save notation, itself can be denoted by $F_1(a_1, a_2, a_3)$. Conversely, for any functions $F_1(a_1, a_2, a_3)$ and $F_3(a_2, a_3)$, the function $F_1(a_1, a_2, a_3)$ can be written as $F_3(a_2, a_3)$ plus some function of $\{a_1, a_2, a_3\}$. Analogous reasoning obviously applies to products of nonnegative functions.

2.8. *Index subset symbolism.* The main result of this paper (Theorem 3) involves functions of a special structure: given n variables $\{x_1, ..., x_n\}$ and n factor subsets $\{\Lambda_1, ...\Lambda_n\}$, these functions are

$$\begin{array}{ll} & f(\Phi) \\ f_i(x_i; \Lambda_i), & i = 1, ..., n \\ f_{ij}(x_i, x_j; \Lambda_i \cap \Lambda_j), & i, j = 1, ..., n; \quad i \neq j \\ f_{ijk}(x_i, x_j, x_k; \Lambda_i \cap \Lambda_j \cap \Lambda_k), & i, j, k = 1, ..., n; \quad i \neq j, i \neq j, j \neq k \\ & \cdots \\ f_{1 \cdots n}(x_1, ..., x_n; \Lambda_1 \cap \cdots \cap \Lambda_n). & \end{array} \tag{11}$$

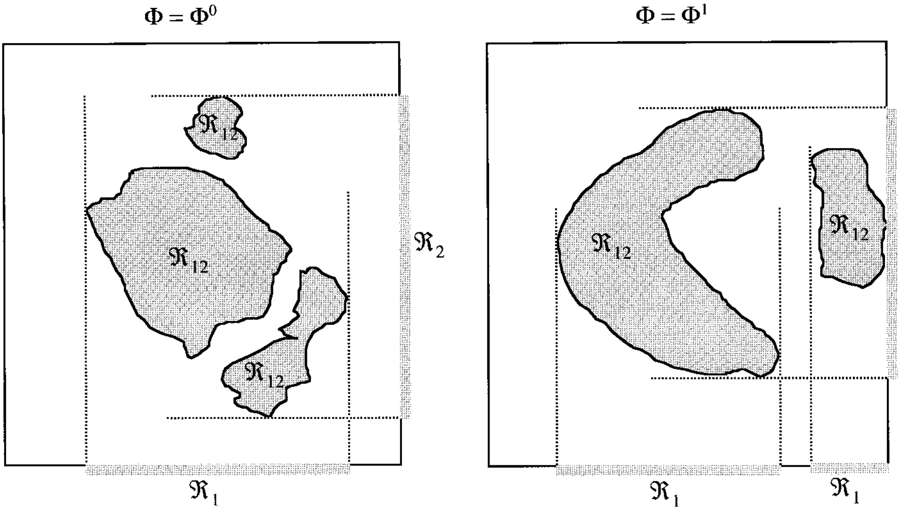

FIG. 1. An illustration for the support $\mathfrak{R}_{12}$ of a density function $\Psi[\mathbf{X}_1 = x_1 \mid \mathbf{X}_2 = x_2]$ and its projections $\mathfrak{R}_1$ and $\mathfrak{R}_2$ on the two axes; shown for two different values of the factor set Φ upon which the density function and its support depend.

Ignoring for the moment the first of these functions, it is easy to notice that the subscripts are formed by all nonempty subsets of the index set $\{1, ..., n\}$; the x-arguments are the corresponding subsets of $\{x_1, ..., x_n\}$; and the factor-arguments are the intersections of the corresponding factor subsets chosen from $\{\Lambda_1, ..., \Lambda_n\}$. A simple way of referring to such functions in general is

$$f_I\left(\{x_i\}_{i \in I}; \bigcap_{i \in I} \Lambda_i\right), \qquad I \subseteq \{1, ..., n\}.$$

To incorporate the first function in (11), $f(\Phi)$, one only has to drop the restriction that the index subset $I \subseteq \{1, ..., n\}$ be nonempty and treat $f(\Phi)$ as $f_\varnothing(\Phi)$. Indeed, from the obvious identities

$$\begin{gathered} \{x_i\}_{i \in I} \cup \{x_i\}_{i \in \varnothing} = \{x_i\}_{i \in I \cup \varnothing} = \{x_i\}_{i \in I} \\ \left(\bigcap_{i \in I} \Lambda_i\right) \cap \left(\bigcap_{i \in \varnothing} \Lambda_i\right) = \bigcap_{i \in I \cup \varnothing} \Lambda_i = \bigcap_{i \in I} \Lambda_i \end{gathered} \tag{12}$$

we conclude that[2]

$$\begin{gathered} \{x_i\}_{i \in \varnothing} = \varnothing \\ \bigcap_{i \in \varnothing} \Lambda_i = \Phi. \end{gathered} \tag{13}$$

[2] The following identities are special cases of the general rule according to which any closed associative operation with an identity element yields this identity element when applied to the empty set of arguments. I am grateful to M. Koppen for pointing out to me this aesthetically pleasing and useful fact.

The index subset symbolism can also be used for other purposes. For instance, any nonempty $I \subseteq \{1, ..., n\}$ defines a hyperplane $\{x_i\}_{i \in I}$, and $\mathfrak{R}_I$ denotes the projection of the support $\mathfrak{R}_{1 \dots n}$ on this hyperplane (see Subsection 2.6). An extensive use of the index subset symbolism is made in Appendix D.

3. RESULTS

All mathematical results are presented below as numbered theorems whose proofs are relegated to Appendices, together with lemmas, definitions, and notation conventions needed in the proofs. To facilitate understanding, the formulations of the theorems in the main text are typically less succinct than their equivalent formulations in the Appendices.

3.1. *Support of joint density.* As stated earlier (Subsection 2.6) and illustrated in Fig. 1, the support $\mathfrak{R}_{1 \dots n}$ of the joint density function $\Psi[\mathbf{X}_1 = x_1 \, \& \cdots \& \, \mathbf{X}_n = x_n]$ generally depends on the values of the factors Φ. It turns out that this dependence imposes an important restriction on the class of subsets $\{\Lambda_1, ..., \Lambda_n\}$ of Φ that may conditionally selectively influence $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$, respectively. Namely, we have the following

THEOREM 1. *In a context system $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, if the support $\mathfrak{R}_{1 \dots n}$ of the joint density function $\Psi[\mathbf{X}_1 = x_1 \, \& \cdots \& \, \mathbf{X}_n = x_n]$ depends on a factor $\lambda \in \Phi$, then $\lambda \in \Lambda(\mathbf{X}_1) \cap \cdots \cap \Lambda(\mathbf{X}_n)$. Put differently, $\mathfrak{R}_{1 \dots n}$ does not depend on factors outside $\Lambda(\mathbf{X}_1) \cap \cdots \cap \Lambda(\mathbf{X}_n)$:*

$$\mathfrak{R}_{1 \dots n} = \mathfrak{R}_{1 \dots n}[\Lambda(\mathbf{X}_1) \cap \cdots \cap \Lambda(\mathbf{X}_n)]. \tag{14}$$

The proof is given in Appendix A. An immediate consequence of this theorem is that all projections $\mathfrak{R}_I$ of the support $\mathfrak{R}_{1 \dots n}$, $I \subseteq \{1, ..., n\}$, also depend on $\Lambda(\mathbf{X}_1) \cap \cdots \cap \Lambda(\mathbf{X}_n)$ only:

$$\mathfrak{R}_I = \mathfrak{R}_I[\Lambda(\mathbf{X}_1) \cap \cdots \cap \Lambda(\mathbf{X}_n)], \qquad I \subseteq \{1, ..., n\}. \tag{15}$$

3.2. *Conditional selectivity for bivariate density.* This section presents the main result of this paper for bivariate context systems, $\{\mathbf{X}_1, \mathbf{X}_2; \Phi\}$: it establishes the structure of the density function $\Psi[\mathbf{X}_1 = x_1 \, \& \, \mathbf{X}_2 = x_2]$ that is necessary and sufficient for $\{\mathbf{X}_1, \mathbf{X}_2\}$ to be conditionally selectively influenced by subsets $\{\Lambda_1, \Lambda_2\}$ of Φ, respectively.

It is convenient to begin by formulating these necessary and sufficient conditions for $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1$ & $\Lambda(\mathbf{X}_2) \subseteq \Lambda_2$, rather than $\Lambda(\mathbf{X}_1) = \Lambda_1$ & $\Lambda(\mathbf{X}_2) = \Lambda_2$, for two reasons. First, the necessary conditions for the inclusion obviously also apply to the equality, whereas the sufficient conditions for the inclusion, as shown below, need only trivial modifications to apply to the equality. Second, the logical essence of the notion of selective influence (conditional or unconditional) is in the *exclusion* of certain factors rather than ascertaining that the factors that have not been excluded are all *effective*. By stating that $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1$ one positively excludes all factors in $\Phi - \Lambda_1$ from the class of those conditionally selectively influencing $\mathbf{X}_1$; the factors within Λ_1 then are treated as those that *may* influence $\mathbf{X}_1$, even if the influence of a given factor from Λ_1 is negligible or even (as a marginal case) non-existent. With this in mind we can formulate the following

THEOREM 2. *In a context system* $\{\mathbf{X}_1, \mathbf{X}_2; \Phi\}$, *the statement*

$$\Lambda(\mathbf{X}_1) \subseteq \Lambda_1 \qquad \& \qquad \Lambda(\mathbf{X}_2) \subseteq \Lambda_2$$

(*i.e.*, $\mathbf{X}_i$ *is not conditionally selectively influenced by factors outside* Λ_i, $i=1,2$) *is true if and only if the joint density function is representable as*

$$\Psi[\mathbf{X}_1 = x_1 \;\&\; \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, f_2(x_2; \Lambda_2)\, f(\Phi), \tag{16}$$

where

(i) $f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)$ *is a nonnegative function whose positive domain coincides with* $\mathfrak{R}_{12}$, *the support of the density function*;

(ii) $f_1(x_1; \Lambda_1)$ *is a function whose positive domain contains* $\mathfrak{R}_1$, *the projection of* $\mathfrak{R}_{12}$ *on the* x_1*-axis*;

(iii) $f_2(x_2; \Lambda_2)$ *is a function whose positive domain contains* $\mathfrak{R}_2$, *the projection of* $\mathfrak{R}_{12}$ *on the* x_2*-axis*;

(iv) $f(\Phi)$ *is a positive function defined as*

$$f(\Phi) = \left(\iint f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, f_2(x_2; \Lambda_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \right)^{-1}. \tag{17}$$

The proof is given in Appendix C; the reader should also consult Appendix B in which a certain "multiple-difference" operator is introduced that is needed in the proof. Recall that, by convention (Subsection 2.6), the support $\mathfrak{R}_{12}$ is a region of non-zero measure at all values of Φ, because of which the double integral in (17) is always non-zero. The statement that $f(\Phi)$ is positive implies that the double integral is finite.

Note that Theorem 2 only states that representation (16), with certain constraints imposed on the functions involved, is equivalent to the statement $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1$ & $\Lambda(\mathbf{X}_2) \subseteq \Lambda_2$. The theorem does not state that (16) is the only possible representation equivalent to this statement. Notably, one can sometimes modify the constraints on the positive domains of the functions involved. For example, one can make the positive domains of f_1 and f_2 strictly coincide with $\mathfrak{R}_1$ and $\mathfrak{R}_2$ while allowing the function f_{12} to be positive in arbitrary regions outside $\mathfrak{R}_1 \times \mathfrak{R}_2$. The constraints listed in Theorem 2, however, are the most economical ones for arbitrary density functions.

The following is an immediate but significant consequence of Theorem 2.

COROLLARY 1 TO THEOREM 2. *In a context system* $\{\mathbf{X}_1, \mathbf{X}_2; \Phi\}$, *if* $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1$ & $\Lambda(\mathbf{X}_2) \subseteq \Lambda_2$, *then* $\Lambda_1 \cup \Lambda_2 = \Phi$. *In other words, every factor conditionally selectively influences at least one of the two random variables.*

Indeed, it is clear from (17) that $f(\Phi) = f(\Lambda_1 \cup \Lambda_2)$, and by applying this to (16) one realizes that the joint density function $\Psi[\mathbf{X}_1 = x_1 \;\&\; \mathbf{X}_2 = x_2]$ cannot depend on any factors outside $\Lambda_1 \cup \Lambda_2$. Since, by convention (Subsection 2.2), all factors in the factor space Φ are effective, in the sense that they all influence the joint density function, one concludes that $\Lambda_1 \cup \Lambda_2 = \Phi$.

Figure 2 shows three possible pairs of $\{\Lambda_1, \Lambda_2\}$ for the context system $\{\mathbf{X}_1, \mathbf{X}_2; \lambda_1, \lambda_2\}$, where λ_1 and λ_2 are two independently manipulated factors. (With obvious notational modifications λ_1 and λ_2 can also be interpreted as disjoint complementary subsets of a factor space.) It is instructive to apply Theorem 2 to these special cases and see the corresponding structures of the joint density function.

A formal application of (16) to panel (a) yields

$$\Psi[\mathbf{X}_1 = x_1 \;\&\; \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2; \lambda_1, \lambda_2)\, f_1(x_1; \lambda_1, \lambda_2)\, f_2(x_2; \lambda_1, \lambda_2)\, f(\lambda_1, \lambda_2),$$

which, by the absorption principle (Subsection 2.7), reduces to

$$\Psi[\mathbf{X}_1 = x_1 \;\&\; \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2; \lambda_1, \lambda_2).$$

The result is a completely uninformative statement: the joint density function is some nonnegative function of $\{x_1, x_2\}$ depending on $\{\lambda_1, \lambda_2\}$. This comes as no surprise, however, because in order to state that $\Lambda(\mathbf{X}_1) \subseteq \{\lambda_1, \lambda_2\}$ & $\Lambda(\mathbf{X}_2) \subseteq \{\lambda_1, \lambda_2\}$ one does not have to know anything except for the context system involved.

The application of (16) to the remaining panels of Fig. 2 (and subsequent reduction of the expressions by means of the absorption principle) yields the results

$$\Psi[\mathbf{X}_1 = x_1 \;\&\; \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2; \lambda_1)\, f_2(x_2; \lambda_2)\, f(\lambda_1, \lambda_2) \qquad \text{for panel (b)},$$

$$\Psi[\mathbf{X}_1 = x_1 \;\&\; \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2)\, f_1(x_1; \lambda_1)\, f_2(x_2; \lambda_2)\, f(\lambda_1, \lambda_2) \qquad \text{for panel (c)}.$$

The latter case is of special interest, as it refers to the classical paradigm in which conditional selective influence ("indirect nonselectivity") has been discussed in the

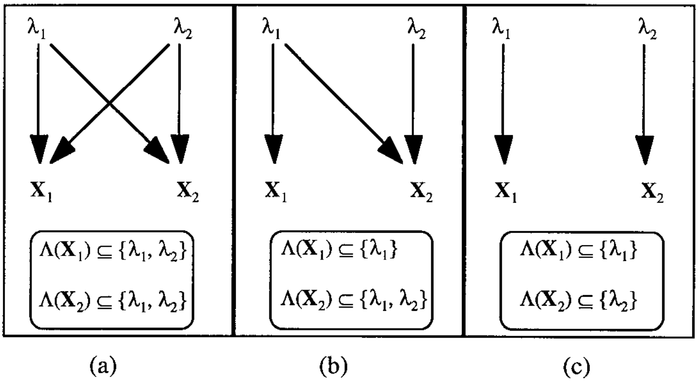

FIG. 2. Three possible pairs of sets containing $\Lambda(\mathbf{X}_1)$ and $\Lambda(\mathbf{X}_2)$ in a context system $\{\mathbf{X}_1, \mathbf{X}_2; \lambda_1, \lambda_2\}$. An arrow from a factor to a random variable indicates that the factor *may* conditionally selectively influence the variable.

literature (Townsend, 1984; Townsend & Thomas, 1994).[3] It deserves, therefore, to be formalized as a special consequence of Theorem 2.

COROLLARY 2 TO THEOREM 2. *In a context system* $\{\mathbf{X}_1, \mathbf{X}_2; \lambda_1, \lambda_2\}$, *the statement*

$$\Lambda(\mathbf{X}_1) \subseteq \{\lambda_1\} \qquad \& \qquad \Lambda(\mathbf{X}_2) \subseteq \{\lambda_2\}$$

(*i.e.*, $\mathbf{X}_1$ *is not conditionally selectively influenced by* λ_2, *and* $\mathbf{X}_2$ *is not conditionally selectively influenced by* λ_1) *is true if and only if the joint density function is representable as*

$$\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2)\, f_1(x_1; \lambda_1)\, f_2(x_2; \lambda_2)\, f(\lambda_1, \lambda_2), \tag{18}$$

where the functions involved have the same properties as the corresponding functions in Theorem 2.

Returning now to conditions that are necessary and sufficient for $\Lambda(\mathbf{X}_1) = \Lambda$ & $\Lambda(\mathbf{X}_2) = \Lambda_2$, rather than merely $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1$ & $\Lambda(\mathbf{X}_2) \subseteq \Lambda_2$, we have the following obvious yet remarkable.

COROLLARY 3 TO THEOREM 2. *Any density function in a context system* $\{\mathbf{X}_1, \mathbf{X}_2; \Phi\}$ *is representable by* (16) *if* Λ_1 *and* Λ_2 *denote* $\Lambda(\mathbf{X}_1)$ *and* $\Lambda(\mathbf{X}_2)$, *respectively. Put differently, any bivariate density has the structure*

$$\begin{aligned}\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2] = {} & f_{12}[x_1, x_2; \Lambda(\mathbf{X}_1) \cap \Lambda(\mathbf{X}_2)]\, f_1[x_1; \Lambda(\mathbf{X}_1)] \\ & \times f_2[x_2; \Lambda(\mathbf{X}_2)]\, f[\Lambda(\mathbf{X}_1) \cup \Lambda(\mathbf{X}_2)],\end{aligned} \tag{19}$$

where the functions have the same properties as the corresponding functions in Theorem 2.

The adaptation of the sufficiency part of Theorem 2 to $\Lambda(\mathbf{X}_1) = \Lambda_1$ & $\Lambda(\mathbf{X}_2) = \Lambda_2$ requires an additional statement that all factors within Λ_1 and Λ_2 are effective in changing the values of the corresponding conditional densities. This purpose is served by

COROLLARY 4 TO THEOREM 2. *In a context system* $\{\mathbf{X}_1, \mathbf{X}_2; \Phi\}$, $\Lambda(\mathbf{X}_1) = \Lambda$ & $\Lambda(\mathbf{X}_2) = \Lambda_2$ *if the density function is representable by* (16), *and if, for* $i = 1, 2$, *any factor* $\lambda \in \Lambda_i$ *possesses two values corresponding to two different values of*

$$\frac{f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_i(x_i; \Lambda_i)}{\int f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_i(x_i; \Lambda_i)\, \mathrm{d}x_i},$$

at some fixed values of $\{x_1, x_2\}$ *and of the remaining factors in* Λ_i.

[3] In the paper presented at the 29th Meeting of the Society for Mathematical Psychology (Dzhafarov, 1996) I mistakenly stated that the situation in Fig. 2c is not possible unless the two random variables are stochastically independent.

The truth of this corollary is apparent from the proof of the necessity part of Theorem 2, in Appendix C: the expression above equals $\Psi[\mathbf{X}_1 = x_1 \mid \mathbf{X}_2 = x_2]$ for $i = 1$ and $\Psi[\mathbf{X}_2 = x_2 \mid \mathbf{X}_1 = x_1]$ for $i = 2$.

3.3. *Conditional selectivity for multivariate density.* It is not immediately obvious how to generalize the results obtained in the previous section to arbitrary context systems, $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$. The correct guess suggests itself, however, if one rewrites representation (16) using the index subset symbolism introduced in Subsection 2.8:

$$\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2] = \prod_{I \subseteq \{1, 2\}} f_I\left(\{x_i\}_{i \in I}; \bigcap_{i \in I} \Lambda_i\right).$$

The multiplication here is across all possible subsets of the index set $\{1, 2\}$, including the empty set and the index set itself.

The following theorem generalizing Theorem 2 is the main result of this paper: it establishes the structure of the density function $\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]$ that is necessary and sufficient for $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1 \,\&\cdots\&\, \Lambda(\mathbf{X}_n) \subseteq \Lambda_n$.

THEOREM 3. *In a context system $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, the statement*

$$\Lambda(\mathbf{X}_1) \subseteq \Lambda_1 \,\&\cdots\&\, \Lambda(\mathbf{X}_n) \subseteq \Lambda_n$$

(*i.e.*, $\mathbf{X}_i$ *is not conditionally selectively influenced by factors outside* Λ_i, $i = 1, ..., n$) *is true if and only if the joint density function is representable as*

$$\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n] = \prod_{I \subseteq \{1, ..., n\}} f_I\left(\{x_i\}_{i \in I}; \bigcap_{i \in I} \Lambda_i\right), \tag{20}$$

where

(i) *the function corresponding to* $I = \{1, ..., n\}$ *is nonnegative, and its positive domain coincides with* $\mathfrak{R}_{1\ldots n}$, *the support of the density function*;

(ii) *any function corresponding to a proper subset* I, $\varnothing \neq I \subset \{1, ..., n\}$, *has a positive domain containing* $\mathfrak{R}_I$, *the projection of* $\mathfrak{R}_{1\ldots n}$ *on the hyperplane* $\{x_i\}_{i \in I}$;

(iii) *the function corresponding to* $I = \varnothing$ *is positive and is related to the other functions as*

$$f(\Phi) = \left(\int \cdots \int \prod_{\varnothing \neq I \subseteq \{1, ..., n\}} f_I\left(\{x_i\}_{i \in I}; \bigcap_{i \in I} \Lambda_i\right) \mathrm{d}x_1 \cdots \mathrm{d}x_n\right)^{-1}. \tag{21}$$

The proof is given in Appendix D, preceded by additional notational conventions and a lemma. (For reasons explained in the appendix the functions denoted above as f_I are denoted there by $f_{I;\,I}$.) The proof also makes use of the multiple-difference operator introduced in Appendix B.

As with Theorem 2, it should be noted that the convention of Subsection 2.6 on the support $\mathfrak{R}_{1\ldots n}$ together with the positiveness of $f(\Phi)$ implies that the multiple integral in (21) is positive and finite.

A detailed illustration for Theorem 3 is given in the concluding section of this paper. All comments and formal consequences of Theorem 2 can now be trivially generalized to the multivariate case.

Corollary 1 to Theorem 3. *In a context system* $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, *if* $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1 \,\&\cdots\&\, \Lambda(\mathbf{X}_n) \subseteq \Lambda_n$, *then* $\Lambda_1 \cup \cdots \cup \Lambda_n = \Phi$. *In other words, every factor conditionally selectively influences at least one of the random variables.*

For n independently manipulated factors $\{\lambda_1, ..., \lambda_n\}$ (or n disjoint subsets of Φ), we have the important

Corollary 2 to Theorem 3. *In a context system* $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, *the statement*

$$\Lambda(\mathbf{X}_1) \subseteq \{\lambda_1\} \,\&\cdots\&\, \Lambda(\mathbf{X}_n) \subseteq \{\lambda_n\}$$

(*i.e.*, $\mathbf{X}_i$ *is not conditionally selectively influenced by any factors except, perhaps, for* λ_i) *is true if and only if the joint density function is representable as*

$$\begin{aligned}\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]\\ = f(\lambda_1, ..., \lambda_n)\, f_{1\cdots n}(x_1, ..., x_n) \prod_{i=1, ..., n} f_i(x_i, \lambda_i),\end{aligned} \tag{22}$$

where the functions have the same properties as the corresponding functions in Theorem 3.

Indeed, the functions containing more than just one x-argument contain no λ-arguments (because all intersections are empty). As a result, all these functions are absorbed by $f_{1\cdots n}(x_1, ..., x_n)$.

The necessary conditions for $\Lambda(\mathbf{X}_1) = \Lambda \,\&\cdots\&\, \Lambda(\mathbf{X}_n) = \Lambda_n$ are given by

Corollary 3 to Theorem 3. *Any density function in a context system* $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$ *is representable by* (20) *if* Λ_i *denotes* $\Lambda(\mathbf{X}_i)$, $i = 1, ..., n$. *Put differently, any multivariate density has the structure*

$$\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n] = \prod_{I \subseteq \{1, ..., n\}} f_I\left[\{x_i\}_{i \in I}; \bigcap_{i \in I} \Lambda(\mathbf{X}_i)\right], \tag{23}$$

where the functions have the same properties as the corresponding functions in Theorem 3.

Finally, the sufficient conditions for $\Lambda(\mathbf{X}_1) = \Lambda_1 \,\&\cdots\&\, \Lambda(\mathbf{X}_n) = \Lambda_n$ are given by

Corollary 4 to Theorem 3. *In a context system* $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, $\Lambda(\mathbf{X}_1) = \Lambda_1 \,\&\cdots\&\, \Lambda(\mathbf{X}_n) = \Lambda_n$ *if the density is representable by* (20), *and if, for* $i = 1, ..., n$, *any factor* $\lambda \in \Lambda_i$ *possesses two values corresponding to different values of*

$$\frac{\prod_{I \subseteq \{1, ..., n\} \,\&\, i \in I} f_I(\{x_j\}_{j \in I}; \bigcap_{j \in I} \Lambda_j)}{\int \prod_{I \subseteq \{1, ..., n\} \,\&\, i \in I} f_I(\{x_j\}_{j \in I}; \bigcap_{j \in I} \Lambda_j)\, \mathrm{d}x_i}, \qquad i = 1, ..., n,$$

for some fixed valued of $\{x_1, ..., x_n\}$ *and of the remaining factors in* Λ_i.

The truth of the corollary immediately follows from observing that the expression above is the conditional density

$$\Psi[\mathbf{X}_i = x_i \mid \mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_{i-1} = x_{i-1} \,\&\, \mathbf{X}_{i+1} = x_{i+1} \,\&\cdots\&\, \mathbf{X}_n = x_n].$$

4. CONCLUSION

4.1. *Reconstructing joint density functions.* Theorem 3 provides a solution for the focal problem of this paper: it establishes the structure of the joint density function for $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ that is necessary and sufficient for $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ to be conditionally selectively influenced by factor subsets $\{\Lambda_1, ..., \Lambda_n\}$, respectively. An important aspect of this theorem is that its necessity part is constructive: it provides an algorithm for recovering the principal structure of the density functions given any diagram of conditionally selective influences.

As an illustration, consider the context system $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ shown in the upper panel of Fig. 3. A straightforward application of (20) to this diagram generates $2^3 = 8$ f-functions (corresponding to the 8 possible subsets of 1, 2, 3). The first step consists in finding the argument sets for all these functions:

name: f_{123}	x-arguments: x_1, x_2, x_3	factors: $\{\lambda_1\} \cap \{\lambda_1, \lambda_2, \lambda_3\} \cap \{\lambda_1, \lambda_3, \lambda_4\} = \{\lambda_1\}$
name: f_{12}	x-arguments: x_1, x_2	factors: $\{\lambda_1\} \cap \{\lambda_1, \lambda_2, \lambda_3\} = \{\lambda_1\}$
name: f_{13}	x-arguments: x_1, x_3	factors: $\{\lambda_1\} \cap \{\lambda_1, \lambda_3, \lambda_4\} = \{\lambda_1\}$
name: f_{23}	x-arguments: x_2, x_3	factors: $\{\lambda_1, \lambda_2, \lambda_3\} \cap \{\lambda_1, \lambda_3, \lambda_4\} = \{\lambda_1, \lambda_3\}$
name: f_1	x-arguments: x_1	factors: $\{\lambda_1\}$
name: f_2	x-arguments: x_2	factors: $\{\lambda_1, \lambda_2, \lambda_3\}$
name: f_3	x-arguments: x_3	factors: $\{\lambda_1, \lambda_3, \lambda_4\}$
name: f	x-arguments: *none*	factors: $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$.

The next step is to eliminate some of the functions by means of the absorption principle. Thus, the functions f_{12}, f_{13}, and f_1 are eliminated because their respective arguments sets, $\{x_1, x_2; \lambda_1\}$, $\{x_1, x_3; \lambda_1\}$, and $\{x_1, \lambda_1\}$, are subsets of $\{x_1, x_2, x_3; \lambda_1\}$, the argument set of f_{123}. The resulting representation is

$$\begin{aligned}\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2 \,\&\, \mathbf{X}_3 = x_3]\\ &= f_{123}(x_1, x_2, x_3; \lambda_1)\, f_{23}(x_2, x_3; \lambda_1, \lambda_3)\, f_2(x_2; \lambda_1, \lambda_2, \lambda_3)\\ &\quad \times f_3(x_3; \lambda_1, \lambda_3, \lambda_4)\, f(\lambda_1, \lambda_2, \lambda_3, \lambda_4),\end{aligned}$$

where $f(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is the reciprocal of

$$\iiint f_{123}(x_1, x_2, x_3; \lambda_1)\, f_{23}(x_2, x_3; \lambda_1, \lambda_3)\, f_2(x_2; \lambda_1, \lambda_2, \lambda_3)$$

$$\times f_3(x_3; \lambda_1, \lambda_3, \lambda_4)\, dx_1\, dx_2\, dx_3.$$

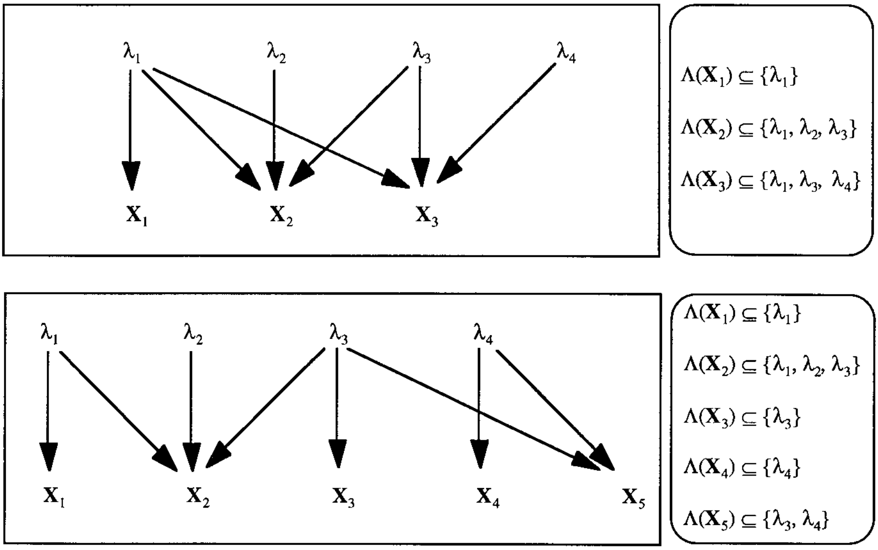

FIG. 3. A diagram of conditional selective influences for a context system $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ and a context system $\{\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4, \mathbf{X}_5; \lambda_1, \lambda_2, \lambda_3, \lambda_4\}$; the arrows have the same meaning as in Fig. 2.

The choice of the functions f_{123}, f_{23}, f_2, and f_3 is arbitrary, except for three considerations. First, f_{123} should have a positive domain of non-zero measure at all values of λ_1 and should equal zero outside this domain (which thereby becomes the support $\mathfrak{R}_{123}$ of the density function being constructed). Second, the positive domains of f_{23}, f_2, and f_3 should include the corresponding projections $\mathfrak{R}_{23}$, $\mathfrak{R}_2$, and $\mathfrak{R}_3$ of $\mathfrak{R}_{123}$. Third, the functions should be chosen so that the triple integral above is finite.

Analogously, the diagram in the lower panel of Fig. 3, after considering $2^5 = 32$ f-functions (corresponding to the 32 possible subsets of 1, 2, 3, 4, 5) and after eliminating most of them by means of the absorption principle, leads to the expression

$$\begin{aligned}\Psi[\mathbf{X}_1 = x_1 \,\&\, \cdots \,\&\, \mathbf{X}_5 = x_5] = {} & f_{12345}(x_1, x_2, x_3, x_4, x_5)\, f_{235}(x_2, x_3, x_5; \lambda_3) \\ & \times f_{12}(x_1, x_2; \lambda_1)\, f_{45}(x_4, x_5; \lambda_4)\, f_2(x_2; \lambda_1, \lambda_2, \lambda_3) \\ & \times f_5(x_5; \lambda_3, \lambda_4)\, f(\lambda_1, \lambda_2, \lambda_3, \lambda_4).\end{aligned}$$

4.2. *Relating the theory to data.* In an experiment, the random variables $\{\mathbf{X}_1, ..., \mathbf{X}_n\}$ can only be known by samples of their joint values, or, more typically, by samples of the values of some functions computed from these random variables, such as $\mathbf{X}_1 + \mathbf{X}_2$, $\min\{\mathbf{X}_1, \mathbf{X}_2\} + \mathbf{X}_3$, or other hypothetical decompositions of the observable random variables (Dzhafarov, 1997). If the experimenter assumes a certain diagram of conditionally selective influences, that is, if the contents of $\Lambda(\mathbf{X}_i)$ for all random variables in the context system are assumed to be known, Theorem 3 provides a general structure of the joint density function for these random variables. This structure, however, is too general to relate it to experimental data directly. As usual, auxiliary assumptions are needed to enable one to relate the data to specific functions with free parameters, rather than the "free functions" of Theorem 3. By their nature, such auxiliary assumptions should be different in different applications, whether they specify the forms of the f-functions directly, or through restrictions imposed on the joint distributions of the random variables or some functions thereof. A systematic discussion of these issues is beyond the scope of this paper, whose aim is exclusively conceptual: to clarify the meanings, represented by (1) and, especially, by (5), in which one can say that certain random variables, while being stochastically interdependent, are selectively influenced by certain factors. Nevertheless, it may be useful to provide simple illustrations for how the concepts developed in this paper can be related to empirical data. The illustrations given below are chosen primarily for their mathematical simplicity and do not imply any particular applied importance.

In our first example, let $\{\mathbf{X}_1, \mathbf{X}_2\}$ be observable random variables (e.g., scores obtained in two performance tests) known to be influenced by certain three factors, $\{\lambda_1, \lambda_2, \lambda_3\}$. Let the researcher be interested in the hypothesis that the diagram of conditionally selective influences is as shown in Fig. 4. In other words, the hypothesis is that $\mathbf{X}_1$ is not conditionally selectively influenced by λ_2, while $\mathbf{X}_2$ is not conditionally selectively influenced by λ_1. To test this hypothesis directly, by analyzing, for each value of $\{\lambda_1, \lambda_2, \lambda_3\}$, the subsamples of $\mathbf{X}_1$-values ($\mathbf{X}_2$-values) corresponding to various fixed values of $\mathbf{X}_2$ (respectively, $\mathbf{X}_1$), one would need unrealistically large samples of $\{\mathbf{X}_1, \mathbf{X}_2\}$-values at all values of $\{\lambda_1, \lambda_2, \lambda_3\}$. The situation simplifies considerably if, as it is often done when dealing with test scores, the auxiliary assumption is made that $\{\mathbf{X}_1, \mathbf{X}_2\}$ are distributed bivariate-normally. Given enough data, this assumption can, of course, be corroborated independently, as a nesting hypothesis with respect to the conjunction of this hypothesis with the diagram of Fig. 4.

Equating the expression for the joint density function derived from (16) to the expression for a bivariate normal density, one can easily prove that the diagram of Fig. 4 holds if and only if there are functions $m_1(\lambda_3)$, $m_2(\lambda_3)$, $c_1(\lambda_1, \lambda_3)$, $c_2(\lambda_2, \lambda_3)$, $c_3(\lambda_3)$ such that the correlation coefficient $\rho(\lambda_1, \lambda_2, \lambda_3)$, the two variances

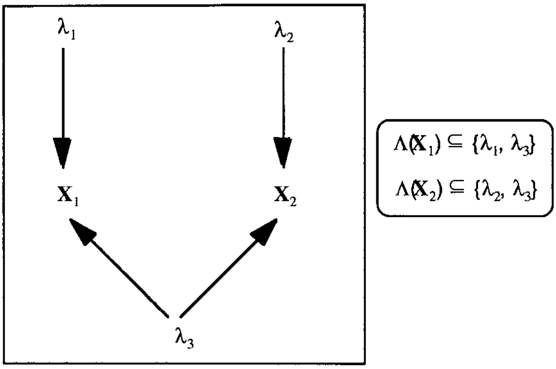

FIG. 4. A diagram of conditionally selective influences for the context system $\{\mathbf{X}_1, \mathbf{X}_2; \lambda_1, \lambda_2, \lambda_3\}$ used in the illustrations involving two test scores and two response time components (see text for details); the arrows have the same meaning as in Fig. 2.

$\sigma_1^2(\lambda_1, \lambda_2, \lambda_3)$ and $\sigma_2^2(\lambda_1, \lambda_2, \lambda_3)$, and the two means $\mu_1(\lambda_1, \lambda_2, \lambda_3)$ and $\mu_2(\lambda_1, \lambda_2, \lambda_3)$ of $\{\mathbf{X}_1, \mathbf{X}_2\}$ can be presented as

$$\begin{aligned} &\rho = c_1(\lambda_1, \lambda_3)\, c_2(\lambda_2, \lambda_3)\, c_3(\lambda_3) \\ &\sigma_1^2 = \frac{c_1^2(\lambda_1, \lambda_3)}{1-\rho^2}, \qquad \sigma_2^2 = \frac{c_2^2(\lambda_2, \lambda_3)}{1-\rho^2} \\ &\mu_1 = m_1(\lambda_3), \qquad \mu_2 = m_2(\lambda_3), \end{aligned} \tag{24}$$

where $m_1(\lambda_3)$, $m_2(\lambda_3)$, $c_1(\lambda_1, \lambda_3)$, $c_2(\lambda_2, \lambda_3)$, $c_3(\lambda_3)$ are arbitrary except for the obvious constraint

$$|\rho| = |c_1(\lambda_1, \lambda_3)\, c_2(\lambda_2, \lambda_3)\, c_3(\lambda_3)| < 1.$$

The validity of the focal hypothesis (i.e., the hypothesis that the diagram of Fig. 4 holds) can be tested in a standard way: by fitting to the data the constrained bivariate normal density,

$$\phi^*_{norm}[x_1, x_2 \,|\, m_1(\lambda_3), m_2(\lambda_3), c_1(\lambda_1, \lambda_3), c_2(\lambda_2, \lambda_3), c_3(\lambda_3)],$$

and comparing the fit with that of the bivariate normal density with unconstrained parameters,

$$\begin{aligned} &\phi_{norm}[x_1, x_2 \,|\, \mu_1(\lambda_1, \lambda_2, \lambda_3), \mu_2(\lambda_1, \lambda_2, \lambda_3), \\ &\qquad \sigma_1^2(\lambda_1, \lambda_2, \lambda_3)\, \sigma_2^2(\lambda_1, \lambda_2, \lambda_3), \rho(\lambda_1, \lambda_2, \lambda_3)]. \end{aligned}$$

Denoting by N_i the number of the levels of factor λ_i ($i = 1, 2, 3$), the addition of the focal hypothesis to the nesting hypothesis (of the bivariate normality) reduces the number of the free parameters from $5N_1N_2N_3$ to $N_3(N_1 + N_2 + 3)$. Observe that the diagram of Fig. 4 can be falsified on the level of marginal distributions alone, by showing that either of the means μ_1 and μ_2 depends on either of the factors λ_1 and λ_2.

In our second example, assume that $\{\mathbf{X}_1, \mathbf{X}_2\}$ are unobservable components of a certain observable function of these random variables, say, $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$. For instance, $\mathbf{X}$ may be a response time recorded in an experiment where the response is assumed to be the termination point for two serially connected processes with individual durations $\mathbf{X}_1$ and $\mathbf{X}_2$. Let this response time be known to depend on three experimental factors, $\{\lambda_1, \lambda_2, \lambda_3\}$, and let the hypothesis be, again, that the diagram of conditionally selective influences is as shown in Fig. 4. To be able to make use of the results already obtained, let the researcher adopt the auxiliary hypothesis (not entirely implausible, given what we know of the empirical properties of response time distributions) that $\{\mathbf{X}_1, \mathbf{X}_2\}$ have a bivariate log-normal distribution, in the sense that $\{\log \mathbf{X}_1, \log \mathbf{X}_2\}$ are distributed bivariate-normally

(Johnson & Kotz, 1976). Making use of (24), the focal hypothesis here can be tested by fitting to the data the equality

$$Prob[X<x]=\iint_{\exp(z_1)+\exp(z_2)<x} \phi^*_{norm}[z_1, z_2 \mid m_1, m_2, c_1, c_2, c_3]\, dz_1\, dz_2,$$

with $N_3(N_1+N_2+3)$ free parameters, and comparing the fit with that of the equality

$$Prob[X<x]=\iint_{\exp(z_1)+\exp(z_2)<x} \phi_{norm}[z_1, z_2 \mid \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho]\, dz_1\, dz_2,$$

with $5N_1N_2N_3$ free parameters.

This example can be modified to apply to the situation when a diagram of conditionally selective influences is not viewed as an empirically testable proposition but rather as a defining property of the hypothetical response time components. In the completely analogous discussion of *unconditionally* selective influences this situation is considered in Cortese and Dzhafarov (1996), Dzhafarov and Cortese (1996), and Dzhafarov (1997). In our case, the unobservable components $\{\mathbf{X}_1, \mathbf{X}_2\}$ of $\mathbf{X}$ can be *defined* as those conditionally selectively influenced by $\{\lambda_1, \lambda_2, \lambda_3\}$ as shown in Fig. 4. The research interest in such a situation shifts from the diagram itself (that is now taken as a given) to recovering either the composition rule by which $\mathbf{X}$ can be obtained from $\{\mathbf{X}_1, \mathbf{X}_2\}$, or the form of the stochastic relationship between $\mathbf{X}_1$ and $\mathbf{X}_2$ (Dzhafarov, 1997; Dzhafarov & Schweickert, 1995). Consider, again, the assumption that $\{\mathbf{X}_1, \mathbf{X}_2\}$ have a bivariate log-normal distribution, and let the researcher be interested in which of the three commonly used operations,

$$\mathbf{X}=\mathbf{X}_1+\mathbf{X}_2, \qquad \mathbf{X}=\min\{\mathbf{X}_1, \mathbf{X}_2\}, \qquad \mathbf{X}=\max\{\mathbf{X}_1, \mathbf{X}_2\},$$

provides a better approximation to the true composition rule (Cortese & Dzhafarov, 1996). Making use of Theorem 2 and (24), this problem is solved by comparing the fits to the data provided by the three equalities

$$Prob[X<x]=\begin{cases} \displaystyle\iint_{\exp(z_1)+\exp(z_2)<x} \phi^*_{norm}[z_1, z_2 \mid m_1, m_2, c_1, c_2, c_3]\, dz_1\, dz_2 \\ \displaystyle\iint_{\min\{z_1, z_2\}<\log x} \phi^*_{norm}[z_1, z_2 \mid m_1, m_2, c_1, c_2, c_3]\, dz_1\, dz_2 \\ \displaystyle\iint_{\max\{z_1, z_2\}<\log x} \phi^*_{norm}[z_1, z_2 \mid m_1, m_2, c_1, c_2, c_3]\, dz_1\, dz_2. \end{cases}$$

It should be noted that none of these or similar illustrations can be used to justify the expectation that the concept of conditionally selective influence is likely to play an important role in response time analysis, psychometrics, or any other applied area. The extent of the empirical usefulness of this concept remains to be seen.

APPENDIX A: SUPPORT OF JOINT DENSITY

THEOREM 1. *In a context system* $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$,

$$\Re_{1\cdots n} = \Re_{1\cdots n}\left\{\bigcap_{i=1,...,n} \Lambda(\mathbf{X}_i)\right\}.$$

Proof. If $\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n]$ is positive for all values of $(x_1, ..., x_n, \Phi)$, then the theorem holds trivially. Assume therefore that $\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n] = 0$ at some values of $(x_1, ..., x_n, \Phi)$. Choose some $\lambda \in \Phi$. If $\Re_{1\cdots n}$ depends on λ, then there are two values $\lambda^{(1)}$ and $\lambda^{(2)}$ of λ such that, for some fixed values of $\Phi - \{\lambda\}$ and some fixed $x_1, ..., x_n$,

$$\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n] = 0 \qquad \text{when } \lambda = \lambda^{(1)}$$

$$\Psi[\mathbf{X}_1 = x_1 \,\&\cdots\&\, \mathbf{X}_n = x_n] > 0 \qquad \text{when } \lambda = \lambda^{(2)}.$$

It follows that for $\lambda = \lambda^{(1)}$,

$$\Psi[\mathbf{X}_i = x_i \mid \mathbf{X}_1 = x_1, ..., \mathbf{X}_{i-1} = x_{i-1}, \mathbf{X}_{i+1} = x_{i+1}, ..., \mathbf{X}_n = x_n]$$
$$= 0 \qquad \text{or } \textit{indef}, \qquad i = 1, ..., n,$$

while for $\lambda = \lambda^{(2)}$,

$$\Psi[\mathbf{X}_i = x_i \mid \mathbf{X}_1 = x_1, ..., \mathbf{X}_{i-1} = x_{i-1}, \mathbf{X}_{i+1} = x_{i+1}, ..., \mathbf{X}_n = x_n]$$
$$> 0, \qquad i = 1, ..., n.$$

Since *indef* is treated as non-removable singularity (Subsection 2.5), all the conditional probabilities above have different values at $\lambda^{(1)}$ and $\lambda^{(2)}$, because of which $\lambda \in \Lambda(\mathbf{X}_i)$ for $i = 1, ..., n$. Hence $\lambda \in \Lambda(\mathbf{X}_1) \cap \cdots \cap \Lambda(\mathbf{X}_n)$.

APPENDIX B: MULTIPLE-DIFFERENCE OPERATOR

The notion of the multiple-difference operator $\delta_{\{1,...,r\}}$ introduced in this appendix is utilized in the proofs of Theorem 2 in Appendix C and Theorem 3 in Appendix D. The reader who is willing to assume that the functions to which this operator is applied are sufficiently smooth may omit this appendix and treat $\delta_{\{1,...,r\}}$ in all proofs as the conventional derivative $\partial^r/\partial x_1 \cdots \partial x_r$.

Let A be a set of arguments of a function $F = F(A)$. Let certain values of all arguments constituting A be chosen, and let the chosen value of $x \in A$ be denoted by x^0. For any vector of arguments $x_1, ..., x_r$, belonging to A, the operator $\delta_{\{1,...,r\}} F(A)$ is defined as follows (by induction):

$$\text{(i)} \qquad \delta_{\{1\}} F(A) = F(x_1; A - \{x_1\}) - F(x_1^0; A - \{x_1\})$$

$$\text{(ii)} \quad \delta_{\{1,...,r\}} F(A) = \delta_{\{1\}} \delta_{\{1,...,r-1\}} F(A).$$

Thus,

$$\delta_{\{1,2\}} F(x_1, x_2, ...) = [F(x_1, x_2, ...) - F(x_1^0, x_2, ...)] - [F(x_1, x_2^0, ...) - F(x_1^0, x_2^0, ...)],$$

$$\delta_{\{1,2,3\}} F(x_1, x_2, x_3) = \{[F(x_1, x_2, x_3, ...) - F(x_1^0, x_2, x_3, ...)] - [F(x_1, x_2^0, x_3, ...) - F(x_1^0, x_2^0, x_3, ...)]\} - \{[F(x_1, x_2, x_3^0, ...) - F(x_1^0, x_2, x_3^0, ...)] - [F(x_1, x_2^0, x_3^0, ...) - F(x_1^0, x_2^0, x_3^0, ...)]\},$$

etc.

It can be easily verified by induction that $\delta_{\{1, ..., r\}} F(A)$ remains the same for all permutations of the indices $\{1, ..., r\}$, and that for any $1 \leqslant i \leqslant r$,

$$\delta_{\{1, ..., r\}} F(A) = \delta_{\{1, ..., i\}} \delta_{\{i+1, ..., r\}} F(A).$$

LEMMA 1. *If, for some* $\vartheta \subseteq \Xi$, $\delta_{\{1, ..., r\}} F(x_1, ..., x_r; \Xi)$ *does not depend on arguments from* $\Xi - \vartheta$, *then*

$$F(x_1, ..., x_r; \Xi) = F_0(x_1, ..., x_r; \vartheta) + \sum_{i=1, ..., r} F_i(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n; \Xi),$$

where

$$F_0(x_1, ..., x_r; \vartheta) = \delta_{\{1, ..., r\}} F(x_1, ..., x_r; \Xi)$$

Proof (by induction on r). For $r = 1$, the premise of the lemma can be written as

$$\delta_{\{1\}} F(x_1; \Xi) = F(x_1; \Xi) - F(x_1^0; \Xi) = F_0(x_1; \vartheta).$$

Denoting $F(x_1^0; \Xi)$ by $F_1(\Xi)$, we get

$$F(x_1; \Xi) = F_0(x_1; \vartheta) + F_1(\Xi),$$

which satisfies the lemma.

Assume that the lemma is satisfied for $r - 1 \geqslant 1$.

For r, the premise of the lemma can be written as

$$\delta_{\{1, ..., r\}} F(x_1, ..., x_r; \Xi) = F_0(x_1, ..., x_r; \vartheta).$$

Applying the induction hypothesis to

$$\delta_{\{1, ..., r\}} F(x_1, ..., x_r; \Xi) = \delta_{\{2, ..., r\}}[\delta_{\{1\}} F(x_1, ..., x_r; \Xi)],$$

we have

$$\delta_{\{1\}} F(x_1, ..., x_r; \Xi) = F_0(x_1, ..., x_r; \vartheta) + \sum_{i=2, ..., r} F_i(x_1, ..., x_{i-1}, x_{i+1}, ..., x_r; \Xi).$$

But

$$\delta_{\{1\}} F(x_1, ..., x_r; \Xi) = F(x_1, ..., x_r; \Xi) - F(x_1^0, ..., x_r; \Xi),$$

and we conclude the proof by combining the last two equations and renaming $F(x_1^0, ..., x_r; \Xi)$ into $F_1(x_2, ..., x_r; \Xi)$.

By trivial renaming of the indices and the corresponding x-arguments one can define the multiple-difference operator δ_I for an arbitrary set of indices I. This generalization is used in Appendix D. For completeness, the multiple-difference operator with the empty set of indices is defined as an identity operator,

$$\delta_{\varnothing} F(A) = F(A).$$

APPENDIX C: BIVARIATE CONTEXT SYSTEMS

The proof below utilizes the bivariate version, $\delta_{\{1,2\}}$, of the multiple-difference operator introduced in Appendix B.

THEOREM 2. *In a context system* $\{\mathbf{X}_1, \mathbf{X}_2; \Phi\}$, $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1 \subseteq \Phi$ *and* $\Lambda(\mathbf{X}_2) \subseteq \Lambda_2 \subseteq \Phi$ *if and only if*

$$\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2] = f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_2(x_1; \Lambda_1)\, f_2(x_2; \Lambda_2)\, f(\Phi), \tag{C1}$$

where f_{12} *is nonnegative with the positive domain* $\mathfrak{R}_{12}$, f_1 *and* f_2 *are positive on* $\mathfrak{R}_1$ *and* $\mathfrak{R}_2$, *respectively, and*

$$f(\Phi) = \left(\iint f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, f_2(x_2; \Lambda_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \right)^{-1}$$

is positive.

Proof. The sufficiency is proved by direct verification. The conditional density

$$\Psi[\mathbf{X}_1 = x_1 \mid \mathbf{X}_2 = x_2] = \frac{\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2]}{\Psi[\mathbf{X}_2 = x_2]}$$

equals

$$\frac{f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, f_2(x_2; \Lambda_2)\, f(\Phi)}{f_2(x_2; \Lambda_2)\, f(\Phi) \int f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, \mathrm{d}x_1}$$

$$= \frac{f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)}{\int f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, \mathrm{d}x_1}.$$

Indeed, if $f_2(x_2; \Lambda_2) > 0$, then it is eliminated algebraically, whereas if $f_2(x_2; \Lambda_2) = 0$, then $x_2 \notin \Re_2$, $(x_1, x_2) \notin \Re_{12}$ for all x_1, and both expressions above are *indef* (under the convention on 0/0 in Subsection 2.5). Since the right-hand expression does not depend on factors outside Λ_1, one concludes that $\Lambda(\mathbf{X}_1) \subseteq \Lambda_1$. Analogously, $\Lambda(\mathbf{X}_2) \subseteq \Lambda_2$.

To prove the necessity, define the following functions on the support $\Re_{12}$:

$$\log(\Psi[\mathbf{X}_1 = x_1]) = u_1(x_1; \Phi)$$

$$\log(\Psi[\mathbf{X}_2 = x_2]) = u_2(x_2; \Phi)$$

$$\log(\Psi[\mathbf{X}_1 = x_1 \mid \mathbf{X}_2 = x_2]) = u_{1\mid 2}(x_1, x_2; \Lambda_1)$$

$$\log(\Psi[\mathbf{X}_2 = x_2 \mid \mathbf{X}_1 = x_1]) = u_{2\mid 1}(x_1, x_2; \Lambda_2).$$

Then $\log(\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2])$ can be decomposed as

$$u_1(x_1; \Phi) + u_{2\mid 1}(x_1, x_2; \Lambda_2) = u_2(x_2; \Phi) + u_{1\mid 2}(x_1, x_2; \Lambda_1). \tag{C2}$$

Applying $\delta_{\{1,2,\}}$ to these two decompositions, we have

$$\delta_{\{1,2\}} u_{1\mid 2}(x_1, x_2; \Lambda_1) = \delta_{\{1,2\}} u_{2\mid 1}(x_1, x_2; \Lambda_2).$$

Clearly, these two expressions equal some function of x_1, x_2 and $\Lambda_1 \cap \Lambda_2$. By Lemma 1 (Appendix B), this means that

$$u_{1\mid 2}(x_1, x_2; \Lambda_1) = \xi_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2) + a_1(x_1; \Lambda_1) + a_2(x_2; \Lambda_1)$$

$$u_{2\mid 1}(x_1, x_2; \Lambda_2) = \xi_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2) + b_1(x_1; \Lambda_2) + b_2(x_2; \Lambda_2).$$

Substituting these representations in (C2), we see that $\log(\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2])$ equals

$$\begin{aligned} &u_1(x_1; \Phi) + \xi_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2) + b_1(x_1; \Lambda_2) + b_2(x_2; \Lambda_2) \\ &\quad = u_2(x_2; \Phi) + \xi_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2) + a_1(x_1; \Lambda_1) + a_2(x_2; \Lambda_1). \end{aligned}$$

Renaming, by the absorption principle, $u_1(x_1; \Phi) + b_1(x_1; \Lambda_2)$ into $u_1(x_1; \Phi)$ and $u_2(x_2; \Phi) + a_2(x_2; \Lambda_1)$ into $u_2(x_2; \Phi)$, we have

$$\begin{aligned} \log(\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2]) &= \xi_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2) + u_1(x_1; \Phi) + b_2(x_2; \Lambda_2) \\ &= \xi_{12}(x_1, x_2; \Lambda \cap \Lambda_2) + a_1(x_1; \Lambda_1) + u_2(x_2; \Phi). \end{aligned} \tag{C3}$$

It follows that

$$u_2(x_2; \Phi) - b_2(x_2; \Lambda_2) = u_1(x_1; \Phi) - a_1(x_1; \Lambda_1),$$

which can only be the case if the functions involved are decomposable as

$$a_1(x_1; \Lambda_1) = \xi_1(x_1; \Lambda_1) + \eta_1(\Lambda_1)$$

$$u_1(x_1; \Phi) = \xi_1(x_1; \Lambda_1) + \upsilon_1(\Phi)$$

$$b_2(x_2; \Lambda_2) = \xi_2(x_2; \Lambda_2) + \eta_2(\Lambda_2)$$

$$u_2(x_2; \Phi) = \xi_2(x_2; \Lambda_2) + \upsilon_2(\Phi).$$

Substituting the expressions for a_1 and u_2 in the right-hand side of (C3) (or b_2 and u_1 in its left-hand side), we get, after renaming $\eta_1(\Lambda_1) + \upsilon_2(\Phi)$ into $\xi(\Phi)$,

$$\begin{aligned}&\log(\Psi[\mathbf{X}_1 = x_1 \,\&\, \mathbf{X}_2 = x_2])\\ &\quad = \xi_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2) + \xi_1(x_1; \Lambda_1) + \xi_2(x_2; \Lambda_2) + \xi(\Phi).\end{aligned}$$

Representation (C1) is now obtained by exponentiating the equality above and observing that $f_{12} \equiv \exp(\xi_{12})$, $f_1 \equiv \exp(\xi_1)$, $f_2 \equiv \exp(\xi_2)$ are all positive (on areas $\mathfrak{R}_{12}$, $\mathfrak{R}_1$, and $\mathfrak{R}_2$, respectively), and that $f(\Phi) = \exp[\xi(\Phi)]$ is also positive and must be reciprocal to

$$\iint f(x_1, x_2; \Lambda_1 \cap \Lambda_2)\, f_1(x_1; \Lambda_1)\, f_2(x_2; \Lambda_2)\, \mathrm{d}x_1\, \mathrm{d}x_2,$$

if the integration is carried out over $\mathfrak{R}_{12}$. This restriction on the integration area can be removed by making the function $f_{12}(x_1, x_2; \Lambda_1 \cap \Lambda_2)$ vanish outside $\mathfrak{R}_{12}$.

APPENDIX D: MULTIVARIATE CONTEXT SYSTEMS

The proofs below utilize the multiple-difference operator of Appendix B and the index subset symbolism introduced in Subsection 2.8. In addition, we need the following notation agreements.

Given an index subset $I \subseteq \{1, ..., n\}$, $N(I)$ denotes the number of the indices in I: $N(I) = 0, 1, ..., n$.

Given two index subsets, I and J, a function with arguments

$$\{x_i\}_{i \in I}; \bigcap_{i \in J} \Lambda_i$$

is abbreviated as $f_{I;\,J}$ (or another symbol with these subscripts). In particular, the abbreviation $f_{I;\,I}$ is used in this appendix for the functions

$$f_I\left(\{x_i\}_{i \in I}; \bigcap_{i \in I} \Lambda_i\right), \qquad I \subseteq \{1, ..., n\},$$

that play a central role in this work. As explained in Subsection 2.8,

$$\{x_i\}_{i \in \varnothing} = \varnothing, \qquad \bigcap_{i \in \varnothing} \Lambda_i = \Phi,$$

because of which

$$f_{\varnothing; J} = f_{\varnothing; J}\left(\bigcap_{i \in J} \Lambda_i\right), \qquad f_{I; \varnothing} = f_{I; \varnothing}(\{x_i\}_{i \in I}; \Phi), \qquad f_{\varnothing; \varnothing} = f_{\varnothing; \varnothing}(\Phi).$$

Sometimes we need superscripts to distinguish different functions of the same type (i.e., with the same subscripts and arguments):

$$f^{(1)}_{I; J}, f^{(2)}_{I; J}, ..., f^{(i)}_{I; J}, ...$$

The proof of Theorem 3 is based on the following

LEMMA 2. *For $n \geqslant 2$, $1 \leqslant r \leqslant n$, if a function $h(x_1, ..., x_n; \Phi)$ can be represented by n identities*

$$h(x_1, ..., x_n; \Phi) = \sum_{N(I) \leqslant r} \phi^{(i)}_{I; I - \{i\}}, \qquad i = 1, ..., n, \tag{D1}$$

then it can also be represented by n identities

$$h(x_1, ..., x_n; \Phi) = \sum_{N(I) = r} \phi_{I; I} + \sum_{N(I) \leqslant r-1} \varphi^{(i)}_{I; I - \{i\}}, \qquad i = 1, ..., n. \tag{D2}$$

Remark on Notation. The superscript in $\phi^{(i)}_{I; I-\{i\}}$ indicates that for $i \neq j$ the functions $\phi^{(i)}_{I; I-\{i\}}$ and $\phi^{(j)}_{I; I-\{j\}}$ are generally different, even if they have the same subscripts and therefore the same arguments (which is the case, for instance, if I does not include either i or j). By contrast, $\phi_{I; I}$ in (D2) are the same for $i = 1, ..., n$.

Proof. Consider an arbitrary index subset I with $N(I) = r$. Applying the operator δ_I to (D1) we observe that

$$\delta_I h(x_1, ..., x_n; \Phi) = \delta_I \phi^{(i)}_{I; I-\{i\}}, \qquad i = 1, ..., n.$$

The set of arguments for $\delta_I \phi^{(i)}_{I; I-\{i\}}$ is the same as for $\phi^{(i)}_{I; I-\{i\}}$,

$$\{x_j\}_{j \in I}; \bigcap_{j \in I - \{i\}} \Lambda_j,$$

and the intersection of all such sets, for $i = 1, ..., n$, is

$$\{x_j\}_{j \in I}; \bigcap_{j \in I} \Lambda_j.$$

Clearly, all $\delta_I \phi_{I;\, I-\{i\}}$, for $i=1, ..., n$, equal one and the same function of this intersection of the argument sets, and this function has the structure of $\phi_{I;\, I}$. Due to Lemma 1 (Appendix B), we have

$$\phi^{(i)}_{I;\, I-\{i\}} = \phi_{I;\, I} + \sum_{k \in I} \phi^{(i)}_{I-\{k\};\, I-\{i\}}, \qquad i=1, ..., n,$$

and (D1) can be rewritten as

$$h(x_1, ..., x_n; \Phi) = \sum_{N(I)=r} \phi_{I;\, I} + \sum_{N(I)=r} \sum_{k \in I} \phi^{(i)}_{I-\{k\};\, I-\{i\}} + \sum_{N(I) \leqslant r-1} \phi^{(i)}_{I;\, I-\{i\}},$$

for $i=1, ..., n$. Observe now that to any J with $N(J)=r$ there corresponds a function $\phi^{(i)}_{J-\{k\};\, J-\{i\}}$ in the middle double-sum and a function $\phi^{(i)}_{J-\{k\};\, J-\{k,\, i\}}$ in the rightmost sum: they have the same set of x-arguments but the latter generally has a larger set of factors:

$$\bigcap_{j \in I-\{i\}} \Lambda_j \subseteq \bigcap_{j \in I-\{k,\, i\}} \Lambda_j.$$

By the absorption principle,

$$\phi^{(i)}_{J-\{k\};\, J-\{i\}} + \phi^{(i)}_{J-\{k\};\, J-\{k,\, i\}} = \varphi^{(i)}_{J-\{k\};\, J-\{k,\, i\}}.$$

As the resulting function has the structure of $\varphi^{(i)}_{I;\, I-\{i\}}$ with $N(I)=r-1$, (D1) transforms into

$$h(x_1, ..., x_n; \Phi) = \sum_{N(I)=r} \phi_{I;\, I} + \sum_{N(I)=r-1} \varphi^{(i)}_{I;\, I-\{i\}} + \sum_{N(I)<r-1} \phi^{(i)}_{I;\, I-\{i\}},$$
$$i=1, ..., n.$$

The statement of the lemma is now obtained by renaming the ϕ-functions in the rightmost sum into φ-functions and combining this sum with the middle one.

THEOREM 3. *In a context system* $\{\mathbf{X}_1, ..., \mathbf{X}_n; \Phi\}$, $\&_{i=1, ..., n}\, \Lambda(\mathbf{X}_i) \subseteq \Lambda_i \subseteq \Phi$ *if and only if*

$$\Psi\left[\mathop{\&}_{i \in \{1, ..., n\}} \mathbf{X}_i = x_i\right] = \prod_{I \subseteq \{1, ..., n\}} f_{I;\, I}, \tag{D3}$$

where $f_{\{1, ..., n\};\, \{1, ..., n\}}$ *is nonnegative with the positive domain* $\mathfrak{R}_{1 \dots n}$, $f_{\mathbf{I};\, \mathbf{I}}$ *is positive on* $\mathfrak{R}_I$ *for all proper index subsets I, and*

$$f_{\varnothing;\, \varnothing} = \left(\int \cdots \int \prod_{\varnothing \neq I \subseteq \{1, ..., n\}} f_{I;\, I}\, \mathrm{d}x_1 \cdots \mathrm{d}x_n\right)^{-1}$$

is positive.

Proof. Sufficiency is proved by direct verification, as in Theorem 2.

Necessity is proved by induction with respect to the number of random variables, n. Theorem 2 provides the induction basis ($n=2$). Assume that representation (D3) holds for any context system with $n-1 \geqslant 2$ random variables. Then, due to the observation on conditional context systems made in Subsection 2.1, this representation holds for n conditional density functions,

$$\Psi\Big[\mathop{\&}_{j \in \{1, ..., n\} - \{i\}} \mathbf{X}_j = x_j \mid \mathbf{X}_i = x_i\Big], \qquad i = 1, ..., n.$$

Namely,

$$\Psi\Big[\mathop{\&}_{j \in \{1, ..., n\} - \{i\}} \mathbf{X}_j = x_j \mid \mathbf{X}_i = x_i\Big] = \prod_{I \subseteq \{1, ..., n\} - \{i\}} v^{(i)}_{I; I}, \qquad i = 1, ..., n,$$

and consequently,

$$\Psi\Big[\mathop{\&}_{j \in \{1, ..., n\}} \mathbf{X}_j = x_j\Big] = \Psi[\mathbf{X}_i = x_i] \prod_{I \subseteq \{1, ..., n\} - \{i\}} v^{(i)}_{I; I}, \qquad i = 1, ..., n.$$

Every function in the right-hand side of this expression generally depends on x_i, because of which we can write

$$v^{(i)}_{I; I} = w^{(i)}_{I \cup \{i\}; I}$$

and

$$\Psi[\mathbf{X}_i = x_i] = w^{(i)}(x_i, \Phi) = w^{(i)}_{\varnothing \cup \{i\}; \varnothing}.$$

It follows that

$$\Psi\Big[\mathop{\&}_{j \in \{1, ..., n\}} \mathbf{X}_j = x_j\Big] = \prod_{I \subseteq \{1, ..., n\} - \{i\}} w^{(i)}_{I \cup \{i\}; I}, \qquad i = 1, ..., n.$$

On the support $\mathfrak{R}_{1 \ldots n}$ one can take logarithms of both sides:

$$h(x_1, ..., x_n; \Phi) = \sum_{I \subseteq \{1, ..., n\} - \{i\}} \phi_{I \cup \{i\}; I}, \qquad i = 1, ..., n.$$

It is easy to see that

$$\sum_{I \subseteq \{1, ..., n\} - \{i\}} \phi_{I \cup \{i\}; I} = \sum_{I \subseteq \{1, ..., n\} \,\&\, i \in I} \phi_{I; I - \{i\}}.$$

The summation area $I \subseteq \{1, ..., n\}$ & $i \in I$ can always be formally extended to $I \subseteq \{1, ..., n\}$ by, say, adding functions identically equal to zero. The n additive decompositions of $h(x_1, ..., x_n; \Phi)$, therefore, can be presented as

$$h(x_1, ..., x_n; \Phi) = \sum_{N(I) \leqslant n} \phi^{(i)}_{I; I - \{i\}}, \qquad i = 1, ..., n.$$

Applying Lemma 2 (with $r = n$) to this expression, we get

$$h(x_1, ..., x_n; \Phi) - \phi_{\{1, ..., n\}; \{1, ..., n\}} = \sum_{N(I) \leqslant n-1} \varphi^{(i)}_{I; I - \{i\}}, \qquad i = 1, ..., n.$$

Applying Lemma 2 once again (with $r = n - 1$, treating the left-hand side as a single function of $x_1, ..., x_n$ and Φ), we get

$$h(x_1, ..., x_n; \Phi) - \phi_{\{1, ..., n\}; \{1, ..., n\}} - \sum_{N(I) = n-1} \phi_{I; I} = \sum_{N(I) \leqslant n-2} \xi^{(i)}_{I; I - \{i\}},$$
$$i = 1, ..., n.$$

Continuing to apply Lemma 2 in this fashion (with $r = n - 2, ..., 1$), we eventually come to

$$h(x_1, ..., x_n; \Phi) - \sum_{1 \leqslant N(I) \leqslant n} \phi_{I; I} = \sum_{N(I) = 0} \zeta^{(i)}_{I; I - \{i\}}, \qquad i = 1, ..., n.$$

Clearly,

$$\sum_{N(I) = 0} \zeta^{(i)}_{I; I - \{i\}} = \zeta^{(i)}_{\varnothing; \varnothing} = \phi_{\varnothing; \varnothing}, \qquad i = 1, ..., n,$$

and therefore,

$$h(x_1, ..., x_n; \Phi) = \sum_{N(I) \leqslant n} \phi_{I; I}.$$

Representation (D3) is now obtained by exponentiation and observations analogous to those ending the proof of Theorem 2.

REFERENCES

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. (1996). *A canonical representation for selectively influenced processes and component times*. Paper presented at the 29th Meeting of the Society for Mathematical Psychology.

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. New York: Erlbaum.

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., & Schweickert, R. (1995). Decompositions of response times: An almost general theory. *Journal of Mathematical Psychology*, **39**, 285–314.

Johnson, N. L., & Kotz, S. (1976). *Distributions in statistics: Continuous multivariate distributions*. New York: Wiley.

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.

Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, **28**, 363–400.

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.

Yermalov, S. M. (1971). *Method of Monte-Carlo and related issues*. Moscow: Nauka. [In Russian]

Received: July 1, 1997