Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Mathematical Psychology 50 (2006) 390–401
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

Notes on selective influence, probabilistic causality, and probabilistic dimensionality

Ehtibar N. Dzhafarov[a,*], Ilya Gluhovsky[b]

[a]*Purdue University, USA*
[b]*Sun Microsystems Laboratories, USA*

Received 23 April 2005; received in revised form 8 March 2006
Available online 2 May 2006

Abstract

The paper provides conceptual clarifications for the issues related to the dependence of jointly distributed systems of random entities on external factors. This includes the theory of selective influence as proposed in Dzhafarov [(2003a). Selective influence through conditional independence. *Psychometrika*, *68*, 7–26] and generalized versions of the notions of probabilistic causality [Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, *48*, 191–199] and dimensionality in the latent variable models [Levine, M. V. (2003). Dimension in latent variable models. *Journal of Mathematical Psychology*, *47*, 450–466]. One of the basic observations is that any system of random entities whose joint distribution depends on a factor set can be represented by functions of two arguments: a single factor-independent source of randomness and the factor set itself. In the case of random variables (i.e., real-valued random entities endowed with Borel sigma-algebras) the single source of randomness can be chosen to be any random variable with a continuous distribution (e.g., uniformly distributed between 0 and 1).

Keywords: Selective influence; Probabilistic causality; Probabilistic dimensionality; Latent variable models; Probability spaces; Random variables; Joint distributions; Conditional independence

1. Informal introduction

When dealing with an issue it occasionally proves useful to "get back to basics," to relate the issue to the fundamental notions of the area to which it belongs. This seems especially true when dealing with problems whose formulations involve random variation, where one's intuitions are notoriously faulty. The aim of this paper is to improve the clarity of and further develop the concept of *selective influence* (as defined in Dzhafarov, 2003a) and, as its special case, the notion of *probabilistic causality* (introduced in Suppes & Zanotti, 1981), by relating them to the basics of Kolmogorov's probability theory. The analysis also places within the context of selective influence and sheds light on the notion of *probabilistic dimensionality*, derived from Levine (2003).

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

0022-2496/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2006.03.003

1.1. Selective influence

The history (dating from Townsend, 1984) and various approaches to the notion of selective influence (Dzhafarov, 1999, 2001; Townsend & Thomas, 1994) are discussed in Dzhafarov (2003a).

To give a simple example of selective influence, let (X_1, X_2, X_3) be random variables representing scores in three performance tests, generally stochastically interdependent. Let (p_1, p_2, p_3, p_4) be four external factors (observable conditions or covariates, say, test duration, prior training level, sex of the examinee, and age of the examinee). The factors are considered deterministic variables: this means that the joint distribution of (X_1, X_2, X_3) is always being conditioned on specific values of (p_1, p_2, p_3, p_4). Equivalently, (X_1, X_2, X_3) is viewed as a family of random vectors indexed by values of (p_1, p_2, p_3, p_4). Consider the following conjunction of statements: X_1 is selectively influenced by p_1 (i.e., it is not influenced by p_2, p_3, or p_4), X_2 is selectively influenced by

p_2 and p_3 (not influenced by p_1 or p_4), and X_3 is selectively influenced by p_4 (not by p_1, p_2, p_3). In Dzhafarov's (2003a) notation this is presented as

$$(X_1, X_2, X_3) \hookleftarrow (\{p_1\}, \{p_2, p_3\}, \{p_4\}).$$

What is the meaning of stating this selective dependence if X_1, X_2, X_3 are not necessarily stochastically independent? According to Dzhafarov (2003a), the meaning is that X_1, X_2, X_3 are representable (not necessarily uniquely) in the form

$$\begin{bmatrix} X_1 = f_1(C, S_1, p_1) \\ X_2 = f_2(C, S_2, p_2, p_3) \\ X_3 = f_3(C, S_3, p_4) \end{bmatrix}, \tag{1}$$

where f_1, f_2, f_3 are measurable functions, and C, S_1, S_2, S_3 are mutually independent "*sources of randomness*" whose distributions do not depend on any of the factors p_1, p_2, p_3, p_4. C is interpreted as a *common source of randomness*, its implied role is to "explain" the stochastic interdependence of (X_1, X_2, X_3). S_i is interpreted as a *specific source of randomness* for X_i ($i = 1, 2, 3$). Given any value c of C, the variables $f_1(c, S_1, p_1)$, $f_2(c, S_2, p_2, p_3)$, $f_3(c, S_3, p_4)$ (i.e., X_1, X_2, X_3 conditioned upon $C = c$) are stochastically independent, each being influenced by "its own" set of external factors.[1]

The sphere of behavioral applications of the notion of selective influence is not, of course, confined to performance scores. Thus, (X_1, X_2, X_3) in the above example could be durations of three processes in a network of mental operations aimed at solving a certain task (e.g., deciding whether a letter being currently presented was or was not within a previously memorized set of letters). Then factors (p_1, p_2, p_3, p_4) would designate experimental manipulations that prolong or shorten these durations (in the sense of monotonically transforming their distribution functions). The necessity to speak of selective influence in this context was, in fact, the reason this notion was historically introduced (Townsend, 1984). The durations (X_1, X_2, X_3) here are unobservable variables, and their very identity is predicated on selective influence: within the hypothetical mental architecture X_1 may be *defined* as the duration of the process which is selectively influenced by p_1, X_2 as the duration of the process which is selectively influenced by p_2 and p_3, etc.

Representations like (1) were used in Dzhafarov et al. (2004) to establish certain properties of mental "parallel–serial" architectures with stochastically interdependent components.

[1] Formally, as explained in Dzhafarov (2003a), this approach to selective influence can be viewed as generalizing the combined use of (nonlinear) regression and factor analyses (with the term *external factor*, however, corresponding to the traditional *regressor*, and the term *source of randomness* to the traditional (*unobservable*) *factor*). In view of the results presented in Section 5, however, this generalization does not lend itself to extensions of the traditional data-analytic techniques, because in the general setting the idea of "interpreting" the sources of variability loses its meaning.

Another, perhaps even more basic application of selective influence is found in the context of modeling comparative judgments, say, deciding whether two stimuli being presented, p_1, p_2, are the same or different. (X_1, X_2) in this case are interpreted as perceptual images of these two stimuli, presumably randomly varying in some perceptual space. The notion of selective influence here,

$$(X_1, X_2) \hookleftarrow (\{p_1\}, \{p_2\}),$$

is needed to express the very fact that X_1 is the image of p_1 (and not of p_2) while X_2 is the image of p_2 (and not of p_1), even though X_1 and X_2 are generally stochastically interdependent. By analogy with (1), this statement is taken to mean the representability of (X_1, X_2) in the form

$$\begin{bmatrix} X_1 = f_1(C, S_1, p_1) \\ X_2 = f_2(C, S_2, p_2) \end{bmatrix}, \tag{2}$$

where f_1, f_2 are measurable functions, while C, S_1, S_2 are mutually independent and do not depend on stimuli p_1, p_2.

As a simple example, let p_1, p_2 be representable by real numbers (e.g., two line segments being compared can be represented by their lengths, in some units), and let (X_1, X_2) be bivariate normally distributed (as it is frequently assumed in models derived from Thurstone, 1927), with respective means $(p_1, 1)$, respective variances $(1 + p_1^2, 1 + p_2^2)$, and covariance p_1. Then, as one can easily check,

$$\begin{bmatrix} X_1 = S_1 + p_1(C + 1) \\ X_2 = p_2 S_2 + C + 1 \end{bmatrix},$$

where C, S_1, S_2 are mutually independent standard normally distributed variables. In this case, therefore, $(X_1, X_2) \hookleftarrow (\{p_1\}, \{p_2\})$, and the meaning of saying that X_1 is the image of p_1 while X_2 is the image of p_2 is well defined. In other cases representability of the form (2) may not be achievable, even if the marginal distribution of X_1 depends only on p_1 and the marginal distribution of X_2 depends only on p_2 (see an example in Dzhafarov, 2003a).

The notion of selective influence and representations (2) in the context of selectively relating stimuli to their perceptual images were used in Dzhafarov (2003c) to investigate the feasibility of Thurstonian-type models for same–different judgments (a nontechnical description of this issue can be found in Dzhafarov & Colonius, 2006).

Returning to (1), this example can be modified to include the cases when the sets of factors selectively influencing (X_1, X_2, X_3) may overlap, coincide, or include empty sets. Thus, the conjunction of statements "X_1 is selectively influenced by p_1 and p_2, X_2 is selectively influenced by p_2, p_3, and p_4, and X_3 is not influenced by any of the factors" translates into the representability of (X_1, X_2, X_3) in the form

$$\begin{bmatrix} X_1 = f_1(C, S_1, p_1, p_2) \\ X_2 = f_2(C, S_2, p_2, p_3, p_4) \\ X_3 = f_3(C, S_3) \end{bmatrix},$$

which can be schematically presented as

$(X_1, X_2, X_3) \looparrowleft (\{p_1, p_2\}, \{p_2, p_3, p_4\}, \emptyset).$

The reason cases like this should be viewed as special cases of selective influence is that it would be inconvenient, artificial, and unnecessary to isolate them. The mathematical theory presented in Dzhafarov (2003a) makes no use of the constraint that the factor sets are nonempty and disjoint: given a set ϕ of factors known to influence the joint distribution of (X_1, X_2, X_3), the treatment of the selective influence relationship

$(X_1, X_2, X_3) \looparrowleft (\phi_1, \phi_2, \phi_3)$

is the same for any three subsets ϕ_1, ϕ_2, ϕ_3 of ϕ. Thus, with $\phi = \{p_1, p_2, p_3, p_4\}$, the treatment of such relations of selective influence as

$(X_1, X_2, X_3) \looparrowleft (\{p_1, p_2\}, \{p_1, p_2\}, \{p_2, p_3, p_4\})$

or

$(X_1, X_2, X_3) \looparrowleft (\{p_1, p_2, p_3, p_4\}, \emptyset, \emptyset)$

or even

$(X_1, X_2, X_3) \looparrowleft (\{p_1, p_2, p_3, p_4\}, \{p_1, p_2, p_3, p_4\}, \{p_1, p_2, p_3, p_4\})$

is no different from that of (1). The notion of selectivity in the latter example loses its intuitive meaning, but this special case is important as it links the notion of selective influence with that of probabilistic causality, discussed next.

1.2. Probabilistic causality

The term *probabilistic causality* has a variety of usages. Ours is based on the formulation given in Suppes and Zanotti (1981), as pertaining to seeking a "probabilistic cause" (in a technical meaning clarified below) for the stochastic interdependence among $X_1, \ldots, X_n$. We generalize, however, Suppes and Zanotti's meaning by including in the issue the dependence of $X_1, \ldots, X_n$ upon a set of external factors ϕ. For simplicity, we continue here to use the example with three random variables (X_1, X_2, X_3).

It follows from a representation like (1) that fixing the common source of randomness C at any value c will make (X_1, X_2, X_3) *conditionally independent*, with the conditional distribution of X_1 being dependent only on factor p_1, the conditional distribution of X_2 only on factors p_2 and p_3, etc. As shown in Dzhafarov (2003a), the possibility of finding a random entity C with this property is *equivalent* to the representability of (X_1, X_2, X_3) in the form (1). Note that specific sources of randomness (S_1, S_2, S_3) in functions (f_1, f_2, f_3) may very well be dummy (removable) arguments. In this case fixing C at any value c will make (X_1, X_2, X_3) *conditionally deterministic* (with the conditional value of X_i being selectively dependent on the corresponding factor set ϕ_i, $i = 1, 2, 3$). Deterministic entities are formally stochastically independent.

Suppes and Zanotti (1981) define the issue of probabilistic causality as that of finding a random entity C such that fixing it at any value c would make (X_1, X_2, X_3) conditionally independent (possibly, conditionally deterministic). Probabilistic causality therefore is a purely technical term, with no "what causes what" interpretations implied. Since Suppes and Zanotti do not consider external factors, their notion of probabilistic causality can be viewed as a degenerate case of selective influence, formally obtained by putting $\phi = \emptyset$ (i.e., the joint distribution of X_1, X_2, X_3 does not depend on any factors) and $\phi_1 = \phi_2 = \phi_3 = \emptyset$. Then (1) is replaced with the representability of (X_1, X_2, X_3) in the form

$$\begin{bmatrix} X_1 = f_1(C, S_1) \\ X_2 = f_2(C, S_2) \\ X_3 = f_3(C, S_3) \end{bmatrix}. \tag{3}$$

This case being too restrictive, we extend the notion of probabilistic causality to include all cases where $\phi_1 = \phi_2 = \phi_3 = \phi$, that is, where (X_1, X_2, X_3) jointly depend on some factor set ϕ,

$(X_1, X_2, X_3) \looparrowleft (\phi, \phi, \phi),$

or equivalently,

$$\begin{bmatrix} X_1 = f_1(C, S_1, \phi) \\ X_2 = f_2(C, S_2, \phi) \\ X_3 = f_3(C, S_3, \phi) \end{bmatrix}. \tag{4}$$

Since relation $(X_1, X_2, X_3) \looparrowleft (\phi_1, \phi_2, \phi_3)$ obviously implies $(X_1, X_2, X_3) \looparrowleft (\phi, \phi, \phi)$ with ϕ denoting the union $\phi_1 \cup \phi_2 \cup \phi_3$ of the factor sets, probabilistic causality (in our generalized sense) may be viewed as pertaining to the dependence of (X_1, X_2, X_3) on a factor set ϕ when the internal structure of this factor set is of no interest. The relation of selective influence, such as (1), then is obtained from (4) by eliminating from each of the functions f_i the components of ϕ that do not affect its value.

1.3. Probabilistic dimensionality

Levine (2003) considers a vector of dichotomous or polytomous items in an aptitude test, say (X_1, X_2, X_3), under the assumption that there is a common source of randomness C, such that given any its value c, (X_1, X_2, X_3) are conditionally independent (a common assumption in the context of aptitude analysis). Levine, in addition, assumes that C has a special form: it is a k-dimensional vector of random variables (interpreted as "k-component aptitude"). Let us call the smallest such k the *probabilistic dimensionality* of (X_1, X_2, X_3). Levine's main result is that if vector (X_1, X_2, X_3) has a probabilistic dimensionality k, then $k = 1$. In other words, if a k-dimensional C with the above property exists, for some $k \geqslant 1$, then one can always find a single (one-dimensional) random variable C^* with the same property. The result, of course, holds for an arbitrary number of random variables, $\{X_1, \ldots, X_n\}$.

In our conceptual setup, the assumption that vector (X_1, X_2, X_3) possesses a probabilistic dimensionality is equivalent to positing a representation of the form (3) for (X_1, X_2, X_3), with C being a k-dimensional vector of random variables. Then, as we show below, Levine's main result (that one can always choose C with $k = 1$) is an immediate consequence of one of the basic theorems in the theory of *standard Borel spaces* (e.g., Kechris, 1995). This holds true for an arbitrary number of arbitrary random variables $X_1, \ldots, X_n$, not necessarily dichotomous or polytomous. Moreover, the result also applies to $\{X_1, \ldots, X_n\}$ whose joint distribution depends on external factors ϕ: if representation (4) holds with C being a k-dimensional vector of random variables, then such a representation also holds with C being a single random variable.[2]

1.4. Plan of the paper

In Section 2, we explain the distinction between *random variables* in the Kolmogorov's (1933) sense and the more general concept of *random entities*. This section also introduces the notation and terminological conventions used throughout this paper. In particular, we explain the meaning of speaking of a system of random entities whose joint distribution depends on external factors.

In Section 3, we present a simple lemma that shows that any such system of random entities can be defined on a probability space with a probability measure that does not depend on external factors. In this section, we also introduce the notion of multiple probability spaces corresponding to *stochastically unrelated* random entities, and we clarify logical distinctions associated with saying that a random entity can be presented as a function of another random entity.

In Section 4, we introduce the most general version of the notion of selective influence and provide a rigorous definition of selective influence with sources of variability classifiable into specific sources and a common one.

In Section 5, we deal with finite systems of random variables depending on factors, and show under what conditions they can be defined on a probability space with a single random variable serving as their source of randomness.

[2]This fact may seem to contradict Levine's own conclusion, based on what he calls "a tentative, interim characterization" of embedding unidimensional models into multidimensional ones. He states that $k = 1$ may not be obtainable when $X_1, \ldots, X_n$ depend on a parameter, such as time ("Multidimensionality is not needed to account for one distribution. Multidimensionality is only needed to account for orderly changes of observed distributions," Levine, 2003, p. 465). The discrepancy is only apparent, however: Levine's "tentative characterization" is very different from representations of the form (4). For a detailed comparison of Levine's treatment with ours (which is not made in this paper) one should also take into account that values of C in Levine's analysis are related to probability distributions rather than random variables per se, and these relations are constrained by continuity and smoothness requirements which do not belong in our treatment.

2. Terminology, notation, and basic notions

2.1. Random entities and random variables

Following the original definition by Kolmogorov (1933), we reserve the familiar term *random variable* to designate functions with values in the set of reals endowed with the usual (Borel) sigma-algebra. Functions with values in probability spaces of more general nature we term *random entities*. Random variables are random entities, and so are random processes, random fields, product spaces involving random processes and random fields, random functionals and operators on a space of functions, etc. The terminology not being firmly established in the mathematical literature, other authors use for this purpose such terms as *random element* (Fréchet, 1948) and *random quantity* (Blank-Lapierre & Fortet, 1967). Nor is it unusual to use the term "random variable" in the generalized sense, coinciding with our usage of "random entity." This, however, would be inconvenient for our purposes, as some of the results presented below are general, while others are formulated for (real valued) random variables and vectors thereof only.

When dealing with observable random outcomes of an experiment, they can usually be represented by random variables. In theoretical considerations, however, we often deal with hypothetical random outcomes or sources of variability and interdependence whose nature is not known and cannot be constrained by theoretical considerations. As an example, Dzhafarov (2003b, 2003c) analyzes in the context of same–different discriminations the hypothesis that perceptual images selectively attributable to stimuli can be viewed as randomly varying, and the decision as to whether two stimuli are different is based on realizations of their randomly varying images in a given trial. As we do not know the "correct" way of describing a perceptual image (a finite or countably infinite number of real-valued attributes? a real-valued function with its domain in $\mathbb{R}^n$? a subset of a set of functions?, etc.), the value of such analysis is comeasurable with its generality. With reference to (2), perceptual images (X_1, X_2) and the sources of variability C, S_1, S_2 in Dzhafarov (2003b, 2003c) are viewed as random entities identified on probability spaces of arbitrary nature. It is proved then that the hypothesis in question cannot account for certain observable properties of discrimination probabilities if the dependence of (X_1, X_2) on C, S_1, S_2 possesses a property called *well-behavedness*, defined in terms of random entities of unspecified nature. The import of this conclusion would be greatly diminished if the perceptual images, their sources of variability, or the notion of well-behavedness were confined to random entities of a specific type only (say, random functions).

This does not mean, of course, that sources of randomness cannot sometimes be proved to be random entities of a particular kind, or constrained to be of a particular kind by hypothesis. Thus, in an example given earlier, if (X_1, X_2, X_3) in (4) are random variables (which they are, if interpreted as three performance scores or durations of three processes in a mental architecture), then according to our Theorem 1, C can be chosen to be a random variable uniformly varying between 0 and 1. In Levine's (2003) work as well as in our Theorem 2, C is constrained by hypothesis to being a vector of random variables.

2.2. Conventions

We denote random entities (including random variables) by capital italics $(B, X, \ldots)$ and their values by the corresponding lowercase italics $(b, x, \ldots)$. We use open letters $(\mathbb{B}, \mathbb{X}, \ldots)$ to denote sets of values for random variables, and script letters $(\mathscr{B}, \mathscr{X}, \ldots)$ to denote sigma-algebras defined on such sets. $\mathbb{R}$ stands for the set of reals, $\mathscr{R}$ for the usual (Borel) sigma-algebra on $\mathbb{R}$.

If the joint distribution of $\{X_1, \ldots, X_n\}$ (or, more generally, $\{X_\lambda\}_{\lambda\in\Lambda}$, with an arbitrary indexing set Λ) depends on some external factors, the set of these factors is denoted by ϕ, and the set of values of this factor set is denoted by Φ. A value of a factor set is understood as the set of its constituting factors each given with its specific value. Thus, if $\phi = \{p_1, p_2\}$, where $p_1 = 1, 2$ and $p_2 = 1, 2, 3$, then the values of the factor set ϕ are $\{p_1 = 1, p_2 = 1\}$, $\{p_1 = 1, p_2 = 2\}$, etc., comprising the six-element set Φ. The situation when the joint distribution does not depend on any factors $(\phi = \emptyset)$ formally corresponds to $\Phi = \{\emptyset\}$, a singleton.

In the context of selective influence we present factor set ϕ as the union of factor sets $\bigcup_{i=1}^{n}\phi_i$. In this case $\Phi = \prod_{i=1}^{n}\Phi_i$, where Φ_i is the set of possible values for ϕ_i. We impose no restrictions on the internal structure of factor sets ϕ_i (they may contain a single factor, infinity of factors, no factors, ϕ_i and ϕ_j may intersect or coincide, etc.).

2.3. Basic notions

We assume that the reader is familiar with the basic notions of measure-based probability theory, such as probability space $(\mathbb{B}, \mathscr{B}, \mu)$, measurable function f from one probability space to another, etc. To avoid unnecessary technicalities, throughout this paper we tacitly assume that in any probability space $(\mathbb{B}, \mathscr{B}, \mu)$ (consisting of a set $\mathbb{B}$, sigma-algebra $\mathscr{B}$, and a probability measure μ) singleton subsets are measurable (i.e., belong to $\mathscr{B}$).

We say that random entity X is *defined* on probability space $(\mathbb{B}, \mathscr{B}, \mu)$ and write $X = f(B)$ if f is a measurable function on $\mathbb{B}$. This means that on the codomain of this function, $\mathbb{X} = f(\mathbb{B})$, we have a sigma-algebra $\mathscr{X}$ and a probability measure α such that

$$\mathbb{P} \in \mathscr{X} \implies f^{-1}(\mathbb{P}) \in \mathscr{B},$$

and, for any $\mathbb{P} \in \mathscr{X}$,

$$\alpha[\mathbb{P}] = \Pr[X \in \mathbb{P}] = \mu[f^{-1}(\mathbb{P})] = \Pr[B \in f^{-1}(\mathbb{P})].$$

Obviously, if $(\mathbb{B}, \mathscr{B}, \mu) = (\mathbb{X}, \mathscr{X}, \alpha)$ and f is identity, then $B = f(B)$. We say in this case that B is *identified* on $(\mathbb{B}, \mathscr{B}, \mu)$ (in the sense of mapping the space onto itself by means of the identity function on $\mathbb{B}$). Thus, any random entity X maps a probability space on which it is defined, $(\mathbb{B}, \mathscr{B}, \mu)$, into the probability space on which it is identified, $(\mathbb{X}, \mathscr{X}, \alpha)$.

Given an arbitrary indexing set Λ, a *system of random entities* $\{X_\lambda\}_{\lambda\in\Lambda}$, where each X_λ is identified on a probability space $(\mathbb{X}_\lambda, \mathscr{X}_\lambda, \alpha_\lambda)$, is a random entity X identified on a probability space $(\mathbb{X}, \mathscr{X}, \alpha)$, such that

$$\mathbb{X} = \prod_{\lambda\in\Lambda}\mathbb{X}_\lambda, \quad \mathscr{X} = \prod_{\lambda\in\Lambda}\mathscr{X}_\lambda,^{3}$$

and α is a probability measure such that for any $\mathbb{P}_\kappa \in \mathscr{X}_\kappa$,

$$\alpha\left[\mathbb{P}_\kappa \times \prod_{\lambda\in\Lambda\setminus\{\kappa\}}\mathbb{X}_\lambda\right] = \alpha_\kappa[\mathbb{P}_\kappa].$$

Subsystems $\{X_\lambda\}_{\lambda\in\Lambda'}$ of $\{X_\lambda\}_{\lambda\in\Lambda}$ (where $\Lambda' \subset \Lambda, \Lambda' \neq \emptyset$) are defined in the obvious way.

In this paper, we primarily deal with finite systems, $\Lambda = \{1, \ldots, n\}$, but the general notion is needed too. In particular, we need it to define the notion of an indexed set of independent random entities for our first lemma below. Given a system of random entities $\{X_\lambda\}_{\lambda\in\Lambda}$ with each X_λ identified on $(\mathbb{X}_\lambda, \mathscr{X}_\lambda, \alpha_\lambda)$, the meaning of saying that $\{X_\lambda\}_{\lambda\in\Lambda}$ are mutually stochastically independent (with either or both adjectives dropped at one's convenience) is as follows: the probability measure α for $\{X_\lambda\}_{\lambda\in\Lambda}$ is such that for any $\{\mathbb{P}_\lambda \in \mathscr{X}_\lambda\}_{\lambda\in\Lambda'}$ with a *finite* $\Lambda' \subset \Lambda$,

$$\alpha\left[\prod_{\lambda\in\Lambda'}\mathbb{P}_\lambda \times \prod_{\lambda\in\Lambda\setminus\Lambda'}\mathbb{X}_\lambda\right] = \prod_{\lambda\in\Lambda'}\alpha_\lambda[\mathbb{P}_\lambda].$$

This is conventionally presented as

$$\Pr[\{X_\lambda \in \mathbb{P}_\lambda\}_{\lambda\in\Lambda'}] = \prod_{\lambda\in\Lambda'}\Pr[X_\lambda \in \mathbb{P}_\lambda].$$

The probability space for such a system will be denoted by $\prod_{\lambda\in\Lambda}(\mathbb{X}_\lambda, \mathscr{X}_\lambda, \alpha_\lambda)$, and measure α by $\prod_{\lambda\in\Lambda}\alpha_\lambda$.

We say that a system of random entities $\{X_\lambda\}_{\lambda\in\Lambda}$ (jointly) depends on factor set ϕ (or, the joint distribution of $\{X_\lambda\}_{\lambda\in\Lambda}$ depends on factor set ϕ) if α depends on ϕ

$$\alpha \equiv \alpha_\phi$$

and correspondingly

$$\alpha_\lambda \equiv \alpha_{\lambda,\phi},$$

for every $\lambda \in \Lambda$. In this case, we may present the system as $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$.[4] Note that $(\mathbb{X}, \mathscr{X})$ does not change with ϕ; this explicates the meaning of viewing $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ as "*the same*"

[3]Product sigma-algebra $\prod_{\lambda\in\Lambda}\mathscr{X}_\lambda$ is defined as the smallest sigma-algebra containing sets of the form $\mathbb{P}_\kappa \times \prod_{\lambda\in\Lambda\setminus\{\kappa\}}\mathbb{X}_\lambda$ $(\mathbb{P}_\kappa \in \mathscr{X}_\kappa)$.

[4]More precise notation would be $\{X_\lambda\}_{\lambda\in\Lambda}(\phi)$, which, however, is less convenient.

random entities whose distributions may be different at different values of ϕ.

3. The factor-independence lemma and multiple unrelated probability spaces

The following lemma shows that $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ identified on $(\mathbb{X}, \mathscr{X}, \alpha_\phi)$ (with a ϕ-dependent measure) can always be presented as a function of ϕ and a ϕ-independent random entity.

Lemma 1. *Any system $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ identified on $(\mathbb{X}, \mathscr{X}, \alpha_\phi)$ can be defined on a probability space with ϕ-independent probability measure.*

Proof. What we have to show is that one can construct a random entity B, identified on $(\mathbb{B}, \mathscr{B}, \omega)$, and functions $\{f_\lambda\}_{\lambda\in\Lambda}$, such that

$$X_\lambda = f_\lambda(B, \phi) \tag{5}$$

for every $\lambda \in \Lambda$ and every value of $\phi \in \Phi$. Take B identified on $\prod_{\phi\in\Phi}(\mathbb{X}, \mathscr{X}, \alpha_\phi)$, and put, for all $\lambda \in \Lambda$,

$$f_\lambda(b, \phi) = \pi_\lambda(\tilde{\pi}_\phi(b)),$$

where $\tilde{\pi}_\phi : \mathbb{X}^\Phi \to \mathbb{X}$ and $\pi_\lambda : \mathbb{X} \to \mathbb{X}_\lambda$ are projection functions.[5] Clearly, $\omega = \prod_{\phi\in\Phi}\alpha_\phi$ is ϕ-independent, $\{f_\lambda(b, \phi)\}_{\lambda\in\Lambda}$ are measurable for every ϕ, and for any value of ϕ and any $\mathbb{P} \in \mathscr{X}$,

$$\begin{aligned}\omega[\{f_\lambda(b, \phi)\}_{\lambda\in\Lambda} \in \mathbb{P}] &= \omega[B \in \tilde{\pi}_\phi^{-1}(\mathbb{P})]\\ &= \alpha_\phi[\{X_\lambda(\phi)\}_{\lambda\in\Lambda} \in \mathbb{P}].\end{aligned}$$

This means (5). $\square$

To illustrate the construction used in the proof, consider the case $\Lambda = \{1, 2, 3\}$, $\phi = \{p\}$, where the external factor p can attain values 1, 2. Then $\Phi = \{\{1\}, \{2\}\}$. We have then system (X_{11}, X_{21}, X_{31}) for $\phi = \{1\}$ (i.e., $p = 1$) and system (X_{12}, X_{22}, X_{32}) for $\phi = \{2\}$ ($p = 2$) identified on set $\mathbb{X} = \mathbb{X}_1 \times \mathbb{X}_2 \times \mathbb{X}_3$ endowed with sigma-algebra $\mathscr{X} = \mathbb{X}_1 \otimes \mathscr{X}_2 \otimes \mathscr{X}_3$ and probability measures α_1 and α_2, respectively. Random entity B is defined as

$$B = ((X_{11}, X_{21}, X_{31}), (X_{12}, X_{22}, X_{32})),$$

identified on space $(\mathbb{X} \times \mathbb{X}, \mathscr{X} \otimes \mathscr{X}, \omega = \alpha_1\alpha_2)$. We have $\tilde{\pi}_1(B) = (X_{11}, X_{21}, X_{31})$, $\tilde{\pi}_2(B) = (X_{12}, X_{22}, X_{32})$, $\pi_1[\tilde{\pi}_1(B)] = X_{11}$, $\pi_2[\tilde{\pi}_1(B)] = X_{21}$, etc. For $\phi = \{1\}$ and any $\mathbb{P} \in \mathscr{X}$,

$$\begin{aligned}&\omega[\{\pi_\lambda(\tilde{\pi}_1(B))\}_{\lambda\in\{1,2,3\}} \in \mathbb{P}]\\ &\quad = \omega[(X_{11}, X_{21}, X_{31}) \in \mathbb{P} \& (X_{12}, X_{22}, X_{32}) \in \mathbb{X}]\\ &\quad = \alpha_1[(X_{11}, X_{21}, X_{31}) \in \mathbb{P}],\end{aligned}$$

and analogously for $\phi = \{2\}$.

Innocuous as it may seem, our lemma has important consequences. We present them in the remainder of this section.

3.1. Probability causality

In relation to the issue of probabilistic causality, the lemma solves, on the most general level possible, the problem posed by Suppes and Zanotti (1981) and supersedes their theorem. We see that given any system $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$, one can always find a common source of randomness B such that given any of its values $b \in \mathbb{B}$, random entities $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ are conditionally deterministic, which is a special case of conditional independence.[6] As shown later, in Section 5 (Theorem 1), this result can be improved on if one additionally assumes (as Suppes and Zanotti do) that $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ is a finite system of random variables: in this case the common source of randomness B can be chosen to be any random variable whose singletons have zero probability (e.g., a variable uniformly distributed between 0 and 1).

3.2. Separation of randomness from dependence on factors

The lemma says that the source of randomness for a system $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ can always be chosen to be ϕ-independent. The influence of a factor set ϕ on $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$, therefore, can always be *structurally separated* from the source of randomness B: ϕ and B are logically orthogonal determinants of the joint distribution of $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$.

3.3. Dependence on a single source of randomness versus stochastic in(ter)dependence

The fact that all components of $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$ are functions of *one and the same* random entity B turns out to hold essentially by definition (of a system of random entities with joint distribution). In particular, (5) does not imply a stochastic interdependence of $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$: it may very well be that, for any $\{\mathbb{P}_i \in \mathscr{X}_i\}_{i\in I}$ with a finite $I \subset \Lambda$,

$$\begin{aligned}&\Pr[X_i(\phi) \in \mathbb{P}_i \text{ for all } i \in I]\\ &\quad = \omega\left[\bigcap_{i\in I} f_i^{-1}(\mathbb{P}_i, \phi)\right] = \prod_{i\in I}\omega[f_i^{-1}(\mathbb{P}_i, \phi)]\\ &\quad = \prod_{i\in I}\alpha_{i,\phi}[\mathbb{P}_\lambda],\end{aligned}$$

in which case the random entities are mutually independent. Our lemma shows that "the reason" for independence or interdependence of random entities X_i is not in the

[5] To help with notation, $\mathbb{X}^\Phi$ is the set of all functions $\Phi \to \mathbb{X}$, i.e., $\mathbb{X}^\Phi = \{\{x_\phi\}_{\phi\in\Phi} : x_\phi \in \mathbb{X} \text{ for all } \phi \in \Phi\}$. For a specific value of ϕ, projection function $\tilde{\pi}_\phi$, when applied to element $\{x_\phi\}_{\phi\in\Phi}$ of this set returns the component x_ϕ. In our case, this component is itself a function (indexed set), $x = \{x_\lambda\}_{\lambda\in\Lambda}$, and projection function π_λ returns its element x_λ.

[6] The theorem by Suppes and Zanotti (which can be viewed as a special case of our lemma, with $\phi = \emptyset$, Λ being finite, and X's being random variables) was proved only for systems consisting of two binary variables. The argument used in that proof was very different from ours.

sources of randomness per se, but in the functions relating these sources to values of X_i.

3.4. Stochastic in(ter)dependence and stochastic unrelatedness

The proof of Lemma 1 essentially consists in redefining system

$$X(\phi) = \{X_\lambda(\phi)\}_{\lambda\in\Lambda}$$

into system

$$B = \{\{\{X_\lambda\}_{\lambda\in\Lambda}\}_\phi\}_{\phi\in\Phi},$$

by treating ϕ as an outer indexing parameter. With respect to probability spaces $(\mathbb{X}, \mathscr{X}, \alpha_\phi)$ on which systems $X(\phi)$ are identified, $X(\phi)$ taken at different values of ϕ are *stochastically unrelated*, that is, they are not defined on a common probability space. Consequently, one cannot speak of their joint distribution. Intuitively, stochastic unrelatedness of random entities means the absence of a coupling or co-occurrence scheme for their realizations, something one routinely encounters in empirical studies (think, e.g., of response times recorded under different experimental conditions, or one's perceptual images on two different days). Our construction shows that unrelated random entities can always be transformed, if needed, into independent (hence stochastically related) ones by introducing "new" probability spaces, products of probability spaces already in play.

This simple observation seems useful in assessing the mathematical practice of introducing a "universal probability space" upon which all random entities are supposed to be defined. Thinking of "all random entities imaginable," this practice is dubious, very much on a par with introducing a "set of all possible things." In fact, set $\mathbb{U}$ in the truly universal probability space $(\mathbb{U}, \mathscr{U}, \omega)$ cannot but coincide with this set of all possible things, something of undefinable cardinality and with an enigmatic sigma-algebra. The reasonable approach illustrated by our lemma is to allow for *freely introducible unrelated probability spaces* (with correspondingly unrelated random entities identified thereon) but to keep in mind that given any set of such spaces (random entities) one can always form a new probability space upon which one can define all the random entities already introduced.

3.5. Stochastic unrelatedness and representability statements

The notion of multiple probability spaces is important for understanding logical ramifications of saying that a random entity X is *representable* as a function of another random entity, $X = f(B)$. If we posit from the outset that X and B are defined on a common probability space, $(\mathbb{U}, \mathscr{U}, \omega)$, then for some measurable functions $g : \mathbb{U} \to \mathbb{X}$ and $h : \mathbb{U} \to \mathbb{B}$, we have $X = g(U)$ and $B = h(U)$. In this context, if we state that for any $\mathbb{P} \in \mathscr{X}$,

$$\alpha[\mathbb{P}] = \Pr[X \in \mathbb{P}] = \mu[f^{-1}(\mathbb{P})] = \Pr[B \in f^{-1}(\mathbb{P})], \tag{6}$$

this would only mean that X and $f(B)$ are identically distributed, but not that they are equal. The statement $X = f(B)$ in this context means $f \equiv g \circ h$, of which (6) is a consequent but not an equivalent.

The situation is different, however, if X is identified on $(\mathbb{X}, \mathscr{X}, \alpha)$ and one introduces a new, unrelated, probability space, $(\mathbb{B}, \mathscr{B}, \mu)$, on which one identifies B. Then saying that X is representable as $X = f(B)$ is *equivalent* to (6). X which was initially identified on $(\mathbb{X}, \mathscr{X}, \alpha)$ in here being redefined, rendered a random entity defined on $(\mathbb{B}, \mathscr{B}, \mu)$. The distinction between the statements "$X = f(B)$" and "X and $f(B)$ have identical distributions" in this context is meaningless, as we do not have a joint distribution of (X, B) before we represent X as $f(B)$.

To prevent any misunderstanding, consider the situation when X and Y are stochastically unrelated random entities (say, the perceptual images of two stimuli presented to two different subjects) identified on, respectively, $(\mathbb{X}, \mathscr{X}, \alpha)$ and $(\mathbb{Y}, \mathscr{Y}, \beta)$. Suppose that with an appropriately chosen B stochastically unrelated to both X and Y (e.g., uniformly distributed between 0 and 1), and with appropriately chosen measurable functions p and q, one can state that X is representable as $X = p(B)$, in the sense of (6), and, separately, that Y is representable as $Y = q(B)$, in the same sense. Both these statements are perfectly meaningful when taken separately. What one cannot conclude from this is that the two-component system (X, Y) is representable as $(p(B), q(B))$, or, in our notation,

$$\begin{bmatrix} X = p(B) \\ Y = q(B) \end{bmatrix}.$$

This would wrongly imply a joint distribution for X and Y. In this situation, if one has to present the two representability statements together, one would have to say that $X = p(B)$ and $Y = q(B')$, where B and B' are stochastically unrelated but identically distributed entities (say, both are uniformly distributed between 0 and 1).

4. Selective influence

We now focus on finite systems of random entities, $\Lambda = \{1, \ldots, n\}$, and put $\phi = \bigcup_{i=1}^n \phi_i$. Notation $\{X_1(\phi), \ldots, X_n(\phi)\}$ is a special case of $\{X_\lambda(\phi)\}_{\lambda\in\Lambda}$, designating dependence of the joint distribution of $\{X_1, \ldots, X_n\}$ on factor set ϕ. Notation $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ means that the joint distribution of $\{X_1, \ldots, X_n\}$ depends on $\bigcup_{i=1}^n \phi_i$ and, in addition, the marginal distribution of each X_i depends on ϕ_i only.

Representation (5) being universal, the general definition of selective influence it leads to is as follows.

Definition 1. If $\{X_1(\phi), \ldots, X_n(\phi)\}$ can be represented as

$$\begin{bmatrix} X_1 = f_1(B, \phi_1) \\ \vdots \\ X_i = f_i(B, \phi_i) \\ \vdots \\ X_n = f_n(B, \phi_n) \end{bmatrix}, \tag{7}$$

with B identified on some probability space $(\mathbb{B}, \mathscr{B}, \omega)$ and $\{f_1, \ldots, f_n\}$ being measurable functions, then we say that $\{X_1(\phi), \ldots, X_n(\phi)\}$ are selectively influenced by (respective) factor sets $\{\phi_1, \ldots, \phi_n\}$, and write

$$\{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\}.$$

In the special case when $\phi_i = \phi$ for all i, relation

$$\{X_1, \ldots, X_n\} \hookleftarrow \{\phi, \ldots, \phi\}$$

means that $\{X_1(\phi), \ldots, X_n(\phi)\}$ can be represented as

$$\begin{bmatrix} X_1 = f_1(B, \phi) \\ \vdots \\ X_i = f_i(B, \phi) \\ \vdots \\ X_n = f_n(B, \phi) \end{bmatrix}$$

which, in view of Lemma 1, is uninformative: this representation can always be achieved. In contrast, if ϕ_i's are not all the same, Definition 1 is a hypothesis that may or may not hold for a given vector $\{X_1(\phi), \ldots, X_n(\phi)\}$ that depends on $\phi = \bigcup_{i=1}^{n} \phi_i$.

In particular, although $\{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\}$ implies

$$\{X_1(\phi), \ldots, X_n(\phi)\} = \{X_1(\phi_1), \ldots, X_n(\phi_n)\}$$

(the property called *marginal selectivity*),[7] the reverse implication does not work. Moreover, as shown in Dzhafarov (2003a), it is possible that subsystems $\{X_i\}_{i \in I}$ depend only on corresponding $\bigcup_{i \in I} \phi_i$ for all possible nonempty subsets $I \subset \{1, \ldots, n\}$ (the property called *complete marginal selectivity*), but (7) does not hold for any random entity B. The reverse, obviously, cannot be true:

$$\text{if } \{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\} \text{ then } \{X_i\}_{i \in I} \hookleftarrow \{\phi_i\}_{i \in I}$$

for all possible nonempty subsets $I \subset \{1, \ldots, n\}$ (this is called the *nestedness property* of selective influence). We will not discuss here these properties in detail, as this discussion would not substantially differ from the one in Dzhafarov (2003a). Instead, we turn to the aspect of selective influence which was not represented in that paper with sufficient clarity.

The definition of selective influence in Dzhafarov (2003a) is given in the form generalizing our example (1): given $\{X_1(\phi), \ldots, X_n(\phi)\}$ and $\{\phi_1, \ldots, \phi_n\}$,

$$\{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\} \iff \begin{bmatrix} X_1 = f_1(C, S_1, \phi_1) \\ \vdots \\ X_i = f_i(C, S_i, \phi_i) \\ \vdots \\ X_n = f_n(C, S_n, \phi_n) \end{bmatrix}, \tag{8}$$

where $C, S_1, \ldots, S_n$ are mutually independent random entities whose distributions do not depend on $\{\phi_1, \ldots, \phi_n\}$. This definition allows one to give "nice and intuitive" interpretations to special cases: if the common source of randomness C in (8) is a dummy argument (i.e., if $f_i(c, s, \phi_i)$ does not depend on c, for all $i = 1, \ldots, n$), then $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ is a system of mutually independent random entities each of which depends on "its own" set of factors; if all S_i's are dummy arguments, and if all f_i's are one-to-one, then any two components of $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ are functions of each other; etc.

In view of Lemma 1, however, it is clear that in the absence of additional constraints, representation (8) with the sources of randomness classified into common (C) and specific ($S_1, \ldots, S_n$), is equivalent to representation (7), with a single source of randomness B. Indeed, given (7) one can always rename B into C and add dummy variables $S_1, \ldots, S_n$. Conversely, given (8), by the definition of jointly distributed random entities, one can always find a random entity B such that

$$C = g(B), \quad S_1 = h_1(B), \ldots, S_n = h_n(B),$$

leading to (7). For instance, one can always put $B = (C, S_1, \ldots, S_n)$. As stated before, "the reason" for independence or interdependence of random entities X_i is not in the sources of randomness per se, but in functions f_i.

There is, however, a clear difference between the two equivalent representations, (7) and (8). Fixing B at some value b in (7) makes $\{f_1(b, \phi_1), \ldots, f_n(b, \phi_n)\}$, conditionally deterministic, whereas fixing C at some value c in (8) makes $\{f_1(c, S_1, \phi_1), \ldots, f_n(c, S_n, \phi_n)\}$ conditionally independent, possibly but not necessarily deterministic. One might be tempted to think that (7) offers greater conceptual simplicity than (8): is it not easier to deal with deterministic entities than with stochastic ones? The answer is that it is not always so. With all $f_i(b, \phi_i)$ being deterministic, one can always find measurable sets $\mathbb{P}$ in $\prod_{i=1}^{n} \mathbb{X}_i$ such that the probability of $\{f_1(b, \phi_1), \ldots, f_n(b, \phi_n)\} \in \mathbb{P}$ would jump from 0 to 1 as one changes the value of $\{\phi_1, \ldots, \phi_n\}$. If some of the factors change continuously, this property might be a serious impediment for analysis. For an example see Dzhafarov (2003b, 2003c) where the property

[7]Strictly speaking, the meaning of "X_i depends on ϕ_i only" is "X_i does not depend on $\phi \backslash \phi_i$": the marginal effectiveness of the factors (see Dzhafarov, 2001) is not important for the present considerations. Thus, $(X_1, X_2) \hookleftarrow (\phi_1, \phi_2)$ implies $(X_1, X_2) = (X_1(\phi_1), X_2(\phi_2))$ in the sense that the distribution of X_1 does not change with ϕ_2 and the distribution of X_2 does not change with ϕ_1. It is possible, however, that the distribution, say, of X_1 does not change with some or even all factors in ϕ_1 either.

of well-behavedness for probability distributions of perceptual images does not hold if these distributions are singular. Another example can be found in Dzhafarov et al. (2004) where the conditional distribution functions for randomly varying durations are assumed to be differentiable. In all such cases we need the assumption that $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ are representable by means of (8) with *nondummy* specific sources of randomness $S_1, \ldots, S_n$. In the remainder of this section we show how such an assumption should be formulated rigorously.

We need two preliminary notions and some notation conventions.

A random entity A identified on $(\mathbb{A}, \mathscr{A}, \alpha)$ is called nondeterministic (with respect to α) if $\alpha[\{a\}] < 1$ for every value $a \in \mathbb{A}$ (recall our convention that all singletons $\{a\}$ belong to $\mathscr{A}$). This implies that if a nondeterministic A is defined on $(\mathbb{B}, \mathscr{B}, \beta)$ by means of $A = f(B)$, then $\beta[\{b : f(b) = a\}] < 1$ for every value $a \in \mathbb{A}$.

Given a finite system of random entities $A = (A_1, \ldots, A_n)$ jointly identified on $(\mathbb{A}, \mathscr{A}, \alpha)$, subsystem A_I for any nonempty $I \subset \{1, \ldots, n\}$ is identified on the corresponding projection space, denoted $(\mathbb{A}_I, \mathscr{A}_I, \alpha_I)$. System $A = (A_1, \ldots, A_n)$ is called *tight*[8] if for some $i \in \{1, \ldots, n\}$, on denoting $I_{(i)} = \{1, \ldots, n\} \backslash \{i\}$,

$$A_i = f(A_{I_{(i)}}) \quad \text{a.s. } (\alpha_{I_{(i)}}).$$

We propose now the following definition as generalizing and clarifying the intended meaning of representation (8). The motivation for the term *classifiable sources of randomness* is derived from the intuition of classifying different "aspects" of a single source of randomness B into common and specific sources.

Definition 2. Let $\{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\}$. We say that random entities $\{X_1, \ldots, X_n\}$ have classifiable sources of randomness if B in Definition 1 can be chosen so that for any nonempty subset $I \subset \{1, \ldots, n\}$ one can find measurable functions $\{f_i\}_{i \in I}$ and stochastically independent random entities $C = g(B), \{S_i = h_i(B)\}_{i \in I}$ identified on, respectively, $(\mathbb{C}, \mathscr{C}, \gamma)$ and $\{(\mathbb{S}_i, \mathscr{S}_i, \delta_i)\}_{i \in I}$, such that

1. for all $i \in I$
$$X_i = f_i(C, S_i, \phi_i) \tag{9}$$
and
2. for any value of $\{\phi_i\}_{i \in I}$ at which $\{X_i\}_{i \in I}$ is not tight, $f_i(c, S_i, \phi_i)$ is nondeterministic (with respect to δ_i) for all $i \in I$ and almost all $c \in \mathbb{C}$ (with respect to γ).

The second condition essentially says that $C, \{S_i\}_{i \in I}$ for a nontight finite subsystem $\{X_i(\phi_i)\}_{i \in I}$ can be chosen so that for all values c of C (except, perhaps, for a subset of probability zero), $\{X_i(\phi_i)\}_{i \in I}$ are conditionally independent without being conditionally deterministic.

The choice of $\{f_i\}_{i \in I}$ and $\{C, \{S_i\}_{i \in I}\}$ may be different for different choices of $I \subset \{1, \ldots, n\}$. It is obvious, however, that if $\{f_i\}_{i \in I}, C, \{S_i\}_{i \in I}$ satisfy Definition 2 for some I, then they satisfy this definition for any $I' \subset I$. In particular, if $\{1, \ldots, n\}$ is nontight for at least some values of $\{\phi_1, \ldots, \phi_n\}$, then a single choice of $\{C, S_1, \ldots, S_n\}$ can be used for all its subsystems.

5. Random variables

We turn now to what is arguably the most important special case to consider: finite systems of *random variables* $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ selectively influenced by factor sets $\{\phi_1, \ldots, \phi_n\}$. For any $i = 1, \ldots, n$, $X_i(\phi_i)$ is identified on probability spaces $(\mathbb{R}, \mathscr{R}, \mu_{\phi_i})$, where $\mathscr{R}$ is the standard *Borel sigma-algebra* on $\mathbb{R}$.[9] System $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ is then identified on $(\mathbb{R}^n, \mathscr{R}^n, \alpha_\phi)$, where $\mathscr{R}^n$ stands for $\prod_{i=1}^n \mathscr{R}$, and probability measure α_ϕ agrees with the coordinate probability measures μ_{ϕ_i} as described earlier in the general definition of a system of random entities.

Every X_i is uniquely described by its *distribution function* $d_i(x, \phi_i) = \Pr[X_i(\phi_i) \leqslant x]$, $x \in \mathbb{R}$. A distribution function can have a countable number of discontinuities corresponding to *atoms* of μ_{ϕ_i}.[10] For reasons to become apparent shortly, it is desirable for us to deal with continuous, or *atomless* coordinate measures (which implies an atomless measure on the product space). To this end we introduce random variables $\{W_1(\phi_1), \ldots, W_n(\phi_n)\}$ defined as

$$\begin{bmatrix} W_1(\phi_1) = d_1(X_1(\phi_1) - 0, \phi_1) + [d_1(X_1(\phi_1), \phi_1) - d_1(X_1(\phi_1) - 0, \phi_1)]V_1 \\ \vdots \\ W_i(\phi_i) = d_i(X_i(\phi_i) - 0, \phi_i) + [d_i(X_i(\phi_i), \phi_i) - d_i(X_i(\phi_i) - 0, \phi_i)]V_i \\ \vdots \\ W_n(\phi_n) = d_n(X_n(\phi_n) - 0, \phi_n) + [d_n(X_n(\phi_n), \phi_n) - d_n(X_n(\phi_n) - 0, \phi_n)]V_n \end{bmatrix}, \tag{10}$$

where $\{V_1, \ldots, V_n\}$ are independent variables Uni$[0, 1]$ (uniformly distributed on unit interval) such that vectors $\{V_1, \ldots, V_n\}$ and $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ are mutually independent.

Lemma 2. *Let* $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ *and* $\{W_1(\phi_1), \ldots, W_n(\phi_n)\}$ *be as above. Then*:

1. $W_i(\phi_i)$ *is* Uni$[0, 1]$, *for all* ϕ_i *and* $i = 1, \ldots, n$;[11]

[8] We choose this term to avoid potentially more confusing terms "interdependent," "overdetermined," etc. The tightness of finite systems of random entities must not be confused with the property of tightness for probability measures.

[9] Recall that this is the sigma-algebra consisting of all countable combinations of intersections and unions involving open and closed subsets of $\mathbb{R}$. In view of Lemma 2, it is sometimes convenient to consider $\mathbb{R}$ as the extended set of reals (with added points ∞ and $-\infty$) and to modify $\mathscr{R}$ accordingly.

[10] An atom is a value x such that $\alpha_\phi[\{x\}] > 0$. A measure on a Borel space is atomless if it equals zero for all singletons.

[11] In view of this result one might be tempted to conclude that notation $\{W_1(\phi_1), \ldots, W_n(\phi_n)\}$ is superfluous. This would not be correct. Even

2.

$$\begin{bmatrix} X_1 = q_1(W_1(\phi_1), \phi_1) \\ \vdots \\ X_i = q_i(W_i(\phi_i), \phi_i) \\ \vdots \\ X_n = q_n(W_n(\phi_n), \phi_n) \end{bmatrix},$$

for some measurable functions $\{q_1, \ldots, q_n\}$;

3. $\{X_1, \ldots, X_n\} \looparrowleft \{\phi_1, \ldots, \phi_n\}$ *if and only if* $\{W_1, \ldots, W_n\} \looparrowleft \{\phi_1, \ldots, \phi_n\}$.

Proof. (Ad 1) For any $p \in (0, 1)$ there is one and only one u such that $d_i(u - 0, \phi_i) \leqslant p \leqslant d_i(u, \phi_i)$ and $x > u \Longrightarrow p < d_i(x, \phi_i)$. If $d_i(u - 0, \phi_i) < d_i(u, \phi_i)$, then

$$\begin{aligned}\Pr[W_i(\phi_i) \leqslant p] = {} & \Pr[X_i(\phi_i) < u] + \Pr[X_i(\phi_i) = u] \\ & \times \Pr\left[V \leqslant \frac{p - d_i(u - 0, \phi_i)}{d_i(u, \phi_i) - d_i(u - 0, \phi_i)}\right] \\ = {} & d_i(u - 0, \phi_i) + [d_i(u, \phi_i) - d_i(u - 0, \phi_i)] \\ & \times \frac{p - d_i(u - 0, \phi_i)}{d_i(u, \phi_i) - d_i(u - 0, \phi_i)} = p.\end{aligned}$$

If $d_i(u - 0, \phi_i) = d_i(u, \phi_i) = p$, then

$$\Pr[W_i(\phi_i) \leqslant p] = \Pr[X(\phi_i) \leqslant u] = p.$$

That $\Pr[W_i(\phi_i) \leqslant 0] = 0$ and $\Pr[W_i(\phi_i) \leqslant 1] = 1$ is obvious.

(Ad 2) Choose $q_i(p, \phi_i)$ to be the quantile function for $X_i(\phi_i)$, defined as

$$q_i(p, \phi_i) = \inf\{x : d_i(x, \phi_i) \geqslant p\}, \quad p \in [0, 1],$$

with the convention allowing the function to attain values ∞ (at $p = 1$) and $-\infty$ (at $p = 0$) (see footnote 9). Quantile functions are obviously measurable.

(Ad 3) $\{W_1, \ldots, W_n\} \looparrowleft \{\phi_1, \ldots, \phi_n\}$ means that for some random entity B, $W_i = g_i(B, \phi_i)$ for all $i \in \{1, \ldots, n\}$. Then

$$X_i(\phi_i) = q_i(g_i(B, \phi_i), \phi_i) = f_i(B, \phi_i),$$

for all $i \in \{1, \ldots, n\}$. Conversely, if $X_i = f_i(B, \phi_i)$, then

$$W_i(\phi_i) = h_i(f_i(B, \phi_i), V_i),$$

where h_i replaces the expression in (10). Since B and $\{V_1, \ldots, V_n\}$ are jointly distributed, one can always construct a random entity B^* so that B and all of $\{V_1, \ldots, V_n\}$ can be presented as measurable functions of B^*. Hence

$$W_i(\phi_i) = g_i(B^*, \phi_i),$$

for all $i \in \{1, \ldots, n\}$. □

(*footnote continued*)
though according to the claim the marginal distribution of W_i does not depend on ϕ_i, the joint distribution of $\{W_1, \ldots, W_n\}$ depends on $\phi_1 \cup \cdots \cup \phi_n$. We could, however, always write $\{W_1(\phi), \ldots, W_n(\phi)\}$ in place of $\{W_1(\phi_1), \ldots, W_n(\phi_n)\}$: the latter is preferable only insofar as it reminds one of how the W's were obtained.

Thus, insofar as the relation of selective influence is concerned we can deal with $\{W_1(\phi_1), \ldots, W_n(\phi_n)\}$ instead of $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$; for any i, $W_i(\phi_i)$ is identified on spaces $([0, 1], \mathscr{I}, \psi_{\phi_i})$, where $\mathscr{I}$ is the Borel sigma-algebra on $[0, 1]$ and ψ_{ϕ_i} is an atomless measure. $\{W_1(\phi_1), \ldots, W_n(\phi_n)\}$ is then identified on $([0, 1]^n, \mathscr{I}^n, \omega_\phi)$ with atomless measure ω_ϕ. The reason for our interest in atomless measures lies in the following important fact.

Lemma 3. *For a* Uni$[0, 1]$ *random variable* U *and any stochastically unrelated to* U *vector of random variables* $(A_1, \ldots, A_n)$ *identified on* $(\mathbb{A}, \mathscr{A}, \alpha)$ *with atomless* α, *there is a function* $h : [0, 1] \to \mathbb{A}$ *such that*

(i) *h is bijective*;
(ii) *h and h^{-1} are measurable* (*Borel*);
(iii) *for any* $\mathbb{P} \in \mathscr{A}$, $\alpha[\mathbb{P}] = \ell[h^{-1}(\mathbb{P})]$, *where* ℓ *indicates the usual Lebesgue measure on* $[0, 1]$.

Proof. Follows from a general result for standard Borel spaces that can be found, e.g., in Kechris (1995, p. 116). □

In accordance with the explanations given in the last subsection of Section 3, one consequence of this lemma is that $(A_1, \ldots, A_n)$ can be represented as $(A_1, \ldots, A_n) = h(U)$, or equivalently,

$$\begin{bmatrix} A_1 = h_1(U) \\ \vdots \\ A_i = h_i(U) \\ \vdots \\ A_n = h_n(U) \end{bmatrix},$$

where h_i is the ith projection of function h ($i = 1, \ldots, n$).

In conjunction with Lemma 2, this mathematical fact yields Levine's (2003) main result as an immediate consequence (see the subsection on probabilistic dimensionality in Introduction). We prefer, however, to consider implications of Lemmas 2 and 3 in a broader context, for the case $\phi_1 = \cdots = \phi_n = \phi$, that is, for our generalizations of both probabilistic causality and probabilistic dimensionality to indexed families of random systems.

Theorem 1. *Any vector of random variables* $\{X_1(\phi), \ldots, X_n(\phi)\}$ *is representable as*

$$\begin{bmatrix} X_1 = g_1(B, \phi) \\ \vdots \\ X_i = g_i(B, \phi) \\ \vdots \\ X_n = g_n(B, \phi) \end{bmatrix},$$

where B is any vector of random variables with atomless distributions, and $\{g_1, \ldots, g_n\}$ are measurable functions. In particular, B can always be chosen to be Uni[0, 1].

Proof. Applying Lemma 3 to $\{(W_1(\phi), \ldots, W_n(\phi))\}$ we get representation

$$W_i = h_i(U, \phi),$$

for all $i = 1, \ldots, n$, where h_i is the ith projection of h (which now depends on ϕ), and U is Uni[0, 1]. Next, we use the second statement of Lemma 2 to obtain, for all $i = 1, \ldots, n$,

$$X_i = q_i(W_i(\phi), \phi) = q_i(h_i(U, \phi), \phi) = g_i(U, \phi).$$

It remains to rename U into B to obtain the last statement of the theorem. The possibility of using any vector of random variables with atomless distribtuions in place of U is obvious from Lemma 3. □

This is a great improvement over Lemma 1: the application of the latter to random vector $\{X_1(\phi), \ldots, X_n(\phi)\}$ yields a representation $\{g_1(B, \phi), \ldots, g_n(B, \phi)\}$ with B being a random entity defined on a potentially very large probability space. Now we know that because $X_1, \ldots, X_n$ are random variables, B can be always be chosen as Uni[0, 1], any other atomless variable, or vector of such variables. This result subsumes and generalizes both Levine's (2003) treatment of probabilistic dimensionality and Suppes and Zanotti's (1981) treatment of probabilistic causality, insofar as the latter is confined to finite systems of random variables.

Unfortunately, we cannot use the same argument as in the previous theorem to conclude that if

$$\{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\}$$

then $\{X_1, \ldots, X_n\}$ are representable as

$$\begin{bmatrix} X_1 = f_1(U, \phi_1) \\ \vdots \\ X_i = f_i(U, \phi_i) \\ \vdots \\ X_n = f_n(U, \phi_n) \end{bmatrix},$$

with U being Uni[0, 1]. The reason for this is that projection h_i of h in Lemma 3 generally depends on entire $\{\phi_1, \ldots, \phi_n\}$, for every $i = 1, \ldots, n$. We can, however, state a weaker but still useful result.

Theorem 2. *If $\{X_1, \ldots, X_n\} \hookleftarrow \{\phi_1, \ldots, \phi_n\}$ and if there is a representation*

$$\begin{bmatrix} X_1 = f_1(C, \phi_1) \\ \vdots \\ X_i = f_i(C, \phi_i) \\ \vdots \\ X_n = f_n(C, \phi_n) \end{bmatrix},$$

with C being a vector of random variables,[12] *then a representation*

$$\begin{bmatrix} X_1 = g_1(B, \phi_1) \\ \vdots \\ X_i = g_i(B, \phi_i) \\ \vdots \\ X_n = g_n(B, \phi_n) \end{bmatrix}$$

can be found with B being any vector of random variables with atomless distribtuions, and $\{g_1, \ldots, g_n\}$ being some measurable functions. In particular, B can always be chosen to be Uni[0, 1].

Proof. Immediately follows from Lemma 3. □

It is worth observing once again, since here it seems even more counterintuitive than in the general case, that the dependence of $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$ on a single random variable uniformly distributed between 0 and 1 does not tell us anything about the stochastic relationship among $\{X_1(\phi_1), \ldots, X_n(\phi_n)\}$: depending on functions g_i, this stochastic relationship may range from stochastic independence to perfect functional interdependence.

6. Conclusion

We have established that any system of random entities whose joint distribution depends on a factor set can be represented by functions of two arguments: a single factor-independent source of randomness and the factor set itself. In the case of random variables (i.e., real-valued random entities endowed with Borel sigma-algebras) the single source of randomness can be chosen to be any random variable (or a vector of random variables) with a continuous distribution. In particular, it can always be chosen to be uniformly distributed between 0 and 1. These results have direct implications for the issue of establishing sources of stochastic interdependence among random entities (the problem of probabilistic causality). As it turns out, finding a source of randomness whose values render the random entities conditionally independent (including the possibility of being conditionally deterministic) does not in any reasonable sense "explain" stochastic interdependence: the true explanation lies in the structure of the functions that relate this source of randomness to the random entities.

We have proposed a refined and very general version of the definition of selective influence. A system of random entities is selectively influenced by a respective system of factor sets if each of the random entities can be presented as a function of two arguments: a single source of randomness (shared by all random entities) and the respective factor set. The definition involving both

[12] The condition can be weakened by allowing B to be defined on any *standard Borel space* (Kechris, 1995).

common (shared by all) and specific (individual) sources of randomness is equivalent to the definition involving a single common source. We have also defined, however, a more restrictive notion of selective influence, with "classifiable" (into common and specific) sources of randomness. The issue of classifiability remains badly underdeveloped: we do not know any useful sufficient conditions for the classifiability to hold. For the case of vectors of (real-valued) random variables we have established that if the common source of randomness in their representation is itself a vector of random variables, then it can be replaced with any other system of random variables with a continuous distribution, including a single variable uniformly distributed between 0 and 1.

Finally, we have presented arguments in favor of what we believe to be a useful approach to random entities and their relationships. This approach involves multiple unrelated to each other probability spaces, with the possibility, if needed, of redefining the unrelated random entities identified on them as each other's functions, or treating them as stochastically independent random entities on products of these spaces. The mathematical tradition of thinking of all random entities under consideration as jointly distributed on a common (usually undefined) sample space is unsuitable for applied probabilistic considerations, where it is common to consider sets of random entities with no reasonable notion of co-occurrence applicable to them, and where it is common to freely introduce unobservable random entities as part of theoretical constructs.

Acknowledgment

The first author's contribution was supported by the NSF Grant SES 0318010, Humboldt Research Award, and Fellowship at the Swedish Collegium for Advanced Studies in Social Sciences.

References

Blank-Lapierre, A., & Fortet, R. (1967). *Theory of random functions*. New York: Gordon & Breach.

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

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

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

Dzhafarov, E. N. (2003b). Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and independent images. *Journal of Mathematical Psychology*, *47*, 208–228.

Dzhafarov, E. N. (2003c). Thurstonian-type representations for "same-different" discriminations: Probabilistic decisions and interdependent images. *Journal of Mathematical Psychology*, *47*, 229–243.

Dzhafarov, E. N., & Colonius, H. (2006). Regular minimality: A fundamental law of discrimination. In H. Colonius, & E. N. Dzhafarov (Eds.), *Measurement and representation of sensations* (pp. 1–46). Mahwah, NJ: Erlbaum.

Dzhafarov, E. N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, *48*, 51–64.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l'Institute H. Poincare*, *10*, 215–310.

Kechris, A. S. (1995). *Classical descriptive set theory*. New York: Springer.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer (English translation (1950): *Foundations of the theory of probability*. Chelsea: New York).

Levine, M. V. (2003). Dimension in latent variable models. *Journal of Mathematical Psychology*, *47*, 450–466.

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

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

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

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