

Noncontextuality with Marginal Selectivity in Reconstructing Mental Architectures

Ru Zhang^{*} and Ehtibar N. Dzhafarov^{**}

Department of Psychological Sciences, Purdue University

^{*}zhang617@purdue.edu

^{**}ehtibar@purdue.edu (corresponding author)

Abstract

We present a general theory of series-parallel mental architectures with selectively influenced stochastically non-independent components. A mental architecture is a hypothetical network of processes aimed at performing a task, of which we only observe the overall time it takes under variable parameters of the task. It is usually assumed that the network contains several processes selectively influenced by different experimental factors, and then the question is asked as to how these processes are arranged within the network, e.g., whether they are concurrent or sequential. One way of doing this is to consider the distribution functions for the overall processing time and compute certain linear combinations thereof (interaction contrasts). The theory of selective influences in psychology can be viewed as a special application of the interdisciplinary theory of (non)contextuality having its origins and main applications in quantum theory. In particular, lack of contextuality is equivalent to the existence of a “hidden” random entity of which all the random variables in play are functions. Consequently, for any given value of this common random entity, the processing times and their compositions (minima, maxima, or sums) become deterministic quantities. These quantities, in turn, can be treated as random variables with (shifted) Heaviside distribution functions, for which one can easily compute various linear combinations across different treatments, including interaction contrasts. This mathematical fact leads to a simple method, more general than the previously used ones, to investigate and characterize the interaction contrast for different types of series-parallel architectures.

KEYWORDS: interaction contrast, mental architectures, noncontextuality, response time, selective influences, series-parallel network.

Running head: Mental Architectures

1 Introduction

The notion of a network of mental processes with components selectively influenced by different experimental factors was introduced to psychology in Saul Sternberg's (1969) influential paper. Sternberg considered networks of processes a, b, c, \dots involved in performing a mental task. Denoting their respective durations by A, B, C, \dots , the hypothesis he considered was that the observed response time T is $A + B + C + \dots$. One cannot test this hypothesis, Sternberg wrote, without assuming that there are some factors, $\alpha, \beta, \gamma, \dots$, that selectively influence the durations A, B, C, \dots , respectively. Sternberg's analysis was confined to stochastically independent A, B, C, \dots , and the consequences of the assumptions of seriality and selective influences were tested on the level of the mean response times only.

Subsequent development of these ideas was aimed at the entire distributions of the response times and at a greater diversity and complexity of mental architectures than just series of "stages." This development prominently includes Roberts and Sternberg (1993), Schweickert, Giorgini, and Dzhafarov (2000), Schweickert and Townsend (1989), Townsend (1984, 1990a, 1990b), Townsend and Nozawa (1995), Townsend and Schweickert (1989), and several other publications, primarily by James Townsend and Richard Schweickert with colleagues. For an overview of these developments see Dzhafarov (2003) and Schweickert, Fisher, and Sung (2012). In the present context we should separately mention the development of the ideas of *stochastic ordering* of processing times in Townsend (1984, 1990a) and Townsend and Schweickert (1989); as well as the idea of *marginal selectivity* (Townsend & Schweickert, 1989).

The notion of selective influences also underwent a significant development, having been generalized from stochastically independent random variables to arbitrarily interdependent ones (Dzhafarov, 2003; Dzhafarov & Gluhovsky, 2006; Kujala & Dzhafarov, 2008; Dzhafarov & Kujala, 2010; in press). The essence of the development is easy to understand using two random variables (e.g., process durations) A, B selectively influenced by two respective factors α, β . In Dzhafarov's (2003) notation, this is written $(A, B) \leftrightarrow (\alpha, \beta)$. According to the definition given in Dzhafarov (2003), this means that there are functions f and g and a random variable R (a *common source of randomness*) such that $f(\alpha, R) = A$ and $g(\beta, R) = B$. If such a choice of (f, g, R) exists, it is not unique. For instance, R can always be chosen to have any distribution that is absolutely continuous with respect to the usual Borel measure on the real line (e.g., a standard uniform, or standard normal distribution, see Dzhafarov & Gluhovsky, 2006). However, a triple (f, g, R) need not exist. It does not exist, e.g., if marginal selectivity (Townsend & Schweickert, 1989) is violated, i.e., if the distribution of, say, A at a given value of α changes in response to changing β . But marginal selectivity is not sufficient for the existence of a triple (f, g, R) . Let, e.g., α and β be binary factors, with values 1, 2 each, and let the correlation ρ between A and B for a treatment (α, β) be denoted $\rho_{\alpha\beta}$. Then the triple in question does not exist if the correlations violated the "cosphericity test" (Kujala & Dzhafarov, 2008), also known in quantum mechanics as Landau's inequality (Landau, 1989):

$$|\rho_{11}\rho_{12} - \rho_{21}\rho_{22}| \leq \bar{\rho}_{11}\bar{\rho}_{12} + \bar{\rho}_{21}\bar{\rho}_{22}, \quad (1)$$

where $\bar{\rho}_{\alpha\beta} = \sqrt{1 - \rho_{\alpha\beta}^2}$. There are many other known conditions that must be satisfied for the existence of a triple (f, g, R) when marginal selectivity is satisfied (Dzhafarov & Kujala, 2010, 2012a,b, 2013, 2014a).

The allusion to quantum mechanics is not accidental: as shown in Dzhafarov and Kujala (2012a,b), the theory of selective influences in psychology can be viewed as a special application of the theory of (non)contextuality. This theory is interdisciplinary (Dzhafarov & Kujala, 2014b-d; Khrennikov, 2009), but its origins are in quantum theory, dating from Kochen and Specker (1967) and John Bell's (1964, 1966) celebrated work. For the modern state of the theory see Dzhafarov, Kujala, and Larsson (2015). A simplified account of the (non)contextuality analysis of the example given above is as follows. One labels each random variable in play *contextually*, i.e., by what property is being measured/recorded under what treatment (context):

$$\left(\begin{array}{c} \underbrace{A}_{\text{value of } \alpha} \\ \text{property: what is measured} \end{array} \right) \overbrace{\text{context: under what treatment}}^{\text{(value of } \alpha, \text{ value of } \beta)}, \left(\begin{array}{c} \underbrace{B}_{\text{value of } \beta} \\ \text{property: what is measured} \end{array} \right) \overbrace{\text{context: under what treatment}}^{\text{(value of } \alpha, \text{ value of } \beta)}. \quad (2)$$

The notation here is, of course, redundant, because the context and property identifiers overlap, but we need now to emphasize the logic rather than achieve notational convenience. Once the labeling is done, one looks at all possible joint distributions imposable on all these random variables, for all properties and all treatments. A system is noncontextual if there exists such a joint distributions in which any two random variables that represent the same property

(“what is measured”) are equal with probability 1. The latter is possible only if the random variables representing the same property always have the same distribution: in our case

$$(A_\alpha)^{(\alpha,\beta)} \sim (A_\alpha)^{(\alpha,\beta')}, (B_\beta)^{(\alpha,\beta)} \sim (B_\beta)^{(\alpha',\beta)} \quad (3)$$

for any values $\alpha, \beta, \alpha', \beta'$ of the two factors. This is called consistent connectedness (Dzhafarov, Kujala, & Larsson, 2015), and in physics is known under a variety of names, including (in certain paradigms) “no-signaling condition” (Cereceda, 2000; Masanes, Acin, & Gisin, 2006; Popescu & Rohrlich, 1994). In psychology, this is marginal selectivity. The definition of noncontextuality just given is not the most general one, as the notion of contextuality can be extended to inconsistently connected (violating marginal selectivity) systems (Dzhafarov, Kujala, & Larsson, 2015), but we do not need this generality in this paper. What is important for us here is that the existence of a joint distribution mentioned in our definition is equivalent to the existence of a random variable R and the functions f, g mentioned in the introductory paragraph.

It is easy to show (Dzhafarov, 2003) that the existence of a triple (f, g, R) for given joint distributions of (A, B) under different treatments (α, β) is equivalent to the existence of a quintuple (f', g', S, S_A, S_B) , where S, S_A, S_B are random variables, such that $f'(\alpha, S, S_A) = A$ and $g'(\beta, S, S_B) = B$. In such a representation, one can speak of a common source of randomness S and specific sources of randomness S_A, S_B . In Dzhafarov, Schweickert, and Sung (2004) this representation was used to investigate different series-parallel arrangements of the hypothetical durations A and B . The reason this representation has been considered convenient is that if one fixes the value $S = s$, then $f'(\alpha, s, S_A) = A_c$ and $g'(\beta, s, S_B) = B_c$ are stochastically independent random variables. One can therefore use theorems proved for stochastically independent selectively influenced components (Schweickert, Giorgini, & Dzhafarov, 2000) to obtain a general result by averaging across possible values of s . For instance, let α, β be binary factors (with values 1, 2 each), and let us assume that the observed duration $T_{\alpha\beta}$ is $\min(A_\alpha, B_\beta)$ for every treatment (α, β) . Then $T_{\alpha\beta s} = \min(A_{\alpha s}, B_{\beta s})$ for every value $S = s$, and it is known that, for the independent $A_{\alpha s}, B_{\beta s}$ (satisfying a prolongation condition, as explained below),

$$\Pr(T_{11s} \leq t) - \Pr(T_{12s} \leq t) - \Pr(T_{21s} \leq t) + \Pr(T_{22s} \leq t) \leq 0. \quad (4)$$

Since this should be true for every value $S = s$, then it should also be true that

$$C(t) = \Pr(T_{11} \leq t) - \Pr(T_{12} \leq t) - \Pr(T_{21} \leq t) + \Pr(T_{22} \leq t) \leq 0. \quad (5)$$

This follows from the fact that

$$\Pr(T_{\alpha\beta} \leq t) = \int \Pr(T_{\alpha\beta s} \leq t) dm(s), \quad (6)$$

where $m(s)$ is the probability measure for S , and the integration is over the space of all possible s . The linear combination $C(t)$ in (5) is called the *interaction contrast of distributions functions*.

The Prolongation Assumption used in Dzhafarov et al. (2004), and derived from Townsend (1984, 1990a) and Townsend and Schweickert (1989), is that, for every $S = s$,

$$\Pr(A_{1s} \leq t) \geq \Pr(A_{2s} \leq t), \Pr(B_{1s} \leq t) \geq \Pr(B_{2s} \leq t). \quad (7)$$

For this particular architecture, $T = \min(A, B)$, this is the only assumption needed. To prove analogous results for more complex mental architectures, however, one needs additional assumptions, such as the existence of density functions for $A_{\alpha s}, B_{\beta s}$ at every s , and even certain ordering of these density functions in some vicinity $[0, \tau]$.

The same results, however, can be obtained without these additional assumptions, if one adopts the other, equivalent definition of selective influences: $f(\alpha, R) = A$ and $g(\beta, R) = B$, for some triple (f, g, R) . If such a representation exists, then

$$a_{\alpha r} = f(\alpha, r), b_{\beta r} = g(\beta, r) \quad (8)$$

are deterministic quantities (real numbers), for every value $R = r$. Any real number x in turn can be viewed as a random variable whose distribution function is a shifted Heaviside function

$$h(t - x) = \begin{cases} 0, & \text{if } t < x, \\ 1, & \text{if } t \geq x. \end{cases} \quad (9)$$

In particular, the quantity $t_{\alpha\beta r} = \min(a_{\alpha r}, b_{\beta r})$ for the simple architecture $T = \min(A, B)$ considered above is distributed according to

$$h(t - t_{\alpha\beta r}) = h_{\alpha\beta r}(t). \quad (10)$$

Let us see how inequality (5) can be derived using these observations.

We first formulate the (*conditional*) *Prolongation Assumption*, a deterministic version of (7): the assumption is that f, g, R can be so chosen that for every $R = r$,

$$a_{1r} \leq a_{2r}, b_{1r} \leq b_{2r}. \quad (11)$$

Without loss of generality, we can also assume, for any given r ,

$$a_{1r} \leq b_{1r} \quad (12)$$

(if not, rename a into b and vice versa).

Remark 1.1. The prolongation Assumption clearly implies (7). Conversely, if (7) holds, one can always find functions f, g, R for which the Prolongation Assumption holds in the form above. For instance, one can choose $R = (S, S_A, S_B)$, take S_A and S_B to be uniformly distributed between 0 and 1, and choose $f(\alpha, \dots), g(\beta, \dots)$ to be the quantile functions for the hypothetical distributions of A and B at the corresponding factor levels.

We next form the conditional (i.e., conditioned on $R = r$) interaction contrast

$$c_r(t) = h_{11r}(t) - h_{12r}(t) - h_{21r}(t) + h_{22r}(t). \quad (13)$$

Notation Convention. When r is fixed throughout a discussion, we omit this argument and write $a_\alpha, b_\beta, t_{\alpha\beta}, h_{\alpha\beta}(t), c(t)$ in place of $a_{\alpha r}, b_{\beta r}, t_{\alpha\beta r}, h_{\alpha\beta r}(t), c_r(t)$. (For binary factors α, β , we also conveniently replace α, β in indexation with i, j .)

Following this convention, there are three different arrangements of a_1, a_2, b_1, b_2 (for a given $R = r$) satisfying (11)-(12):

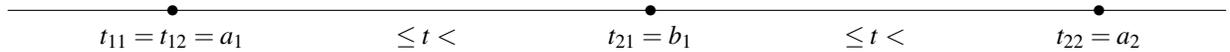
$$\begin{aligned} \text{(i)} \quad & a_1 \leq b_1 \leq a_2 \leq b_2 \\ \text{(ii)} \quad & a_1 \leq a_2 \leq b_1 \leq b_2 \\ \text{(iii)} \quad & a_1 \leq b_1 \leq b_2 \leq a_2 \end{aligned} \quad (14)$$

In all three cases,

$$t_{11} = \min(a_1, b_1) = a_1 = \min(a_1, b_2) = t_{12}. \quad (15)$$

For arrangement (i) we have

$$\begin{array}{ll} +h_{11}(t) = 1 & +h_{11}(t) = 1 \\ -h_{12}(t) = -1 & -h_{12}(t) = -1 \\ -h_{21}(t) = -0 & -h_{21}(t) = -1 \\ +h_{22}(t) = 0 & +h_{22}(t) = 0 \\ = c(t) = 0 & = c(t) = -1 \end{array}$$



This diagram shows the values of $h_{ijr}(t)$ and the resulting values of $c_r(t)$ as t changes with respect to the fixed positions of t_{ijr} (with index r dropped everywhere). Analogously, for arrangements (ii) and (iii), we have, respectively

$$\begin{array}{c} c(t) = 0 \\ \text{---} \bullet \text{---} \leq t < \text{---} \bullet \text{---} \\ t_{11} = t_{12} = a_1 \qquad \qquad \qquad t_{21} = t_{22} = a_2 \end{array}$$

and

$$\begin{array}{c} c(t) = 0 \qquad \qquad \qquad c(t) < 0 \\ \text{---} \bullet \text{---} \leq t < \text{---} \bullet \text{---} \leq t < \text{---} \bullet \text{---} \\ t_{11} = t_{12} = a_1 \qquad \qquad \qquad t_{21} = b_1 \qquad \qquad \qquad t_{22} = b_2 \end{array}$$

In all three cases, $c(t)$ is obviously zero for $t < t_{11}$ and $t \geq t_{22}$. We see that $c(t) = c_r(t) \leq 0$ for all t and every $R = r$. It follows that $C(t) \leq 0$, because

$$\Pr(T_{ij} \leq t) = \int h_{ijr}(t) d\mu(r), \quad (16)$$

for $i, j \in \{1, 2\}$, and

$$C(t) = \int c_r(t) d\mu(r) \leq 0, \quad (17)$$

where μ is the probability measure associated with R and the integration is over all possible r . We obtain the same result as in (5), but in a very different way.

In this paper we extend this approach to other mental architectures belonging to the class of series-parallel networks, those involving other composition operations and possibly more than just two selectively influenced processes. In doing so we follow a long trail of work mentioned earlier. When dealing with multiple processes we follow Yang, Fific, and Townsend (2013) in using high-order interaction contrasts. All our results are replications or straightforward generalizations of the results already known: the primary value of our work therefore is not in characterizing mental architectures, but rather in demonstrating a new theoretical approach and a new proof technique.

1.1 Definitions, Terminology, and Notation

Since we deal with the durations of processes rather than the processes themselves, we use the term *composition* to describe a function that relates the durations of the components of a network to the overall (observed) duration. Formally, a composition is a real-valued function $t = t(a, b, \dots, z)$ of an arbitrary number of real-valued arguments. The arguments a, b, \dots, z are referred to as *durations* or *components*. In this article, we will use $X \wedge Y \wedge \dots \wedge Z$ to denote $\min(X, Y, \dots, Z)$, and $X \vee Y \vee \dots \vee Z$ to denote $\max(X, Y, \dots, Z)$.

A *series-parallel composition* (SP) is defined as follows.

Definition 1.2. (1) A single duration is an SP composition. (2) If X and Y are SP compositions with disjoint sets of arguments, then $X \wedge Y$, $X \vee Y$, and $X + Y$ are SP compositions. (3) There are no other SP compositions than those construable by Rules 1 and 2.

Remark 1.3. The requirement that X and Y in Rule 2 have disjoint sets of arguments prevents expressions like $X \wedge X$ or $X + X \vee Y$. But if the second X in $X \wedge X$ is renamed into X' , or $X \vee Y$ in $X + X \vee Y$ is renamed into Z , then the resulting $X \wedge X'$ and $X + Z$ are legitimate SP compositions. This follows from the generality of our treatment, in which different components of an SP composition may have arbitrary joint distributions: e.g., X and X' in $X \wedge X'$ may very well be jointly distributed so that $\Pr[X = X'] = 1$. One should, however, always keep in mind the pattern of selective influences: thus, if X is influenced by α , then Z is also influenced by α in $X + Z$ above.

Any SP composition is obtained by a successive application of Rules 1 and 2 (the sequence being generally non-unique), and at any intermediate stage of such a sequence we also have an SP composition that we can term a *subcomposition*.

Definition 1.4. Two durations X, Y in an SP composition are said to be *parallel* or *concurrent* if there is a subcomposition of this SP composition of the form $\text{SP}^1(X, X', \dots) \wedge \text{SP}^2(Y, Y', \dots)$ (in which case X, Y are said to be *min-parallel*) or $\text{SP}^1(X, X', \dots) \vee \text{SP}^2(Y, Y', \dots)$ (X, Y are *max-parallel*). X, Y in an SP composition are said to be *sequential* or *serial* if there is a subcomposition of this SP composition of the form $\text{SP}^1(X, X', \dots) + \text{SP}^2(Y, Y', \dots)$.

Definition 1.5. An SP composition is called *homogeneous* if it does not contain both \wedge and \vee in it; if it does not contain \wedge , it is denoted SP_\vee ; if it does not contain \vee , it is denoted SP_\wedge .

The only SP composition that is both SP_\wedge and SP_\vee is a purely serial one: $a + b + \dots + z$. Most of the results previously obtained for mental networks are confined to homogeneous compositions. We will not need this constraint for the most part.

Since we will be dealing with compositions of more than just two components, we need to extend the definition of selective influences mentioned above. In the formulation below, \sim stands for “has the same distribution as.” A treatment $\phi = (\lambda_{i_1}^1, \dots, \lambda_{i_n}^n)$ is a vector of values of the factors $\lambda^1, \dots, \lambda^n$, the values of λ^k ($k = 1, \dots, n$) being indicated by subscripts, $\lambda_{i_k}^k$.

Definition 1.6. Random variables (X^1, \dots, X^n) are *selectively influenced* by factors $(\lambda^1, \dots, \lambda^n)$, respectively,

$$(X^1, \dots, X^n) \leftarrow_P (\lambda^1, \dots, \lambda^n), \quad (18)$$

if for some random variable R , whose distribution does not depend on $(\lambda^1, \dots, \lambda^n)$, and for some functions g_1, \dots, g_n ,

$$(X_\phi^1, \dots, X_\phi^n) \sim (g_1(\lambda_{i_1}^1, R), \dots, g_n(\lambda_{i_n}^n, R)), \quad (19)$$

for any treatment $\phi = (\lambda_{i_1}^1, \dots, \lambda_{i_n}^n)$.

In the subsequent discussion we assume that all non-dummy factors involved are binary in a completely crossed design (i.e., the overall time T is recorded for all 2^n vectors of values for ϕ). When we have random variables not influenced by any of these factors, we will say they selectively influenced by an empty set of factors (we could also, equivalently, introduce for them dummy factors, with one value each).

2 SP Compositions Containing Two Selectively Influenced Processes

Consider two processes, with durations A and B in an SP composition. The overall duration of this SP composition can be written as a function of A, B and other components: $T = T(A, B, \dots)$. We assume that A, B , and all other components are selectively influenced by α, β , and empty set, respectively: $(A, B, \dots) \leftarrow_P (\alpha, \beta, \emptyset)$. Let each factor has two levels: $\alpha = 1, 2$ and $\beta = 1, 2$, with four allowable treatments $(1, 1)$, $(1, 2)$, $(2, 1)$, and $(2, 2)$. The corresponding overall durations (random variables) are written as T_{11}, T_{12}, T_{21} , and T_{22} .

By Definition 1.6 of selective influences, each process duration (a random variable) is a function of some random variable R and the corresponding factor: $A = a(\alpha, R)$, $B = b(\beta, R)$. For any given value $R = r$, the component durations are fixed numbers,

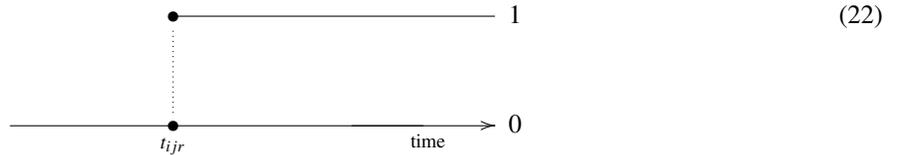
$$\begin{aligned} a(\alpha = 1, r) &= a_{1r}, & a(\alpha = 2, r) &= a_{2r}, \\ b(\beta = 1, r) &= b_{1r}, & b(\beta = 2, r) &= b_{2r}, \\ x(\emptyset, r) &= x_r, \end{aligned} \quad (20)$$

where x is the value of any duration X in the composition other than A and B . We assume that R is chosen so that the Prolongation Assumption (11) holds, with the convention (12).

The overall duration T at $R = r$ is also a fixed number, written as (recall that we replace α, β in indexation with i, j)

$$T(a_{ir}, \beta_{jr}, \dots) = t_{ijr}, i, j \in \{1, 2\}. \quad (21)$$

The distribution function for t_{ijr} is the shifted Heaviside function $h_{ijr}(t) = h(t - t_{ijr})$,



The *conditional interaction contrast* $c_r(t)$ is defined by (13). Denoting by $H_{ij}(t)$ the distribution function of T_{ij} , we have

$$H_{ij}(t) = \int_{\mathcal{R}} h_{ijr}(t) d\mu_r, \quad (23)$$

with \mathcal{R} denoting the set of possible values of R . For the observable (i.e., estimable from data) interaction contrast

$$C(t) = H_{11}(t) - H_{12}(t) - H_{21}(t) + H_{22}(t), \quad (24)$$

we have then

$$C(t) = \int_{\mathcal{R}} c_r(t) d\mu_r. \quad (25)$$

Note that it follows from our Prolongation Assumption that

$$H_{11}(t) \geq H_{12}(t), H_{21}(t) \geq H_{22}(t), H_{11}(t) \geq H_{21}(t), H_{12}(t) \geq H_{22}(t). \quad (26)$$

We also define two conditional cumulative interaction contrasts (conditioned on $R = r$):

$$c(0, t) = \int_0^t c(\tau) d\tau. \quad (27)$$

$$c(t, \infty) = \int_t^\infty c(\tau) d\tau = \lim_{u \rightarrow \infty} \int_t^u c(\tau) d\tau. \quad (28)$$

The corresponding observable cumulative interaction contrasts are

$$C(0, t) = \int_{\mathcal{R}} c(0, t) d\mu_r = \int_{\mathcal{R}} \left(\int_0^t c(\tau) d\tau \right) d\mu_r = \int_0^t \left(\int_{\mathcal{R}} c(\tau) d\mu_r \right) d\tau = \int_0^t C(\tau) d\tau. \quad (29)$$

$$C(t, \infty) = \int_{\mathcal{R}} c(t, \infty) d\mu_r = \int_{\mathcal{R}} \left(\int_t^\infty c(\tau) d\tau \right) d\mu_r = \int_t^\infty \left(\int_{\mathcal{R}} c(\tau) d\mu_r \right) d\tau = \int_t^\infty C(\tau) d\tau. \quad (30)$$

In these formulas we could switch the order of integration by Fubini's theorem, because, for any interval of reals I ,

$$\int_{I \times \mathcal{R}} |c(\tau)| d(\tau \times \mu_r) \leq \int_{I \times \mathcal{R}} 2 d(\tau \times \mu_r) \leq 2. \quad (31)$$

2.1 Four lemmas

Recall the definition of $c_r(t)$ in (13). We follow our Notation Convention and drop the index r in $c_r(t)$ and all other expressions for a fixed r .

Lemma 2.1. *In any SP architecture, for any r ,*

$$t_{11} \leq t_{12} \wedge t_{21} \leq t_{12} \vee t_{21} \leq t_{22}.$$

Proof. Follows from the (nonstrict) monotonicity of the SP composition in all arguments. \square

Lemma 2.2. *In any SP architecture, for any r , $c(t)$ equals 0 for all values of t except for two cases:*

(Case⁺) *if $t_{11} \leq t < t_{12} \wedge t_{21}$, then $c(t) = 1 - 0 - 0 + 0 > 0$,*

and

(Case⁻) *if $t_{12} \vee t_{21} \leq t < t_{22}$, then $c(t) = 1 - 1 - 1 + 0 < 0$.*

Proof. By direct computation. \square

Lemma 2.3. *In any SP architecture, for any r , $c(t) \leq 0$ for all values of t if and only if $t_{11} = t_{12} \wedge t_{21}$; $c(t) \geq 0$ for all values of t if and only if $t_{12} \vee t_{21} = t_{22}$.*

Proof. Immediately follows from Lemma 2.2. \square

Lemma 2.4. *In any SP architecture, for any r ,*

(i) $c(0, t) = \int_0^t c(\tau) d\tau \geq 0$ for any t if and only if $-t_{11} + t_{12} + t_{21} - t_{22} \geq 0$,

(ii) $c(t, \infty) = \int_t^\infty c(\tau) d\tau \leq 0$ for any t if and only if $-t_{11} + t_{12} + t_{21} - t_{22} \leq 0$,

(iii) $\lim_{t \rightarrow \infty} c(0, t) = 0$ if and only if $-t_{11} + t_{12} + t_{21} - t_{22} = 0$.

(iv) $\lim_{t \rightarrow 0} c(t, \infty) = 0$ if and only if $-t_{11} + t_{12} + t_{21} - t_{22} = 0$.

Proof. Without loss of generality, put $t_{12} \leq t_{21}$. We have

$$c(0, t) = \begin{cases} 0 & \text{if } t < t_{11} \\ (t - t_{11}) & \text{if } t_{11} \leq t < t_{12} \\ (t - t_{11}) - (t - t_{12}) = t_{12} - t_{11} & \text{if } t_{12} \leq t < t_{21} \\ (t - t_{11}) - (t - t_{12}) - (t - t_{21}) \\ \quad = -t_{11} + t_{12} + t_{21} - t & \text{if } t_{21} \leq t < t_{22} \\ (t - t_{11}) - (t - t_{12}) - (t - t_{21}) + (t - t_{22}) \\ \quad = -t_{11} + t_{12} + t_{21} - t_{22} & \text{if } t \geq t_{22} \end{cases}$$

The expressions for the first three cases are obviously nonnegative. If $-t_{11} + t_{12} + t_{21} - t_{22} \geq 0$, then $c(0, t) \geq 0$ for all t in the last case ($t \geq t_{22}$). With $-t_{11} + t_{12} + t_{21} - t_{22} \geq 0$, we have $-t_{11} + t_{12} + t_{21} - t \geq t_{22} - t \geq 0$ for the fourth case ($t_{21} \leq t < t_{22}$). Hence $c(0, t) \geq 0$ for all t if $-t_{11} + t_{12} + t_{21} - t_{22} \geq 0$. Conversely, if $c(0, t) \geq 0$ for all t , then it is also true for $t \geq t_{22}$, whence $-t_{11} + t_{12} + t_{21} - t_{22} \geq 0$.

The proof for $c(t, \infty) = \int_t^\infty c(\tau) d\tau$ requires replacing it first with $\int_t^u c(\tau) d\tau \leq 0$ for some $u > t_{22}$. We have

$$\int_t^u c(\tau) d\tau = \begin{cases} (u - t_{11}) - (u - t_{12}) - (u - t_{21}) + (u - t_{22}) \\ \quad = -t_{11} + t_{12} + t_{21} - t_{22} & \text{if } t < t_{11} \\ (u - t) - (u - t_{12}) - (u - t_{21}) + (u - t_{22}) \\ \quad = -t + t_{12} + t_{21} - t_{22} & \text{if } t_{11} \leq t < t_{12} \\ (u - t) - (u - t) - (u - t_{21}) + (u - t_{22}) \\ \quad = t_{21} - t_{22} & \text{if } t_{12} \leq t < t_{21} \\ (u - t) - (u - t) - (u - t) + (u - t_{22}) \\ \quad = t - t_{22} & \text{if } t_{21} \leq t < t_{22} \\ (u - t) - (u - t) - (u - t) + (u - t) \\ \quad = 0 & \text{if } t \geq t_{22} \end{cases}$$

The expressions for the last three cases are obviously nonpositive. If $-t_{11} + t_{12} + t_{21} - t_{22} \leq 0$, then $\int_t^u c(\tau) d\tau \leq 0$ for all t in the first case ($t < t_{11}$). With $-t_{11} + t_{12} + t_{21} - t_{22} \leq 0$, we have $-t + t_{12} + t_{21} - t_{22} \leq t_{11} - t < 0$ for the second case ($t_{11} \leq t < t_{12}$). Hence $\int_t^u c(\tau) d\tau \leq 0$ for all t if $-t_{11} + t_{12} + t_{21} - t_{22} \leq 0$. Since in all expressions u is algebraically eliminated, they remain unchanged as $u \rightarrow \infty$. Conversely, if $c(t, \infty) \leq 0$ for all t , then it is also true for $t < t_{11}$, whence $-t_{11} + t_{12} + t_{21} - t_{22} \leq 0$.

The statements (iii) and (iv) follow trivially. \square

2.2 Parallel Processes

2.2.1 Simple Parallel Architectures of Size 2

A simple parallel architecture corresponds to one of the two compositions: $T = A \wedge B$ or $T = A \vee B$, with $(A, B) \leftrightarrow_P (\alpha, \beta)$. Recall the definition of $C(t)$ in (24).

Theorem 2.5. *For $T = A \wedge B$, we have $c(t) \leq 0$ for any r, t ; consequently, $C(t) \leq 0$ for any t . For $T = A \vee B$, we have $c(t) \geq 0$ for any r, t ; consequently, $C(t) \geq 0$ for any t .*

Proof. For $T = A \wedge B$ with the Prolongation Assumption (11)-(12), we have

$$t_{11} = a_1 \wedge b_1 = a_1, t_{12} = a_1 \wedge b_2, t_{21} = a_2 \wedge b_1.$$

It follows that

$$t_{12} \wedge t_{21} = a_1 \wedge b_2 \wedge a_2 \wedge b_1 = a_1 = t_{11}.$$

By Lemma 2.3, $c(t) \leq 0$. As $C(t)$ in (25) preserves the sign of $c(t)$, we have $C(t) \leq 0$. For $T = A \vee B$, we have

$$t_{22} = a_2 \vee b_2, t_{12} = a_1 \vee b_2, t_{21} = a_2 \vee b_1.$$

It follows that

$$t_{12} \vee t_{21} = a_1 \vee b_2 \vee a_2 \vee b_1 = t_{22},$$

whence, by Lemma 2.3, $c(t) \geq 0$ and therefore $C(t) \geq 0$. \square

2.2.2 Two Parallel Processes in an Arbitrary SP Network

Consider now a composition $SP(A, B, \dots)$ with $(A, B, \dots) \leftrightarrow (\alpha, \beta, \emptyset)$.

Lemma 2.6. *If A, B in $SP(A, B, \dots)$ are parallel, then $SP(A, B, \dots)$ can be presented as $A' \wedge B'$ if they are min-parallel, or as $A' \vee B'$ if they are max-parallel, so that $(A', B') \leftrightarrow (\alpha, \beta)$ and, for any fixed $R = r$, the Prolongation Assumption holds.*

Proof. By Definitions 1.2 and 1.4, if A, B are min-parallel, then $SP_{\wedge}(A, B, \dots)$ can be presented either as

$$SP^1(A, \dots) \wedge SP^2(B, \dots)$$

or

$$(SP^1(A, \dots) \wedge SP^2(B, \dots) + X) \wedge Y,$$

or else

$$(SP^1(A, \dots) \wedge SP^2(B, \dots) \wedge X) + Y,$$

where B does not enter in SP^1 and A does not enter in SP^2 . On renaming

$$\underbrace{SP^1(A, \dots)}_{=A'} \wedge \underbrace{SP^2(B, \dots)}_{=B'},$$

$$(SP^1(A, \dots) \wedge SP^2(B, \dots) + X) \wedge Y = \underbrace{(SP^1(A, \dots) + X)}_{=A'} \wedge \underbrace{(SP^2(B, \dots) + X)}_{=B'} \wedge Y,$$

and

$$(SP^1(A, \dots) \wedge SP^2(B, \dots) \wedge X) + Y = \underbrace{(SP^1(A, \dots) + Y)}_{=A'} \wedge \underbrace{(SP^2(B, \dots) \wedge X + Y)}_{=B'},$$

we have, obviously, $(A', B') \leftrightarrow (\alpha, \beta)$. Fixing $R = r$, by the (nonstrict) monotonicity of SP compositions,

$$a'_1 = SP^1(a_1, \dots) \leq SP^1(a_2, \dots) = a'_2$$

and

$$b'_1 = SP^2(b_1, \dots) \leq SP^2(b_2, \dots) = b'_2$$

We can also put $a'_1 = SP^1(a_1, \dots) \leq SP^2(b_1, \dots) = b'_1$ (otherwise we can rename the variables). The proof for the max-parallel case is analogous. \square

Theorem 2.7. *If A, B in $SP(A, B, \dots)$ are min-parallel, then $c(t) \leq 0$ for any r, t ; consequently, $C(t) \leq 0$ for any t . If A, B in $SP(A, B, \dots)$ are max-parallel, then $c(t) \geq 0$ for any r, t ; consequently, $C(t) \geq 0$ for any t .*

Proof. Immediately follows from Lemma 2.6 and Theorem 2.5. \square

2.3 Sequential Processes

2.3.1 Simple Serial Architectures of Size 2

Simple serial architectures of size 2 corresponds to the SP composition $T = A + B$, with $(A, B) \leftrightarrow (\alpha, \beta)$. Recall the definitions of the two cumulative interaction contrasts: (27)-(28) and (29)-(30).

Theorem 2.8. *If $T = A + B$, then $c(0, t) \geq 0$ and $c(t, \infty) \leq 0$ for any r, t ; moreover,*

$$\lim_{t \rightarrow \infty} c(0, t) = \lim_{t \rightarrow 0} c(t, \infty) = 0,$$

for any r, t . Consequently, $C(0, t) \geq 0$, $C(t, \infty) \leq 0$ for any t , and

$$\lim_{t \rightarrow \infty} C(0, t) = \lim_{t \rightarrow 0} C(t, \infty) = 0$$

Proof. Follows immediately from Lemma 2.4, since

$$-t_{11} + t_{12} + t_{21} - t_{22} = -(a_1 + b_1) + (a_1 + b_2) + (a_2 + b_1) - (a_2 + b_2) = 0.$$

□

2.3.2 Two Sequential Processes in an Arbitrary SP Network

Consider now a composition $\text{SP}(A, B, \dots)$ with $(A, B, \dots) \leftarrow (\alpha, \beta, \emptyset)$.

Theorem 2.9. *If A and B are sequential in an $\text{SP}(A, B, \dots)$ composition, then one or both of the following statements hold:*

- (i) $c(0, t) \geq 0$ for any r, t , and $C(0, t) \geq 0$ for any t ,
- (ii) $c(t, \infty) \leq 0$ for any r, t , and $C(t, \infty) \leq 0$ for any t .

Proof. In accordance with Definitions 1.2 and 1.4, $\text{SP}(A, B, \dots)$ with sequential A, B can be presented as either

$$(\text{SP}^1(A, \dots) + \text{SP}^2(B, \dots)) \wedge X + Y \quad (32)$$

or

$$(\text{SP}^1(A, \dots) + \text{SP}^2(B, \dots)) \vee X + Y \quad (33)$$

(note that any Z in $\text{SP}^1(A, \dots) + \text{SP}^2(B, \dots) + Z$ can be absorbed by either of the first two summands). For both cases, by the monotonicity of SP compositions, for any $R = r$, $\text{SP}^1(a_1, \dots) \leq \text{SP}^1(a_2, \dots)$, $\text{SP}^2(b_1, \dots) \leq \text{SP}^2(b_2, \dots)$, and we can always assume $\text{SP}^1(a_1, \dots) \leq \text{SP}^2(b_1, \dots)$. Denoting the durations of $\text{SP}^1(a_i, \dots) + \text{SP}^2(b_j, \dots)$ by t'_{ij} , we have therefore, by Theorem 2.8, $-t'_{11} + t'_{12} + t'_{21} - t'_{22} = 0$. Denoting the durations of X and Y by t' and t'' , respectively, in the case (32) we have

$$t_{ij} = t'_{ij} \wedge t' + t''.$$

By Lemma 2.4, all we have to show is that $-t_{11} + t_{12} + t_{21} - t_{22} \geq 0$. It is easy to see that t'' does not affect this linear combination, and its value is (assuming $t'_{12} \leq t'_{21}$, without loss of generality)

$$\begin{cases} 0 & \text{if } t' < t'_{11} \\ -t'_{11} + t' & \text{if } t'_{11} \leq t' < t'_{12} \\ -t'_{11} + t'_{12} & \text{if } t'_{12} \leq t' < t'_{21} \\ -t'_{11} + t'_{12} + t'_{21} - t' & \text{if } t'_{21} \leq t' < t'_{22} \\ -t'_{11} + t'_{12} + t'_{21} - t'_{22} & \text{if } t' \geq t'_{22}. \end{cases}$$

The nonnegativity of the first three expressions is obvious, the fifth one is zero, and the fourth expression is larger than the fifth because $t' < t'_{22}$.

The proof for the case (33) is analogous. □

If the SP composition with sequential A, B is homogeneous (Definition 1.5), the statement of theorem can be made more specific.

Theorem 2.10. *If A and B are sequential in an $\text{SP}_{\wedge}(A, B, \dots)$ composition, then $c(0, t) \geq 0$ for any r, t , and $C(0, t) \geq 0$ for any t ; if the composition is $\text{SP}_{\vee}(A, B, \dots)$, then $c(t, \infty) \leq 0$ for any r, t , and $C(t, \infty) \leq 0$ for any t .*

3 Multiple Processes

We now turn to networks containing $n \geq 2$ processes with durations (X^1, \dots, X^n) , selectively influenced by factors $(\lambda^1, \dots, \lambda^n)$. In other words, we deal with compositions $\text{SP}(X^1, \dots, X^n, \dots)$ such that $(X^1, \dots, X^n, \dots) \leftarrow \rho (\lambda^1, \dots, \lambda^n, \emptyset)$, where each λ^k is binary, with values 1,2. There are 2^n allowable treatments and 2^n corresponding overall durations, $T_{11\dots 1}, T_{11\dots 2}, \dots, T_{22\dots 2}$. According to Definition 1.6 of selective influences, each process duration here is a function of some random variable R and of the corresponding factor, $X^k = x^k(R, \lambda^k)$. For any fixed value $R = r$, these durations are fixed numbers for any given treatment, and so is the overall, observed value of the SP composition. We denote them

$$x^k(r, \lambda^k = 1) = x_{1r}^k, x^k(r, \lambda^k = 2) = x_{2r}^k, \quad (34)$$

and

$$T(x_{1r}^1, x_{2r}^2, \dots, x_{i_n r}^n, \dots), \dots = t_{i_1 i_2 \dots i_n r}, \quad (35)$$

where $i_1, i_2, \dots, i_n \in \{1, 2\}$. The distribution function for $t_{i_1 i_2 \dots i_n r}$ is a shifted Heaviside function

$$h_{i_1 i_2 \dots i_n r}(t) = \begin{cases} 0, & \text{if } t < t_{i_1 i_2 \dots i_n r} \\ 1, & \text{if } t \geq t_{i_1 i_2 \dots i_n r} \end{cases}. \quad (36)$$

Denoting by $H_{i_1 i_2 \dots i_n}(t)$ the distribution function of $T_{i_1 i_2 \dots i_n}$, we have

$$H_{i_1 i_2 \dots i_n}(t) = \int_{\mathcal{R}} h_{i_1 i_2 \dots i_n r}(t) d\mu_r. \quad (37)$$

Conditioned on $R = r$, the n -th order interaction contrast is defined in terms of mixed finite differences as

$$c_r^{(n)}(t) = \Delta_{i_1} \Delta_{i_2} \dots \Delta_{i_n} h_{i_1 i_2 \dots i_n r}(t), \quad (38)$$

which, with some algebra can be shown to be equal to

$$c_r^{(n)}(t) = \sum_{i_1, i_2, \dots, i_n} (-1)^{n + \sum_{k=1}^n i_k} h_{i_1 \dots i_n r}(t). \quad (39)$$

Thus,

$$c_r^{(1)}(t) = \Delta_{i_1} h_{i_1 r}(t) = h_{1r}(t) - h_{2r}(t) = \sum_{i_1} (-1)^{1+i_1} h_{i_1 r}(t), \quad (40)$$

$$\begin{aligned} c_r^{(2)}(t) &= \Delta_{i_1} \Delta_{i_2} h_{i_1 i_2 r}(t) = [h_{11r}(t) - h_{12r}(t)] - [h_{21r}(t) - h_{22r}(t)] \\ &= h_{11r}(t) - h_{12r}(t) - h_{21r}(t) + h_{22r}(t) = \sum_{i_1, i_2} (-1)^{2+i_1+i_2} h_{i_1 i_2 r}(t), \end{aligned} \quad (41)$$

$$\begin{aligned} c_r^{(3)}(t) &= \Delta_{i_1} \Delta_{i_2} \Delta_{i_3} h_{i_1 i_2 i_3 r}(t) \\ &= \{[h_{111r}(t) - h_{112r}(t)] - [h_{121r}(t) - h_{122r}(t)]\} - \{[h_{211r}(t) - h_{212r}(t)] - [h_{221r}(t) - h_{222r}(t)]\} \\ &= h_{111r}(t) - h_{112r}(t) - h_{121r}(t) - h_{211r}(t) + h_{122r}(t) + h_{212r}(t) + h_{221r}(t) - h_{222r}(t) \\ &= \sum_{i_1, i_2, i_3} (-1)^{3+i_1+i_2+i_3} h_{i_1 i_2 i_3 r}(t), \end{aligned} \quad (42)$$

etc. The observable distribution function interaction contrast of order n is defined as

$$C^{(n)}(t) = \int_{\mathcal{R}} c_r^{(n)}(t) d\mu_r. \quad (43)$$

By straightforward calculus this can be written in extenso as

$$C^{(n)}(t) = \sum_{i_1, i_2, \dots, i_n} (-1)^{n + \sum_{k=1}^n i_k} H_{i_1 \dots i_n}(t), \quad (44)$$

or, in terms of finite differences,

$$C^{(n)}(t) = \Delta_{i_1} \Delta_{i_2} \dots \Delta_{i_n} H_{i_1 i_2 \dots i_n}(t). \quad (45)$$

This is essentially the high-order interaction contrast used by Yang, Fific, and Townsend (2013), the only difference being that they use survivor functions $1 - H(t)$ rather than the distribution functions $H(t)$. We see that $c_r(t)$ and $C(t)$ in the preceding analysis correspond to $c_r^{(2)}(t)$ and $C^{(2)}(t)$, respectively.

We also introduce *n-th order cumulative contrasts*. Conditioned on $R = r$, we define

$$c_r^{[1]}(0, t) = c_r^{[1]}(t, \infty) = h_{1r}(t) - h_{2r}(t), \quad (46)$$

$$c_r^{[2]}(0, t) = \int_0^t c_r^{(2)}(t_1) dt_1, \quad c_r^{[2]}(t, \infty) = \int_t^\infty c_r^{(2)}(t_1) dt_1, \quad (47)$$

$$c_r^{[3]}(0, t) = \int_0^t \int_0^{t_1} c_r^{(3)}(t_2) dt_2 dt_1, \quad c_r^{[3]}(t, \infty) = \int_t^\infty \int_{t_1}^\infty c_r^{(3)}(t_2) dt_2 dt_1, \quad (48)$$

etc. Generalizing,

$$c_r^{[n]}(0, t) = \int_0^t \left(\int_0^{t_1} \dots \int_0^{t_{n-2}} c_r^{(n)}(t_{n-1}) dt_{n-1} \dots dt_2 \right) dt_1, \quad (49)$$

$$c_r^{[n]}(t, \infty) = \int_t^\infty \left(\int_{t_1}^\infty \dots \int_{t_{n-2}}^\infty c_r^{(n)}(t_{n-1}) dt_{n-1} \dots dt_2 \right) dt_1. \quad (50)$$

The corresponding unconditional cumulative contrasts of the *n-th* order are, as always, defined by integration of the conditional ones:

$$C^{[n]}(0, t) = \int_{\mathcal{R}} c_r^{[n]}(0, t) d\mu_r = \int_0^t \left(\int_0^{t_1} \dots \int_0^{t_{n-2}} C^{(n)}(t_{n-1}) dt_{n-1} \dots dt_2 \right) dt_1, \quad (51)$$

$$C^{[n]}(t, \infty) = \int_{\mathcal{R}} c_r^{[n]}(t, \infty) d\mu_r = \int_t^\infty \left(\int_{t_1}^\infty \dots \int_{t_{n-2}}^\infty C^{(n)}(t_{n-1}) dt_{n-1} \dots dt_2 \right) dt_1. \quad (52)$$

In the proofs below we will make use of the recursive representation of the conditional cumulative contrasts $c_r^{[n]}$. It is verified by straightforward calculus. Denoting

$$c_{i_w r}^{(n-1)}(t) = \sum_{i_1, \dots, i_{w-1}, i_{w+1}, \dots, i_n} (-1)^{n-1-i_w+\sum_{k=1}^n i_k} h_{i_1 \dots i_{w-1} i_w i_{w+1} \dots i_n r}(t), \quad (53)$$

where $w \in \{1, \dots, n\}$ and i_w is fixed at 1 or 2, we have:

$$c_r^{[1]}(0, t) = c_r^{[1]}(t, \infty) = h_{1r}(t) - h_{2r}(t), \quad (54)$$

$$\begin{aligned} c_r^{[2]}(0, t) &= \int_0^t c_r^{(2)}(\tau) d\tau \\ &= \int_0^t (h_{11r}(\tau) - h_{12r}(\tau) - h_{21r}(\tau) + h_{22r}(\tau)) d\tau \\ &= \int_0^t [c_{i_w=1,r}^{(1)}(\tau) - c_{i_w=2,r}^{(1)}(\tau)] d\tau \\ &= \int_0^t c_{i_w=1,r}^{[1]}(0, \tau) d\tau - \int_0^t c_{i_w=2,r}^{[1]}(0, \tau) d\tau, \end{aligned} \quad (55)$$

$$\begin{aligned}
c_r^{[2]}(t, \infty) &= \int_t^\infty c_r^{(2)}(\tau) d\tau \\
&= \int_t^\infty (h_{11r}(\tau) - h_{12r}(\tau) - h_{21r}(\tau) + h_{22r}(\tau)) d\tau \\
&= \int_t^\infty [c_{i_w=1,r}^{(1)}(\tau) - c_{i_w=2,r}^{(1)}(\tau)] d\tau \\
&= \int_t^\infty c_{i_w=1,r}^{[1]}(\tau, \infty) d\tau - \int_t^\infty c_{i_w=2,r}^{[1]}(\tau, \infty) d\tau,
\end{aligned} \tag{56}$$

$$\begin{aligned}
c_r^{[3]}(0, t) &= \int_0^t \int_0^{t_1} c_r^{(3)}(t_2) dt_2 dt_1 \\
&= \int_0^t \int_0^{t_1} [c_{i_w=1,r}^{(2)}(t_2) - c_{i_w=2,r}^{(2)}(t_2)] dt_2 dt_1 \\
&= \int_0^t \left[\int_0^{t_1} c_{i_w=1,r}^{(2)}(t_2) dt_2 - \int_0^{t_1} c_{i_w=2,r}^{(2)}(t_2) dt_2 \right] dt_1 \\
&= \int_0^t c_{i_w=1,r}^{[2]}(0, \tau) d\tau - \int_0^t c_{i_w=2,r}^{[2]}(0, \tau) d\tau,
\end{aligned} \tag{57}$$

$$\begin{aligned}
c_r^{[3]}(t, \infty) &= \int_t^\infty \int_{t_1}^\infty c_r^{(3)}(t_2) dt_2 dt_1 \\
&= \int_t^\infty \int_{t_1}^\infty [c_{i_w=1,r}^{(2)}(t_2) - c_{i_w=2,r}^{(2)}(t_2)] dt_2 dt_1 \\
&= \int_t^\infty \left[\int_{t_1}^\infty c_{i_w=1,r}^{(2)}(t_2) dt_2 - \int_{t_1}^\infty c_{i_w=2,r}^{(2)}(t_2) dt_2 \right] dt_1 \\
&= \int_t^\infty c_{i_w=1,r}^{[2]}(\tau, \infty) d\tau - \int_t^\infty c_{i_w=2,r}^{[2]}(\tau, \infty) d\tau,
\end{aligned} \tag{58}$$

and generally, for $n > 1$,

$$c_r^{[n]}(0, t) = \int_0^t c_{i_w=1,r}^{[n-1]}(0, \tau) d\tau - \int_0^t c_{i_w=2,r}^{[n-1]}(0, \tau) d\tau, \tag{59}$$

$$c_r^{[n]}(t, \infty) = \int_t^\infty c_{i_w=1,r}^{[n-1]}(\tau, \infty) d\tau - \int_t^\infty c_{i_w=2,r}^{[n-1]}(\tau, \infty) d\tau. \tag{60}$$

Also we have, by substitution of variables under integral,

$$c_{i_w r}^{[n-1]}(0, t) = c_r^{[n-1]}(0, t - x_{i_w r}^w), \tag{61}$$

$$c_{i_w r}^{[n-1]}(t, \infty) = c_r^{[n-1]}(t - x_{i_w r}^w, \infty). \tag{62}$$

The Prolongation Assumption generalizing (11)-(12) is formulated as follows.

Prolongation Assumption. R and functions x^1, \dots, x^n in (34) can be chosen so that $x_{1r}^k \leq x_{2r}^k$ for all $R = r$ and for all $k = 1, \dots, n$. Without loss of generality, we can also assume $x_{1r}^1 \leq x_{1r}^2 \leq \dots \leq x_{1r}^n$ (if not, rearrange $x_{1r}^1, \dots, x_{1r}^n$).

Notation Convention. As we did before for $n = 2$, when r is fixed throughout a discussion, we omit this argument and write $x_{i_1}^1, \dots, x_{i_n}^n, t_{i_1 i_2 \dots i_n}, h_{i_1 i_2 \dots i_n}(t), c^{(n)}(t)$ in place of $x_{i_1 r}^1, \dots, x_{i_n r}^n, t_{i_1 i_2 \dots i_n r}, h_{i_1 i_2 \dots i_n r}(t), c_r^{(n)}(t)$.

3.1 Parallel Processes

3.1.1 Simple Parallel Architectures of Size n

Theorem 3.1. *If $T = X^1 \wedge \dots \wedge X^n$, then for any r, t , $c^{(n)}(t) \leq 0$ if n is even and $c^{(n)}(t) \geq 0$ if n is odd. Consequently, for any t , $C^{(n)}(t) \leq 0$ if n is even and $C^{(n)}(t) \geq 0$ if n is odd.*

Proof. By induction on n , the case $n = 1$ being true by the Prolongation Assumption:

$$c^{(1)}(t) = h_1(t) - h_2(t) \geq 0.$$

Let the statement of the theorem be true for $c^{(n-1)}(t)$, with $n - 1 \geq 1$. By the Prolongation Assumption,

$$t_{i_1 i_2 \dots i_n} = x_1^1 \wedge x_{i_2}^2 \wedge \dots \wedge x_{i_n}^n = x_1^1,$$

for any $i_2 \dots i_n$, whence

$$h_{i_1 i_2 \dots i_n}(t) = \begin{cases} 0, & \text{if } t < x_1^1 \\ 1, & \text{if } t \geq x_1^1 \end{cases}.$$

Therefore $c_{i_1=1}^{(n-1)}(t) = 0$, and, applying the induction hypothesis to $c_{i_1=2}^{(n-1)}(t)$,

$$c^{(n)}(t) = c_{i_1=1}^{(n-1)}(t) - c_{i_1=2}^{(n-1)}(t) = -c_{i_1=2}^{(n-1)}(t) = \begin{cases} \leq 0, & \text{if } n \text{ is even} \\ \geq 0, & \text{if } n \text{ is odd} \end{cases}.$$

That $C^{(n)}(t) \leq 0$ if n is even and $C^{(n)}(t) \geq 0$ if n is odd follows by the standard argument. \square

Theorem 3.2. *If $T = X^1 \vee \dots \vee X^n$, then for any r, t , $c^{(n)}(t) \geq 0$. Consequently, for any t , $C^{(n)}(t) \geq 0$.*

Proof. By induction on n , the case $n = 1$ being true by the Prolongation Assumption:

$$c^{(1)}(t) = h_1(t) - h_2(t) \geq 0.$$

Let the theorem be true for $c^{(n-1)}(t)$, where $n - 1 \geq 1$. Let

$$x_2^1 \vee x_2^2 \vee \dots \vee x_2^n = x_2^m,$$

where $1 \leq m \leq n$. We have then

$$t_{i_1 i_2 \dots i_{m-1} i_{m+1} \dots i_n} = x_2^m,$$

and

$$h_{i_1 \dots i_{m-1} i_{m+1} \dots i_n}(t) = \begin{cases} 0, & \text{if } t < x_2^m \\ 1, & \text{if } t \geq x_2^m \end{cases},$$

for all $i_1 \dots i_{m-1}, i_{m+1} \dots i_n$. Then $c_{i_m=2}^{(n-1)}(t) = 0$, and

$$c^{(n)}(t) = c_{i_m=1}^{(n-1)}(t) - c_{i_m=2}^{(n-1)}(t) = c_{i_m=1}^{(n-1)}(t) \geq 0.$$

Consequently, $C^{(n)}(t) \geq 0$, for any t . \square

3.1.2 Multiple Parallel Processes in Arbitrary SP Networks

In a composition $SP(X^1, \dots, X^n, \dots)$, the components X^1, \dots, X^n are considered parallel if any two of them are parallel. We assume selective influences $(X^1, \dots, X^n, \dots) \leftarrow \varphi(\lambda^1, \dots, \lambda^n, \emptyset)$. We do not consider the complex situation when some of the selectively influenced processes X^1, \dots, X^n are min-parallel and some are max-parallel. However, if they are all (pairwise) min-parallel or all max-parallel, we have essentially the same situation as with a simple parallel arrangement of n durations.

Lemma 3.3. *If X^1, \dots, X^n are all min-parallel or max-parallel in an SP composition, this composition can be presented as $T = A^1 \wedge \dots \wedge A^n$ or $T = A^1 \vee \dots \vee A^n$, respectively. In either case, $(A^1, \dots, A^n) \leftrightarrow (\lambda^1, \dots, \lambda^n)$ and the Prolongation Assumption holds for any $R = r$.*

Proof. For the min-parallel case, by a minor modification of the proof of Lemma 2.6 we present the SP composition as

$$\underbrace{\text{SP}^1(X^1, \dots)}_{=A^1} \wedge \text{SP}^2(X^2, \dots, X^n, \dots),$$

or

$$(\text{SP}^1(X^1, \dots) \wedge \text{SP}^2(X^2, \dots, X^n, \dots) + X) \wedge Y = \underbrace{(\text{SP}^1(X^1, \dots) + X)}_{=A^1} \wedge (\text{SP}^2(X^2, \dots, X^n, \dots) + X) \wedge Y,$$

or else

$$(\text{SP}^1(X^1, \dots) \wedge \text{SP}^2(X^2, \dots, X^n, \dots) \wedge X) + Y = \underbrace{(\text{SP}^1(X^1, \dots) + Y)}_{=A^1} \wedge (\text{SP}^2(X^2, \dots, X^n, \dots) \wedge X + Y).$$

Then we analogously decompose $\text{SP}^2(X^2, \dots, X^n, \dots)$ achieving $A^1 \wedge A^2 \wedge \text{SP}^3(X^3, \dots, X^n, \dots)$, and proceed in this fashion until we reach the required $A^1 \wedge \dots \wedge A^n$. The pattern of selective influences is seen immediately, and the Prolongation Assumption follows by the monotonicity of the SP compositions. The proof for the max-parallel case is analogous. \square

Theorem 3.4. *If X^1, \dots, X^n are min-parallel in an SP composition, then for any r, t , $c^{(n)}(t) \leq 0$ if n is even and $c^{(n)}(t) \geq 0$ if n is odd. Consequently, for any t , $C^{(n)}(t) \leq 0$ if n is even and $C^{(n)}(t) \geq 0$ if n is odd. If X^1, \dots, X^n are max-parallel, then for any r, t , $c^{(n)}(t) \geq 0$, and for any t , $C^{(n)}(t) \geq 0$.*

Proof. Follows from Lemma 3.3 and Theorems 3.1 and 3.2. \square

3.2 Sequential Processes

3.2.1 Simple Serial Architectures of Size n

Theorem 3.5. *If $T = X^1 + \dots + X^n$, then for any r, t , $c^{[n]}(0, t) \geq 0$, while $c^{[n]}(t, \infty) \leq 0$ if n is even and $c^{[n]}(t, \infty) \geq 0$ if n is odd; moreover, $c^{[n]}(0, \infty) = 0$ for any r . Consequently, for any t , $C^{[n]}(0, t) \geq 0$, while $C^{[n]}(t, \infty) \leq 0$ if n is even and $C^{[n]}(t, \infty) \geq 0$ if n is odd; moreover, $C^{[n]}(0, \infty) = 0$.*

Proof. By induction on n , the case $n = 1$ being true by the Prolongation Assumption:

$$c^{[1]}(0, t) = c^{[1]}(t, \infty) = h_1(t) - h_2(t) \geq 0,$$

and

$$\lim_{t \rightarrow \infty} c^{[1]}(0, t) = \lim_{t \rightarrow 0} c^{[1]}(t, \infty) = 0.$$

Let the statement of the theorem hold for all natural numbers up to and including $n - 1 \geq 1$. Using the recursive representations (59)-(60),

$$\begin{aligned} c^{[n]}(0, t) &= \int_0^t c_{i_w=1}^{[n-1]}(0, \tau) d\tau - \int_0^t c_{i_w=2}^{[n-1]}(0, \tau) d\tau \\ &= \int_0^{t-x_1^w} c^{[n-1]}(0, \tau) d\tau - \int_0^{t-x_2^w} c^{[n-1]}(0, \tau) d\tau, \\ &= \int_{t-x_2^w}^{t-x_1^w} c^{[n-1]}(0, \tau) d\tau \end{aligned} \quad (63)$$

which is ≥ 0 since $c^{[n-1]}(0, \tau) \geq 0$ and $t - x_2^w \leq t - x_1^w$. Analogously,

$$\begin{aligned} c^{[n]}(t, \infty) &= \int_t^\infty c_{i_w=1}^{[n-1]}(\tau, \infty) d\tau - \int_t^\infty c_{i_w=2}^{[n-1]}(\tau, \infty) d\tau \\ &= \int_{t-x_1^w}^\infty c^{[n-1]}(\tau, \infty) d\tau - \int_{t-x_2^w}^\infty c^{[n-1]}(\tau, \infty) d\tau, \\ &= - \int_{t-x_2^w}^{t-x_1^w} c^{[n-1]}(\tau, \infty) d\tau \end{aligned} \quad (64)$$

which is ≤ 0 if n is even and ≥ 0 if n is odd. Applying the mean value theorem to the results of (63) and (64), we get, for some $t - x_2^w < t', t'' < t - x_1^w$

$$\int_{t-x_2^w}^{t-x_1^w} c^{[n-1]}(0, \tau) d\tau = c^{[n-1]}(0, t')(-x_1^w + x_2^w),$$

$$\int_{t-x_2^w}^{t-x_1^w} c^{[n-1]}(\tau, \infty) d\tau = c^{[n-1]}(t'', \infty)(-x_1^w + x_2^w),$$

and, as $c^{[n-1]}(0, \infty) = 0$, both expressions tend to zero as, respectively, $t \rightarrow \infty$ (implying $t' \rightarrow \infty$) and $t \rightarrow 0$ (implying $t'' \rightarrow 0$). \square

3.2.2 Multiple Sequential Processes in Arbitrary SP Networks

In a composition $SP(X^1, \dots, X^n, \dots)$, the components X^1, \dots, X^n are considered sequential if any two of them are sequential. By analogy with Theorem 2.9 for two sequential processes and with Theorem 3.4 for parallel X^1, \dots, X^n , one might expect that the result for the simple sequential arrangement $X^1 + \dots + X^n$ will also extend to n sequential components of more complex compositions $SP(X^1, \dots, X^n, \dots)$. However, this is not the case, as one can see from the following counterexample.

Consider the composition

$$SP(X^1, X^2, X^3, Y) = (X^1 + X^2 + X^3) \wedge (Y = 2), \quad (65)$$

with (X^1, X^2, X^3) selectively influenced by binary factors, so that

$$\begin{aligned} x_1^1 &= x_1^2 = x_1^3 = 0, \\ x_2^1 &= x_2^2 = x_2^3 = 1. \end{aligned} \quad (66)$$

It follows that

$$\begin{aligned} t_{111} &= 0, \\ t_{112} &= t_{121} = t_{211} = 1, \\ t_{122} &= t_{212} = t_{221} = t_{222} = 2. \end{aligned} \quad (67)$$

This is clearly a sequential arrangement of the three durations X^1, X^2, X^3 , but one can easily check that $c^{[3]}(0, t)$ here is not nonnegative for all t . For instance, at $t=3$ we have, by direct computation, $c^{[3]}(0, t) = -1$. We conclude that there is no straightforward generalization of Theorems 3.5 to arbitrary SP compositions.

4 Conclusion

The work presented in this paper is summarized in the abstract. By proving and generalizing most of the known results on the interaction contrast of distribution functions, we have demonstrated a new way of dealing with SP mental architectures. It is based on conditioning all hypothetical components of a network on a fixed value of a common source of randomness R (the ‘‘hidden variable’’ of the contextuality analysis in quantum theory), which renders these components deterministic quantities, and then treating these deterministic quantities as random variables with shifted Heaviside distribution functions.

The potential advantage of this method can be seen in the fact that the shifted Heaviside functions have the simplest possible arithmetic among distribution functions: for every time moment it only involves 0’s and 1’s. As a result, the complexity of this arithmetic does not increase with nonlinearity of the relations involved. Thus, Dzhafarov and Schweickert (1995), Cortese and Dzhafarov (1996), and Dzhafarov and Cortese (1996) argued that composition rules for mental architectures need not be confined to $+$, \wedge , \vee . They analyzed architectures involving other associative and commutative operations, such as multiplication. Due to mathematical complexity, however, this work was confined to networks consisting of two components that are either stochastically independent or monotone functions of each other. It remains to be seen whether the approach presented here, *mutatis mutandis*, will lead to significant generalizations in this line of work.

The limitations of the approach, however, are already apparent. Thus, we were not able to achieve any progress over known results in applying it to Wheatstone bridges (Schweickert & Giorgini, 1999; Dzhafarov et al., 2004).

The possibility that the “architecture” (composition rule) itself changes as one changes experimental factors makes the perspective of a general theory based on our approach even more problematic (e.g., Townsend & Fific, 2004). It seems, however, that these problems are not specific for just our approach.

Acknowledgments

This work is supported by NSF grant SES-1155956 and AFOSR grant FA9550-14-1-0318.

REFERENCES

- Bell, J. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics*, 1, 195-200.
- Bell, J. (1966). On the problem of hidden variables in quantum mechanics. *Review of Modern Physics*, 38, 447-453.
- Cereceda, J. (2000). Quantum mechanical probabilities and general probabilistic constraints for Einstein–Podolsky–Rosen–Bohm experiments. *Foundations of Physics Letters*, 13, 427–442.
- 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. (2003). Selective influence through conditional independence. *Psychometrika*, 68, 7–26.
- 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., & Gluhovsky, I. (2006). Notes on selective influence, probabilistic causality, and probabilistic dimensionality. *Journal of Mathematical Psychology*, 50, 390-401.
- Dzhafarov, E. N., & Kujala, J. V. (2010). The Joint Distribution Criterion and the Distance Tests for selective probabilistic causality. *Frontiers in Psychology* 1:151 doi: 10.3389/fpsyg.2010.00151.
- Dzhafarov, E. N., & Kujala, J. V. (2012a). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology*, 56, 54-63.
- Dzhafarov, E. N., & Kujala, J. V. (2012b). Quantum entanglement and the issue of selective influences in psychology: An overview. *Lecture Notes in Computer Science* 7620, 184-195.
- Dzhafarov, E. N., & Kujala, J. V. (2013). Order-distance and other metric-like functions on jointly distributed random variables. *Proceedings of the American Mathematical Society*, 141, 3291-3301.
- Dzhafarov, E. N., & Kujala, J. V. (2014a). On selective influences, marginal selectivity, and Bell/CHSH inequalities. *Topics in Cognitive Science* 6, 121-128.
- Dzhafarov, E. N., & Kujala, J. V. (2014b). Embedding quantum into classical: contextualization vs conditionalization. *PLoS One* 9(3):e92818. doi:10.1371/journal.pone.0092818.
- Dzhafarov, E. N., & Kujala, J. V. (2014c). A qualified Kolmogorovian account of probabilistic contextuality. *Lecture Notes in Computer Science* 8369, 201–212.
- Dzhafarov, E. N., & Kujala, J. V. (2014d). Contextuality is about identity of random variables. *Physica Scripta* T163, 014009.
- Dzhafarov, E.N., & Kujala, J.V. (in press). Probability, random variables, and selectivity. In W.Batchelder, H. Colonius, E.N. Dzhafarov, J. Myung (Eds). *The New Handbook of Mathematical Psychology*. Cambridge, UK: Cambridge University Press.
- Dzhafarov, E. N., Kujala, J. V., & Larsson, J.-Å. (2015). Contextuality in three types of quantum-mechanical systems. *Foundations of Physics* 2015, doi: 10.1007/s10701-015-9882-9.
- Dzhafarov, E. N., & Schweickert, R. (1995). Decompositions of response times: An almost general theory. *Journal of Mathematical Psychology*, 39, 285-314.
- Dzhafarov, E. N., Schweickert, R., & Sung, K. (2004). Mental architectures with selectively influenced but stochastically interdependent components. *Journal of Mathematical Psychology*, 48, 51-64.
- Khrennikov, A. Yu. (2009). *Contextual Approach to Quantum Formalism*. New York: Springer.
- Kochen, S., & Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17, 59–87.
- Kujala, J. V., & Dzhafarov, E. N. (2008). Testing for selectivity in the dependence of random variables on external factors. *Journal of Mathematical Psychology*, 52, 128–144.
- Landau, L. J. (1988). Empirical two-point correlation functions. *Foundations of Physics*, 18, 449–460.

- Masanés, L.I., Acín, A., & Gisin, N. (2006). General properties of nonsignaling theories. *Physical Review A* 73, 012112.
- Popescu, S. & Rohrlich, D. (1994). Quantum nonlocality as an axiom. *Foundations of Physics*, 24, 379–385.
- Roberts, S., & Sternberg, S. (1993). The meaning of additive reaction-time effects: Tests of three alternatives. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 611–654). Cambridge, MA: MIT Press.
- Schweickert, R., Fisher, D.L., & Sung, K. (2012). *Discovering Cognitive Architecture by Selectively Influencing Mental Processes*. New Jersey: World Scientific.
- Schweickert, R., & Giorgini, M. (1999). Response time distributions: Some simple effects of factors selectively influencing mental processes. *Psychonomic Bulletin & Review*, 6, 269–288.
- Schweickert, R., Giorgini, M., & Dzhafarov, E. N. (2000). Selective influence and response time cumulative distribution functions in serial-parallel task networks. *Journal of Mathematical Psychology*, 44, 504–535.
- Schweickert, R., & Townsend, J. T. (1989). A trichotomy: Interactions of factors prolonging sequential and concurrent mental processes in stochastic discrete mental (PERT) networks. *Journal of Mathematical Psychology*, 33, 328–347.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), *Attention and Performance II. Acta Psychologica*, 30, 276–315.
- Suppes, P., & Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese* 48:191–199.
- Townsend, J. T. (1984). Uncovering mental processes with factorial experiments. *Journal of Mathematical Psychology*, 28, 363–400.
- Townsend, J. T. (1990a). The truth and consequences of ordinal differences in statistical distributions: Toward a theory of hierarchical inference. *Psychological Bulletin*, 108, 551–567.
- Townsend, J. T. (1990b). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science*, 1, 46–54.
- Townsend, J. T., & Fific, M. (2004). Parallel & serial processing and individual differences in high-speed scanning in human memory. *Perception & Psychophysics*, 66, 953–962.
- Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, 39, 321–359.
- Townsend, J. T., & Schweickert, R. (1989). Toward the trichotomy method of reaction times: Laying the foundation of stochastic mental networks. *Journal of Mathematical Psychology*, 33, 309–327.
- Yang, H., Fific, M., & Townsend, J. T. (2013). Survivor interaction contrast wiggle predictions of parallel and serial models for an arbitrary number of processes. *Journal of Mathematical Psychology*, 58, 21–32.