

# Consistent Aggregation of Simply Scalable Families of Choice Probabilities

János Aczél<sup>a,\*</sup>, Attila Gilányi<sup>b</sup>, Gyula Maksa<sup>b</sup>, Anthony A.J. Marley<sup>c</sup>

<sup>a</sup>*Faculty of Mathematics, University of Waterloo  
Waterloo, Ontario N2L 3G1, Canada*

<sup>b</sup>*Institute of Mathematics and Informatics, Lajos Kossuth University  
H-4010 Debrecen, Hungary*

<sup>c</sup>*Department of Psychology, McGill University  
Montréal, Québec H3A 1B1, Canada*

## Abstract

We characterize probabilistic choice models that have two (or more) fixed ratio scales associated with each choice option. A version of Luce's choice model with fixed scales and exponents, which we also characterize, is included. The characterizing properties are: Simple scalability, i.e. the choice probabilities depend upon a priori arbitrary combinations of the two scale values attached to each choice option. Aggregation equations express that both the probabilities and the scales are aggregated from individual to overall values. Generalized homogeneity type equations appear because we deal with ratio scales. As regards regularity, we suppose only local boundedness, non-constancy and an injectivity condition.

*Keywords:* Aggregation; scales; choice probabilities; choice models; equations.

---

Version: April 15, 1999; Accepted by Mathematical Social Sciences.

This research has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) collaborative project grant no. CPG 0164211 and in part by the Hungarian National Foundation for Scientific Research (OTKA), grant No. T-030082 and by the Hungarian High Educational Research and Development Found (FKFP), grant No. 0310/1997. The authors are grateful to the referee and to the consulting editor for helpful suggestions.

\*Corresponding author: János Aczél, 97 McCarron Crescent, Waterloo, Ontario N2L 3G1, Canada, e-mail: [jdaczal@math.uwaterloo.ca](mailto:jdaczal@math.uwaterloo.ca) .

# 1. Introduction

Aczél, Maksa, Marley and Moszner [1] characterized Luce’s choice model and a considerably more general class of choice models by scaling and aggregation properties. The representations studied there were ‘uni-dimensional’ in that each object of choice was assumed to be evaluated in terms of a single (ratio scale) value. (These were called in [1] ‘selection models’ to emphasize that the objects of choice could be sets; this interpretation is implicitly permitted also here, for our ‘choice models’ and ‘choice probabilities’). In this paper we generalize those results to ‘two-dimensional’ (‘multi-dimensional’) representations. Without our aggregation conditions, our representation for choice between two-dimensional objects would be similar to that of [2] (Theorems 7 and 8), which deals with the binary ( $n = 2$ ) case. The assumptions were different in [2], which also contains more general models, with the relation to Luce’s choice model less visible. We use one of their motivating examples to illustrate our approach:

Consider an experiment in which a person hears a sequence of pure tone bursts which are all of the same frequency but of different intensities and durations, and the person has to decide which burst is the loudest. For instance, if each pure tone sequence consisted of 3 bursts, the person could indicate which burst was “loudest” by saying “first” or “second” or “third”, referring thereby to the relevant component. Assume that there are *scales*  $v, w$  such that each tone burst  $d$  can be specified by the two (nonnegative) scale values  $v(d), w(d)$ ; for instance,  $v(d)$  might measure the tone’s psychological intensity (‘loudness’), and  $w(d)$  the tone’s psychological (perceived) duration. (Of course, the intensity and duration of two different tone bursts *may* occasionally be perceived the same). Since such comparisons are frequently ‘noisy’, let  $P(e: E)$  denote the probability that tone burst  $e$  is judged the loudest in the set  $E$ . A special case of the representation that we derive in this paper has the following form: there are non-zero constants  $a, b$  such that

$$P(e: E) = \frac{v(e)^a w(e)^b}{\sum_{d \in E} v(d)^a w(d)^b}. \quad (1)$$

We will call a choice model with (1) as choice probability by the inelegant but descriptive name “two-scale Luce model with exponents” (for the “original” Luce model cf. [1] and (13)). Going one step further, in order to introduce our aggregation ideas,

now suppose that we have data from  $m$  subjects for each of whom the choice probability satisfies (1), but with different scales. That is,  $P_i(e : E)$ , the probability that tone burst  $e$  is ‘chosen’ as the loudest in the set  $E$  by subject  $i$ , is of the form given in (1) with different scales  $v_i, w_i$  ( $i = 1, \dots, m$ ). Now suppose that we wish to create a representation of the performance of an ‘average’ person in this task, perhaps for use in a clinical setting for deciding whether a particular person is performing better or worse than one might expect ‘on average’. What form should we assume for the (probabilistic) performance of this ‘average’ person? Given that each individual satisfies (1), it seems reasonable to assume a parallel representation for the ‘average’ person. In particular, consider the case where the ‘aggregate’ (average) choice probability  $P(e : E)$  is of the form (1) with scale values given, for each  $d$  in  $E$ , by

$$v(d) = \prod_{i=1}^m v_i(d)^{c_i}, \quad w(d) = \prod_{i=1}^m w_i(d)^{c_i} \quad \left( \sum_{i=1}^m c_i = c \neq 0 \right). \quad (2)$$

For ‘averaging’,  $\sum_{i=1}^m c_i = 1$  seems appropriate but we want to keep our considerations open for possible wider interpretations of ‘aggregation’. Note on the other hand that, since the aggregate scale values ‘collapse’ to zero if any of the component scale values are zero, which seems unreasonable, we assume that this is not the case, that is, the scale values and thus the choice probabilities in (1) are nonzero. This we will suppose also for other choice probabilities.

The question seems interesting whether the choice probabilities of this ‘average’ person can be ‘aggregated’ from the choice probabilities of the  $m$  individual subjects. To be exact, our question (which we answer affirmatively) is: can this ‘aggregate’ choice probability  $P(e : E)$  also be written as an ‘aggregate’ of the component choice probabilities  $P_i(e : E)$  ( $i = 1, \dots, m$ ). We can do this by choosing the aggregating formula for the  $P$ ’s as

$$P(e : E) = \frac{\prod_{i=1}^m P_i(e : E)^{c_i}}{\sum_{d \in E} \prod_{i=1}^m P_i(d : E)^{c_i}}. \quad (3)$$

Indeed we get from (1) and (2)

$$P(e : E) = \frac{\prod_{i=1}^m v_i(e)^{c_i^a} w_i(e)^{c_i^b}}{\sum_{d \in E} \prod_{i=1}^m v_i(d)^{c_i^a} w_i(d)^{c_i^b}}.$$

On the other hand, by formulas similar to (1) but for  $P_i$ ,

$$\begin{aligned} \frac{\prod_{i=1}^m P_i(e:E)^{c_i}}{\sum_{d \in E} \prod_{i=1}^m P_i(d:E)^{c_i}} &= \frac{\prod_{i=1}^m \frac{[v_i(e)^a w_i(e)^b]^{c_i}}{[\sum_{d' \in E} v_i(d')^a w_i(d')^b]^{c_i}}}{\sum_{d \in E} \prod_{i=1}^m \frac{[v_i(d)^a w_i(d)^b]^{c_i}}{[\sum_{d' \in E} v_i(d')^a w_i(d')^b]^{c_i}}} \\ &= \frac{\prod_{i=1}^m v_i(e)^{c_i a} w_i(e)^{c_i b}}{\sum_{d \in E} \prod_{i=1}^m v_i(d)^{c_i a} w_i(d)^{c_i b}}. \end{aligned}$$

Thus we have (3).

So this ‘aggregation’ rule is plausible in that it ‘works’ *both* at the level of the choice probabilities *and* at the level of the scale values. A major contribution of this paper is to study ‘natural’ conditions under which the above result holds, and to consider generalizations to other choice probabilities besides (1).

It is interesting to observe that we can define new ‘scales’ for each option  $e$  in  $E$ , with  $aA_i + bB_i = c_i$  ( $i = 1, \dots, m$ ), by

$$\tilde{v}(e:E) = \prod_{i=1}^m P_i(e:E)^{A_i}, \quad \tilde{w}(e:E) = \prod_{i=1}^m P_i(e:E)^{B_i}. \quad (4)$$

Then, by (3),

$$P(e:E) = \frac{\tilde{v}(e:E)^a \tilde{w}(e:E)^b}{\sum_{d \in E} \tilde{v}(d:E)^a \tilde{w}(d:E)^b},$$

which is ‘of the same form’ as (1) and as the component choice probabilities, but in terms of ‘scales’  $\tilde{v}, \tilde{w}$  that are functions of the component choice probabilities  $P_i(e:E)$  ( $i = 1, \dots, m$ ).

In the above example, we assumed that we had data from  $m$  individuals, which we wanted to ‘aggregate’ in some sensible fashion. We now briefly present an example where it might be plausible to assume that all the data are from a single person, and that the ‘aggregation’ is carried out by the person (either consciously or unconsciously). In a *complete identification task*, there is a (finite) set of stimuli  $E$ , and each  $f \in E$  is assigned a unique ‘correct’ response  $r[f]$ . On each trial of the experiment, the participant’s task is to attempt to ‘correctly’ identify a presented stimulus  $f \in E$  by making the response  $r[f]$ . Then paralleling (1), the probability  $P(r[e]|f:E)$  ( $e, f \in E$ ) is interpreted as the probability that an option (stimulus)  $f \in E$  is identified as the option (stimulus)  $e \in E$ , i.e., that the response  $r[e]$  is made to the stimulus  $f$ . In a

*similarity choice model* or biased Luce model ([3], [4]) these probabilities are of the form

$$P(r[e]|f:E) = \frac{\eta(f,e)\beta(r[e])}{\sum_{d \in E} \eta(f,d)\beta(r[d])} \quad (5)$$

(often the notation  $P_E(r[e]|f)$  is used), with the constraints

$$0 \leq \eta(f,e) \leq 1, \quad \eta(f,e) = \eta(e,f), \quad \eta(e,e) = 1, \quad 0 \leq \beta(r[e]) \leq 1, \quad \sum_{d \in E} \beta(r[d]) = 1. \quad (6)$$

The usual interpretation is that  $\eta$  is the stimulus scale or *similarity scale* and  $\beta$  the response scale or *bias scale* (it is usually denoted by  $b$ ). We see that, for each *fixed*  $f$  and for given scales  $\eta$  and  $\beta$ ,  $P(r[e]|f:E)$  is a two-scale Luce choice probability (that is, a two-scale Luce choice probability with exponents  $a = b = 1$ ) – at least within the bounds usually imposed upon values of the scales  $\eta$  and  $\beta$ . (Our mathematical results apply also if the scales for  $\eta$  and  $\beta$  are assumed to be ratio scales even for very large scale values. The issue of fitting our results to the usual formulation of similarity choice models will be considered in section 5). A natural generalization of this representation to the multi-component case involves  $m$  similarity and bias functions  $\eta_i, \beta_i$  ( $i = 1, \dots, m$ ), where each  $i$  is associated with a different ‘component’ of the options – for instance, the (visual) objects being categorized may differ on the ‘independent’ dimensions of size, shape, and color ([5] discusses in detail the complexities caused by the fact that even though these dimensions can be varied ‘independently’, each such visual object must have a value on each of these dimensions). The various concepts and results developed for the previous example then apply naturally for this example. Of course, as in the first example, the similarity scales  $\eta_i$  and the bias scale  $\beta_i$  ( $i = 1, \dots, m$ ) may *not* depend on the components  $i$  ( $i = 1, \dots, m$ ), but deciding whether this is the case or not is an empirical and statistical issue. In fact, there are quite complex issues related to fitting two-scale Luce choice probabilities, their aggregates, and our later more general representations, to empirical data (see Remark 7 below); [6] is an excellent introduction to these issues for the similarity choice model.

Of course, one could again assume that the above model holds for each of a group of  $m$  individuals or for a single individual on  $m$  different occasions (rather than across  $m$  dimensions for a single individual) and consider appropriate ways to ‘aggregate’ this data. Our earlier discussion illustrated our results concerning the ‘correct’ way to

perform such aggregation when one wishes the component and the aggregate forms to be related in certain systematic ways.

In summary, we have illustrated representations that can be aggregated *both* at the level of the scales  $v_i, w_i$  (or  $\eta_i, \beta_i$ ) *and* at the level of the component choice probabilities  $P_i$  ( $i = 1, \dots, m$ ). In this paper we characterize the class of probabilistic models for choice between ‘two-dimensional’ options that have similar aggregation properties, where the *forms* of the aggregation and probabilistic choice functions are *derived*, rather than assumed as they have been in the above examples.

We now discuss the relevance of the introductory examples to the remainder of the paper and the relation of the paper to earlier work. Let  $E = \{e_1, \dots, e_n\}$ , so  $|E| = n$ . If we define functions  $h, K$  and  $F_j$  ( $j = 1, \dots, n$ ) on the positive reals by

$$h(x, y) = x^a y^b, \quad (7)$$

$$K_j(z_1, \dots, z_n) = \frac{z_j}{\sum_{k=1}^n z_k} \quad (j = 1, \dots, n), \quad (8)$$

$$F_j(x_1, y_1, \dots, x_n, y_n) = \frac{x_j^a y_j^b}{\sum_{k=1}^n x_k^a y_k^b} \quad (j = 1, \dots, n) \quad (9)$$

then (1) in the earlier example takes the form

$$\begin{aligned} P(e_j : E) &= F_j[v(e_1), w(e_1), \dots, v(e_n), w(e_n)] \\ &= K_j[h(v(e_1), w(e_1)), \dots, h(v(e_n), w(e_n))] \quad (j = 1, \dots, n). \end{aligned} \quad (10)$$

Equation (8) is of the form of the expression for the ‘original’ Luce choice model (see [1] and (13) below).

Notice the following properties (i), (ii), (iii) of these functions, with terminology for properties (i) and (ii) generalized from their  $n = 2$  form in [2] ((ii) further generalized from (10) to allow  $n$  functions  $h_1, \dots, h_n$  which, as we will later prove, have to be equal). These equations are meant for arbitrary  $\mu, \nu, \lambda, x_j, y_j, z_j > 0$  ( $j = 1, \dots, n$ ), though smaller intervals would do.

(i) ‘Conjoint Weber Type II’:

$$F_j(\mu x_1, \nu y_1, \dots, \mu x_n, \nu y_n) = F_j(x_1, y_1, \dots, x_n, y_n) \quad (j = 1, \dots, n);$$

(ii) “Simple Scalability”:

$$F_j(x_1, y_1, \dots, x_n, y_n) = K_j(h_1(x_1, y_1), \dots, h_n(x_n, y_n)) \quad (j = 1, \dots, n);$$

(iii) “Homogeneity of Degree 0”:

$$K_j(\lambda z_1, \dots, \lambda z_n) = K_j(z_1, \dots, z_n) \quad (j = 1, \dots, n).$$

The homogeneity (iii) looks similar to (i) and will turn out to be related to it through (ii).

In this paper we develop aggregation conditions which, with (i), (ii), (iii) and under some simple regularity conditions, imply a choice model which includes (1) as particular case. Then we give conditions which characterize just (1). This characterization will also show why powers and not other functions figure in (1). This is related to considering  $v$  and  $w$  as ratio scales.

## 2. Definitions, Assumptions, Equations

We will call choice probability and denote by  $P(e : E)$  the probability of choosing (selecting) an element (option)  $e$  from an  $n$ -element ( $1 < n < \infty$ ) subset  $E$  of a global set  $R$  of options. We exclude elements which are impossible to choose. Thus,

$$0 < P(e : E) < 1 \tag{11}$$

and, of course,

$$\sum_{e \in E} P(e : E) = 1. \tag{12}$$

We now formally define this and further concepts required in this paper (interspersed with remarks).

**Definition 1.** Let  $R$  be a set and  $S$  the set of its finite subsets, where both  $R$  and every  $E \in S$  has at least two elements, and  $D = \{(e : E) | E \in S, e \in E\} \subseteq R \times S$ . The pair  $(R, P)$  consisting of the set  $R$  and of the function  $P : D \rightarrow ]0, 1[$  is a *choice model*. The function value  $P(e : E)$  ( $e \in E, E \in S$ ) is a *choice probability*,  $e$  is an *option*.

**Remark 1.** While here we required that  $R$  have at least two elements, a later assumption will imply that it has uncountably infinite elements (see Definition 3, Remark 4).

**Definition 2.** A choice model, two functions  $v, w : R \rightarrow ]0, \infty[$  and two nonzero numbers  $a, b$  form a *two-scale Luce model with exponents* if (1), that is

$$P(e : E) = \frac{v(e)^a w(e)^b}{\sum_{d \in E} v(d)^a w(d)^b}$$

holds. If  $a = b = 1$  then we have a *two-scale Luce model*. The functions  $v$  and  $w$  are the *scales*,  $v(e), w(e)$  are *scale values*.

**Remark 2.** Here and throughout the paper *the scales  $v$  and  $w$  are given*; otherwise not only could the exponents be immersed into  $v$  and  $w$  but the two scales could be replaced by a single scale  $u = v^a w^b$  and we would have the original (one-scale) Luce model (cf. [1] and (8))

$$P(e : E) = \frac{u(e)}{\sum_{d \in E} u(d)}. \quad (13)$$

Notice, however, that, if the choice probabilities can be represented in the form (1) with scales  $v, w$ , they can equally well be represented by scales  $\gamma v, \delta w$  with arbitrary constants  $\gamma > 0, \delta > 0$ .

As we said in the Introduction, our purpose in this paper is to determine those scale-dependent (two-scale, see Definition 4) families of choice probabilities for which formulae like (10)

$$\begin{aligned} P(e_j : E) &= F_j[v(e_1), w(e_1), \dots, v(e_n), w(e_n)] \\ &= K_j[h_1(v(e_1), w(e_1)), \dots, h_n(v(e_n), w(e_n))] \quad (j = 1, \dots, n) \end{aligned} \quad (14)$$

hold, where both the scales and the probabilities can be *aggregated*. This means that, in analogy to (2) and (3), equations of the form

$$\begin{aligned} v(e_j) &= G_1[v_1(e_j), \dots, v_m(e_j)], \quad w(e_j) = G_2[w_1(e_j), \dots, w_m(e_j)], \\ P(e_j : E) &= H_j[P_1(e_1 : E), \dots, P_m(e_1 : E), \dots, P_1(e_n : E), \dots, P_m(e_n : E)], \end{aligned}$$



( $j = 1, \dots, n$ ) are satisfied, where the component probabilities satisfy equations analogous to (14) with the *same*  $F_j$  ( $j = 1, \dots, n$ ):

$$\begin{aligned} P_i(e_j : E) &= F_j[v_i(e_1), w_i(e_1), \dots, v_i(e_n), w_i(e_n)] \\ &= K_j[h_1(v_i(e_1), w_i(e_1)), \dots, h_n(v_i(e_n), w_i(e_n))] \\ &\quad (i = 1, \dots, m; j = 1, \dots, n). \end{aligned} \tag{15}$$

These equations, with assumptions linked to  $v_i, w_i$  and their combinations  $h_j(v_i, w_i)$  being ratio scales, will lead to functional equations, from which we will endeavor to determine the functions  $h_j, K_j$  (and thus  $F_j$ ;  $j = 1, \dots, n$ ) obtaining in passing also the general forms of  $G_1, G_2$  and, under further restrictions,  $H_j$  ( $j = 1, \dots, n$ ). The variables in these equations will go through the set  $\mathbb{R}_{++}$  of positive numbers, because we will deal with *nonatomic* scales (in analogy to nonatomic Kolmogorov probability spaces, see e.g. in [7]). These are such that the scales can assume *any* positive value. (We will make further assumptions in Definitions 3 and 4).

A *nonatomic pair of scales* is a triple  $(R, v, w)$ , where  $R$  is a set and  $v, w : R \rightarrow \mathbb{R}_{++}$  are scales (functions) such that for every pair  $(x, y) \in \mathbb{R}_{++}^2$  there exists an  $e \in R$  for which  $x = v(e), y = w(e)$ .

**Remark 3.** Nonatomicity has the following consequence. For each  $(2n)$ -tuple  $(x_1, y_1, \dots, x_n, y_n) \in \mathbb{R}_{++}^{2n}$  for which  $(x_j, y_j) \neq (x_k, y_k)$  if  $j \neq k$  (that is, *the same*  $(x_j, y_j)$  *pair does not occur more than once*) there exist  $(e_1, \dots, e_n) \in R^n$  such that  $x_j = v(e_j), y_j = w(e_j)$  ( $j = 1, \dots, n$ ). The italicized exclusion was necessary because otherwise  $E$  may not have  $n$  *distinct* elements; in particular it may have just one element – but we excluded singletons (cf. (11)). Therefore we will need here (as in [1]) a somewhat stronger condition, the  $n$ -nonatomicity which postulates the existence of different (distinct) elements  $e_j \neq e_k$  even if  $x_j = x_k, y_j = y_k$  thus  $v(e_j) = x_j = x_k = v(e_k), w(e_j) = y_j = y_k = w(e_k)$  ( $j \neq k$ ). This is not unexpected: For instance,  $e_j, e_k$  might be two different cars that a person considers equal in safety ( $x_j = v(e_j) = v(e_k) = x_k$ ) and equal in road handling (second dimension  $y_j = w(e_j) = w(e_k) = y_k$ ); see also the example in the Introduction about equal perceived intensity and duration for different tone bursts.

We define now the  $n$ -nonatomicity formally.

**Definition 3.** An  $n$ -nonatomic pair of scales is a quadruple  $(R, v, w, n)$ , where  $R$  is a set,  $v, w : R \rightarrow \mathbb{R}_{++}$  are scales (functions) and  $n > 1$  is an integer such that for every  $(2n)$ -tuple  $(x_1, y_1, \dots, x_n, y_n)$  of positive numbers there exist  $n$  *distinct* elements  $e_1, \dots, e_n$  of  $R$  for which  $x_j = v(e_j)$ ,  $y_j = w(e_j)$  ( $j = 1, \dots, n$ ).

**Remark 4.** While the  $n$ -nonatomicity assumption clearly implies that the global set  $R$  is uncountably infinite, the choice is always done from *finite* subsets  $E$  of  $R$  (as is appropriate for applications in psychology and elsewhere). The papers [8] and [9] discuss the use (and misuse) of nonatomicity in related problems without explicitly defining this concept.

**Remark 5.** Of course, there may exist not only  $e_j \neq e_k$  of the same set  $E$  with equal scale values but to a  $2n$ -tuple of positive reals there may also exist two (or more)  $n$ -tuple(s)  $(e_1, \dots, e_n)$ ,  $(e'_1, \dots, e'_n)$  in  $R^n$  such that, for the same  $(x_1, y_1, \dots, x_n, y_n) \in \mathbb{R}_{++}^{2n}$  simultaneously  $x_j = v(e_j) = v(e'_j)$ ,  $y_j = w(e_j) = w(e'_j)$  ( $j = 1, \dots, n$ ) hold. The last part of the next definition refers to this situation.

**Definition 4.** Let  $n > 1$  be an integer. An  $n$ -nonatomic two-scale family of choice probabilities (a *two-scale model* for short) is a 5-tuple  $(R, P, v, w, n)$ , where  $(R, P)$  is a choice model with only  $n$  element subsets  $E$  of  $R$  considered and  $(R, v, w, n)$  is an  $n$ -nonatomic pair of scales such that, whenever for two subsets  $E = \{e_1, \dots, e_n\}$  and  $E' = \{e'_1, \dots, e'_n\}$  of  $R$  we have  $v(e_j) = v(e'_j)$  and  $w(e_j) = w(e'_j)$  ( $j = 1, \dots, n$ ) then also  $P(e_j : E) = P(e'_j : E')$  ( $j = 1, \dots, n$ ).

**Remark 6.** The last requirement, where  $n$  has an essential role, implies that *there exist functions*  $F_j : \mathbb{R}_{++}^{2n} \rightarrow ]0, 1[$  such that

$$P(e_j : E) = F_j(v(e_1), w(e_1), \dots, v(e_n), w(e_n)) \quad (j = 1, \dots, n)$$

for every  $n$ -element subset  $E$  of  $R$ . We need the last assumption and  $n$ -nonatomicity so that equations like (14) make sense and that our later functional equations hold on suitable domains. We will call  $(F_1, \dots, F_n)$  the *generating system* of the choice probability  $P$ . As already mentioned, we will suppose that  $P$  and  $P_i$  ( $i = 1, \dots, m$ ) have the same generating system.

We generalize (1), (2) and (3) to the following aggregation equations (cf. [1]):

$$P(e_j : E) = F_j [v(e_1), w(e_1), \dots, v(e_n), w(e_n)], \quad (16)$$

$$P_i(e_j : E) = F_j [v_i(e_1), w_i(e_1), \dots, v_i(e_n), w_i(e_n)], \quad (17)$$

$$v(e_j) = G_1 [v_1(e_j), \dots, v_m(e_j)], \quad w(e_j) = G_2 [w_1(e_j), \dots, w_m(e_j)], \quad (18)$$

$$P(e_j : E) = H_j [P_1(e_1 : E), \dots, P_m(e_1 : E), \dots, P_1(e_n : E), \dots, P_m(e_n : E)], \quad (19)$$

here  $i = 1, \dots, m$  and  $j = 1, \dots, n$ . Substituting (18) into (16) on one hand and (17) into (19) on the other gives, in view of the nonatomicity, a system of  $n$  functional equations. The fact that the scales and the probabilities depending on them by (16) and (17) are invariant (covariant) under linear transformations (ratio scales) give  $2n+2$  others. Finally (11) and (12) generate one more equation and  $n$  inequalities.

As in [1], however, there would be far too many solutions for practical use and this system is far too weak to characterize the “two-scale Luce model with exponents” or even halfway reasonable generalizations. So two further specifications are needed. The first links  $H_j$  more closely to  $F_j$ , at the same time introducing two more scale values depending on the choice probabilities:

$$\begin{aligned} \tilde{v}(e_j : E) = \Phi_1 [P_1(e_j : E), \dots, P_m(e_j : E)], \quad \tilde{w}(e_j : E) = \Phi_2 [P_1(e_j : E), \dots, P_m(e_j : E)] \\ (j = 1, \dots, n), \quad (20) \end{aligned}$$

(cf. (4) and [1]) to which (16) should also apply:

$$P(e_j : E) = F_j [\tilde{v}(e_1 : E), \tilde{w}(e_1 : E), \dots, \tilde{v}(e_n : E), \tilde{w}(e_n : E)] \quad (j = 1, \dots, n) \quad (21)$$

(the more specific Eq. (21) with (20) will replace (19)). The second (“simple scalability” (ii)) we know already (see (14)). We repeat it here to point out that it establishes a link between the two original sets of scales within  $F_j$ :

$$\begin{aligned} P(e_j : E) = F_j [v(e_1), w(e_1), \dots, v(e_n), w(e_n)] \\ = K_j (h_1 [v(e_1), w(e_1)], \dots, h_n [v(e_n), w(e_n)]) \quad (j = 1, \dots, n). \quad (22) \end{aligned}$$

We list now the assumptions linked to  $v_i, w_i$  and the combinations  $h_j(v, w)$  of  $v$  and  $w$  being ratio scales ( $i = 1, \dots, m; j = 1, \dots, n$ ). There exist functions  $M_1, M_2$

such that

$$G_1[\gamma_1 v_1(e_1), \dots, \gamma_m v_m(e_m)] = M_1(\gamma_1, \dots, \gamma_m) G_1[v_1(e_1), \dots, v_m(e_m)], \quad (23)$$

$$G_2[\delta_1 w_1(e_1), \dots, \delta_m w_m(e_m)] = M_2(\delta_1, \dots, \delta_m) G_2[w_1(e_1), \dots, w_m(e_m)]. \quad (24)$$

Also

$$F_j[\mu v(e_1), \nu w(e_1), \dots, \mu v(e_n), \nu w(e_n)] = F_j[v(e_1), w(e_1), \dots, v(e_n), w(e_n)] \\ (j = 1, \dots, n),$$

(“conjoint Weber type II” [2]), which translates for  $K_j$  and  $h_j$  into

$$K_j(h_1[\mu v(e_1), \nu w(e_1)], \dots, h_n[\mu v(e_n), \nu w(e_n)]) \\ = K_j(h_1[v(e_1), w(e_1)], \dots, h_n[v(e_n), w(e_n)]) \quad (j = 1, \dots, n). \quad (25)$$

Furthermore there exists a function  $N$  such that

$$K_j(\lambda h_1[v(e_1), w(e_1)], \dots, \lambda h_n[v(e_n), w(e_n)]) \\ = N(\lambda) K_j(h_1[v(e_1), w(e_1)], \dots, h_n[v(e_n), w(e_n)]) \quad (j = 1, \dots, n). \quad (26)$$

holds. Finally, the obvious equation (12) and inequalities (11) translate, in view of (16), into

$$\sum_{j=1}^n F_j[v(e_1), w(e_1), \dots, v(e_n), w(e_n)] = 1, \quad (27)$$

$$0 < F_j[v(e_1), w(e_1), \dots, v(e_n), w(e_n)] < 1 \quad (j = 1, \dots, n). \quad (28)$$

We introduce, for the sake of brevity, the notation (permitted by the  $n$ -nonatomicity)

$$x_j = v(e_j), \quad y_j = w(e_j), \quad z_j = h_j(x_j, y_j), \quad x_{ij} = v_i(e_j), \quad y_{ij} = w_i(e_j) \\ (i = 1, \dots, m; j = 1, \dots, n). \quad (29)$$

(The scale values and thus these variables are supposed to be *positive*). So, by (22), (27), (28) and (29),

$$\sum_{j=1}^n K_j(z_1, \dots, z_n) = 1, \quad (30)$$

$$0 < K_j(z_1, \dots, z_n) < 1 \quad (j = 1, \dots, n). \quad (31)$$

From (26) and (30)  $N(\lambda) \equiv 1$ , thus

$$K_j(z_1, \dots, z_n) = K_j(\lambda z_1, \dots, \lambda z_n) = K_j\left(\frac{z_1}{\sum_{k=1}^n z_k}, \dots, \frac{z_n}{\sum_{k=1}^n z_k}\right) \quad (j = 1, \dots, n). \quad (32)$$

Also, by (19), (20) and (21), with  $s_{ij} = P_i(e_j : E) \in \mathbb{R}_{++}$  (not all values in  $\mathbb{R}_{++}$  need to be assumed):

$$\begin{aligned} & H_j(s_{11}, \dots, s_{m1}, \dots, s_{1n}, \dots, s_{mn}) \\ &= F_j[\Phi_1(s_{11}, \dots, s_{m1}), \Phi_2(s_{11}, \dots, s_{m1}), \dots, \Phi_1(s_{1n}, \dots, s_{mn}), \Phi_2(s_{1n}, \dots, s_{mn})] \\ & \quad (j = 1, \dots, n). \end{aligned} \quad (33)$$

We conclude by stating our regularity assumptions. About  $G_1$  and  $G_2$  we suppose only that they are *locally bounded* (on an  $m$ -dimensional interval, no matter how small) and that  $G_1(x, \dots, x)$  is *not constant*. With (23) and (24) this gives (cf. [10] and [1])

$$G_1(x_1, \dots, x_m) = C \prod_{i=1}^m x_i^{c_i}, \quad \sum_{i=1}^m c_i \neq 0, \quad G_2(y_1, \dots, y_m) = D \prod_{i=1}^m y_i^{d_i}.$$

Also  $h_1, \dots, h_n$  will be supposed *locally bounded and nonconstant*. We may assume, without loss of generality, that

$$G_1(1, \dots, 1) = G_2(1, \dots, 1) = 1, \quad h_j(1, 1) = 1 \quad (j = 1, \dots, n). \quad (34)$$

So  $C = D = 1$ , and we have

$$G_1(x_1, \dots, x_m) = \prod_{i=1}^m x_i^{c_i}, \quad \sum_{i=1}^m c_i \neq 0, \quad G_2(y_1, \dots, y_m) = \prod_{i=1}^m y_i^{d_i}, \quad (35)$$

In addition to these very weak conditions, we assume (cf. [1]) that  $K_j$  is *injective* in the following sense:

$$\begin{aligned} & \left[ K_j(z_1, \dots, z_n) = K_j(z'_1, \dots, z'_n) \quad (j = 1, \dots, n) \text{ and } \sum_{k=1}^n z_k = \sum_{k=1}^n z'_k \right] \\ & \quad \text{imply } z_j = z'_j \quad (j = 1, \dots, n). \end{aligned} \quad (36)$$

Notice that (32) allows us to bring the sum of variables in  $K_j$  to 1, which we need so that the number of equations on both sides of the implication (36) be equal (to  $n - 1$ , remember (30)).

Substituting (33) and (17) into (19) and (18) into (16), and equating the two expressions thus obtained for  $P(e_j: E)$  ( $j = 1, \dots, n$ ) we get, in terms of the notation (29), the system of functional equations

$$\begin{aligned}
& F_j[G_1(x_{11}, \dots, x_{m1}), G_2(y_{11}, \dots, y_{m1}), \dots, G_1(x_{1n}, \dots, x_{mn}), G_2(y_{1n}, \dots, y_{mn})] \\
&= F_j(\Phi_1[F_1(x_{11}, y_{11}, \dots, x_{1n}, y_{1n}), \dots, F_1(x_{m1}, y_{m1}, \dots, x_{mn}, y_{mn})], \\
&\quad \Phi_2[F_1(x_{11}, y_{11}, \dots, x_{1n}, y_{1n}), \dots, F_1(x_{m1}, y_{m1}, \dots, x_{mn}, y_{mn})], \\
&\quad \vdots \\
&\quad \Phi_1[F_n(x_{11}, y_{11}, \dots, x_{1n}, y_{1n}), \dots, F_n(x_{m1}, y_{m1}, \dots, x_{mn}, y_{mn})], \\
&\quad \Phi_2[F_n(x_{11}, y_{11}, \dots, x_{1n}, y_{1n}), \dots, F_n(x_{m1}, y_{m1}, \dots, x_{mn}, y_{mn})]) \\
&\hspace{20em} (j = 1, \dots, n), \quad (37)
\end{aligned}$$

which looks at first sight quite formidable.

The question is, whether (cf. (1) and (9))

$$F_j(x_1, y_1, \dots, x_n, y_n) = \frac{x_j^a y_j^b}{\sum_{k=1}^n x_k^a y_k^b} \quad (j = 1, \dots, n) \quad (38)$$

with (35) and with aptly chosen  $\Phi_1$  and  $\Phi_2$  are the only solutions of the above equations. The answer is negative but we determine the general solutions  $F_1, \dots, F_n$  under these assumptions and also the auxiliary functions  $G_1, G_2, H_1, \dots, H_n, \Phi_1, \Phi_2, K_1, \dots, K_n, h_1, \dots, h_n$  and  $M_1, M_2$  (we have already  $N = 1$ ). Then we characterize the “two-scale Luce model with exponents” (38) by one additional supposition.

### 3. Solutions

**Theorem 1.** *If (25), (32), (36) hold, the  $h_j$ 's are locally bounded but not constant in either variable and  $h_j(1, 1) = 1$  ( $j = 1, \dots, n$ ) then, for all  $j \in \{1, \dots, n\}$ ,*

$$h_j(x, y) = x^a y^b \quad (x > 0, y > 0) \quad (39)$$

for some nonzero constants  $a, b$ .

Eq. (39) is, of course, the same as (7). Accordingly, we will write

$$h(x, y) = h_j(x, y) = x^a y^b. \quad (40)$$

**Proof:** Eqs. (25) and (32) imply that

$$\begin{aligned} K_j & \left( \frac{h_1(\mu x_1, \nu y_1)}{\sum_{k=1}^n h_k(\mu x_k, \nu y_k)}, \dots, \frac{h_n(\mu x_n, \nu y_n)}{\sum_{k=1}^n h_k(\mu x_k, \nu y_k)} \right) \\ & = K_j \left( \frac{h_1(x_1, y_1)}{\sum_{k=1}^n h_k(x_k, y_k)}, \dots, \frac{h_n(x_n, y_n)}{\sum_{k=1}^n h_k(x_k, y_k)} \right) \quad (j = 1, \dots, n). \end{aligned}$$

Thus it follows from (36) that

$$\frac{h_j(\mu x_j, \nu y_j)}{\sum_{k=1}^n h_k(\mu x_k, \nu y_k)} = \frac{h_j(x_j, y_j)}{\sum_{k=1}^n h_k(x_k, y_k)} \quad (j = 1, \dots, n),$$

that is,

$$\frac{h_j(\mu x_j, \nu y_j)}{h_j(x_j, y_j)} = \frac{\sum_{k=1}^n h_k(\mu x_k, \nu y_k)}{\sum_{k=1}^n h_k(x_k, y_k)} = Q(x_1, \dots, x_n, y_1, \dots, y_n, \mu, \nu) \quad (j = 1, \dots, n).$$

On the right end  $Q$  carries no subscript because the middle term does not depend on  $j$ . On the other hand, the term on the left does not depend upon  $x_l, y_l$  ( $l \neq j$ ). Writing the same string of equations for different  $j$ 's, the middle term and thus  $Q$  remains the same but  $j$  varies on the left. So  $Q$  cannot depend on any  $x_j, y_j$  and we get

$$\frac{h_j(\mu x_j, \nu y_j)}{h_j(x_j, y_j)} = \tilde{Q}(\mu, \nu) \quad (j = 1, \dots, n),$$

that is,

$$h_j(\mu x, \nu y) = \tilde{Q}(\mu, \nu) h_j(x, y) \quad (j = 1, \dots, n).$$

The general locally bounded nonconstant solution of this equation (see e. g. [10]) is given by  $\tilde{Q}(\mu, \nu) = \mu^a \nu^b$  and by (39). This concludes the proof of Theorem 1.

From (22) and (39) we have

$$F_j(x_1, y_1, \dots, x_n, y_n) = K_j(x_1^a y_1^b, \dots, x_n^a y_n^b) \quad (j = 1, \dots, n). \quad (41)$$

Notice that, while we used (32) to get (41), conversely, the (41) form of “simple scalability” ((ii) in section 1) and the “conjoint Weber type II”, that is (i), imply homogeneity of degree 0 ((32), (iii)):

$$\begin{aligned} K_j(\mu^a \nu^b x_1^a y_1^b, \dots, \mu^a \nu^b x_n^a y_n^b) & = F_j(\mu x_1, \nu y_1, \dots, \mu x_n, \nu y_n) \\ & = F_j(x_1, y_1, \dots, x_n, y_n) = K_j(x_1^a y_1^b, \dots, x_n^a y_n^b) \quad (j = 1, \dots, n). \end{aligned}$$

Now, substituting (41) into (37) and applying (35) and (36), we get

$$\frac{\prod_{i=1}^m x_{ij}^{ac_i} y_{ij}^{bd_i}}{\sum_{k=1}^n \prod_{i=1}^m x_{ik}^{ac_i} y_{ik}^{bd_i}} = \frac{\Phi[K_j(x_{11}^a y_{11}^b, \dots, x_{1n}^a y_{1n}^b), \dots, K_j(x_{m1}^a y_{m1}^b, \dots, x_{mn}^a y_{mn}^b)]}{\sum_{k=1}^n \Phi[K_k(x_{11}^a y_{11}^b, \dots, x_{1n}^a y_{1n}^b), \dots, K_k(x_{m1}^a y_{m1}^b, \dots, x_{mn}^a y_{mn}^b)]} \quad (j = 1, \dots, n), \quad (42)$$

where (cf. (40))

$$\begin{aligned} \Phi(t_1, \dots, t_m) &= h(\Phi_1(t_1, \dots, t_m), \Phi_2(t_1, \dots, t_m)) \\ &= \Phi_1(t_1, \dots, t_m)^a \Phi_2(t_1, \dots, t_m)^b. \end{aligned} \quad (43)$$

Since on the right hand side of (42) every  $x_{il}$  figures with exponent  $a$ , multiplied by  $y_{il}^b$  to give  $x_{il}^a y_{il}^b$ , also the products of powers of  $x_{il}$  and  $y_{il}$  on the left hand side can only be  $x_{il}^a y_{il}^b$  or its powers. Thus *we have to have*

$$d_i = c_i \quad (i = 1, \dots, m), \quad d = \sum_{i=1}^m d_i = \sum_{i=1}^m c_i = c \neq 0 \quad (44)$$

and, by (35),

$$G_1(z_1, \dots, z_m) = G_2(z_1, \dots, z_m) = G(z_1, \dots, z_m) = \prod_{i=1}^m z_i^{c_i}, \quad c = \sum_{i=1}^m c_i \neq 0, \quad (45)$$

(cf. (2)) which defines  $G$ .

In view of  $d_i = c_i$ , with the notation  $z_{ij} = x_{ij}^a y_{ij}^b$  ( $i = 1, \dots, m; j = 1, \dots, n$ ), (42) becomes simpler:

$$\frac{\prod_{i=1}^m z_{ij}^{c_i}}{\sum_{k=1}^n \prod_{i=1}^m z_{ik}^{c_i}} = \frac{\Phi[K_j(z_{11}, \dots, z_{1n}), \dots, K_j(z_{m1}, \dots, z_{mn})]}{\sum_{k=1}^n \Phi[K_k(z_{11}, \dots, z_{1n}), \dots, K_k(z_{m1}, \dots, z_{mn})]} \quad (j = 1, \dots, n). \quad (46)$$

The following has been proved in [1].

**Lemma.** *Equations (30), (31), (32), (36), (46) and  $c = \sum_{i=1}^m c_i \neq 0$  are satisfied and  $\Phi(t, \dots, t)$  is continuous and strictly monotonic if, and only if, the function  $\varphi$ , given by*

$$\varphi(t) = \Phi(t, \dots, t)^{\frac{1}{c}} \quad (47)$$

*is the inverse of a continuous, strictly monotonic function  $\psi$  satisfying*

$$0 \leq \inf_{0 < z < \infty} \psi(z) < \frac{1}{n} < \sup_{0 < z < \infty} \psi(z) \quad (48)$$



and  $K_j$  is of the form

$$K_j(z_1, \dots, z_n) = \psi[z_j L(z_1, \dots, z_n)] \quad (j = 1, \dots, n), \quad (49)$$

where  $w = L(z_1, \dots, z_n)$  is the unique solution of

$$\sum_{j=1}^n \psi(z_j w) = 1,$$

thus

$$\sum_{j=1}^n \psi(z_j L(z_1, \dots, z_n)) = 1, \quad (50)$$

and  $L$  is homogeneous of degree  $(-1)$ :

$$L(\lambda z_1, \dots, \lambda z_n) = \lambda^{-1} L(z_1, \dots, z_n). \quad (51)$$

We reformulate equation (46) of this Lemma as follows. Up to notation, (46) is the same as (42) [even (44)  $d_i = c_i$ ; ( $i = 1, \dots, m$ ) followed from (42)]. We put both sides of (42) into  $K_j$ , apply (44) and (32), and choose

$$\Phi_1(t_1, \dots, t_m)^a \Phi_2(t_1, \dots, t_m)^b = \Phi(t_1, \dots, t_m) = \prod_{i=1}^m \varphi(t_i)^{c_i}. \quad (52)$$

This transforms (46), that is (42), into

$$\begin{aligned} & K_j \left[ \prod_{i=1}^m (x_{i1}^a y_{i1}^b)^{c_i}, \dots, \prod_{i=1}^m (x_{in}^a y_{in}^b)^{c_i} \right] \\ &= K_j \left( \prod_{i=1}^m \varphi[K_1(x_{i1}^a y_{i1}^b, \dots, x_{in}^a y_{in}^b)]^{c_i}, \dots, \prod_{i=1}^m \varphi[K_n(x_{i1}^a y_{i1}^b, \dots, x_{in}^a y_{in}^b)]^{c_i} \right) \\ & \hspace{15em} (j = 1, \dots, n). \end{aligned} \quad (53)$$

We trust that the reader noticed that the strange looking Eq. (53) is really (37) with (41), (52) and (45) taken into account, the latter originating from (23) and (24) through (35) and (44).

It may seem that we chose (52) arbitrarily. But it has been proved in [1] (Appendix B) that, in (46) for  $n > 2$  this is the only possible  $\Phi$ , while for  $n = 2$  the general  $\Phi$  is given by

$$\Phi(t_1, \dots, t_m) = T(t_1, \dots, t_m) \prod_{i=1}^m \varphi(t_i)^{c_i}$$

where  $T$  is an arbitrary function satisfying

$$T(1 - t_1, \dots, 1 - t_m) = T(t_1, \dots, t_m) \quad (t_i \in ]0, 1[; i = 1, \dots, m) \quad (54)$$

(for  $T \equiv 1$  we get (52) again), cf. also Remark 8.

Thus we have proved the following.

**Theorem 2.** *The general solution of the system (30), (31), (32), (36) and (53) of functional equations with continuous strictly monotonic  $\varphi$  is given by (49), where  $\psi = \varphi^{-1}$  has to satisfy (48) and  $L$  given by (50) has the property (51). For  $n = 2$  we get the same solution if, in (53) each  $\prod_{i=1}^m \varphi[K_j]^{c_i}$  term is multiplied by a  $T$  which satisfies (54) but is otherwise arbitrary.*

From the Lemma and from (41) we get

$$F_j(x_1, y_1, \dots, x_n, y_n) = \psi[x_j^a y_j^a L(x_1^a y_1^b, \dots, x_n^a y_n^b)] \quad (j = 1, \dots, n), \quad (55)$$

which, by (16) and (17), gives explicit expressions for  $P(e_j : E)$  and  $P_i(e_j : E)$  ( $i = 1, \dots, m; j = 1, \dots, n$ ).

All these results can be summarized, as they concern the probabilities and scales, in the following main theorem.

**Theorem 3.** *Assume the following. [A]:  $(R, P, v, w, n)$  and  $(R, P_i, v_i, w_i, n)$  ( $i = 1, \dots, m; m > 1$ ) are nonatomic two-scale  $n$ -families of choice probabilities with the same generating system  $(F_1, \dots, F_n)$  (i. e. (16), (17) and (27), (28) hold). [B]: There exist functions  $H_1, \dots, H_n, \tilde{v}, \tilde{w}, \Phi_1, \Phi_2, M_1, M_2, N$ , injective functions  $K_1, \dots, K_n$ , in the sense (36) and locally bounded nonconstant functions  $h_1, \dots, h_n, G_1, G_2$  with*

$$h_1(1, 1) = \dots = h_n(1, 1) = G_1(1, \dots, 1) = G_2(1, \dots, 1) = 1$$

such that  $x \mapsto G_1(x, \dots, x)$  is nonconstant and equations (18)–(26) are satisfied.

Then the following statements hold.

(a) *There exist nonzero real numbers  $a, b$ , a continuous strictly monotonic function  $\psi$  which satisfies (48) and a function  $L$  satisfying (50) and (51) such that, for all*

$$E = \{e_1, \dots, e_n\} \in S,$$

$$P(e_j : E) = \psi(v(e_j)^a w(e_j)^b L[v(e_1)^a w(e_1)^b, \dots, v(e_n)^a w(e_n)^b]), \quad (56)$$

$$P_i(e_j : E) = \psi(v_i(e_j)^a w_i(e_j)^b L[v_i(e_1)^a w_i(e_1)^b, \dots, v_i(e_n)^a w_i(e_n)^b]) \quad (57)$$

$$(i = 1, \dots, m; j = 1, \dots, n).$$

(b) Furthermore, there exist real numbers  $c_1, \dots, c_n$  with nonzero sum such that

$$v(e_j) = \prod_{i=1}^m v_i(e_j)^{c_i}, \quad w(e_j) = \prod_{i=1}^m w_i(e_j)^{c_i}, \quad (j = 1, \dots, n).$$

We have determined also the auxiliary functions:

**Corollary.** Assumptions [A] and [B] imply also the following:

(c) For the functions  $N, M_1, M_2, G_1, G_2, h_1, \dots, h_n, K_1, \dots, K_n$  we have  $N(\lambda) \equiv 1$  ( $\lambda > 0$ ),

$$M_1(\tau_1, \dots, \tau_m) = M_2(\tau_1, \dots, \tau_m) = \prod_{i=1}^m \tau_i^{c_i} \quad (\tau_1, \dots, \tau_m > 0),$$

$$G_1(z_1, \dots, z_m) = G_2(z_1, \dots, z_m) = \prod_{i=1}^m z_i^{c_i},$$

$$h_j(x, y) = x^a y^b \quad (x, y > 0; j = 1, \dots, n),$$

and

$$K_j(z_1, \dots, z_n) = \psi[z_j L(z_1, \dots, z_n)] \quad (z_j > 0; j = 1, \dots, n).$$

(d) Finally, the functions  $H_j$  ( $j = 1, \dots, n$ ) are given by (33), where for  $\Phi_1$  and  $\Phi_2$  we have

$$\Phi_1(t_1, \dots, t_m)^a \Phi_2(t_1, \dots, t_m)^b = \prod_{i=1}^m \varphi(t_i)^{c_i}, \quad (58)$$

if  $n > 2$  and

$$\Phi_1(t_1, \dots, t_m)^a \Phi_2(t_1, \dots, t_m)^b = T(t_1, \dots, t_m) \prod_{i=1}^m \varphi(t_i)^{c_i}, \quad (59)$$

if  $n = 2$ . Here  $\varphi$  is the inverse function of  $\psi$ ,  $t_j \in ]0, 1[$  ( $j = 1, \dots, m$ ) and  $T$  is an arbitrary function satisfying (54).

Somewhat intricate calculations show that the converse of Theorem 3 and Corollary holds too:

**Proposition.** *If (a), (b), (c), (d) hold then also [A], [B], in particular equations (16)–(28) are satisfied.*

**Remark 7.** One may ask the following question. If an experiment has been done, in which various choice probabilities  $P_i$  ( $i = 1, \dots, m$ ) have been measured, how can it be ascertained whether there *exist* appropriate functions  $\psi$  and  $L$ , and real numbers  $a, b$  such that (57) hold. Theorem 3 shows that in principle this could be done by recognizing from the experimental data that each probability depends only upon the corresponding scale values, more exactly upon their power products, and that the  $P_1, \dots, P_m$  can be combined (aggregated) into a probability which has the same property. This aggregation should be implemented in two steps: first taking a function of  $P_1(e_j : E), \dots, P_m(e_j : E)$  (the same function for  $j = 1, \dots, n$ ), then combining the results into what will be the aggregated probability [this is done with the functions  $\Phi$  and  $K_j$  cf. (43), (33), (41), (40), and (19)]. A method of *finding*  $a, b, \psi$  and  $L$  may be based again on noticing that  $P_i(e_j : E)$  depends only upon  $v_i(e_k)^a w_i(e_k)^b$  ( $k = 1, \dots, n$ ), that is

$$P_i(e_j : E) = K_j[v_i(e_1)^a w_i(e_1)^b, \dots, v_i(e_n)^a w_i(e_n)^b] = K_j(z_1, \dots, z_n),$$

which yields  $a, b$ . One could possibly also notice that, with an external function  $\varphi$ , the expression  $z_j = v(e_j)^a w(e_j)^b$  may be extracted from  $\varphi \circ K_j$  (same  $j$ ) so:

$$\varphi[K_j(z_1, \dots, z_n)] = z_j L(z_1, \dots, z_n) \quad (j = 1, \dots, n).$$

The requirement, that (57) should hold, gives

$$\psi = \varphi^{-1} \quad \text{and} \quad L(z_1, \dots, z_n) = \frac{\varphi[K_j(z_1, \dots, z_n)]}{z_j}.$$

If the last expression is the same for all  $j \in \{1, \dots, n\}$  and  $L$  is homogeneous of degree  $-1$  then we have the representations (57), (56) and we know also what  $a, b, \psi$  and  $L$  in it are.

The above process may require, however, too much guesswork. It will therefore probably be as (or, possibly, more) feasible to test these ideas by attempting to fit data with the representation (57). We only touch on the relevant model fitting issues here; as indicated earlier, related issues are discussed in detail in [4], [6], [11], [12],

and in [13]. In the present paper (also in this Remark) we have assumed and continue to assume that the scales  $v_i, w_i$  ( $i = 1, \dots, m$ ) are known; the model fitting issues are similar, though more complex, when the scales have also to be estimated from data. We also continue to assume that only the  $P_i$  ( $i = 1, \dots, m$ ) are observable, i.e., we want to fit the data with a representation that satisfies (57) and thus can be aggregated into a  $P$  as represented in (56). Again, the model fitting issues are similar, though more complex, when  $P$  is also observable.

First we consider the case where we have data only for one set  $E$ . Note that (57) implies that, for every  $e_j, e_k$  in  $E$  and for  $i = 1, \dots, m$ ,

$$\frac{\psi^{-1}[P_i(e_j : E)]}{\psi^{-1}[P_i(e_k : E)]} = \frac{v_i(e_j)^a w_i(e_j)^b}{v_i(e_k)^a w_i(e_k)^b}.$$

Thus one should search for the ‘best fitting’ function  $\psi$  and parameters  $a, b$ , using traditional and newly developing techniques. (For instance, one might compare the fit of  $\psi(t) = t$  (corresponding to a two-scale Luce model with exponents) to that of  $\psi(t) = t + t^{1/2}$ ). Once one has the ‘best fitting’ function  $\tilde{\psi}$ , the estimate  $\tilde{L}$  of  $L$  can be taken to be the solution of

$$\sum_{j=1}^n \tilde{\psi}[z_j \tilde{L}(z_1, \dots, z_n)] = 1.$$

Our task is easier when we have data for (at least) two sets  $E, E'$  with two (or more) common elements, say  $e_j, e_k$ . Then, for (57) to hold, we must have

$$\frac{\psi^{-1}[P_i(e_j : E)]}{\psi^{-1}[P_i(e_k : E)]} = \frac{\psi^{-1}[P_i(e_j : E')]}{\psi^{-1}[P_i(e_k : E')]}.$$

When  $\psi(t) = t$ , we have an *observable property* of the representation. In general, we can first determine the function  $\psi$  that gives the ‘best fit’ to equality for these (and similar) ratios, then turn to the estimation of the exponents  $a, b$  and of the function  $L$ .

**Remark 8.** On the other hand, as mentioned, while (58), (59) and (43) determine  $\Phi$ , even ignoring  $T$  there is some arbitrariness left for  $\Phi_1$  and  $\Phi_2$ : even if  $\Phi_1(t_1, \dots, t_m) = \prod_{i=1}^m t_i^{A_i}$ ,  $\Phi_2(t_1, \dots, t_m) = \prod_{i=1}^m t_i^{B_i}$ , we have only  $aA_i + bB_i = c_i$  ( $i = 1, \dots, m$ ) as restriction. Notice, however, that already in the calculations in section 1 we used the function  $\Phi = \Phi_1^a \Phi_2^b$  for aggregation, rather than the individual functions  $\Phi_1, \Phi_2$ .

Actually, we could write (37) as (cf. (53))

$$\begin{aligned}
& K_j(h_1[G_1(x_{11}, \dots, x_{m1}), G_2(y_{11}, \dots, y_{m1}), \dots, h_n[G_1(x_{1n}, \dots, x_{mn}), G_2(y_{1n}, \dots, y_{mn})]) \\
& = K_j[\Phi(K_1[h_1(x_{11}, y_{11}), \dots, h_n(x_{1n}, y_{1n})], \dots, K_1[h_1(x_{m1}, y_{m1}), \dots, h_n(x_{mn}, y_{mn})]), \\
& \quad \vdots \\
& \quad \Phi(K_n[h_1(x_{11}, y_{11}), \dots, h_n(x_{1n}, y_{1n})], \dots, K_n[h_1(x_{m1}, y_{m1}), \dots, h_n(x_{mn}, y_{mn})])] \\
& \qquad \qquad \qquad (j = 1, \dots, n), \qquad (60)
\end{aligned}$$

which has the advantage of containing  $\Phi$  rather than  $\Phi_1$  and  $\Phi_2$ , although it does not look simpler than (37). But the symmetry in having both  $\Phi_1$  and  $\Phi_2$  as counterparts of  $G_1$  and  $G_2$  may be attractive, even though it turns out eventually that  $G_1 = G_2$  (but in general  $\Phi_1 \neq \Phi_2$ ). Ultimately the choice between (37) and (60) is a matter of taste.

**Remark 9.** The above Proposition shows that *without adding further assumptions to [A] and [B] one cannot restrict  $\Phi_1, \Phi_2, T$  and  $\psi, L$  more closely than by (58), (59), (54) and by (48), (50), (51), respectively.* One further restriction will complete the characterization of two-scale Luce models with exponents in the next section.

## 4. Characterization of the “Two-Scale Luce Model with Exponents”

In order to get (38), that is (9), characterized, we aim to obtain  $\varphi(t) = t$  because then  $\psi(s) = s$ , so that (55) reduces to

$$F_j(x_1, y_1, \dots, x_n, y_n) = x_j^a y_j^b L(x_1^a y_1^b, \dots, x_n^a y_n^b) \quad (j = 1, \dots, n). \quad (61)$$

Summing from 1 to  $n$  we get, in view of (27),

$$1 = \sum_{k=1}^n x_k^a y_k^b L(x_1^a y_1^b, \dots, x_n^a y_n^b),$$

that is,

$$L(x_1^a y_1^b, \dots, x_n^a y_n^b) = \frac{1}{\sum_{k=1}^n x_k^a y_k^b},$$

so that (61) indeed becomes (38):

$$F_j(x_1, y_1, \dots, x_n, y_n) = \frac{x_j^a y_j^b}{\sum_{k=1}^n x_k^a y_k^b}, \quad (j = 1, \dots, n).$$

If  $\varphi(t) = t$  then  $\Phi(t_1, \dots, t_m)$  given by (52) will be, cf. (45),

$$\Phi(t_1, \dots, t_m) = \prod_{i=1}^m t_i^{c_i} = G(t_1, \dots, t_m). \quad (62)$$

Conversely, if (62) holds then, by (47) and (44),

$$\varphi(t) = \Phi(t, \dots, t)^{\frac{1}{c}} = t^{\Sigma c_i / \Sigma c_i} = t.$$

[As we saw before Theorem 2, (52) and thus (62) is a possible choice for  $n = 2$  and the only choice for  $n > 2$ ]. So we have the following.

**Theorem 4.** *The choice probabilities in the “two-scale Luce model with exponents”, that is those given by (1) or equivalently by (9), and only these satisfy, in addition to the conditions in Theorem 2 or Theorem 3, also  $\Phi = G$ .*

**Remark 10.** Till now we supposed that  $G_1(x, \dots, x)$  is not constant, that is, see (45),  $\sum_{i=1}^m c_i \neq 0$ . Conversely,  $\sum_{i=1}^m c_i = 0$  means that  $G_1(x, \dots, x)$  is constant. But  $G_1(x_1, \dots, x_m)$  should *not* be constant, that is,  $c_p \neq 0$  should hold for at least one  $p$ . In this case, everything else in Theorem 4 being unchanged, one gets instead of (38) the equation

$$F_j(x_1, y_1, \dots, x_n, y_n) = \frac{C_j x_j^a y_j^b}{\sum_{k=1}^n C_k x_k^a y_k^b} \quad (j = 1, \dots, n), \quad (63)$$

where  $C_1, \dots, C_n$  are positive constants (the proof uses a method similar to that in section 4 of [1]).

The representation in (63) has interesting psychological interpretations. The form indicates that the ordinal position of an option  $e$  in the set  $E$  affects its representation – in particular, option  $e_j$  in position  $j$  has a ‘weighting’ parameter  $C_j$  placed in front of its (e. g. overall) scale value  $v(e_j)^a w(e_j)^b$ . Returning to the noise burst example in section 1, one could think of the stimuli  $e_j$  ( $j = 1, \dots, n$ ), as being presented in temporal order, with  $e_j$  being presented before  $e_{j+1}$  for  $j = 1, \dots, n-1$ . The weight  $C_j$

then might be interpreted as a *bias* to report the stimulus in position  $j$  as louder than its scale values would indicate; alternatively,  $C_j$  might be interpreted as a *memory* effect, with stimuli at different positions in the sequence not being equally represented relative to their scale values.

## 5. Conclusion. Open problem

In this paper we have extended results in [1] to the aggregation of two-scale homogeneous, simply scalable, conjoint Weber type II families of choice probabilities. The further extension to more than two components is routine. The proofs required us to assume that the representations are *simply scalable* (Eq. (22)) in order that we can use an appropriate *injectivity condition* (Eq. (36)). The essence of the injectivity condition used in this paper (and in [1]) is that (certain combinations of) the parameters of the model are *identifiable* from the data. It would be of interest to develop results that do not require the use of simple scalability as a means to get to an appropriate injectivity condition.

Consider, for instance, the following generalization of the “similarity choice model” (5):

$$P(r[e_j]|e_k:E) = F_j[\eta(e_k, r[e_1]), \beta(r[e_1]), \dots, \eta(e_k, r[e_n]), \beta(r[e_n])] \quad (j, k = 1, \dots, n) \quad (64)$$

( $E = \{e_1, \dots, e_n\}$ ), where the scales  $\eta$  and  $\beta$  satisfy (cf. (6))

$$\begin{aligned} 0 \leq \eta(e_j, e_k) \leq 1, \quad \eta(e_j, e_k) &= \eta(e_k, e_j), \quad \eta(e_j, e_j) = 1, \\ 0 \leq \beta(r[e_j]) \leq 1, \quad \sum_{k=1}^n \beta(r[e_k]) &= 1 \quad (j, k = 1, \dots, n). \end{aligned} \quad (65)$$

Assume that the model is “*identifiable*” (cf. [12]) in the sense that, if

$$\begin{aligned} &F_j[\eta(e_k, r[e_1]), \beta(r[e_1]), \dots, \eta(e_k, r[e_n]), \beta(r[e_n])] \\ &= F_j[\eta'(e_k, r[e_1]), \beta'(r[e_1]), \dots, \eta'(e_k, r[e_n]), \beta'(r[e_n])] \end{aligned}$$

for all  $j, k = 1, \dots, n$  and for the scales  $\eta, \beta$  and  $\eta', \beta'$  satisfying (65), then

$$\eta(e_j, e_k) = \eta'(e_j, e_k) \text{ and } \beta(r[e_j]) = \beta'(r[e_j]) \quad (j, k = 1, \dots, n).$$



Notice that this property resembles, but also differs from, the injectivity (36). The ‘similarity choice model’ (5) is a model “identifiable” in the sense just stated (cf. e.g. [4]). It is an open problem what the general “identifiable” solutions of a system of equations for  $F_j$  ( $j = 1, \dots, n$ ) in (64), corresponding to (37) (or to (60)), are – and what should be supposed, for adjusting the number of equations to the number of parameters (scale values), in place of an analogue of the simple scalability (22) and such properties of the functions  $h_j$  ( $j = 1, \dots, n$ ) therein as (25) and (26). Closely related questions are considered in [12].

## References

- [1] J. Aczél, Gy. Maksa, A.A.J. Marley, Z. Moszner, Consistent aggregation of scale families of selection probabilities, *Mathematical Social Sciences* **33**(1997), 227-250.
- [2] J.-C. Falmagne, G. Iverson, Conjoint Weber laws and additivity, *Journal of Mathematical Psychology* **20**(1979), 164-183.
- [3] R.D. Luce, Detection and Recognition, in: R.D. Luce, R.R. Bush, E. Galanter, (Eds.), *Handbook of Mathematical Psychology*, Vol. I, Wiley, New York, 1963, 103-189, in partic. 113-116.
- [4] J.T. Townsend, D.E. Landon, An experimental and theoretical investigation of the constant-ratio rule and other models of visual letter confusion, *Journal of Mathematical Psychology* **25**(1982), 119-162.
- [5] A.A.J. Marley, Developing and characterizing multidimensional Thurstonian and Luce models for identification and preference, in: F. G. Ashby, (Ed.), *Multidimensional models of perception and cognition*, Erlbaum, Hillsdale, NJ, 1992, 299-333.
- [6] F.G. Ashby, W.T. Maddox, W.W. Lee, On the dangers of averaging across subjects when using multidimensional scaling or the similarity choice model, *Psychological Science* **5**(1994), 144-151.

- [7] A. Rényi, Probability Theory, North-Holland – Elsevier, Amsterdam/New York, 1970, 81-84.
- [8] J.-C. Falmagne, On a recurrent misuse of a classical unctional equation result, *Journal of Mathematical Psychology* **3** (1981), 190-193.
- [9] A.A.J. Marley, Random utility models with all choice probabilities expressible as ‘functions’ of the binary choice probabilities, *Mathematical Social Sciences* **3**(1982), 39-56.
- [10] J. Aczél, A Short Course on Functional Equations Based Upon Recent Applications in the Social and Behavioral Sciences, Reidel-Kluwer, Dordrecht/Boston, 1987, 52-56.
- [11] W.K. Estes, Some reflections on the role of the choice model in theories of categorization, identification, and learning, in: A.A.J. Marley, (Ed.), *Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce*, Erlbaum, Mahwah, NJ, 1997, 321-328.
- [12] D. Bamber, J.P.H. van Santen, How to tell whether a model is testable: The example of the discrete-state selective-influence model, to appear in *Journal of Mathematical Psychology* **43**(1999).
- [13] I.J. Myung, M. Browne, M. Foster, (Eds.), Special issue on model selection, *Journal of Mathematical Psychology* **43**(1999).