Available online at www.sciencedirect.com

SCIENCE DIRECT

Journal of Mathematical Psychology 49 (2005) 464–480

Journal of Mathematical Psychology

www.elsevier.com/locate/jmp

Some probabilistic models of best, worst, and best–worst choices

A.A.J. Marley[a,b,*], J.J. Louviere[c]

[a]Department of Psychology, University of Victoria, Victoria, BC, Canada V8W 3P5
[b]Department of Economics, University of Groningen, The Netherlands
[c]School of Marketing, University of Technology, Sydney, NSW Australia

Received 26 October 2004; received in revised form 15 May 2005
Available online 19 August 2005

Abstract

Over the past decade or so, a choice design in which a person is asked to select both the best and the worst option in an available set of options has been gaining favor over more traditional designs, such as where the person is asked, for instance, to: select the best option; select the worst option; rank the options; or rate the options. In this paper, we develop theoretical results for three overlapping classes of probabilistic models for best, worst, and best–worst choices, with the models in each class proposing specific ways in which such choices might be related. The models in these three classes are called random ranking and random utility, joint and sequential, and ratio scale. We include some models that belong to more than one class, with the best known being the maximum-difference (*maxdiff*) model, summarize estimation issues related to the models, and formulate a number of open theoretical problems.

Keywords: Best–worst choice; Preference; Choice; Probabilistic modeling; Discrete choice

1. Introduction

Finn and Louviere (1992) proposed a discrete choice task in which a person is asked to select both the best and the worst option in an available (sub)set of options. They presented an analytical model for data from that task, and considered its evaluation using an experimental design (2^j fractional factorial) that ensures that each option, and each pair of distinct options, is presented equally often across the selected subsets of size j. Since the publication of that paper, interest in, and use of, such best-worst choice tasks has been increasing, with two recent empirical applications receiving "best paper" awards (Cohen, 2003; Cohen and Neira, 2003). In principle, best–worst tasks have a number of advantages over traditional discrete choice tasks: (1) a single pair of best–worst choices contains a great deal of information about the person's ranking of options (e.g., if there are 3 items in a set, one obtains the entire ranking of that set; if there are 4 items in a set, one obtains information on the implied best option in 9 of the 11 possible non-empty, non-singleton subsets;[1] and if there are 5 items in a set, one obtains information on the implied best option in 18 of the 26 possible non-empty, non-singleton subsets); (2) best–worst tasks take advantage of a person's propensity to identify and respond more consistently to extreme options; and (3) best–worst tasks seem to be easy for people. Despite increasing use of the approach, the underlying models have not been axiomatized, leaving practitioners

*Corresponding author. Department of Psychology, University of Victoria, P.O. Box 3050 STN CSC, Victoria, BC, Canada V8W 3P5. Fax: +1 250 721 8929.

E-mail addresses: ajmarley@uvic.ca (A.A.J. Marley), jordan.louviere@uts.edu.au (J.J. Louviere).

[1]Let $\{a, b, c, d\}$ be the set of options and suppose that we know that a is selected as best and d as worst. If we now check each (sub)set of size 2, 3, 4 of $\{a, b, c, d\}$ in turn, then we see that the subsets $\{b, c\}$ and $\{b, c, d\}$ are the only ones where the best element is not determined either by the information that a is best or by the information that d is worst.

0022-2496/$ - see front matter
doi:10.1016/j.jmp.2005.05.003

without clear guidelines on appropriate experimental designs, data analyses, and interpretation of results.

This paper develops theoretical results for three overlapping classes of probabilistic models for best, worst, and best–worst choices, with the models in each class proposing specific relationships between such choices. The model classes are called random ranking and random utility, joint and sequential, and ratio scale. We consider models that belong simultaneously to one or more of these classes, with the best known being the maximum-difference (maxdiff) model, which is introduced later in this section, and we formulate a number of open theoretical problems.

We now illustrate the framework, and the basic models, through the maximum-difference (maxdiff) model of best–worst choice. To do so we require some basic notation. Let T with $|T|\geqslant 2$ denote the finite set of potentially available choice options, and for any subset $X\subseteq T$, with $|X|\geqslant 2$, let $B_X(x)$ denote the probability that the alternative x is chosen as best in X, $W_X(y)$ the probability that the alternative y is chosen as worst in X, and $BW_X(x,y)$ the probability that, jointly, the alternative x is chosen as best in X and the alternative $y\neq x$ is chosen as worst in X. Thus

$$0\leqslant B_X(x), W_X(y), BW_X(x,y)\leqslant 1$$

and

$$\sum_{x\in X} B_X(x)=\sum_{y\in X} W_X(y)=\sum_{\substack{x,y\in X\\ x\neq y}} BW_X(x,y)=1.$$

We assume throughout the paper that for each $x\in T$, $B_{\{x\}}(x)=W_{\{x\}}(x)=1$.

For motivational purposes only, we assume now that best and worst choices are in some sense more basic than simultaneous best–worst choices, and develop a model of the latter based on the former. So suppose that when asked to choose the best and the worst option in a (finite) set X, the person simultaneously, but independently, chooses the best, respectively, the worst, option in X. If the resulting options are distinct, these are reported as the best–worst pair of options in X, otherwise the person re-samples. As we show later in detail, such a process gives rise to the following representation for the best–worst choice probabilities in terms of the best and the worst choice probabilities:[2] for $x,y\in X, x\neq y$

$$BW_X(x,y)=\frac{B_X(x)W_X(y)}{\sum_{\substack{r,s\in X\\ r\neq s}} B_X(r)W_X(s)}. \tag{1}$$

Now, suppose that the multinomial logit (MNL or Luce's choice) model holds separately for the best and the worst choice probabilities, i.e., there exist ratio scales b and w such that for $x,y\in X$,

$$B_X(x)=\frac{b(x)}{\sum_{r\in X} b(r)},\quad W_X(y)=\frac{w(y)}{\sum_{s\in X} w(s)}. \tag{2}$$

Then direct substitution of (2) in (1) yields that for $x\neq y$,

$$BW_X(x,y)=\frac{b(x)w(y)}{\sum_{\substack{r,s\in X\\ r\neq s}} b(r)w(s)}. \tag{3}$$

Notice that this combined set of representations has three interesting properties—the best–worst choice probabilities are represented in (1) directly in terms of the best and worst choice probabilities, and are represented in (3) in terms of the ratio scale values that determine the best and the worst choice probabilities, with the same functional form in both representations. Thus, this aggregation method is plausible, and interesting, in that it works *both* at the level of the choice probabilities *and* at the level of the scale values. We are immediately lead to the first theoretical question, namely, are there other (ratio scale) models that satisfy this type of aggregation property, and, if so, how large is this class. The detailed formulation of this problem requires extensive further notation, and its solution the use of complex functional equation techniques. The conjectured final result is that the class of solutions is large, but in an interesting sense the individual solutions do not differ greatly from the above model.

It is important to note that the above example was for choices from a given *fixed* set X. Often probabilistic models of choice (or, briefly, probabilistic choice models) are assumed to be consistent over all the subsets of a finite master set T. If that is assumed in the present context, and we consider the binary choice probabilities $B_{\{x,y\}}(x)$ and $W_{\{x,y\}}(y)$, then it is reasonable to assume that, and to test whether,

$$B_{\{x,y\}}(x)=W_{\{x,y\}}(y)$$

which with (2) gives that[3] for each for $z\in T$,

$$b(z)=\frac{1}{w(z)} \tag{4}$$

and so in particular any scale transform α of b is linked to the scale transform $\frac{1}{\alpha}$ of w. In the following work, we consider separately the cases where b and w are independent ratio scales, and where they are subject to common scale transforms.

The paper has the following structure. Each of the three main sections develops one of three classes of models, gives examples of that class, and presents solved, and open, theoretical problems about that class. Section 2 introduces the terminology and basic conditions, Section 3 presents random ranking and random utility models, Section 4 joint and sequential choice

[2] We discuss later the theoretical and empirical reasonableness of this representation when $|X|=2$.

[3] (2) gives actually that $b(z)=c/w(z)$ for some constant $c>0$, but no generality is lost by assuming that $c=1$.

models, and Section 5 ratio scale models. Section 6 summarizes the results and restates the main open theoretical problems.

2. General terminology and conditions

Given a finite master set T, and a particular set X, $X \subseteq T$, we refer to the set $\{B_X(x), x \in X\}$, $\{W_X(y), y \in X\}$, $\{BW_X(x,y), x, y \in X, x \neq y\}$, respectively, as a *set of best*, *worst*, *best–worst choice probabilities* (*on X*). We have a *complete set of best*, *worst*, *best–worst choice probabilities*, *respectively*, (*on a master set T*) when we have a set of best, worst, best–worst choice probabilities on each X, $X \subseteq T$. Unless stated otherwise, we assume that we have a complete set of best, worst, and best–worst choice probabilities on a finite master set T with $|T| \geqslant 2$.

One can, of course, study distinct models for each of the three types of choice paradigm—that is, for best, worst, and best–worst. However, this creates a very large number of model types. For instance, as indicated above, in the paper we discuss three classes of models—random ranking and random utility, joint and sequential choice, and ratio scale. If the model for each type of choice paradigm—best, worst, best–worst—is assumed to belong independently to each of these classes, then we have already 27 different patterns of assumptions, and, in fact, there will be even more possible patterns because of model subtypes within each model class. Partly because of this proliferation of models, but also because it makes conceptual, and one would hope, empirical, sense, we will assume normally that the three types of choices satisfy a common type of model. We explore also certain natural consistency conditions for linking the best, worst, and best–worst choices on single or multiple sets. The following is one such condition:

Definition 1. A set of best, worst, and best–worst choice probabilities on a set X has *consistent margins* iff for each $x, y \in X$,

$$B_X(x) = \sum_{z \in X - \{x\}} BW_X(x, z)$$

and

$$W_X(y) = \sum_{z \in X - \{y\}} BW_X(z, y).$$

Perhaps surprisingly, we present plausible models that do not satisfy the above condition—and others that do. Also, note that the best and worst choice probabilities will not in general determine a unique set of best–worst choice probabilities, but if we have consistent margins, the converse in true. Even if we do not have consistent margins, best–worst choices likely focus the person's attention on the task in a way that separate best and worst choices do not.

Random ranking and random utility models are the most commonly studied probabilistic choice models, and frequently models in the other classes have alternative descriptions as models in this class. Thus, we include the general terminology related to random ranking and random utility models in the present section.

First, we need the idea of a ranking of a set X from its best (most preferred) to its worst (least preferred) element, and similarly of a ranking of X from its worst (least preferred) to its best (most preferred) element. We refer to these as *best to worst*, and *worst to best*, rankings, respectively. Such rankings may be empirical, i.e., resulting from a person's judgments, but of at least equal importance in this paper is the role such rankings play in the development of probabilistic choice models. For any set X, $X \subseteq T$, let $R(X)$ denote the set of rank orders of X. Then, with $|X| = n$, and a given rank order $\rho = \rho_1 \rho_2 \ldots \rho_{n-1} \rho_n$ of X, let $B_X(\rho)$ denote the probability that ρ occurs as a best to worst rank order, and $W_X(\rho)$ the probability that ρ occurs as a worst to best rank order—thus, in the former case ρ_1 is the best element in the rank order, in the latter case it is the worst element.[4] The assumption that $B_X(\rho)$ and $W_X(\rho)$ are probabilities and that a ranking occurs at each choice opportunity is summarized by: for each $\rho \in R(X)$,

$$0 \leqslant B_X(\rho), W_X(\rho) \leqslant 1$$

and

$$\sum_{\rho \in R(X)} B_X(\rho) = 1 = \sum_{\rho \in R(X)} W_X(\rho).$$

A set of best and worst ranking probabilities on a set X is any such set $\{B_X(\rho), W_X(\rho) : \rho \in R(X)\}$. *A complete set of best and worst ranking probabilities on a set T* is any set $\{B_X(\rho), W_X(\rho) : \rho \in R(X), X \subseteq T\}$.

When we have a complete set of rankings, we can ask what the relation is between the probabilities of the rankings on the master set T and on a given subset $X \subseteq T$. To state possible relations, consider rankings $\rho \in R(X)$ and $\sigma \in R(T)$, let $\sigma | X = \rho$ mean that the ranking σ on T agrees with the ranking ρ on X, and let

$$R_T(\rho, X) = \{\sigma | \sigma \in R(T) \text{ with } \sigma | X = \rho\}.$$

Then a pair of plausible, though not necessary, assumptions is that for each $\rho \in R(X)$,

$$B_X(\rho) = \sum_{\sigma \in R_T(\rho, X)} B_T(\sigma)$$

[4]A more precise notation would be $B_{R(X)}(\rho)$ and $W_{R(X)}(\rho)$, rather than $B_X(\rho)$ and $W_X(\rho)$. However, the simpler notation should not lead to any confusion.

and

$$W_X(\rho) = \sum_{\sigma \in R_T(\rho, X)} W_T(\sigma).$$

3. Random ranking and random utility models

We are naturally interested in relations between these various choice and ranking probabilities. Since the focus of this paper is best–worst choice, we concentrate on how ranking probabilities might determine such best–worst choice probabilities, though we do consider one result in the other direction. There is extensive related work on how best choice probabilities determine, or are determined by, ranking probabilities (see Falmagne, 1978; Fishburn, 1994, 2002).

For each $z \in T$ and rank order $\sigma \in R(T)$, let $\sigma^{-1}(z)$ denote the rank position of z in σ, so, for instance, $\sigma^{-1}(z) = 1$ means that z is in the first position in the rank order. Now let

$$R_T(x, X - \{x\}) \\ = \{\sigma | \sigma \in R(T), \sigma^{-1}(x) < \sigma^{-1}(z), z \in X - \{x\}\},$$

$$R_T(X - \{y\}, y) \\ = \{\sigma | \sigma \in R(T), \sigma^{-1}(z) < \sigma^{-1}(y), z \in X - \{y\}\},$$

$$R_T(x, X - \{x, y\}, y) \\ = \{\sigma | \sigma \in R(T), \sigma^{-1}(x) < \sigma^{-1}(z) < \sigma^{-1}(y), z \in X - \{x, y\}\}.$$

For instance, $R_T(x, X - \{x, y\}, y)$ is the set of rank orders of T for which, relative to only the elements of X, x is first (best), and y is last (worst). A first observation is that a set of best–worst choice probabilities on a single finite set X is always consistent with such a set of rank orders on that set. To see this, consider the equalities that must be satisfied for this result to hold: there is a probability measure p_X on the set of rank orders of X such that for $x, y \in X, x \neq y$,

$$BW_X(x, y) = \sum_{\sigma \in R_X(x, X - \{x, y\}, y)} p_X(\sigma). \tag{5}$$

Note that if $|X| = 2$, then for any $x, y \in X$, $x \neq y$, we can set $p_X(xy) = BW_X(x, y)$, which determines the rank order probabilities $p_X(\rho)$. For any X with $|X| > 2$, each rank order $\rho \in R(X)$ belongs to exactly one of the sets $R_X(x, X - \{x, y\}, y)$, and so, for each $x, y \in X$, $x \neq y$, we can arbitrarily set the probability of one of the rank orders in the set $R_X(x, X - \{x, y\}, y)$ equal to $BW_X(x, y)$, and all the others to zero, leading to (5).

The following two definitions are natural extensions to best–worst choices of a standard definition.

Definition 2. A complete set of best choice probabilities on a finite set T satisfies a *best random ranking model* provided there is a probability measure b_T on the set of rank orders of T, such that for each $x \in X \subseteq T$,

$$B_X(x) = \sum_{\sigma \in R_T(x, X - \{x\})} b_T(\sigma).$$

A complete set of worst choice probabilities on a finite set T satisfies a *worst random ranking model* provided there is a probability measure w_T on the set of rank orders of T, such that for each $y \in X \subseteq T$,

$$W_X(y) = \sum_{\sigma \in R_T(X - \{y\}, y)} w_T(\sigma).$$

A complete set of best–worst choice probabilities on a finite set T satisfies a *best–worst random ranking model* provided there is a probability measure p_T on the set of rank orders of T, such that for each $x, y \in X \subseteq T, x \neq y$,

$$BW_X(x, y) = \sum_{\sigma \in R_T(x, X - \{x, y\}, y)} p_T(\sigma).$$

Note that the above definition assumes no relation between the ranking probability measures underlying the best, worst, and best–worst choice probabilities. In line with our earlier observations, we consider next the case where they agree.

Definition 3. A complete set of best, worst, and best–worst choice probabilities on a finite set T satisfies a (*best, worst, and best–worst*) *consistent random ranking model* provided there is a probability measure q_T on the set of rank orders of T, such that for each $x, y \in X \subseteq T$,

$$B_X(x) = \sum_{\sigma \in R_T(x, X - \{x\})} q_T(\sigma), \tag{6}$$

$$W_X(y) = \sum_{\sigma \in R_T(X - \{y\}, y)} q_T(\sigma) \tag{7}$$

and

$$BW_X(x, y) = \sum_{\sigma \in R_T(x, X - \{x, y\}, y)} q_T(\sigma) \quad (x \neq y). \tag{8}$$

As we see now, the above assumptions are eminently testable. Should they fail we can turn to more general cases where the best, worst, and best–worst choice probabilities, respectively, each satisfy possibly distinct random ranking models.

We now summarize various theoretical issues associated with the definition of a consistent random ranking model, Definition 3. The first result is immediate from the relevant definitions.

Proposition 4. *Consider a complete set of best, worst, and best–worst choice probabilities such that the best–worst choice probabilities satisfy a (best–worst) random ranking model, Definition 2. Then the set satisfies a consistent (best, worst, and best–worst) random ranking model, Definition 3, iff it has consistent margins, Definition 1.*

Thus, if we know that a set of best–worst choice data satisfies a random ranking model, then we can test whether or not an additional set of best (vis, worst) choice data is consistent with the same random ranking model by comparing the relevant margins of the best–worst data with the best (vis, worst) choice data. Thus, an important open theoretical question is what are necessary and sufficient conditions for a complete set of best–worst choice probabilities to satisfy a random ranking model. The following are necessary conditions:

(i) The marginal choice probabilities $\sum_{z\in X-\{x\}} BW_X(x,z)$ satisfy a best random ranking model, and the marginal choice probabilities $\sum_{z\in X-\{y\}} BW_X(z,y)$ satisfy a worst random ranking model (with common rank order probabilities).

(ii) The best–worst choice probabilities satisfy *regularity*: for distinct $x,y\in X\subseteq Y\subseteq T$, $x\neq y$,

$$BW_X(x,y)\geqslant BW_Y(x,y). \tag{9}$$

(iii) For distinct $\{x,y,z\}=X\subseteq T$,

$$BW_{\{x,y\}}(x,y)=BW_X(x,y)+BW_X(x,z)+BW_X(z,y). \tag{10}$$

When $|T|=3$, i.e., we have only a 3 element set and its 2 element subsets, the above condition, (10), is necessary and sufficient for the best–worst choice probabilities to satisfy a random ranking model. This is easily seen by setting, for each $\rho=\rho_1\rho_2\rho_3\in R(T)$, $p_T(\rho)=BW_T(\rho_1,\rho_3)$. Then the best–worst choice probabilities on T are compatible with the random ranking model, and the compatibility of the best–worst choices for the 2 element subsets follows by substituting the rank order probabilities associated with the terms in the right hand side of (10) with $X=T$ in those terms.

One would hope that an approach similar to that used to solve the parallel general case for best (vis., worst) choice probabilities (see Falmagne, 1978; Barbera and Pattanaik, 1986; Fiorini, 2004) would give a general set of necessary and sufficient conditions in the present case, though we do not currently see how to do so. It is quite possible that the relevant conditions will have a similar structure to those in the earlier literature—for instance, the first such condition beyond regularity that occurs in the characterization of best choice probabilities yields a parallel necessary condition in the present situation, though it does not seem to have an obvious interpretation. The condition is: for a master set $T=\{x,y,z,w\}$,

$$BW_{(x,y)}(x,y)-BW_{\{x,y,z\}}(x,y)-BW_{\{x,y,w\}}(x,y)+BW_{\{x,y,z,w\}}(x,y)\geqslant 0,$$

which is easily checked to be a necessary condition by substituting in the relevant rank orders from (8) and checking that the above expression reduces to one involving only nonnegative (rank order) terms.

We have been assuming so far that the best–worst choice probabilities are generated by a random ranking model. We can consider also the possibility that the rankings are generated by a sequence of best–worst choices to yield the following representation: for $\rho\in R(X), |X|=n$,

$$p_X(\rho)=\begin{cases} BW_X(\rho_1,\rho_n) \\ BW_X(\rho_1,\rho_n)BW_{X-\{\rho_1,\rho_n\}}(\rho_2,\rho_{n-1}) \\ BW_X(\rho_1,\rho_n)BW_{X-\{\rho_1,\rho_n\}}(\rho_2,\rho_{n-1})\ldots BW_{\{\rho_j,\rho_{j+1}\}}(\rho_j,\rho_{j+1}) \\ BW_X(\rho_1,\rho_n)BW_{X-\{\rho_1,\rho_n\}}(\rho_2,\rho_{n-1})\ldots BW_{\{\rho_j,\rho_{j+1},\rho_{j+2}\}}(\rho_j,\rho_{j+2}) \end{cases}$$

$$\text{if}\begin{cases} n=2,3, \\ n=4,5, \\ n=2j,\ j\geqslant 3, \\ n=2j+1,\ j\geqslant 3. \end{cases}$$

The distinction between the cases with an even versus an odd number of elements arises because in the latter case, with $2j+1$ elements, the rank position of the final element is determined after j best–worst choices.

We have the open problem of determining whether there are any complete sets of best–worst choice probabilities on (all the subsets of) a set T that satisfy a random ranking model with the above relation between the probabilities of the rank orders in that random ranking model and the complete set of best–worst choice probabilities on the subsets $X\subseteq T$. This question parallels a solved classic one regarding Luce's choice model (Luce and Suppes, 1965).

We now introduce a general class of random utility models for best, worst, and best–worst choice probabilities, specialize it in a way that produces representations formally equivalent to the earlier random ranking framework, and then study related random utility models that do not retain this equivalence. Throughout the paper, we implicitly assume that there is a zero probability of two distinct random variables being equal, which ensures that a single option is selected at each choice opportunity.

Definition 5. A complete set of best, worst, and best–worst choice probabilities on a finite set T satisfies a (*best, worst, best–worst*) *random utility model* provided there are random variables $\mathbf{B}_z, \mathbf{W}_z, \mathbf{BW}_{r,s}, r,s,z\in T$, $r\neq s$, such that for each $x,y\in X\subseteq T$,

$$B_X(x)=\Pr(\mathbf{B}_x=\max_{z\in X}\mathbf{B}_z),$$

$$W_X(y)=\Pr\left(\mathbf{W}_y=\max_{z\in X}\mathbf{W}_z\right)$$

and

$$BW_X(x,y)=\Pr(\mathbf{BW}_{x,y}=\max_{\substack{r,s\in X\\ r\neq s}}\mathbf{BW}_{s,t})\quad(x\neq y).$$

Note that this definition does not imply any relations between the best, worst, and best–worst choice probabilities (other than that each set separately satisfies a random utility model). As we have argued earlier, such generality is unwarranted at this time, so we now look at specializations that involve assumed relations between the three sets of random variables.

One might think that we should require that, when $|X| = 2$,

$$BW_{\{x,y\}}(x, y) = B_{\{x,y\}}(x) = W_{\{x,y\}}(y)$$

or, more generally, that for some α, $0 \leqslant \alpha \leqslant 1$,

$$BW_{\{x,y\}}(x, y) = \alpha B_{\{x,y\}}(x) + (1 - \alpha) W_{\{x,y\}}(y).$$

Neither property holds generally for random utility models that satisfy Definition 5. It is a theoretical and empirical issue to decide the appropriate resolution of this apparent inconsistency. Shafir (1993) presents data, in an accept versus reject design, that can be interpreted as showing that $B_{\{x,y\}}(x)$ may differ from $W_{\{x,y\}}(y)$.

Definition 6. A complete set of best, worst, and best–worst choice probabilities on a finite set T satisfies a (*best, worst, best–worst*) *consistent random utility model* provided it satisfies a random utility model, Definition 5, with a common set of random variables $\mathbf{U}_z$, $z \in T$, such that for $r, s \in T$, $r \neq s$,

$$\mathbf{B}_z = -\mathbf{W}_z = \mathbf{U}_z$$

and

$$\mathbf{BW}_{r,s} = \mathbf{U}_r - \mathbf{U}_s.$$

It is important to note that the random variable notation above is intended to mean that the best–worst choice probabilities are derived on the basis of a *single common* set of sample values $\mathbf{U}_z, z \in T$. Combining the various assumptions, a consistent random utility model becomes

$$B_X(x) = \Pr\left(\mathbf{U}_x = \max_{z \in X} \mathbf{U}_z\right), \tag{11}$$

$$W_X(y) = \Pr\left(\mathbf{U}_y = \min_{z \in X} \mathbf{U}_z\right), \tag{12}$$

$$BW_X(x, y) = \Pr(\mathbf{U}_x > \mathbf{U}_z > \mathbf{U}_y, z \in X - \{x, y\}) \quad (x \neq y). \tag{13}$$

We have the following equivalence:

Proposition 7. *A complete set of best, worst, and best–worst choice probabilities on a finite set T satisfies a consistent random utility model, Definition* 6, *iff it satisfies a consistent random ranking model, Definition* 3.

This result follows routinely, using classic methods for showing such equivalences (Luce and Suppes, 1965; Falmagne, 1978). Note that such models have consistent margins, Definition 1, and so in particular satisfy $BW_{\{x,y\}}(x) = B_{\{x,y\}}(x) = W_{\{x,y\}}(y)$. This class of models has been studied extensively for best (equivalently, worst) choice probabilities, with its strengths and weaknesses quite well understood, particularly for Thurstone and Luce (MNL) models. We now present the random utility versions of these two (classes of) models, with extensions to best–worst choices.

Definition 8. A complete set of *best, worst, and best–worst choice probabilities* on a finite set T satisfies a consistent (best, worst, and best–worst) Thurstone random utility model iff it satisfies a consistent random utility model, Definition 6, such that there exist interval scale values $v_z, z \in T$, and independent[5] random variables $\boldsymbol{\varepsilon}_z, z \in T$ with

$$\mathbf{U}_z = v_z + \boldsymbol{\varepsilon}_z.$$

A consistent extreme value random utility model is a Thurstone random utility model where each $\boldsymbol{\varepsilon}_z, z \in T$, has the extreme value distribution.[6]

Rewriting the random variables in a consistent extreme value random utility model in the form of Definition 6, we have

$$\mathbf{B}_z = v_z + \boldsymbol{\varepsilon}_z,$$
$$\mathbf{W}_z = -v_z - \boldsymbol{\varepsilon}_z,$$
$$\mathbf{BW}_{r,s} = v_r - v_s + \boldsymbol{\varepsilon}_r - \boldsymbol{\varepsilon}_s \tag{14}$$

with $\boldsymbol{\varepsilon}_z$, $z \in T$, having extreme value distributions. It is important to note again that all the random variable notation above is intended to mean that the best–worst choice probabilities are derived on the basis of a *single common* set of sample values $\mathbf{U}_z, z \in T$, equivalently $\boldsymbol{\varepsilon}_z, z \in T$.

It follows from standard results, based on (11), that, given a consistent extreme value random utility model, Definition 8, the best choice probabilities satisfy the Luce (MNL) choice model, with scale values $\exp v_z$ (McFadden, 1974). However, and this is important, neither the worst or the best–worst choice probabilities given by that model will then, in general, satisfy the Luce (MNL) choice model when the scale values $v_z, z \in T$, are not all equal. Nonetheless, the relationship between the best and the worst choice probabilities in this model is well known, and has lead to much fascinating research (Yellott, 1977, 1980, 1997). Here, we add a result about the representation of best–worst choice probabilities that follows from that earlier work,

[5]We include independence as part of the definition as we only consider that case in this paper.

[6]This means that

$$\Pr(\mathbf{U}_z \leqslant t) = \exp -e^{-t} \quad (-\infty < t < \infty)$$

but has not been noted previously as others have not studied best–worst choice.[7]

It is known (e.g., Critchlow et al., 1991) that if a complete set of best, worst and best–worst choice probabilities on a finite set T satisfies a consistent extreme value random utility model, Definition 8, then for each $Y \subseteq T, |Y| = m$, and each $\rho = \rho_1\rho_2 ... \rho_{m-1}\rho_m \in R(Y)$,

$$\begin{aligned}&\Pr(\mathbf{U}_{\rho_1} > \mathbf{U}_{\rho_2} > \cdots > \mathbf{U}_m)\\&\quad = B_Y(\rho_1)B_{Y-\{\rho_2\}}(\rho_2)...B_{\{\rho_{m-1},\rho_m\}}(\rho_m)\end{aligned} \tag{15}$$

and so for each $x, y \in T$, $x \neq y$

$$BW_{\{x,y\}}(x,y) = \Pr(\mathbf{U}_x > \mathbf{U}_y) = B_{\{x,y\}}(x),$$

and, using (13) and (15), for $x, y \in X \subseteq T, |X| = n > 2$, $x \neq y$, and $\eta = \eta_2 ... \eta_{n-1} \in R(X - \{x,y\})$,

$$\begin{aligned}&BW_X(x,y)\\&\quad = \Pr(\mathbf{U}_x > \mathbf{U}_z > \mathbf{U}_y, z \in X - \{x,y\})\\&\quad = \sum_{\eta \in R(X-\{x,y\})} B_X(x)B_{X-\{x\}}(\eta_2)...B_{\{\eta_{n-1},y\}}(\eta_{n-1})\\&\quad = B_X(x) \sum_{\eta \in R(X-\{x,y\})} B_{X-\{x\}}(\eta_2)...B_{\{\eta_{n-1},y\}}(\eta_{n-1})\\&\quad = B_X(x)\Pr(\mathbf{U}_z > \mathbf{U}_y, z \in X - \{x,y\})\\&\quad = B_X(x)W_{X-\{x\}}(y),\end{aligned}$$

i.e.,

$$BW_X(x,y) = B_X(x)W_{X-\{x\}}(y), \tag{16}$$

an example of a (mixed) sequential best–worst choice model (Section 4, Definition 13).

On the other hand, suppose that we have a complete set of best, worst and best–worst choice probabilities on a finite set T, $|T| > 2$, that satisfies a Thurstone random utility model, Definition 8, and also satisfies (16) for all $X \subseteq T$. Then, in particular, for every $u, v \in T, u \neq v$, $W_{\{u,v\}}(u) = B_{\{u,v\}}(v)$, and so for any $X \subseteq T$, $|X| = 3$, and $\rho = \rho_1\rho_2\rho_3 \in R(X)$, (16) gives

$$\begin{aligned}p(\rho_1\rho_2\rho_3) = {}& \Pr(\mathbf{U}_{\rho_1} > \mathbf{U}_{\rho_2} > \mathbf{U}_{\rho_3})\\&= BW_X(\rho_1,\rho_3) = B_X(\rho_1)B_{\{\rho_2,\rho_3\}}(\rho_2).\end{aligned}$$

However, it is known that if a Thurstone random utility model[8] (for a complete set of best choice probabilities) satisfies the above "ranking postulate" (Luce, 1959) for a set X with $|X| = 3$, then the complete set of best choice probabilities on X satisfy the Luce choice model (Theorem 50, Luce and Suppes, 1965). Additionally, it is known that if a complete set of best choice probabilities on a finite set T satisfies a Thurstone random utility model, Definition 8, and the best choice probabilities satisfy Luce's choice model on all subsets $X \subseteq T$ with $|X| \leqslant 3$, then the best choice probabilities satisfy Luce's choice model for all subsets $X \subseteq T$ (Theorem 5, Yellott, 1977).

Combining all of the above results we have the following proposition.

Proposition 9. *Assume that a complete set of best, worst, and best–worst choice probabilities on a finite set T satisfies a Thurstone random utility model. Then the following conditions are equivalent*:

(i) *The best choice probabilities satisfy Luce's choice model.*
(ii) *For all $x, y \in X \subseteq T$, $x \neq y$, $BW_X(x,y) = B_X(x)W_{X-\{x\}}(y)$.*

Note that the argument that (ii) implies (i) only uses subsets $X \subseteq T$ with $|X| \leqslant 3$.

Also, given that the best choice probabilities satisfy the Luce (MNL) choice model, it follows from standard results (e.g., Yellott, 1977) that the worst choice probabilities do not satisfy a Luce (MNL) choice model, except in special cases such as when the best choice probabilities are equal for each $x \in X$. Nonetheless, for distinct $x, y \in X$ with $\eta = \eta_2 ... \eta_{n-1} \in R(X - \{x,y\})$ and so $\rho = x\eta y \in R(X)$, we have, by (15),

$$\begin{aligned}W_{X-\{x\}}(y) = {}& \Pr(\mathbf{U}_z > \mathbf{U}_y, z \in X - \{x,y\})\\&= \Pr\left(\mathbf{U}_y = \min_{z \in X - x} \mathbf{U}_z\right)\\&= \sum_{\eta \in R(X-\{x,y\})} \Pr(\mathbf{U}_{\eta_2} > \cdots > \mathbf{U}_{\eta_{n-1}} > \mathbf{U}_y)\\&= \sum_{\eta \in R(X-\{x,y\})} B_{X-\{x\}}(\eta_2) \ldots B_{\{\eta_{n-1},y\}}(\eta_{n-1}),\end{aligned}$$

and so the worst choice probabilities can be (relatively) easily computed from the best choice probabilities. Also, econometricians or psychometricians might attempt to fit the model through the integrals implicit in the formulae: for $X \subseteq T$,

$$B_X(x) = \Pr\left(\mathbf{U}_x = \max_{z \in X} \mathbf{U}_z\right),$$

$$W_X(y) = \Pr\left(\mathbf{U}_y = \min_{z \in X} \mathbf{U}_z,\right).$$

We now summarize the details of a parallel consistent Thurstone random utility model where the roles of best and worst choice probabilities are interchanged. So consider now a consistent Thurstone random utility model, Definition 8, where $\mathbf{U}_z = v_z - \boldsymbol{\varepsilon}_z$, $z \in T$, and each $\boldsymbol{\varepsilon}_z$ has the extreme value distribution (note the term $-\boldsymbol{\varepsilon}_z$, whereas in the previous example we have $+\boldsymbol{\varepsilon}_z$). Then,

[7]We thank the anonymous referee for observations that lead us to complete the following example, and Harry Joe for discussing it with us.

[8]In fact, this part of the argument goes through under the weaker assumption of a consistent random ranking model, Definition 3.

using (12), we have

$$\begin{aligned} W_X(y) &= \Pr\left(\mathbf{U}_y = \min_{z\in X} \mathbf{U}_z\right) \\ &= \Pr\left(v_x - \boldsymbol{\varepsilon}_x = \min_{z\in X}[v_z - \boldsymbol{\varepsilon}_z]\right) \\ &= \Pr\left(-v_x + \boldsymbol{\varepsilon}_x = \max_{z\in X}[-v_z + \boldsymbol{\varepsilon}_z]\right) = \frac{e^{-v_x}}{\sum e^{-v_z}}, \end{aligned}$$

where the final equality corresponds to the standard result regarding the representation of the Luce (MNL) choice model in terms of random variables $\boldsymbol{\varepsilon}_z$, $z \in T$, with each $\boldsymbol{\varepsilon}_z$ having the extreme value distribution. If we now proceed in a manner exactly paralleling the previous example, but using (11)–(13) with the worst choice probabilities satisfying the Luce (MNL) choice model and (15) stated in terms of the worst choice probabilities, we obtain the following result.

Proposition 10. *Assume that a complete set of best, worst, and best–worst choice probabilities on a finite set T satisfies a Thurstone random utility model. Then the following conditions are equivalent*:

(i) *The worst choice probabilities satisfy Luce's choice model.*
(ii) *For all* $x, y \in X \subseteq T$, $x \neq y$, $BW_X(x, y) = W_X(x)B_{X-\{x\}}(y)$.

Note that the representation in (ii) is an example of a (mixed) sequential best–worst choice model (Section 4, Definition 13) where the roles of best and worst choices are interchanged relative to the first example.

Now we present a variant on the assumptions of the above examples that leads to a random utility model where all three sets of choice probabilities—best, worst, and best–worst—satisfy the Luce (MNL) model. In fact, the resulting choice probabilities are those of the example in Section 1.

Definition 11. A complete set of best, worst, and best–worst choice probabilities satisfies an inverse (*best, worst, best–worst*) *Thurstone random utility model* iff it satisfies a random utility model, Definition 5, for which there exist interval scale values $u_z, v_z, z \in T$, and independent random variables $\boldsymbol{\varepsilon}_{r,s}, r, s, z \in T$, $r \neq s$, such that

$$\begin{aligned} &\mathbf{B}_z = u_z + \boldsymbol{\varepsilon}_z, \\ &\mathbf{W}_z = -v_z + \boldsymbol{\varepsilon}_z, \\ &B\mathbf{W}_{r,s} = u_r - v_s + \boldsymbol{\varepsilon}_{r,s}. \end{aligned} \tag{17}$$

An inverse extreme value random utility model is an inverse Thurstone random utility model where each $\boldsymbol{\varepsilon}_z$ and $\boldsymbol{\varepsilon}_{r,s}, r, s, z \in T$, $r \neq s$, has the extreme value distribution.

An inverse extreme value random utility model is *not* a consistent random utility model, Definition 6, even when $u_z = v_z$ (see below).

If we let, for $z \in T$,

$$b(z) = \exp u_z, w(z) = \exp -v_z,$$

then standard techniques show that the choice probabilities given by an inverse extreme value random utility model, Definition 11, yield the Luce (MNL) choice model for each of B_X, W_X, BW_X, $X \subseteq T$, with, for $x, y \in X \subseteq T$,

$$\begin{aligned} &B_X(x) = \frac{b(x)}{\sum_{z\in X} b(z)}, \\ &W_X(y) = \frac{w(y)}{\sum_{z\in X} w(z)}, \\ &BW_X(x, y) = \frac{b(x)w(y)}{\sum_{\substack{r,s\in X \\ r\neq s}} b(r)w(s)} \quad (x \neq y). \end{aligned} \tag{18}$$

We now make several observations about the similarities and differences between the two extreme value models of (14) and (17):

(i) Each random variable $\mathbf{BW}_{r,s}$ depends on a difference of extreme value random variables, $\boldsymbol{\varepsilon}_r - \boldsymbol{\varepsilon}_s$, in (14) and on a single extreme value random variable, $\boldsymbol{\varepsilon}_{r,s}$, in (17). Obviously, this difference leads to differences in the predictions of each model for the best–worst choice probabilities. In particular, (14) predicts that the three types of binary choice probabilities agree, i.e., that $BW_{\{x,y\}}(x, y) = B_{\{x,y\}}(x) = W_{\{x,y\}}(x)$, whereas the general case of (17) predicts that they all differ. It is an empirical question as to which pattern of results—or neither—is correct.
(ii) The extreme value random variables $\mathbf{B}_z$ and $\mathbf{W}_z$ in (14), but not in (17), satisfy $\mathbf{B}_z \overset{\mathrm{d}}{=} -\mathbf{W}_z$, where $\overset{\mathrm{d}}{=}$ means equal in distribution. Thus, a consistent extreme value model, Definition 8, is a consistent random utility model, Definition 6, but an inverse extreme value random utility model, Definition 11, is not—the latter result follows from the fact that the extreme value distribution is not symmetric. If the extreme value distributed random variables $\boldsymbol{\varepsilon}_z$ are replaced by random variables with symmetric distributions (for instance, normals), we have $\mathbf{B}_z \overset{\mathrm{d}}{=} -\mathbf{W}_z$ for both (14) and (17).

As noted above, the extreme value random variables $\boldsymbol{\varepsilon}_z$ in Definition 8 and Definition 11 are not symmetric. Rephrasing the above results, we have the following results for consistent Thurstone and inverse Thurstone random utility models, i.e., those where the extreme value random variables are replaced by general random variables $\boldsymbol{\varepsilon}_z$ with equal means and variances:

1. When $\boldsymbol{\varepsilon}_z$ is symmetric, the representations of the best and worst probabilities given by (14) are of the same

form as are the representations of the best and worst probabilities given by (17), and (thus) (14) and (17) agree with each other for the best and the worst choice probabilities.
2. When $\boldsymbol{\varepsilon}_z$ is not symmetric, the representations of the best and worst probabilities given by (14) are not of the same form, but the representations of the best and worst probabilities given by (17) are of the same form, and (thus) (14) and (17) do not agree with each other for the best and the worst choice probabilities.
3. It is an empirical question as to which pattern of results—(1) or (2) or neither—is correct.

4. Joint and sequential best–worst choice models

We now motivate models through the idea that a person, in selecting the best–worst pair of options, approaches the selection of the best and the worst option independently, and follows up on these choices, as needed, to ensure that the same option is not selected as both best and worst. We first consider models where the best and worst choices are made jointly, then models where these choices are made sequentially.

A referee asked whether such an independence assumption is reasonable and suggested that we should also consider models with some kind of correlation. The question and suggestion are eminently reasonable. We introduce the "independence" assumption mainly as a way to introduce the two models of this subsection, the first of which can be interpreted as a quasi-independent (log-linear) model[9] of a type that has been extensively studied, and successfully applied, in the analysis of other types of categorical data, including best choices (see, e.g., Agresti, 2002; Louviere et al., 2000) and best–worst choices (Cohen, 2003; Cohen and Neira, 2003, Finn and Louviere (1992)). The second model involves a slight generalization of the first model. These two models have not received detailed theoretical analysis previously and, given the length of the present paper, we leave the study of correlated versions for future work. And the models of Section 4.2 involve more subtle uses of the idea of "independence" between best and worst choices.

4.1. Simultaneous best–worst choice between options

The example of Section 1 is now generalized to a broader class of models, and then reconsidered in the light of that class.

Suppose that when a person is asked to choose the best and the worst option in a choice set X, then the person simultaneously, but independently, chooses a best and a worst option in X. If the resulting chosen options are distinct, these are reported as the best–worst pair in X. Otherwise, further decisions are required to select the best–worst pair. We now consider two possible processes for these later decisions, leaving other more complex possibilities for future study should such be required by data. The two cases we consider are:

1. The person chooses with equal probability amongst the possible distinct best–worst pairs. Thus, each pair $x, y \in X$, $x \neq y$, is chosen with probability $\frac{1}{|X|(|X|-1)}$.
2. The person re-samples for the best–worst pair.

We now study the details of models generated by each of these processes.

4.1.1. Case 1: Any choices after the first are equally probable

First we consider the case where, if no decision is reached at the first stage, then at the second stage the person chooses with equal probability amongst the possible distinct best–worst pairs. Then the best–worst choice probabilities are given by: for all $x, y \in X, x \neq y$,

$$BW_X(x,y) = B_X(x)W_X(y) + \frac{1}{|X|(|X|-1)}\sum_{z\in X} B_X(z)W_X(z). \tag{19}$$

We have the following interesting proposition:

Proposition 12. *A set of best–worst choice probabilities on a finite set X that satisfies* (19) *has consistent margins*, *Definition* 1, *iff for every* $r, s \in X$, $r \neq s$,

$$B_X(r)W_X(r) = B_X(s)W_X(s). \tag{20}$$

Proof. From (19), we obtain

$$\begin{aligned}\sum_{y\in X-\{x\}} BW_X(x,y) &= \sum_{y\in X-\{x\}} B_X(x)W_X(y) + \sum_{y\in X-\{x\}} \frac{1}{|X|(|X|-1)}\sum_{z\in X} B_X(z)W_X(z)\\ &= B_X(x)[1-W_X(x)] + \frac{1}{|X|}\sum_{z\in X} B_X(z)W_X(z)\\ &= B_X(x) + \left(\frac{1}{|X|}\sum_{z\in X} B_X(z)W_X(z) - B_X(x)W_X(x)\right)\end{aligned} \tag{21}$$

and a similar argument shows that

$$\sum_{x\in X-\{y\}} BW_X(x,y) = W_X(y) + \left(\frac{1}{|X|}\sum_{z\in X} B_X(z)W_X(z) - B_X(y)W_X(y)\right). \tag{22}$$

---

[9] We thank Tom Wickens for reminding us of this fact.

It then follows from (21) and (22) that the best–worst choice probabilities have consistent margins iff, for every $x \in X$,

$$\frac{1}{|X|}\sum_{z\in X} B_X(z)W_X(z) = B_X(x)W_X(x). \tag{23}$$

Clearly, (20) is sufficient for (23) to hold. We now show, by contradiction, that (20) is necessary for (23) to hold. So assume that (23) holds for all $z \in X$, but that (20) does not hold for all $r, s \in X$, $r \neq s$. Then, using the properties of an arithmetic average, there must be two distinct elements of X that we denote by $x_{\min}, x_{\max}$, with

$$\begin{aligned} &B_X(x_{\max})W_X(x_{\max}) \\ &\quad > \frac{1}{|X|}\sum_{z\in X} B_X(z)W_X(z) \\ &\quad > B_X(x_{\min})W_X(x_{\min}), \end{aligned}$$

which contradicts (23). Hence (20) must hold for all $x \in X$. $\square$

We now illustrate Proposition 12 with the Luce (MNL) model holding for the best and worst choices. For each $x, y \in X$, we have

$$\begin{aligned} B_X(x) &= \frac{b(x)}{\sum_{z\in X} b(z)}, \\ W_X(y) &= \frac{w(y)}{\sum_{z\in X} w(z)}. \end{aligned} \tag{24}$$

Then with the set of best–worst choice probabilities given by (19), Proposition 12 shows that this model has consistent margins iff there is a constant c such that for every $z \in X$,

$$b(z)w(z) = c,$$

i.e.,

$$w(z) = \frac{c}{b(z)}. \tag{25}$$

Note that the resulting representation for W_X in (24) will not depend on c. Then it follows from a routine calculation that (19) becomes: for $x, y \in X, x \neq y$,

$$BW_X(x,y) = \frac{\frac{b(x)}{b(y)} + \frac{1}{|X|-1}}{\sum_{\substack{r,s\in X\\ r\neq s}}\left(\frac{b(r)}{b(s)} + \frac{1}{|X|-1}\right)}. \tag{26}$$

For reassurance, one can check that the representation (26) satisfies

$$\sum_{z\in X-\{x\}} BW_X(x,z) = \frac{b(x)}{\sum_{z\in X} b(z)}$$

and

$$\sum_{z\in X-\{y\}} BW_X(z,y) = \frac{1/b(y)}{\sum_{z\in X} 1/b(z)},$$

which with (24) and (25) shows that it has consistent margins. This is quite fascinating as the best and worst choice probabilities each satisfy a Luce (MNL) choice model, and the best–worst choice probabilities do not, yet the model has consistent margins. Later, we show that the second proposed decision procedure has the 'opposite' property, namely that the best, worst and the best–worst choice probabilities each satisfy a Luce (MNL) choice model, but these sets of choice probabilities do not have consistent margins.

Since the above model has consistent margins, we have that for two element sets $X = \{x, y\}$,

$$BW_{\{x,y\}}(x,y) = B_{\{x,y\}}(x) = W_{\{x,y\}}(y).$$

We have already discussed issues related to these equalities, and mentioned that Shafir (1993) presents data, in an accept versus reject design, that can be interpreted as showing that the second and third probabilities may be unequal.

An alternative way of writing (26) sets $u(z) = \log b(z)$, $z \in X$, giving, for $x, y \in X, x \neq y$,

$$BW_X(x,y) = \frac{\exp[u(x)-u(y)] + \frac{1}{|X|-1}}{\sum_{\substack{r,s\in X\\ r\neq s}}\left(\exp[u(r)-u(s)] + \frac{1}{|X|-1}\right)}. \tag{27}$$

which is a biased form of the maximum-difference (maxdiff) best–worst choice model (see Section 4.1.2) that converges to the latter model for 'large' sets X.

Now we consider some estimation issues related to the representation (26), equivalently (27). An appropriate experimental design for testing this model, called a 2^j fractional factorial, ensures that each option and each pair of distinct options, is presented equally often across the selected subsets of size j of the master set T (Finn and Louviere, 1992). We assume such a design, and thus we lose no generality by developing the results for a fixed set $X, X \subseteq T$, with $|X| = j$.

Suppose that we have a sample of N best–worst choices from the set X. Denote these choices by $(x_i, y_i), i = 1, \ldots, N$. These may be N best–worst choices for a single individual, or one best–worst choice for each of N individuals. For the samples $i = 1, \ldots, N$ and any $x, y \in X$, $x \neq y$, let

$$\widehat{bw}_i(x,y) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \textit{if} \begin{Bmatrix} x \text{ is best, } y \text{ is worst, for sample } i \\ \text{otherwise} \end{Bmatrix}$$

and for each $z \in X$, let

$$\widehat{b}(z) = \sum_{i=1}^{N} \sum_{y \in X-\{z\}} \widehat{bw_i}(z, y),$$

$$\widehat{w}(z) = \sum_{i=1}^{N} \sum_{x \in X-\{z\}} \widehat{bw_i}(x, z).$$

Note that the likelihood of obtaining the data $(x_i, y_i), i = 1, \ldots, N$, is

$$\prod_{i=1}^{N} \prod_{\substack{x,y \in X \\ x \neq y}} [BW_X(x, y)]^{\widehat{bw_i}(x,y)}$$

and it is clear from (26), or equivalently from (27), that neither $\widehat{b}(z)$ or $\widehat{w}(z)$, $z \in X$, or both together, are sufficient statistics for this likelihood. However, we know that this model has consistent margins that satisfy the Luce (MNL) choice model—in fact, it is routine to show, with E denoting expectation, that for each $x \in X$,

$$E[\widehat{b}(x)] = N.B(X)b(x), \tag{28}$$

where

$$B(X) = \frac{1}{\sum_{z \in X} b(z)}.$$

Thus, since b is a ratio scale, the scores $\widehat{b}(x), x \in X$, or rather $\frac{1}{N}\widehat{b}(x)$, give unbiased estimates of the scale values for $b(x), x \in X$.

Similarly, with

$$W(X) = \frac{1}{\sum_{z \in X} \frac{1}{b(z)}},$$

we obtain, for each $x \in X$,

$$E[\widehat{w}(x)] = N.W(X)\frac{1}{b(x)}, \tag{29}$$

and so the scores $\widehat{w}(x), x \in X$, or rather $\frac{1}{N}\widehat{w}(x)$, give unbiased estimates of the scale values $1/b(x)$.

However, it would be preferable to estimate each scale value $b(x), x \in X$, from some combination of the scores $\widehat{b}(x)$ and $\widehat{w}(x)$. Louviere, Burgess, Street, and Marley (2004) explore the properties of related estimation procedures for the model of the following Section 4.1.2, and it will be useful in the future to study whether or not it is possible to discriminate between the model just presented and that of the next section on the basis of data.

4.1.2. Case 2: Any choices after the first involve re-sampling

Now we consider the case where, if no decision is reached at the first stage, then the person re-samples for the best–worst choice pair from the set X of available choice options. In order for such a resampling process to terminate, it is necessary that there are $r, s \in X$, $r \neq s$, with $B_X(r) \neq 0$ and $W_X(s) \neq 0$,[10] in which case $\sum_{\substack{r,s \in X \\ r \neq s}} B_X(r)W_X(s) > 0$. The formula for the best–worst choice probabilities is then, with $x \neq y$,

$$\begin{aligned}
BW_X(x, y) &= \sum_{k=o}^{\infty} \left[\sum_{r \in X} B_X(r)W_X(r)\right]^k B_X(x)W_X(y) \\
&= B_X(x)W_X(y) \sum_{k=o}^{\infty} \left[1 - \sum_{\substack{r,s \in X \\ r \neq s}} B_X(r)W_X(s)\right]^k \\
&= \frac{B_X(x)W_X(y)}{\sum_{\substack{r,s \in X \\ r \neq s}} B_X(r)W_X(s)}.
\end{aligned}$$

As discussed earlier, if the best and the worst choice probabilities each satisfy the Luce (MNL) choice model, then the above representation of the best–worst choice probabilities is a Luce (MNL) choice model representation with scale values $b(x)w(y), x, y \in X, x \neq y$. As with Case 1, when we impose the constraint that for all $z \in X, b(z)w(z) = c$, i.e., $w(z) = c/b(z)$, an alternative version of this representation is given by: for $z \in X$, let $u(z) = \log b(z)$, then for $x, y \in X, x \neq y$,

$$BW_X(x, y) = \frac{\exp[u(x) - u(y)]}{\sum_{\substack{r,s \in X \\ r \neq s}} \exp[u(r) - u(s)]}, \tag{30}$$

a representation that is usually called maximum-difference (maxdiff).

When $X = \{x, y\}$, we have

$$BW_{\{x,y\}}(x, y) = \frac{b(x)/b(y)}{b(x)/b(y) + b(y)/b(x)},$$

which does not in general equal $b(x)/[b(x) + b(y)]$, which is the value of each of $B_{\{x,y\}}(x)$ and $W_{\{x,y\}}(y)$. As we have discussed already, we can replace the prediction of the maxdiff model for the case $|X| = 2$ with the representation

$$BW_{\{x,y\}}(x, y) = \alpha B_{\{x,y\}}(x) + (1 - \alpha)W_{\{x,y\}}(y).$$

However, this is now a sequential process and there is no reason why it should not be applied for all set sizes, leading to a different model that is discussed in Section 4.2.

We have just seen that the unmodified model does not have consistent margins when $|X| = 2$, and the fact that

[10] An equivalent condition is that there is no $r \in X$ with $B_X(r) = W_X(r) = 1$.

it does not have consistent margins in general can be seen by noting that for each $x, y \in X$,

$$\sum_{z\in X-\{x\}} BW_X(x,z) = \frac{1}{\sum_{\substack{r,s\in X\\ r\neq s}} b(r)/b(s)} \left(b(x) \sum_{z\in X} \frac{1}{b(z)} - 1 \right) \neq B_X(x)$$

and

$$\sum_{z\in X-\{y\}} BW_X(z,y) = \frac{1}{\sum_{\substack{r,s\in X\\ r\neq s}} b(r)/b(s)} \left(\frac{1}{b(y)} \sum_{z\in X} b(z) - 1 \right) \neq W_X(y).$$

Turning to estimation issues, and proceeding as for Case 1, it follows from the above formulae that neither $E[\widehat{b}(x)]$ or $E[\widehat{w}(x)]$ or, in fact, $E[\widehat{b}(x) - \widehat{w}(x)]$, are unbiased estimates of $b(x)$. The latter fact is of interest since, as we show now, $\widehat{b}(x) - \widehat{w}(x), x \in X$, is a set of sufficient statistics for the scale values $b(x), x \in X$.

Turning to the likelihood for the maxdiff model, and using the notation as before, we have that

$$\begin{aligned}
\prod_{i=1}^{N} \prod_{\substack{x,y\in X\\ x\neq y}} [BW_X(x,y)]^{\widehat{bw}_i(x,y)}
&= \prod_{\substack{x,y\in X\\ x\neq y}} \prod_{i=1}^{N} \frac{[b(x)/b(y)]^{\widehat{bw}_i(x,y)}}{\sum_{\substack{r,s\in X\\ r\neq s}} b(r)/b(s)} \\
&= \frac{1}{\left[\sum_{\substack{r,s\in X\\ r\neq s}} b(r)/b(s)\right]^N} \cdot \prod_{\substack{x,y\in X\\ x\neq y}} \left(\left[\frac{b(x)}{b(y)}\right]^{\sum_{i=1}^{n} \widehat{bw}_i(x,y)} \right) \\
&= \frac{1}{\left[\sum_{\substack{r,s\in X\\ r\neq s}} b(r)/b(s)\right]^N} \cdot \prod_{x\in X} b(x)^{[\widehat{b}(x) - \widehat{w}(x)]} \qquad (31)
\end{aligned}$$

and so the difference scores $\widehat{b}(x) - \widehat{w}(x), x \in X$, are sufficient statistics for the scale values $b(x), x \in X$. However, as pointed out above, they are biased estimates of the scale values. Again, we are exploring the properties of these sufficient statistics for data, and the extent of their bias (Louviere et al., 2004). Thus far, it appears that the biases are hard to detect even for choice sets with relatively few items, say $N > 3$. Also, we conjecture that the likelihood, (31), depends on the data through the difference scores $\widehat{b}(x) - \widehat{w}(x), x \in X$, iff the maxdiff model holds—we have shown above that the if part of this conjecture holds.

4.2. Sequential best–worst choice processes

The following definition is based on the idea that a person, in choosing the best–worst pair from a set X, first decides in which order to choose the best and the worst option (best–worst order with probability α, worst–best order with probability $1 - \alpha$), then proceeds to the actual choices. Note that the question format in the experiment will likely affect (though not necessarily completely determine) the value of α.

Definition 13. A set of best, worst, and best–worst choice probabilities on a set X satisfies a *mixed sequential* (*best, worst, and*) *best–worst choice model* iff there is a constant $\alpha, 0 \leqslant \alpha \leqslant 1$, such that for all $x, y \in X, x \neq y$,

$$BW_X(x,y) = \alpha B_X(x) W_{X-\{x\}}(y) + (1-\alpha) W_X(y) B_{X-\{y\}}(x).$$

Marley (1968) introduced the following special case of the above class of models, though he did not interpret his model as being for best–worst choice. This model assumes that the elements of the best–worst pair are chosen sequentially, and that the probability that a particular best–worst pair is selected is unaffected by the order in which the best and worst criteria are applied.

Definition 14. A set of best, worst, and best–worst choice probabilities on a set X satisfies a *concordant* (*best, worst, and*) *best–worst choice model* iff for all $x, y \in X, x \neq y$,

$$BW_X(x,y) = B_X(x) W_{X-\{x\}}(y) = W_X(y) B_{X-\{y\}}(x). \qquad (32)$$

Note that a concordant best–worst choice model has consistent margins, Definition 1, and thus so does any mixture of a concordant best–worst choice model. Such a mixture is also a mixed sequential best–worst choice model, Definition 13. It is an open question as to whether it is possible to characterize the class of mixed sequential best–worst choice models, Definition 13, that have consistent margins.

We need a technical condition, and various additional notation, to state a result from Marley (1968) that leads to a concordant best–worst choice model. First, for notational simplicity, given an arbitrary set Y, for distinct $x, y \in Y$, let $b(x,y) = B_{\{x,y\}}(x)$, $w(x,y) = W_{\{x,y\}}(y)$. A set of binary best and worst choice probabilities on a set X is *transitive* if for distinct $x, y, z \in Y$, $b(x,y) = b(y,z) = 1$ implies that $b(x,z) = 1$, and $w(x,y) = w(y,z) = 1$ implies that $w(x,z) = 1$. Now, as previously, let $R(Y)$ denote the set of rank orders of the set Y. When $|Y| = n, n \geqslant 2$, we can denote

$\rho = \rho_1\rho_2...\rho_{n-1}\rho_n \in R(Y)$ by $\rho = \rho_1\sigma$, where $\sigma = \rho_2\rho_3...\rho_{n-1}\rho_n \in R(Y - \{\rho_1\})$. Also, for any $\rho \in R(Y)$, let

$$b_Y(\rho) = \prod_{1 \leqslant i < j \leqslant n} b(\rho_i, \rho_j),$$
$$w_Y(\rho) = \prod_{1 \leqslant i < j \leqslant n} w(\rho_i, \rho_j). \tag{33}$$

Using this notation, we state a variant of Marley's (1968) result, and later in this section we give a process interpretation of his result.

Proposition 15. *Let a transitive set of binary choice probabilities on a finite set X be such that for all distinct $r, s \in X$, $b(r,s) = w(s,r)$. For $x, y \in Y$, $Y \subseteq X$, let*

$$B_Y(x) = \frac{\sum_{\sigma \in R(y-\{x\})} b_Y(x\sigma)}{\sum_{\eta \in R(Y)} b_Y(\eta)},$$
$$W_Y(y) = \frac{\sum_{\sigma \in R(Y-\{y\})} w_X(y\sigma)}{\sum_{\eta \in R(Y)} w_X(\eta)}. \tag{34}$$

Then for distinct $x, y \in X$,

$$B_X(x)W_{X-\{x\}}(y) = W_X(y)B_{X-\{y\}}(x).$$

Now define a set of best–worst choice probabilities by, for distinct $x, y \in X$,

$$BW_X(x,y) = B_X(x)W_{X-\{x\}}(y) = W_X(y)B_{X-\{y\}}(x), \tag{35}$$

where the best and worst choice operabilities are as given in Proposition 15. Then by that theorem we have an example of a concordant best–worst choice model, Definition 14. In fact, Marley (1968) shows that, under the conditions of Proposition 15, this is the unique representation of a concordant best–worst-choice model. However, note that this is really a "class" of models as it depends on the assumed representations of the binary choice probabilities.

Marley (1968) shows also that, when the binary choice probabilities are transitive and for all distinct $r, s \in X$, $b(r,s) = w(s,r)$, the above concordant best–worst choice model is the only one that is compatible with a structure of "best to worst" and "worst to best" ranking probabilities that he called a *reversible ranking model* (Marley, 1968 Theorem 8). However, we see next that a concordant best–worst choice model, Definition 14, is not in general compatible with a consistent random ranking model, Definition 3.

Proposition 16. *Assume that a complete set of best, worst and best–worst choice probabilities on a three-element set T satisfies a concordant best–worst choice model, Definition 14. Then the following are equivalent.*

(i) *The complete set of best, worst and best–worst choice probabilities satisfies a consistent random ranking model, Definition 3.*

(ii) *For each $x, y \in X \subseteq T, |X| \geqslant 2, x \neq y$, we have $B_X(x) = W_X(x) = \frac{1}{|X|}$ and $BW_X(x,y) = \frac{1}{|X|} \cdot \frac{1}{|X-1|}$.*

Proof. Clearly, when (ii) holds, the following consistent random ranking model holds: for each $\rho = \rho_1\rho_2\rho_3 \in R(T)$, set $p(\rho) = \frac{1}{|T|} \cdot \frac{1}{|T-1|}$. So it remains to show that (i) implies (ii)

From (i), we have that for each $x, y \in X \subseteq T$, $|X| \geqslant 2, x \neq y$,

$$BW_{\{x,y\}}(x,y) = B_{\{x,y\}}(x) = W_{\{x,y\}}(y),$$

which with (i) and the assumption that the complete set of best, worst and best–worst choice probabilities on the three-element set T satisfies a concordant best–worst choice model, Definition 14, gives that for $\rho = \rho_1\rho_2\rho_3 \in R(T)$,

$$p(\rho_1\rho_2\rho_3) = BW_T(\rho_1, \rho_3)$$
$$= B_T(\rho_1)W_{\{\rho_2,\rho_3\}}(\rho_3) = B_T(\rho_1)B_{\{\rho_2,\rho_3\}}(\rho_2), \tag{36}$$

and

$$p(\rho_1\rho_2\rho_3) = BW_T(\rho_1, \rho_3)$$
$$= W_T(\rho_3)B_{\{\rho_1,\rho_2\}}(\rho_1) = W_T(\rho_3)W_{\{\rho_1,\rho_2\}}(\rho_2), \tag{37}$$

i.e.,

$$p(\rho_1\rho_2\rho_3) = B_X(\rho_1)B_{\{\rho_2,\rho_3\}}(\rho_2)$$
$$= W_T(\rho_3)W_{\{\rho_1,\rho_2\}}(\rho_2). \tag{38}$$

However, (i) with (38) implies that both the best and the worst choice probabilities satisfy Luce's choice model (Theorem 50, Luce and Suppes, 1965) and then, using (38) again, we obtain that, for each $x, y \in X \subseteq T$, $|X| \geqslant 2, x \neq y$, we have $B_X(x) = W_X(x) = \frac{1}{|X|}$ (Yellott, 1980). It then follows from the fact that the best, worst and best–worst choice probabilities on T satisfy a concordant best–worst choice model, Definition 14, that $BW_X(x,y) = \frac{1}{|X|} \cdot \frac{1}{|X-1|}$. Combining these results, we have that ii. holds. □

Returning to Proposition 15, note that, using (33), the best and worst probabilities in (34) can be rewritten in the form

$$B_X(x) = \frac{\prod_{z \in X-\{x\}} b(x,z) \sum_{\sigma \in R(X-\{x\})} b_{X-\{x\}}(\sigma)}{\sum_{r \in X} \prod_{z \in X-\{r\}} b(r,z) \sum_{\sigma \in R(X-\{r\})} b_{X-\{r\}}(\sigma)},$$
$$W_X(y) = \frac{\prod_{z \in X-\{y\}} w(y,z) \sum_{\sigma \in R(X-\{y\})} w_{X-\{y\}}(\sigma)}{\sum_{r \in X} \prod_{z \in X-\{r\}} w(r,z) \sum_{\sigma \in R(X-\{r\})} w_{X-\{r\}}(\sigma)}, \tag{39}$$

which can be given the following process interpretation: the person makes all possible paired comparisons according to the best paired comparison probabilities b. If the options end up as rank ordered by this process, then the person selects as best the (necessarily unique) option that beats every other option in these comparisons—that is, the option that is first (best) in the resulting rank order; otherwise the process starts over.

The process for worst choices is parallel, with the best binary choice probabilities b replaced by the worst binary choice probabilities w.

It is interesting to compare the above process with a related process model for best choices (with a parallel interpretation for worst choices): the person makes all possible paired comparisons according to the best-paired comparison probabilities b; if some (necessarily unique) option beats every other option in these comparisons, then that option is selected as best; otherwise the process starts over.[11] Note that, in this case, in contrast to the previous process, we do not require the results of the binary choices to be consistent with a rank order, only that they are consistent with the existence of a best option. This process, with a parallel one for worst choices, gives the following representation for the best and the worst choice probabilities:

$$B_X(x) = \frac{\prod_{z \in X-\{x\}} b(x,z)}{\sum_{r \in X} \prod_{s \in X-\{r\}} b(r,s)},$$
$$W_X(y) = \frac{\prod_{z \in X-\{y\}} w(y,z)}{\sum_{r \in X} \prod_{s \in X-\{r\}} w(r,s)}.$$

By comparing these equations with those of (39), it is clear that the two processes give different representations when $|X|>3$. It is routine, though tedious, to show also that the above representation does not satisfy a concordant best–worst choice model, Definition 14, when $|X|>3$—one simple counter-example takes the set $X = \{x,y,z,w\}$ with the binary choice probabilities satisfying Luce's choice model with x,y,z,w having scale values $1,2,3,4$, respectively.

These are only two of a vast array of possible best–worst choice models based on binary comparisons. There are many alternate ways to combine a sequence of (probabilistic) best and worst choices that will lead to a final best–worst pair. Also, the best (respectively, worst) choices in a mixed sequential choice process can be assumed to satisfy any "standard" discrete choice model, not necessarily one motivated by a sequence of binary comparisons. Of course, for parsimony, we would expect some links between the best and worst choice processes, such as those that arise as a result of assuming a consistent extreme value model, Definition 8. Given the vast diversity of possible models, we leave their further systematic study for the future, at which time we expect to have data with which to challenge them.

[11] The following is an alternate description of the process. Some element $x \in X$ is chosen at random and compared pairwise with every other element of X, with the best option in each comparison being determined by the paired comparison probabilities b. If x wins every comparison, it is selected as best, otherwise the process starts over.

5. Ratio scale models

We now return to the maxdiff model, i.e., the example in Section 1, and use it to motivate a theoretical question concerning when a set of best, worst and best–worst choice probabilities are "of the same form" with the best–worst choice probabilities determined by some "functions" of the best and worst choice probabilities. Unfortunately, the notation required for the general formulation is complex, so we use the earlier example to illustrate the ideas, and include details in the Appendix. We concentrate on a fixed choice set X, though the ideas can be extended to cover all subsets $X, X \subseteq T$, of a master set T.

We begin with the restatement of the formulae of the example: for $x, y \in X$,

$$B_X(x) = \frac{b(x)}{\sum_{r \in X} b(r)},$$
$$W_X(y) = \frac{w(y)}{\sum_{s \in X} w(s)} \tag{40}$$

and

$$BW_X(x,y) = \frac{B_X(x)W_X(y)}{\sum_{\substack{r,s \in X \\ r \neq r}} B_X(r)W_X(s)} \quad (x \neq y). \tag{41}$$

Then direct substitution of (40) in (41) yields: for $x, y \in X, x \neq y$,

$$BW_X(x,y) = \frac{b(x)w(y)}{\sum_{\substack{r,s \in X \\ r \neq s}} b(r)w(s)}. \tag{42}$$

Thus, combining (41) with (42), we have

$$BW_X(x,y) = \frac{B_X(x)W_X(y)}{\sum_{\substack{x,y \in X \\ x \neq y}} B_X(x)W_X(y)} = \frac{b(x)w(y)}{\sum_{\substack{r,s \in X \\ r \neq s}} b(r)w(s)}, \tag{43}$$

that is, the best–worst choice probabilities are functions of the best and worst choice probabilities, as well as functions of the ratio scale values that determine those choice probabilities, and, in fact, the two functional forms have identical structure. We wish to know what other functional forms, if any, have parallel properties. The full notation required to formulate, and the techniques to solve, this question are complex (see the appendix). Nonetheless, based on previous work on closely related aggregation problems (Aczel et al., 1997; Aczel et al., 2000), we conjecture that the representations given by (43) are the only ones that satisfy the full set of constraints, which include the important requirements that the two functional forms have identical structure and that b and w are independent ratio scales. If we generalize the formulation by allowing the two functional forms to have somewhat different structures, then, again based on the previous work, we conjecture that the class of solutions is larger, but still involves relatively simple functions of the best and worst choice probabilities (vis., the best and worst ratio scales). Finally, as in the maxdiff model, we need to consider the cases, where for each $z \in X$,

$$b(z) = \frac{1}{w(z)}.$$

Nothing similar to this case has been studied in the previous work, and it raises several complexities in generalizing the earlier results to the present situation.

6. Summary and conclusions

We derived and discussed theoretical results for a number of best, worst and best–worst choice models, with the focus on the latter. Our results include a number of interesting theoretical relationships between these types of models, which in turn suggest a variety of tests to determine which model is most consistent with choice data. For example, the maximum-difference model is of the well-known Luce (1959), equivalently Multinomial Logit (McFadden, 1974), form with ratios of scale values, and has difference scores for best versus worst choices that are sufficient statistics for the parameters of the model, with the separate best and worst scores having biases that decrease with increasing choice set size. It will be interesting and important to undertake empirical studies for different set sizes, which can be achieved by constructing the sets using balanced incomplete block designs (see, e.g., Street and Street, 1987). One can then compare theoretically correct estimates with the obtained best, worst, and best minus worst scores to learn how much the biases actually matter in real applications.

General classes of best–worst choice models have not been studied previously in a systematic way, and, as far as we know, our results constitute the first formal presentation of the properties of some of these models. The results suggest that best–worst choice tasks and the associated models are both theoretically and empirically interesting, with the potential to provide important insights into preference and choice processes. We noted a number of open problems, the most pressing of which demand the axiomatization of important best–worst choice models, such as the best–worst ranking model, the maximum-difference model, and the concordant best–worst choice model, each without reference to representations of best or worst choice probabilities. It is also important to characterize the class of ratio scale models of best, worst, and best–worst choice probabilities where all three sets of choice probabilities are "of the same form." Finally, as already indicated, there are interesting practical questions regarding the usefulness of "simple" (sufficient) statistics in analyzing and summarizing the data obtained using best–worst choice in discrete choice designs.

Acknowledgments

This research has been supported by Australian Research Council Discovery Grant DP034343632 to the University of Technology, Sydney, for Louviere, Street and Marley, and Natural Science and Engineering Research Council Discovery Grant 8124-98 to the University of Victoria for Marley. It was begun while Marley was a Visiting Researcher at Systems, Organizations and Management of the University of Groningen and supported by the Netherlands' Organization for Scientific Research for the period July 1, 2003, to June 30, 2004. We appreciate useful feedback from Tom Wickens and J. C. Falmagne on earlier versions of this work, and the detailed evaluation by two referees (one anonymous, the other Reinhard Suck) on the original version of the paper.

Appendix A. The functional equations associated with ratio scale models of best, worst, and best–worst choice probabilities

We use a slightly different notation than previously for the options and the choice probabilities, namely we let $X = \{x_1, \ldots, x_n\}$, and let $B_X(x_i)$, $W_X(x_j)$, $BW_X(x_i, x_j)$, respectively, $i \neq j$, $i,j \in \{1, \ldots, n\}$, denote the best, worst, best–worst, respectively, choice probabilities. Thus, as before, we have the constraints

$$0 \leqslant B_X(x_i), W_X(x_j), BW_X(x_i, x_j) \leqslant 1 \quad (i,j = 1, \ldots, n)(i \neq j)$$

and

$$\sum_{i=1}^{n} B_X(x_i) = \sum_{j=1}^{n} W_X(x_j) = \sum_{\substack{x_i, x_j \in X \\ i \neq j}} BW_X(x_i, x_j) = 1.$$

Also, b and w will denote ratio scales defined over possible options, and so their values are nonnegative real numbers. For mathematical simplicity we eliminate options that have scale values of zero.

We now assume that there are functions B_i and W_j $(i,j = 1, \ldots, n)$, BW_{ij}, $H_{i,j}$ and $K_{i,j}$ $(i,j = 1, \ldots, n)$ $(i \neq j)$, such that

$$B_X(x_i) = B_i(b(x_1), \ldots, b(x_n)) \quad (i = 1, \ldots, n), \tag{44}$$

$$W_X(x_j) = W_j(w(x_1), \ldots, w(x_n)) \quad (j = 1, \ldots, n). \tag{45}$$

We need a matrix notation in order to formulate the form of the functional equations that we need to solve. Let $\mathscr{R}_{++} =]0, \infty[$. For any pair of n-component vectors $\mathbf{r} = (r_1, \ldots, r_n), \mathbf{s} = (s_1, \ldots, s_n)$ with $r_i, s_i \in \mathscr{R}_{++}$, and a function $\Omega : \mathscr{R}_{++} \times \mathscr{R}_{++} \longrightarrow \mathscr{R}_{++}$, define the matrix $\|\Omega(r_i, s_j)\|$ by

$$\|\Omega(r_k, s_l)\| = \begin{bmatrix} 0 & \Omega(r_1, s_2) & . & . & . & \Omega(r_1, s_{n-1}) & \Omega(r_1, s_n) \\ \Omega(r_2, s_1) & 0 & . & . & . & \Omega(r_2, s_{n-1}) & \Omega(r_2, s_n) \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ \Omega(r_{n-1}, s_1) & \Omega(r_{n-1}, s_2) & . & . & . & 0 & \Omega(r_{n-1}, s_n) \\ \Omega(r_n, s_1) & \Omega(r_n, s_2) & . & . & . & \Omega(r_n, s_{n-1}) & 0 \end{bmatrix}.$$

Thus, $\|\Omega(r_k, s_l)\|$ is simply the matrix of pairs $\Omega(r_k, s_l), k, l \in \{1, \ldots, n\}, k \neq l$, with the diagonal elements being 0.

With this notation, generalizing the example in the text, we assume that the functional relation for $BW_X(x_i, x_j)$ is: there are functions Φ and G, such that for $i,j = 1, \ldots, n$, $i \neq j$,

$$\begin{aligned} BW_X(x_i, x_j) &= H_{i,j}[\|\Phi(B_X(x_k), W_X(x_l))\|] \\ &= K_{i,j}[\|G(b(x_i), w(x_j))\|]. \end{aligned} \quad (46)$$

Notice that if we substitute (44) and (45) into (46), we obtain a set of functional equations involving the scale values b and w, and the general functions $B_j, W_j, H_{i,j}, K_{i,j}, \Phi$, and G. The additional constraints that follow from the fact that b and w are distinct ratio scales are that we have for μ, v in $\mathscr{R}_{++}, j = 1, \ldots, n$

$$B_j[\mu b(x_1), \ldots, \mu b(x_n)] = B_j[b(x_1), b(x_2), \ldots, b(x_n)], \quad (47)$$

$$W_j[vw(x_1), \ldots, vw(x_n)] = W_j[w(x_1), \ldots, w(x_n)], \quad (48)$$

and

$$G([\|\mu b(x_k), vw(x_l)\|) = M(\mu, v).G([\|b(x_k), w(x_l)\|). \quad (49)$$

Clearly, the model that we introduced in the text, namely (40)–(43), is the special case of (46) where, for $j = 1, \ldots, n$,

$$B_X(x_j) = B_j[b(x_1), b(x_2), \ldots, b(x_n)] = \frac{b(x_j)}{\sum_{i=1}^{n} b(x_i)},$$

$$W_X(x_j) = W_j[w(x_1), w(x_2), \ldots, w(x_n)] = \frac{w(x_j)}{\sum_{i=1}^{n} w(x_i)}$$

and where $\Phi(u, v) = uv$, $G(r, s) = rs$, and for $i,j = 1, \ldots, n$, $i \neq j$,

$$\begin{aligned} &H_{i,j}[\|\Phi(B_X(x_k), W_X(x_l))\|] \\ &= \frac{\Phi(B_X(x_i), W_X(x_j))}{\sum_{\substack{k=1 \\ k \neq l}}^{n} \sum_{l=1}^{n} \Phi(B_X(x_k), W_X(x_l))} \\ &= \frac{B_X(x_i) W_X(x_j)}{\sum_{\substack{k=1 \\ k \neq l}}^{n} \sum_{l=1}^{n} B_X(x_k) W_X(x_l)} \end{aligned}$$

and

$$\begin{aligned} &K_{i,j}[\|G(b(x_k), w(x_l))\|] \\ &= \frac{G(b(x_i), w(x_j))}{\sum_{\substack{k=1 \\ k \neq l}}^{n} \sum_{l=1}^{n} G(b(x_k), w(x_l))} \\ &= \frac{b(x_i) w(x_j)}{\sum_{\substack{k=1 \\ k \neq l}}^{n} \sum_{l=1}^{n} b(x_k) w(x_l)} \quad (i,j = 1, \ldots, n)\ (i \neq j). \end{aligned}$$

Thus, the problem before us is, given reasonable technical assumptions, find all solutions of the set of functional equations (44)–(49).

References

Aczel, J., Maksa, G., Marley, A. A. J., & Moszner, Z. (1997). Consistent aggregation of scale families of selection probabilities. *Mathematical Social Sciences*, *33*, 227–250.

Aczel, J., Gilanyi, G., Maksa, G., & Marley, A. A. J. (2000). Consistent aggregation of simply scalable families of choice probabilities. *Mathematical Social Sciences*, *39*, 241–262.

Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley.

Barbera, S., & Pattanaik, P. K. (1986). Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, *54*, 707–715.

Cohen, S. (2003). Maximum difference scaling: Improved measures of importance and preference for segmentation. Sawtooth software conference proceedings, Sawtooth Software, Inc., 530 W. Fir St., Sequim, WA (www.sawtoothsoftware.com), (pp. 61–74).

Cohen, S., & Neira, L. (2003). Measuring preference for product benefits across countries: Overcoming scale usage bias with maximum difference scaling. Paper presented at the Latin American conference of the European society for opinion and marketing research, Punta del Este, Uruguay (pp. 1–22). Reprinted in Excellence in International Research: 2004. ESOMAR, Amsterdam, Netherlands, 2004.

Critchlow, D. E., Fligner, M. A., & Verducci, J. S. (1991). Probability models on rankings. *Journal of Mathematical Psychology*, *35*, 294–318.

Falmagne, J. C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, *18*, 52–72.

Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: The case of food safety. *Journal of Public Policy and Marketing*, *11*(1), 12–25.

Fiorini, S. (2004). A short proof of a theorem of Falmagne. *Journal of Mathematical Psychology*, *48*, 80–82.

Fishburn, P. (1994). On "choice" probabilities derived from ranking distributions. *Journal of Mathematical Psychology*, *38*, 274–285.

Fishburn, P. C. (2002). Stochastic utility. In S. Barbera, P. J. Hammond, & C. Seidl (Eds.). *Handbook of utility theory*, Vol. I, pp. 275–319. Dordrecht, Boston, London: Kluwer Academic Publishers.

Louviere, J. J., Burgess, L., Street, D., & Marley, A. A. J. (2004). Modelling the choices of single individuals by combining efficient choice experiment designs with extra preference information. Working paper no. 04-003, Centre for the Study of Choice, University of Technology, Sydney.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice models*. Cambridge: Cambridge University Press.

Luce, R. D. (1959). *Individual choice behavior*. New York, NY: Wiley.

Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.). *Handbook of mathematical psychology* (Vol III, pp. 235–406). New York, NY: Wiley.

Marley, A. A. J. (1968). Some probabilistic models of simple choice and ranking. *Journal of Mathematical Psychology*, *5*, 311–332.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Econometrics* (pp. 105–142). New York, NY: Academic Press.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse. *Memory and Cognition*, *21*, 546–556.

Street, A. P., & Street, D. J. (1987). *Combinatorics of experimental design*. New York: Oxford University Press.

Yellott, J. I., Jr. (1977). The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, *15*, 109–144.

Yellott, J. I., Jr. (1980). Generalized Thurstone models for ranking: Equivalence and reversibility. *Journal of Mathematical Psychology*, *22*, 48–69.

Yellott, J. I., Jr. (1997). Preference models and reversibility. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 131–151). Mahwah, NJ: Lawrence Erlbaum Associates.