Available online at www.sciencedirect.com
ScienceDirect
Journal of Mathematical Psychology 52 (2008) 281–296
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

Probabilistic models of set-dependent and attribute-level best–worst choice

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

[a] *Department of Psychology, University of Victoria, PO Box 3050 STN CSC, Victoria BC V8W 3P5, Canada*
[b] *Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, England, United Kingdom*
[c] *School of Marketing, University of Technology Sydney, PO Box 123, Broadway, Sydney, NSW, Australia*

Received 23 April 2007; received in revised form 16 January 2008
Available online 1 April 2008

Abstract

We characterize a class of probabilistic choice models where the choice probabilities depend on two scales, one with a value for each available option and the other with a value for the set of available options. Then, we develop similar results for a task in which a person is presented with a profile of attributes, each at a pre-specified level, and chooses the best or the best and the worst of those attribute-levels. The latter design is an important variant on previous designs using best–worst choice to elicit preference information, and there is various evidence that it yields reliable interpretable data. Nonetheless, the data from a single such task cannot yield separate measures of the "importance" of an attribute and the "utility" of an attribute-level. We discuss various empirical designs, involving more than one task of the above general type, that may allow such separation of importance and utility.

Keywords: Attribute choice; Best–worst choice; Importance; Probabilistic choice; Profile

1. Introduction

Over the past decade or so, a choice task in which a person is asked to select both the best option and the worst option in an available set of options has seen increased applications over more traditional choice tasks, such as asking a person to choose (a) the best option; (b) choose the worst option; (c) rank the options; or (d) rate the options. Marley and Louviere (2005) summarized that work and developed an integrative theoretical approach to 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.

In this paper, we discuss a different task, and associated experimental design(s), that involves several profiles of attributes, with the levels of the attributes varying between profiles. This type of task is commonly referred to as a "conjoint task" (e.g. Louviere, 1988). When the profiles are generated by a suitable experimental design, such as a factorial design or an orthogonal fractional factorial design, one can view each profile generated in this way as a choice set of attribute-levels.[1] When each profile is viewed as a choice set, and if a person is asked to simultaneously choose the best and the worst attribute-level (most and least attractive level or the attribute-levels that, respectively, matter most and least) in each profile, the task parallels that considered in Marley and Louviere (2005). However, it differs empirically and theoretically in that in the profile case each choice set is constrained to contain an attribute-level for each attribute, whereas in the previously studied best–worst task, any combination of choice options can occur as a choice set.

Although this type of choice task may sound unusual, evidence exists that it yields reliable, interpretable data if one constructs the profiles using a suitable experimental design (see Coast et al., 2006; Louviere, 1994; McIntosh & Louviere, 2002). We demonstrate that choice data from a single such task cannot meaningfully separate the "importance" of each attribute and the "utility" associated with each level of an attribute.

* Corresponding author.
E-mail addresses: ajmarley@uvic.ca (A.A.J. Marley), terry.flynn@bristol.ac.uk (T.N. Flynn), jordan.louviere@uts.edu.au (J.J. Louviere).

[1] In the remainder of the paper, we use "attribute-level" for a specific level of an attribute.

0022-2496/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2008.02.002

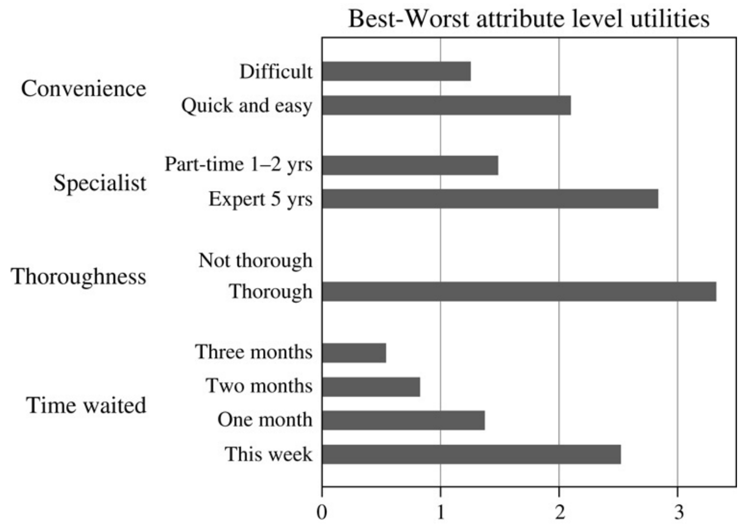

Fig. 1. Estimated impact of each attribute-level, rescaled with the lowest estimated value (for 'not thorough') set to zero. Figure 2 of Coast et al. (2006). Reproduced with permission of Blackwell Publishing.

Towards the end of the paper, we summarize prior research on separating "weights" and "values" and then discuss various discrete choice tasks that may allow one to separate importance and utility if repeated for different substantive choice contexts. The general issues that arise are similar, but not identical, to those that arise when one studies the identifiability of scale factors (viz. variance parameters) in underlying random utility representations (see Louviere, Hensher, and Swait (2003), Chapters 8 and 13; Swait and Louviere (1993)).

Section 2 provides an illustration of the type of design that interests us in this paper. Section 3 characterizes a class of probabilistic choice models where the choice probabilities depend on two scales, one with a value for each available option and the other with a value for the set of available options. These characterizations guide us in the development in Section 4 of parallel models where each option is specified in terms of a vector of attribute-levels. We study the special case where a person has to select the best (and, possibly, the worst) attribute-level of the presented option. Section 5 presents special cases of these general models where the impact of each attribute-level is determined by a product of an importance weight for the attribute and a utility value for the specific level on that attribute. Section 6 summarizes prior research on separating "weights" and "values", followed by a discussion of how one might achieve such a separation using discrete choice tasks.

Finally, we note that the notation and derivations are quite complex, but we believe that they are necessary to reach the suggestions in Section 6.2 and the conclusions in Section 7 regarding issues surrounding the measurement of "importance."

2. An illustration of best-worst attribute-level choice

Coast et al. (2006) present a study of patients' preferences for attributes of care using best–worst attribute-level choice (the latter is discussed below and defined formally in Section 4.2). The context is the fact that, in the UK, general practitioners with additional training in a specialist service are being used increasingly to provide service in a primary, rather than a hospital, setting. Coast et al.'s discrete choice experiment (DCE) involved varying levels of attributes of dermatology care that were found to be important to patients (Table 1). The attributes and levels were combined using a fractional factorial design to construct 16 scenarios; the scenarios were presented to each of 60 participants who provided complete data. The design used in the study enabled all main effects to be estimated and would have allowed all two-way interactions to be estimated had the utility of waiting time been linear in real time, which was not. Table 2 shows a typical scenario, with a participant instructed to mark the "best thing" and the "worst thing" in that scenario, plus say whether or not they would attend an appointment that exhibited this combination of attribute-levels. The choice data were analyzed using the attribute-level maxdiff model presented in Section 4.2.

Using the term impact in the sense of "influence on the choices made" (discussed in more detail in Sections 4 and 5), and assuming that the attribute-level maxdiff model (Section 4.2) holds, a researcher can measure[2] the impact of each attribute on the choices (Table 3) as well as the impact of each attribute-level on the choices (Fig. 1). In particular, the estimated impact of an attribute in Table 3 is the average of the impacts for the attribute-levels of that attribute in Fig. 1. Note that in this example the attribute "expertise" of the doctor has the most impact and the attribute "time waited" the least impact (Table 3). Also note that attribute-levels "thorough" and "expert 5 yrs" each have a large impact, while "not thorough" has little (least) impact (Fig. 1).[3] Please note that we have been careful to use the term "impact"

[2] These measurements are on a common ratio scale (see Section 4.2), with the results in Fig. 1 and Table 3 based on the log of that scale. This scale restriction is an important property of the attribute-level maxdiff model.

[3] Coast et al. (2006) restrict the term "impact" to the average of the scale values on an attribute and use the term "utility" for the scale value of an attribute at a specific level. We believe that "impact" is the appropriate term for both contexts.