OPERATIONS RESEARCH
Vol. 63, No. 3, May–June 2015, pp. 512–526
ISSN 0030-364X (print) | ISSN 1526-5463 (online)
http://dx.doi.org/10.1287/opre.2015.1373

Error Theory for Elimination by Aspects

Rajeev Kohli, Kamel Jedidi
Graduate School of Business, Columbia University, New York, New York 10027
{rk35@columbia.edu, kj7@columbia.edu}

Elimination by aspects (EBA) is a random utility model that is considered to represent the choice process used by consumers more faithfully than logit and probit models. One limitation of the model is that it does not have a known error theory. We show that EBA can be derived by assuming that aspects have random utilities with independent, extreme value distributions. Multinomial logit and rank-ordered logit models are special cases of EBA.

Subject classifications: elimination by aspects; Luce axiom; multinomial logit model; rank-ordered logit model; independence from irrelevant alternatives; lexicographic choice; choice models.
Area of review: Marketing Science.
History: Received June 2014; revisions received September 2014, December 2014, January 2015; accepted February 2015. Published online in *Articles in Advance* May 13, 2015.

1. Introduction

Elimination by aspects (EBA) is a theory of choice that was proposed by Tversky (1972a, b) more than 40 years ago. It is considered to reflect the choice process used by consumers more faithfully than traditional choice models. EBA does not assume that a consumer chooses an alternative to maximize utility. Instead, it views choice to be the result of a probabilistic process that eliminates alternatives in stages. Tversky (1972a, p. 281) described EBA as follows:

> In this theory, each alternative is viewed as a set of aspects. At each stage in the process, an aspect is selected (with probability proportional to its weight), and all the alternatives that do not include the selected aspect are eliminated. The process continues until all alternatives but one are eliminated.

A key feature of EBA is that it does not assume order independence, a condition that is probably the weakest form of independence from irrelevant alternatives (IIA). Unlike (strong) IIA, order independence only requires that the ordering—not necessarily the ratio—of the choice probabilities for two alternatives should be independent of any other alternatives in a choice set.[1] But as the following classic example by Debreu (1960) illustrates, even this seemingly mild condition can be violated. Consider a choice set with two alternatives, A and B, and a person who chooses A with a higher probability than B. Now suppose one or more alternatives that are essentially identical to A are added to the choice set. Then order independence is violated if their addition has no effect on the choice probability for B but lowers the choice probability of A to less than $1/2$. Experiments demonstrating violations of order independence were reported in the 1960s by Becker et al. (1963), Chipman (1960), Coombs (1958), Krantz (1967) and Tversky and Russo (1969). Tversky (1972a, b) considered the ability to accommodate violations of order independence to be a significant advantage of EBA over other choice models. McFadden (1981) endorsed the view and remarked that EBA had significant potential for econometric applications because it allowed complex patterns of substitutability among alternatives. Still, EBA has seen little practical use. One possible reason is that it does not have an error theory. Another is that it appears to be too far removed from the utility maximization paradigm that has come to dominate the literature on empirical choice models over the last half century.

The objectives of this paper are to (i) derive an error theory for EBA, (ii) examine its relation with choice models that assume utility maximization, and (iii) consider its implications for model estimation. We discuss each objective below.

(i) We show that EBA can be obtained when each aspect has random utility with an independent, extreme value distribution. We develop the theory in three steps. The first step considers a lexicographic choice rule that uses aspects arranged in decreasing order of their utilities. If each aspect has random utility with extreme value distribution, then the probability of using a particular aspect ordering in a lexicographic rule is specified by the rank-ordered logit model (Beggs et al. 1981). The second step partitions the set of all possible aspect orderings into mutually exclusive and collectively exhaustive subsets. Aspect orderings in each such subset are equivalent in the following sense: when used in a lexicographic rule, they each eliminate alternatives in any choice set in the same sequence, using the same subset of aspects in the same order. The third step shows that there is a one-to-one mapping between the probabilities associated with each subset of equivalent aspect orderings and the EBA choice probabilities.

(ii) The error theory implies that EBA is a type of logit model. It replaces the assumption that a consumer maximizes utility when selecting an alternative by the assumption that a consumer maximizes utility when selecting the next aspect to use for eliminating alternatives. In other words, EBA changes the choice process used by consumers (from trade-offs among aspects to sequential elimination by aspects), but it is consistent with utility maximization when considering the order in which aspects are used to eliminate alternatives. An EBA choice probability reduces to the probability obtained from a rank-ordered logit model when EBA selects an alternative using as many stages as there are aspects. It reduces to a multinomial logit choice probability if we define each alternative to have a unique aspect.

(iii) Our analysis implies that the parameters of an EBA model may be estimated by modifying maximum likelihood procedures that are used for estimating logit models. Maximum likelihood estimates have well-known properties, and their statistical significance can be assessed using parametric methods. It also implies that EBA can be extended in ways analogous to standard random utility models. Examples include allowing the deterministic aspect utilities to be functions of covariates and developing latent class and random-effects models to reflect consumer heterogeneity in aspect utilities.

Background. Although Tversky did not obtain an error model for EBA, he showed that it satisfies the conditions for a random utility model. These conditions were described by Block and Marschak (1960) and require that a random utility model must (i) specify a probability distribution over the set of ordered sequences of alternatives in a choice set, and (ii) choose an alternative from a choice set with a probability that is obtained by adding the probabilities of occurrence of those sequences in which it has the highest rank. Tversky considered associating random utilities with aspects but concluded that this would not lead to a characterization of EBA by an additive utility function over aspects (Tversky and Sattah 1979). McFadden (1981, p. 226) observed:

> The EBA functional form has considerable potential for econometric applications. When the scale functions V are log-linear in parameters, $\ln V(z_A) = \beta'_A z_A$, the choice probabilities can be written as sums of products of MNL forms. Maximum likelihood estimation could be carried out for such systems with relatively minor modifications of current MNL computer programs. One drawback of EBA for econometric applications is that the motivation for the model provides little guidance for the parametric specification of the scale function V.

Even as he noted that the Luce choice rule for selecting an aspect could be represented by the multinomial logit formula, McFadden refrained from proposing that the aspects be considered to have independent random utilities with extreme value distributions. The EBA choice probabilities would then have followed trivially upon the further assumption that aspect choices are independent across stages—that is, by assuming that an aspect's utility changes as a person moves from one stage of elimination to the next. But this assumption is inappropriate. It implies, for example, that an aspect (e.g., low price) could have lowest utility at one stage of elimination and the highest utility at the next stage of elimination. Such an assumption is inconsistent with Tversky's view, who stated (Tversky 1972a, p. 296):

> The EBA model accounts for choice in terms of a covert elimination process based on sequential selection of aspects. Any such sequence of aspects can be regarded as a particular state of mind which leads to a unique choice. In light of this interpretation, the choice mechanism at any given moment in time is entirely deterministic; the probabilities merely reflect the fact that at different moments in time different states of mind (leading to different choices) may prevail.

In other words, Tversky's view was that the ordering of aspects is fixed on a given choice occasion but can change across choice occasions.[2] The error model we consider is consistent with this view. We show that an extreme value distribution for aspect utilities is appropriate even when aspect choices are not independent across stages.

Although no previous research has considered an error theory for EBA, researchers have (i) proposed models that are related to EBA (Corbin and Marley 1974, Manrai and Sinha 1989, Pihlens 2008), (ii) considered the relation between nested logit models and EBA (Batley and Daly 2006), (iii) examined the use of EBA for modeling consumer response to promotions (Fader and McAlister 1990), (iv) proposed nonparametric estimation procedures (Görür et al. 2006, Gilbride and Allenby 2006), and (v) discussed issues concerning the degrees of freedom in EBA models (Batsell et al. 2003, Park and Choi 2013).

Specifically, Corbin and Marley (1974) introduced a random utility model that generalizes EBA by relaxing the assumption that choice probabilities satisfy the multiplicative inequality. Manrai and Sinha (1989) proposed elimination-by-cutoffs, a model that considers choice to be the result of a consumer choosing cutoffs in a multidimensional perceptual space. Pihlens (2008) proposed a multi-attribute elimination by aspects (MEBA) model in which the alternatives are described by attributes, and the aspects correspond to main effects, and selected interaction effects, among the attributes. Batley and Daly (2006) considered a special case of EBA with two aspects and three alternatives and showed that, in this case, there is a one-to-one mapping between the EBA parameters and the parameters of a nested logit model. Görür et al. (2006) used pairwise comparisons of alternatives, and Gilbride and Allenby (2006) used a nonparametric Bayesian approach, to estimate EBA models. Batsell et al. (2003) and Park and Choi (2013) examined issues concerning degrees of freedom in the estimation of EBA. The former also showed that there is a linear relationship between the EBA parameters and the differences in choice probabilities across choice sets.

Organization of the paper. Section 2 describes EBA and introduces a probabilistic lexicographic rule obtained when aspects have independent, extreme value distributions. Section 3 shows that EBA is equivalent to the probabilistic lexicographic rule. The equivalence establishes the error structure for EBA. Section 4 describes a probabilistic utility function for EBA. Section 5 discusses the implications for estimating EBA parameters. Section 6 concludes the paper.

2. EBA and Probabilistic Lexicographic Rule

We describe EBA and a probabilistic lexicographic rule. Let C denote a choice set. Let $A = \{1, \ldots, n\}$ denote the set of n aspects that are used to describe the alternatives in C. An aspect can be a nominal attribute (e.g., color) or a threshold value associated with a continuous attribute (e.g., a price cutoff). Let $x_{ik} = 1$ if alternative $i \in C$ has aspect $k \in A$; otherwise, $x_{ik} = 0$. Let $x_i = (x_{i1}, \ldots, x_{in})$ denote the *profile* of alternative $i \in C$. We assume that all alternatives in C are distinct; that is, $x_i \neq x_{i'}$ for any two alternatives $i, i' \in C$.

2.1. Elimination by Aspects

We begin with an informal description of EBA and then provide a more precise description.

EBA assumes that a person associates an importance weight $\alpha_k > 0$ with each aspect $k \in A$. It uses an iterative procedure to eliminate alternatives from the choice set C. The first stage (step) identifies a subset of aspects $A_1 \subseteq A$ that appear in at least one, but not all, alternatives in C. An aspect $k_1 \in A_1$ is selected with probability proportional to α_{k_1}. Alternatives in which aspect k_1 does not appear (that is, for which $x_{ik_1} = 0$) are eliminated. Let $C_1 \subset C$ denote the set of remaining alternatives. The elimination of alternatives stops if C_1 has only one alternative. Otherwise, the procedure advances to a second stage, where it identifies a subset of aspects $A_2 \subset A_1$ that appear in at least one, but not all, alternatives in C_1. An aspect $k_2 \in A_2$ is selected with probability proportional to α_{k_2}. Alternatives in which aspect k_2 does not appear (that is, for which $x_{ik_2} = 0$) are eliminated. The procedure iterates until all alternatives except one are eliminated.

We will need the following (precedence) conditions to show that EBA is equivalent to the probabilistic lexicographic rule discussed in §2.2. For each step j, define $\pi(k_j) < \pi(k)$ to mean that aspect k_j is selected for eliminating alternatives before any other aspect $k \neq k_j$, $k \in A_j$. For example, if $A = \{1, 2, 3\}$ and $k_1 = 1$ is selected at step $j = 1$, then $\pi(1) < \pi(2)$ and $\pi(1) < \pi(3)$. These conditions mean that aspect 1 is used to eliminate alternatives before aspects 2 and 3 (which may or may not be used in a later elimination step). If there are $m \leqslant n$ elimination steps, then $\pi(k_1) < \pi(k_2) < \cdots < \pi(k_m)$, because aspects that are used to eliminate alternatives at a later step are also available for eliminating alternatives at earlier steps. But the precedence conditions also imply that $\pi(k_j) < \pi(k)$ for any aspect $k \in A_j$ that is distinct from $k_{j+1}, \ldots, k_m$ and is not used to eliminate alternatives at any step, because (i) either EBA terminates before using all remaining aspects in A_j, or (ii) aspect k appears in all or none of the alternatives in at least one of the sets $C_j, \ldots, C_m$.

The following algorithm gives a formal description of EBA.

Initialization step: Let $A_0 = A = \{1, \ldots, n\}$, $C_0 = C$, $x_i = (x_{i1}, \ldots, x_{in})$, for all $i \in C_0$, and $\alpha = (\alpha_1, \ldots, \alpha_n)$.

Iteration step j ($\geqslant$1): Let

$$A_j = \left\{ k \in A_{j-1} \;\middle|\; 0 < \sum_{i \in C_{j-1}} x_{ik} < |C_{j-1}| \right\}$$

denote the subset of aspects in A_{j-1} that appear in at least one, but not all, of the alternatives in C_{j-1}. Select aspect $k_j \in A_j$ with probability

$$q(k_j, A_j) = \frac{\alpha_{k_j}}{\sum_{k \in A_j} \alpha_k}. \tag{1}$$

Set $C_j = \{i \in C_{j-1} \mid x_{ik_j} = 1\}$. Let $\pi(k_j) < \pi(k)$, for all $k \in A_j \backslash \{k_j\}$, to indicate that aspect k_j is used to eliminate alternatives before any other aspect in A_j.

Termination step: Stop if $|C_j| = 1$.

Since each step eliminates at least one alternative, EBA terminates in $m \leqslant n$ steps. We refer to a particular sequence $k_1, \ldots, k_m$ of aspects used by EBA as an *instance*. Observe that $C_j \equiv C_j(k_j)$—that is, C_j is a function of k_j—because its composition depends on which aspect $k_j \in A_j$ is selected to eliminate alternatives from C_{j-1}. Unless necessary, we simplify notation by writing C_j instead of the more explicit $C_j(k_j)$ in the rest of the paper.

Tversky (1972a) noted that the EBA choice probability for an alternative can be computed by using the following recursive formula. Let $A_{ij} \subseteq A_j$ denote the subset of aspects in A_j that appear in alternative $i \in C_{j-1}$. As defined previously, let $C_j(k_j)$ denote the subset of alternatives in C_{j-1} that have aspect k_j. Let $p(i, C_j(k_j))$ denote the choice probability for alternative i, conditional on the selection of aspect $k_j \in A_{ij}$ at step j. Then the probability that EBA chooses alternative $i \in C_{j-1}$ is given by

$$p(i, C_{j-1}) = \sum_{k_j \in A_{ij}} \left(\frac{\alpha_{k_j}}{\sum_{k \in A_j} \alpha_k} \right) p(i, C_j(k_j)), \quad j \geqslant 1,$$

where $\alpha_{k_j} / \sum_{k \in A_j} \alpha_k$ is the probability of selecting aspect $k_j \in A_j$ at step j. The unconditional choice probability for alternative $i \in C$ is $p(i, C_0)$, obtained by setting $C = C_0$.

The following example describes an EBA problem, illustrates the precedence conditions, computes the choice probabilities for alternatives, and shows how EBA allows violations of order independence.

EXAMPLE 1 (BEETHOVEN AND DEBUSSY RECORD ALBUMS). Consider the following choice problem due to Debreu (1960). A person has to choose an alternative from the set $C = \{x_1, x_2, x_3\}$. Alternatives x_1 and x_2 are different but equally good recordings of the same Beethoven symphony. Alternative x_3 is a recording of a Debussy suite.[3] Let $A = \{1, 2, 3, 4\}$ denote the set of four aspects that are used to describe the albums. Aspect 1 is "Beethoven symphony," aspect 2 is "Debussy suite," and aspects 3 and 4 are features that distinguish between the two Beethoven albums, such as two different orchestras or recording labels. Let α_k denote the importance weight that the person associates with aspect k, for each $k = 1, 2, 3, 4$. Let $x_i = (x_{i1}, \ldots, x_{i4})$ denote the profile of alternative i, where $x_{ik} = 1(0)$ if alternative i has (does not have) aspect k. Then $x_1 = (1\ 0\ 1\ 0)$ and $x_2 = (1\ 0\ 0\ 1)$ represent the two Beethoven albums, and $x_3 = (0\ 1\ 0\ 0)$ represents the Debussy album.

We describe the steps for one instance of EBA in detail and then give a summary description of the other instances.

Initialization step. Set $A_0 = A = \{1, 2, 3, 4\}$, $C_0 = C = \{1, 2, 3\}$, $x_1 = (1\ 0\ 1\ 0)$, $x_2 = (1\ 0\ 0\ 1)$, $x_3 = (0\ 1\ 0\ 0)$ and $\alpha = (\alpha_1, \ldots, \alpha_4)$.

Step 1. Set $A_1 = A_0$. Select aspect $k_1 = 1$ with probability

$$q(1, A_1) = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}.$$

Set $\pi(1) < \pi(j)$, for $j = 2, 3, 4$, to indicate that this EBA instance selects aspect 1 before aspects 2, 3 and 4. Eliminate alternative 3 because it does not have aspect 1. Set $C_1 = C_0 \backslash \{3\} = \{1, 2\}$.

Step 2. Do not consider aspect 2 since it does not appear in alternatives 1 and 2. Set $A_2 = \{3, 4\}$. Select aspect $k_2 = 3$ with probability

$$q(3, A_2) = \frac{\alpha_3}{\alpha_3 + \alpha_4}.$$

Set $\pi(3) < \pi(4)$ because this EBA instance selects aspect 3 before aspect 4.

Eliminate alternative 2 because it does not have aspect 3. Set $C_2 = C_1 \backslash \{2\} = \{1\}$.

Termination step. Stop because $|C_2| = 1$. Choose alternative 1.

The preceding EBA instance selects alternative 1 with probability

$$q(1, A_1) \cdot q(3, A_2) = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \cdot \frac{\alpha_3}{\alpha_3 + \alpha_4}.$$

The precedence conditions are $\pi(1) < \pi(2)$, $\pi(1) < \pi(3)$, $\pi(1) < \pi(4)$, and $\pi(3) < \pi(4)$.

Next, consider the EBA instance in which Step 1 still selects aspect $k_1 = 1$ with probability $q(1, A_1)$, but Step 2 selects aspect $k_2 = 4$ from the set $A_2 = \{3, 4\}$ with probability

$$q(4, A_2) = \frac{\alpha_4}{\alpha_3 + \alpha_4}.$$

Then $C_2 = C_1 \backslash \{1\} = \{2\}$. As C_2 has a single alternative, EBA stops after Step 2 and chooses alternative 2 with probability

$$q(1, A_1) \cdot q(4, A_2) = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \cdot \frac{\alpha_4}{\alpha_3 + \alpha_4}.$$

Since aspect 1 is chosen from $A_1 = \{1, 2, 3, 4\}$ and aspect 4 is chosen from $A_2 = \{3, 4\}$, the precedence conditions are $\pi(1) < \pi(2)$, $\pi(1) < \pi(3)$, $\pi(1) < \pi(4)$, and $\pi(4) < \pi(3)$.

Finally, consider the three other instances, in which aspect 2, 3, or 4 is selected in the first stage. In each instance, EBA eliminates two of the three alternatives in Step 1 and then stops. Specifically

(i) Step 1 selects aspect $k_1 = 2$ with probability

$$q(2, A_1) = \frac{\alpha_2}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}.$$

It sets $\pi(2) < \pi(k)$, for $k = 1, 3, 4$; eliminates alternatives 1 and 2; sets $C_1 = C_0 \backslash \{1, 2\} = \{3\}$; and chooses alternative 3 with probability $q(2, A_1)$.

(ii) Step 1 selects aspect $k_1 = 3$ with probability

$$q(3, A_1) = \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}.$$

It sets $\pi(3) < \pi(k)$, for $k = 1, 2, 4$; eliminates alternatives 2 and 3; sets $C_1 = C_0 \backslash \{2, 3\} = \{1\}$; and chooses alternative 1 with probability $q(3, A_1)$.

(iii) Step 1 selects aspect $k_1 = 4$ with probability

$$q(4, A_1) = \frac{\alpha_4}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}.$$

It sets $\pi(4) < \pi(k)$, for $k = 1, 2, 3$; eliminates alternatives 1 and 3; sets $C_1 = C_0 \backslash \{1, 3\} = \{2\}$; and chooses alternative 2 with probability $q(4, A_1)$.

Across instances, EBA selects the alternatives x_1, x_2, and x_3 with the following probabilities:

$$\begin{aligned} p(x_1, C) &= q(1, A_1) \cdot q(3, A_2) + q(3, A_1) \\ &= \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \cdot \frac{\alpha_3}{\alpha_3 + \alpha_4} + \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \\ &= \frac{1}{1 + \alpha_4/\alpha_3}\left(1 - \frac{\alpha_2}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}\right), \end{aligned} \tag{2}$$

$$\begin{aligned} p(x_2, C) &= q(1, A_1) \cdot q(4, A_2) + q(4, A_1) \\ &= \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \cdot \frac{\alpha_4}{\alpha_3 + \alpha_4} + \frac{\alpha_4}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4} \\ &= \frac{1}{1 + \alpha_3/\alpha_4}\left(1 - \frac{\alpha_2}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}\right), \end{aligned} \tag{3}$$

$$p(x_3, C) = q(2, A_1) = \frac{\alpha_2}{\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4}. \tag{4}$$

Observe that $p(x_1, C) = a\{1 - p(x_3, C)\}$, $p(x_2, C) = (1-a)\{1-p(x_3, C)\}$, where $a = \alpha_3/(\alpha_3+\alpha_4)$. That is, the choice probability $1-p(x_3, C)$ of not selecting the Debussy album is split between the two Beethoven albums in direct proportion to the importance weights α_3 and α_4. Suppose the person is indifferent between the two Beethoven albums and associates the same weight $\alpha_3 = \alpha_4 = \alpha$. Then $a = 1/2$ and

$$p(x_1, C) = p(x_2, C) = \frac{1-p(x_3, C)}{2},$$
$$p(x_3, C) = \frac{\alpha_2}{\alpha_1+\alpha_2+2\alpha}.$$

As α becomes smaller, the choice probabilities approach the limiting values

$$\lim_{\alpha\to 0} p(x_1, C) = \lim_{\alpha\to 0} p(x_2, C) = \frac{1}{2}\left(1-\frac{\alpha_2}{\alpha_1+\alpha_2}\right),$$
$$\lim_{\alpha\to 0} p(x_3, C) = \frac{\alpha_2}{\alpha_1+\alpha_2}.$$

Thus, unlike the logit model, EBA considers the essential choice to be between Beethoven and Debussy.

To see how EBA allows violations of order independence, suppose that $\alpha_1/\alpha_2 = 1+\epsilon$, where $\epsilon > 0$ is a small, positive value. Then

$$p(x_3, C) = \frac{\alpha_2}{\alpha_1+\alpha_2} = \frac{1}{1+\alpha_1/\alpha_2} = \frac{1}{2+\epsilon} = \frac{1}{2} - 2\delta,$$

where $\delta = \epsilon/(8+4\epsilon) > 0$. Thus,

$$p(x_1, C) = p(x_2, C) = \frac{1}{2}\left(1-\frac{\alpha_2}{\alpha_1+\alpha_2}\right)$$
$$= \frac{1}{2}\left(\frac{1}{2}+2\delta\right) = \frac{1}{4}+\delta.$$

On the other hand, if x_2 were not available, then alternative x_1 would be selected with probability $1-p(x_3, C) = 1/2+2\delta$. Thus, x_1 has a higher choice probability than x_3 when x_2 is not available, but it has a lower choice probability than x_3 when x_2 is available. This is a violation of order independence, which is captured by EBA but not by the multinomial logit choice model.

Equation (1) gives the probability of selecting an aspect at stage j of an EBA instance. The probability of an EBA instance across its m stages is given by

$$\prod_{j=1}^{m} q(k_j, A_j) = \prod_{j=1}^{m} \frac{\alpha_{k_j}}{\sum_{k\in A_j} \alpha_k}. \tag{5}$$

McFadden (1981) noted that since $q(k_j, A_j)$ has the form of the Luce choice rule, the probability of selecting an aspect at stage j can be written as

$$q(k_j, A_j) = \frac{e^{v_{k_j}}}{\sum_{k\in A_j} e^{v_k}},$$

where $v_k = \ln(\alpha_k)$ for all $k \in A$. Thus, if we restrict attention to a single stage of EBA, then the values of the aspect choice probabilities can be trivially obtained by assuming that aspect $k \in A$ has random utility $u_k = v_k + \epsilon_k$, where v_k is a deterministic component and ϵ_k is a stochastic component with extreme value distribution. The expression in Equation (5) follows immediately if we are willing to assume that the utility of an aspect is an independent draw from its distribution at each stage of EBA. That is, if k_1 is selected in the first stage with probability $q(k_1, A_1)$, then k_2 will be selected in the second stage with probability $q(k_2, A_2 \mid k_1, A_1) = q(k_2, A_2)$. However, as noted in the introduction, Tversky's view was that the aspect ordering can change only across choice occasions, but that it is fixed on a given choice occasion.

Since any of the $n!$ possible orderings of the n aspects is feasible, the derivation of EBA choice probabilities requires considering each such ordering and showing that an extreme value distribution of aspect utilities still yields the EBA choice probabilities. We obtain this result in two steps. The first step considers a probabilistic lexicographic rule in which (i) each aspect has an independent, random utility with an extreme value distribution, (ii) a single draw from the utility distribution is obtained for each aspect on a given choice occasion, and (iii) the aspects are ordered in decreasing order of their utilities and then used in a lexicographic rule to select an alternative from a choice set. This yields an expression for the choice probabilities associated with a subset of aspect sequences that eliminate the same alternatives in the same order when used in a lexicographic rule. The second step shows that each such subset of sequences corresponds to an EBA instance and is selected with a probability given by Equation (5).

2.2. Probabilistic Lexicographic Rule

A lexicographic rule uses a sequence of aspects to eliminate alternatives until a single alternative is selected. A probabilistic lexicographic rule differs from a deterministic lexicographic rule in only one way: it uses a probabilistic mechanism to generate the sequence of aspects used in a lexicographic rule. We first describe the probabilistic mechanism for generating sequences of aspect and then provide a formal description of a lexicographic rule.

Let $u_k = v_k + \epsilon_k$ denote the utility a person associates with aspect k, where v_k is a deterministic component, and ϵ_k is a stochastic component, for all $k \in A = \{1, \ldots, n\}$. The stochastic component represents uncertainty in a person's evaluation of an aspect's utility. It can occur because the same aspect can appear more or less important, depending on the choice context or a person's frame of mind. For example, sometimes a person might have the taste for meat, at other times the taste for a vegetarian meal. Tversky and Sattah (1979) observed that such uncertainty can persist even when the effects of learning, satiation, or changes in tastes are taken into account. They noted that even in unique choice situations, where a choice is essentially made only once, people can express doubt about the importance they associate with one or another aspect of the alternatives. For example, a person might not be sure how much he or she values a larger screen size in a smart phone or television set. Such uncertainty cannot be reflected by a standard error model, in which the total utility of an alternative is allowed to be uncertain, but not the utilities of individual aspects.

We consider a model in which ϵ_k, the stochastic component of aspect k's utility, has an independent, extreme value distribution, with density function $f_k(\epsilon_k) = \exp(-\epsilon_k e^{-\epsilon_k})$ and cumulative distribution function $F_k(\epsilon_k) = \exp(-e^{-\epsilon_k})$. On a given choice occasion, a single value of u_k is obtained as an independent random draw from its distribution, for all $k \in A$. Let $u_{k_1} > \cdots > u_{k_n}$, where $k_j \in A$ for all $j = 1, \ldots, n$. Let $s = (k_1, \ldots, k_n)$ denote a sequence representing a preference ordering of the aspects for a person who prefers aspect k_j to aspect $k_{j'}$, for all $1 \leqslant j < j' \leqslant n$. We say that sequence s has jth element k_j, or equivalently, that it assigns aspect k_j to position j, for all $j = 1, \ldots, n$.

Let S denote the set of $n!$ distinct sequences. A lexicographic rule uses a sequence $s \in S$ to choose an alternative from C by first eliminating alternatives that do not have aspect k_1, then eliminating those remaining alternatives that do not have aspect k_2, and so on, until only one alternative remains. An implementation of the rule terminates in at most n steps and uses an aspect only if it appears in some, but not all, of the remaining alternatives at a given step. For brevity, we say that a sequence selects an alternative from a choice set C, although strictly speaking it is a lexicographic rule that selects an alternative using the ordering of aspects in a sequence. The following algorithm gives a formal description of a lexicographic rule. The input to the algorithm is a choice set C and a particular preference ordering $s = (k_1, \ldots, k_n)$ over the aspects.

Initialization step: Let $s = (k_1, \ldots, k_n)$, $C_0 = C$ and $x_i = (x_{i1}, \ldots, x_{in})$ for all $i \in C$.

Iteration step j ($\geqslant 1$): Set $B_j = \{i \in C_{j-1} \mid x_{ik_j} = 1\}$,

$$C_j = \begin{cases} C_{j-1} & \text{if } B_j = \varnothing \text{ or } B_j = C_{j-1}, \\ B_j & \text{otherwise,} \end{cases}$$

and

$$t_j = \begin{cases} 0 & \text{if } B_j = \varnothing \text{ or } B_j = C_{j-1}, \\ 1 & \text{otherwise.} \end{cases}$$

Termination step: Stop if $|C_j| = 1$. Set $t_{j+1} = \cdots = t_n = 0$ if the algorithm stops at the end of step $j < n$.

In the preceding algorithm, (i) B_j denotes the subset of alternatives in C_{j-1} that have aspect k_j; and (ii) $t_j = 1$ indicates that aspect k_j appears in some, but not all, of the alternatives in C_{j-1}; otherwise, $t_j = 0$. We specify the values of t_j because they will be useful for establishing the equivalence between EBA and the probabilistic lexicographic rule. We illustrate the use of aspect sequences to implement a lexicographic rule and the assignment of the t_j values with an example.

Example 2. We reconsider Example 1 in which $C = \{x_1, x_2, x_3\}$, where x_1 and x_2 denote the two Beethoven albums, and x_3 denotes the Debussy album. Albums have the following aspects: x_1 has aspects 1 and 3. x_2 has aspects 1 and 4, and x_3 has aspect 2. Consider a lexicographic rule using the aspect sequence $s = (2\ 1\ 3\ 4)$. It uses aspects $k_1 = 2$, $k_2 = 1$ and $k_3 = 3$ one after another to eliminate alternatives until only a single alternative remains. Since there is only one Debussy album, this sequence stops after one step. Thus, $t_1 = 1$, $t_2 = t_3 = t_4 = 0$, where

(i) $t_1 = 1$ because aspect 2 (Debussy suite) is used to eliminate both Beethoven albums from the set $C_0 = \{x_1, x_2, x_3\}$; and

(ii) $t_2 = t_3 = t_4 = 0$ because the algorithm stops after step 1 ($C_1 = \{x_3\}$).

Next, consider a lexicographic rule using the aspect sequence $s = (1234)$. It eliminates the Debussy album in the first step and then eliminates the second Beethoven album. Thus, $t_1 = 1$, $t_2 = 0$, $t_3 = 1$, $t_4 = 0$, where

(i) $t_1 = 1$ because aspect 1 (Beethoven symphony) appears in two of the three alternatives in $C_0 = \{x_1, x_2, x_3\}$;

(ii) $t_2 = 0$ because aspect 2 (Debussy suite) appears in neither of the alternatives in $C_1 = \{x_1, x_2\}$ (the two Beethoven recordings);

(iii) $t_3 = 1$ because aspect 3 (which is unique to the first Beethoven recording) appears in one of the two alternatives in $C_1 = \{x_1, x_2\}$; and

(iv) $t_4 = 0$ because the algorithm stops after step 2 ($C_2 = \{x_1\}$).

Table 1 shows the lexicographic choices obtained by each of the $4! = 24$ aspect sequences. Note that an alternative can be chosen by more than one sequence.

To conclude, we note that although the aspect utilities have independent distributions in the probabilistic lexicographic model, they are perfectly correlated across elimination stages. This is because the values of the aspect utilities can only vary across choice occasions, but they are fixed across elimination stages on a choice occasion. Similarly, although the aspect utilities are independent, the choice probabilities of the alternatives are not independent. The reason is that if a (proper) subset of alternatives in a choice set has common aspects, then the choice probabilities of each alternative in the subset depends on the utilities of these common aspects.

3. Equivalence of EBA and Probabilistic Lexicographic Rule

We now establish the equivalence between the probabilistic lexicographic rule and EBA. First we partition the set of all possible aspect orderings into mutually exclusive and collectively exhaustive subsets. Table 1 shows this partitioning for the Beethoven-Debussy example. Aspect orderings in each such subset are equivalent in the following sense: when used by an individual in a lexicographic rule, they each eliminate alternatives in any choice set in the same

Table 1. Choices associated with lexicographic sequences.

k_1	k_2	k_3	k_4	Choice
1	2	3	4	
1	3	2	4	x_1
1	3	4	2	
1	2	4	3	
1	4	2	3	x_2
1	4	3	2	
2	1	3	4	
2	1	4	3	
2	3	1	4	x_3
2	3	4	1	
2	4	1	3	
2	4	3	1	
3	1	2	4	
3	1	4	2	
3	2	1	4	x_1
3	2	4	1	
3	4	1	2	
3	4	2	1	
4	1	2	3	
4	1	3	2	
4	2	1	3	x_2
4	2	3	1	
4	3	1	2	
4	3	2	1	

Notes. x_1 and x_2 denote two different recordings of a Beethoven symphony; x_3 denotes a recording of a Debussy suite. All aspect orderings in a partition correspond to a set of sequences that satisfy the same precedence conditions resulting in the choice of the same alternative.

sequence, using the same subset of aspects in the same order. Then we show that there is a one-to-one mapping between the probabilities associated with each subset of equivalent aspect orderings and the EBA choice probabilities. The mapping establishes the desired equivalence and implies that EBA can be characterized by random aspect utilities with independent, extreme value distributions.

Recall that $t_j = 1(0)$ if aspect k_j is used (not used) by the probabilistic lexicographic rule for eliminating alternatives at stage j. Let $t_j = 1$ for m aspects, where $1 \leqslant m \leqslant n$. If $m = n$, then the choice set has at least n alternatives, and sequence $s = (k_1, \ldots, k_n)$ uses all n aspects to choose alternative $i \in C$; another sequence $s' = (k'_1, \ldots, k'_n)$ may also select $i \in C$, but only by eliminating alternatives in a different order than sequence s. But if $t_j = 0$ for some $j = 1, \ldots, n$, then there can be other sequences s' that choose alternative $i \in C$ by eliminating alternatives in the same order as sequence s. Suppose such a sequence s' exists. Let aspect k_j, which appears in position j in sequence s, appear in position $\sigma(k_j)$ in sequence s', for all $j = 1, \ldots, n$ (for example, suppose $s = (3\ 1\ 2)$ and $s' = (2\ 3\ 1)$; then $\sigma(3) = 2$, $\sigma(1) = 3$ and $\sigma(2) = 1$). Lemma 1 gives the condition under which two sequences, s and s', use the same aspects, in the same order, to eliminate alternatives from a choice set. A proof for Lemma 1, and proofs for subsequent lemmas and theorems, are provided in the appendix.

Lemma 1. *Let sequence $s = (k_1, \ldots, k_n)$ select alternative $i \in C$ using m steps, where $1 \leqslant m < n$. Let $t_j = 1$ for $j = j_1, \ldots, j_m$ and $t_j = 0$ otherwise, where $0 < j_1 \cdots < j_m$. Then sequence s', which assigns position $\sigma(k_j)$ to aspect k_j, also selects alternative $i \in C$ using m steps, eliminating alternatives in the same order as sequence s, if $\sigma(k_{j_r}) < \sigma(k_j)$, for all $j_r + 1 \leqslant j \leqslant n$ and $1 \leqslant r \leqslant m$.*

Lemma 1 is a key result for obtaining the correspondence between EBA and a probabilistic lexicographic rule. We illustrate the result with an example.

Example 3. Returning to Example 2, recall that the sequence $s = (2134)$ chooses the Debussy album (x_3) after one step, so that $t_1 = 1$, $t_2 = t_3 = t_4 = 0$. Thus, all sequences in which aspect 2 precedes all other aspects (that is, $\sigma(2) < \sigma(j)$, for $j = 1, 3, 4$) also select the Debussy album. The set of all such sequences is given by

$$K_1 = \{(2134), (2143), (2314), (2341), (2413), (2431)\}.$$

Similarly, recall that the sequence $s = (1234)$ chooses the first Beethoven album (x_1) after two steps, and that $t_1 = 1$, $t_2 = 0$, $t_3 = 1$, $t_4 = 0$. Thus, all sequences in which aspect 1 precedes all other aspects (that is, $\sigma(1) < \sigma(j)$, for $j = 2, 3, 4$), and in which aspect 3 also precedes aspect 4 (that is, $\sigma(3) < \sigma(4)$), select the same Beethoven album. The set of all such sequences is given by

$$K_2 = \{(1234), (1324), (1342)\}.$$

In Table 1, the third grouping of sequences corresponds to the set K_1, and the first grouping of sequences corresponds to the set K_2. Other groupings in Table 1 correspond to similar sets of sequences, each sequence in a set eliminating the same alternatives, using the same aspects in the same order.

As Example 3 illustrates, the precedence conditions $\sigma(k_j) < \sigma(k)$ collectively specify the subset of sequences $K \subset S$ that eliminate the same alternatives, in the same order, as sequence $s \in S$. We derive the probability of obtaining the sequences in K and show that it is equal to the probability of a corresponding EBA instance.

Consider a lexicographic rule that uses sequence $s = (k_1, \ldots, k_n)$ and terminates in $m \leqslant n$ steps. By definition, $u_{k_1} > \cdots > u_{k_n}$, and

$$K = \{s \in S \mid \sigma(k_{j_r}) < \sigma(k_j), \text{ for all } j_r + 1 \leqslant j \leqslant n \text{ and } 1 \leqslant r \leqslant m\},$$

where each sequence $s \in K$ assigns position $\sigma(k_j)$ to aspect k_j and satisfies the condition $\sigma(k_{j_r}) < \sigma(k_{j_r+1}) < \cdots < \sigma(k_n)$, for each $1 \leqslant r \leqslant m$. From Lemma 1, a lexicographic rule using any sequence $s \in K$ eliminates alternatives in the same order, terminates in m steps, and selects the same alternative. The probability of obtaining a sequence in K is given by

$$p(K)=\sum_{s\in K}p(s)=p(\sigma(k_{j_r})<\sigma(k_j),\\ \text{for all } j_r+1\leqslant j\leqslant n \text{ and } r=1,\ldots,m),$$

where $p(s)$ is the probability of obtaining sequence $s\in K$. Recall that $u_k=v_k+\epsilon_k$ denotes the utility a person associates with aspect k, where v_k is a deterministic component, and ϵ_k is a stochastic component with an extreme value distribution. Following Beggs et al. (1981), if $u_{k_1}>\cdots>u_{k_n}$, then sequence $s=(k_1,\ldots,k_n)$ is obtained with probability

$$p(s)=\prod_{j=1}^{n-1}\frac{e^{v_{k_j}}}{e^{v_{k_j}}+\cdots+e^{v_{k_n}}}. \tag{6}$$

The following example illustrates the computation of $p(K)$ and shows how it relates to the EBA choice probabilities in Example 1.

Example 4. In Example 3, each sequence in the set

$$K_1=\{(2134),(2143),(2314),(2341),(2413),(2431)\},$$

selected the Debussy album (x_3), and each sequence in the set

$$K_2=\{(1234),(1324),(1342)\},$$

selected the first Beethoven album (x_1). The probability of obtaining a sequence in K_1 is given by the sum of the probabilities in Equation (6) for each sequence $s\in K_1$:

$$\begin{aligned}p(K_1)={}&\frac{e^{v_2}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_1}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_3}+e^{v_4}}\\&+\frac{e^{v_2}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_1}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_4}}{e^{v_3}+e^{v_4}}\\&+\frac{e^{v_2}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_1}}{e^{v_1}+e^{v_4}}\\&+\frac{e^{v_2}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_4}}{e^{v_1}+e^{v_4}}\\&+\frac{e^{v_2}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_4}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_1}}{e^{v_1}+e^{v_3}}\\&+\frac{e^{v_2}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_4}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_1}+e^{v_3}}.\end{aligned}$$

Similarly, the probability of obtaining a sequence in K_2 is given by

$$\begin{aligned}p(K_2)={}&\frac{e^{v_1}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_2}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_3}+e^{v_4}}\\&+\frac{e^{v_1}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_1}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_2}}{e^{v_2}+e^{v_4}}\\&+\frac{e^{v_1}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_3}}{e^{v_3}+e^{v_3}+e^{v_4}}\cdot\frac{e^{v_4}}{e^{v_2}+e^{v_4}}.\end{aligned}$$

Simplifying the above expressions gives

$$p(K_1)=\frac{e^{v_2}}{e^{v_2}+e^{v_3}+e^{v_4}+e^{v_5}},$$

$$p(K_2)=\frac{e^{v_1}}{e^{v_1}+e^{v_2}+e^{v_3}+e^{v_4}}\frac{e^{v_3}}{e^{v_3}+e^{v_4}}.$$

Thus, (i) $p(K_1)$ is the multinomial logit probability of choosing aspect 2 from aspects 1–4, and (ii) $p(K_2)$ is the product of two multinomial logit probabilities, one corresponding to the choice of aspect 1 from aspects 1–4, and the other corresponding to the choice of aspect 3 from aspects 3 and 4. If we set $e^{v_k}=\alpha_k$ in the above expressions, we obtain a value of $p(K_1)$ that is identical to the probability $p(x_3,C)$ in Equation (4) for an EBA instance that terminates after a single step upon choosing aspect 2. Similarly, we obtain a value of $p(K_2)$ that is identical to the probability $p(x_2,C)$ in Equation (3) for an EBA instance which (i) uses aspect 1 in step 1, (ii) ignores aspect 2, and (iii) uses aspect 3 in step 2, after which it terminates.

The equivalence between the probabilities of EBA instances and sets of lexicographic sequences in Example 4 is valid in general. We show the equivalence without the cumbersome enumeration of equivalent sequences characterizing a set K. We do so by integrating over the distributions of aspect utilities in such a way that all precedence constraints of the form $\sigma(k_{j_r})<\sigma(k_j)$ that characterize a set K are satisfied.

Since $u_k\geqslant u_{k'}$ when $\sigma(k)<\sigma(k')$, we can write $p(K)$ as

$$\begin{aligned}p(K)&=p(\sigma(k_{j_r})<\sigma(k_j),\text{ for all } j_r+1\leqslant j\leqslant n\\&\qquad\text{and } r=1,\ldots,m)\\&=p(u_{k_{j_r}}\geqslant u_{k_j},\text{ for all } j_r+1\leqslant j\leqslant n\\&\qquad\text{and } r=1,\ldots,m).\end{aligned} \tag{7}$$

For example, if $m=2$, then Equation (7) can be written as

$$p(K)=p(u_{k_{j_1}}\geqslant u_{k_j},\text{ for all } j_1+1\leqslant j\leqslant n \text{ and } u_{k_{j_2}}\geqslant u_{k_j},\\ \text{for all } j_2+1\leqslant j\leqslant n). \tag{8}$$

Since $u_{k_{j_1}}\geqslant u_{k_{j_2}}$, the condition $u_{k_{j_2}}\geqslant u_{k_j}$ implies $u_{k_{j_1}}\geqslant u_{k_j}$ for all $j_2+1\leqslant j\leqslant n$. Thus, we can write Equation (8) as

$$p(K)=p(u_{k_{j_1}}\geqslant u_{k_j},\text{ for all } j_1+1\leqslant j\leqslant j_2 \text{ and } u_{k_{j_2}}\geqslant u_{k_j},\\ \text{for all } j_2+1\leqslant j\leqslant n). \tag{9}$$

More generally, $u_{k_{j_1}}\geqslant u_{k_{j_2}}\geqslant\cdots\geqslant u_{k_{j_m}}$ implies that Equation (7) can be written as

$$p(K)=p(u_{k_{j_r}}\geqslant u_{k_j},\text{ for all } j_r+1\leqslant j\leqslant j_{r+1}\\ \text{and } r=1,\ldots,m), \tag{10}$$

where, by definition, $j_{m+1}=n$.

Let $u_{k_{j_0}} = \infty$. Then for any $r = 1, \ldots, m$,

$$p(u_{k_{j_r}} \geqslant u_{k_j}, \text{ for all } j_r + 1 \leqslant j \leqslant j_{r+1}) = \int_{-\infty}^{u_{k_{j_{r-1}}}} \int_{-\infty}^{u_{k_{j_r}}} \cdots \int_{-\infty}^{u_{k_{j_r}}} f_{k_{j_r}}(u_{k_{j_r}}) f_{k_{j_r+1}}(u_{k_{j_r+1}}) \cdots f_{k_{j_{r+1}}}(u_{k_{j_{r+1}}})\, du_{k_{j_r}}\, du_{k_{j_r+1}} \cdots du_{k_{j_{r+1}}},$$

where the first term integrates $u_{k_{j_r}}$ over $(-\infty, u_{k_{j_{r-1}}})$, and each subsequent term integrates u_{k_j} over $(-\infty, u_{k_{j_r}})$, for each $j = j_r + 1, \ldots, j_{r+1}$. Since the aspect utilities are independent, we can rewrite the above expression as

$$p(u_{k_{j_r}} \geqslant u_{k_j}, \text{ for all } j_r + 1 \leqslant j \leqslant j_{r+1}) = \int_{-\infty}^{u_{k_{j_{r-1}}}} f_{k_{j_r}}(u_{k_{j_r}}) \prod_{j=j_r+1}^{j_{r+1}} F_{k_j}(u_{k_{j_r}})\, du_{k_{j_r}}, \quad (11)$$

where

$$F_{k_j}(u) = \int_{-\infty}^{u} f_{k_j}(u_{k_j})\, du_{k_j}$$

is the cumulative distribution function for $u_{k_j} = v_{k_j} + \epsilon_{k_j}$, the utility of aspect k_j. Let

$$G_r = f_{k_{j_r}}(u_{k_{j_r}}) \prod_{j=j_r+1}^{j_{r+1}-1} F_{k_j}(u_{k_{j_r}}), \quad \text{for all } r = 1, \ldots, m-1,$$

and

$$G_m = f_{k_{j_m}}(u_{k_{j_m}}) \prod_{j=j_m+1}^{j_{m+1}} F_{k_j}(u_{k_{j_r}}),$$

where $j_{m+1} = n$. Then Equation (11) can be written as

$$p(u_{k_{j_r}} \geqslant u_{k_j}, \text{ for all } j_r + 1 \leqslant j \leqslant j_{r+1}) = \int_{-\infty}^{u_{k_{j_{r-1}}}} G_r \int_{-\infty}^{u_{k_{j_r}}} f_{k_{j_{r+1}}}(u_{k_{j_{r+1}}})\, du_{k_{j_{r+1}}}\, du_{k_{j_r}}, \quad (12)$$

for any $r = 1, \ldots, m$. Thus,

$$\begin{aligned} p(K) &= p(u_{k_{j_r}} \geqslant u_{k_j}, \text{for all } j_r+1 \leqslant j \leqslant j_{r+1} \text{ and } r = 1, \ldots, m) \\ &= \int_{-\infty}^{u_{k_{j_0}}} \left(f_{k_{j_1}}(u_{k_{j_1}}) \prod_{j=j_1+1}^{j_2-1} F_{k_j}(u_{k_{j_1}}) \right) \cdot \int_{-\infty}^{u_{k_{j_1}}} \left(f_{k_{j_2}}(u_{k_{j_2}}) \prod_{j=j_2+1}^{j_3-1} F_{k_j}(u_{k_{j_2}}) \right) \\ &\quad \cdots \int_{-\infty}^{u_{k_{j_{m-1}}}} \left(f(u_{k_{j_m}}) \prod_{j=j_m+1}^{j_{m+1}} F_{k_j}(u_{k_{j_m}}) \right) du_{k_{j_m}} \cdots du_{k_{j_2}}\, du_{k_{j_1}} \\ &= \int_{u_{k_{j_1}}=-\infty}^{u_{k_{j_0}}} G_1 \int_{u_{k_{j_2}}=-\infty}^{u_{k_{j_1}}} G_2 \cdots \int_{u_{k_{j_m}}=-\infty}^{u_{k_{j_{m-1}}}} G_m\, du_{k_{j_m}} \cdots du_{k_{j_2}}\, du_{k_{j_1}}. \end{aligned} \quad (13)$$

The following lemma is useful for proving Theorem 1, which gives a closed-form expression for $p(K)$.

Lemma 2.

$$I_r = \int_{u_{k_{j_r}}=-\infty}^{u_{k_{j_{r-1}}}} G_r \exp\left(-e^{-u_{k_{j_r}}} \sum_{j=j_{r+1}}^{n} e^{v_{k_j}} \right) du_{k_{j_r}} = \frac{e^{v_{k_{j_r}}}}{\sum_{j=j_r}^{n} e^{v_{k_j}}} \exp\left(-e^{-u_{k_{j_{r-1}}}} \sum_{j=j_r}^{n} e^{v_{k_j}} \right),$$

for all $r = 1, \ldots, m-1$.

Theorem 1.

$$p(K) = \prod_{r=1}^{m} \frac{w_{k_{j_r}}}{w_{k_{j_r}} + \cdots + w_{k_{j_n}}}, \quad (14)$$

where $w_k = e^{v_k}$, *for all* $k = 1, \ldots, n$.

Observe that if $m = n$, then K contains only one sequence, $s = (k_1, \ldots, k_n)$. In this case, the expression in Theorem 1 becomes identical to the probability in Equation (6), which is obtained by Beggs et al. (1981) for the rank-ordered logit model.

Theorem 2. *Elimination by aspects is equivalent to a probabilistic lexicographic rule.*

Theorem 2 implies that the weights in an EBA model can be specified using a random utility model for aspects. Aspect j has weight w_j if it has random utility $u_j = \ln(w_j) + \epsilon_j$, where ϵ_j is an independent observation with extreme value distribution, for all $j = 1, \ldots, n$. Note that the logit model assumes independence from irrelevant alternatives. In the present instance, the rank-ordered logit model selects aspects (not alternatives). In other words, aspect choice is independent of irrelevant aspects. This property allows EBA to violate order independence when selecting an alternative from a choice set.

A probabilistic lexicographic rule describes a hierarchical model with n levels and $n!$ branches. A path from the root to a terminal node characterizes a lexicographic sequence of aspects. Each nest is probabilistically constructed. For example, aspect j_1 is selected at the first level of the nesting with probability $p_{j_1} = w_{j_1} / \sum_{r=1}^{n} w_{j_r}$, for all $j_1 = 1, \ldots, n$; then aspect j_2 is selected at the second level of nesting with probability $p_{j_2 \mid j_1} = w_{j_2} / \sum_{r=1,\, r \neq j_1}^{n} w_r$, for all $j_2 \neq j_1$, $j_2 = 1, \ldots, n$. The nesting sequence concludes when each aspect has been considered. Elimination by aspects prunes this tree by eliminating those aspects that, given the preceding levels of the tree, are no longer relevant for choosing an alternative from a choice set. However, this is equivalent to adding the occurrence probabilities for those sequences that lead to the same choice, by the same sequence of eliminations by aspect, in the probabilistic lexicographic rule.

Example 5. We reconsider Example 1 to illustrate the equivalence of EBA and the probabilistic lexicographic rule. Recall that we used four aspects to characterize each of three alternatives: $x_1 = (1\ 0\ 1\ 0)$, $x_2 = (1\ 0\ 0\ 1)$, $x_3 =$

(0 1 0 0), where x_1 and x_2 are the two recordings of a Beethoven symphony, and x_3 is a recording of a Debussy suite.

Table 1 shows the alternative chosen by each sequence using the probabilistic lexicographic rule. We explain these choices below for alternative x_1.

(i) Alternative x_1 is chosen by the set of sequences satisfying the conditions $\sigma(1) < \sigma(j)$, for $j = 2, 3, 4$; and $\sigma(3) < \sigma(4)$—that is, the sequences in which aspect 1 is selected before any other aspect, and aspect 3 is selected before aspect 4. This set of sequences is

$$K_1 = \{(1234), (1324), (1342)\}.$$

From Theorem 1, the probability of choosing x_1 by using a sequence in K_1 is

$$p(K_1) = \frac{w_1}{w_1 + w_2 + w_3 + w_4} \cdot \frac{w_3}{w_3 + w_4},$$

where $w_k = e^{v_k}$ and v_k is the deterministic component of the utility for aspect k. Note that $p(K_1)$ is also the probability of choosing x_1 using an instance of EBA in which aspect 1 is first selected (with probability $w_1/(w_1 + w_2 + w_3 + w_4)$), aspect 2 is ignored (because it appears in neither of the two Beethoven albums), and then aspect 3 is selected (with probability $w_3/(w_3 + w_4)$).

(ii) Alternative x_1 is also chosen by the set of sequences satisfying the conditions $\sigma(3) < \sigma(j)$, for $j = 1, 2, 4$. This set is

$$K_2 = \{(3124), (3142), (3214), (3241), (3412), (3421)\}.$$

From Theorem 1, the probability of choosing x_1 by using a sequence in K_2 is

$$p(K_2) = \frac{w_3}{w_1 + w_2 + w_3 + w_4}.$$

Thus, the probabilistic lexicographic rule selects alternative x_1 with probability

$$\begin{aligned} p(x_1) = p(K_1) + p(K_2) &= \frac{w_1}{w_1 + w_2 + w_3 + w_4} \cdot \frac{w_3}{w_3 + w_4} \\ &\quad + \frac{w_3}{w_1 + w_2 + w_3 + w_4} \\ &= \frac{w_3}{w_1 + w_2 + w_3 + w_4}\left(1 + \frac{w_1}{w_3 + w_4}\right) \\ &= \frac{w_3}{w_3 + w_4}\left(1 - \frac{w_2}{w_1 + w_2 + w_3 + w_4}\right). \end{aligned}$$

Similarly, the probabilistic lexicographic rule selects alternative x_2 with probability

$$\begin{aligned} p(x_2) &= \frac{w_1}{w_1 + w_2 + w_3 + w_4} \cdot \frac{w_4}{w_3 + w_4} \\ &\quad + \frac{w_4}{w_1 + w_2 + w_3 + w_4} \\ &= \frac{w_4}{w_1 + w_2 + w_3 + w_4}\left(1 + \frac{w_1}{w_3 + w_4}\right) \\ &= \frac{w_4}{w_3 + w_4}\left(1 - \frac{w_2}{w_1 + w_2 + w_3 + w_4}\right). \end{aligned}$$

Finally, all six sequences satisfying the conditions $\sigma(2) < \sigma(j)$, for $j = 1, 3, 4$, select alternative x_3. Thus, the choice probability for alternative x_3 is given by

$$p(x_3) = 1 - p(x_1) - p(x_2) = \frac{w_2}{w_1 + w_2 + w_3 + w_4}.$$

Let $w_j = \alpha_j = \ln(v_j)$. Then $p(x_j) = p(x_j, C)$, where $p(x_j, C)$ is the EBA probability of choice for alternative j in Example 1, given by Equations (2)–(4).

4. Probabilistic Utility Function for EBA

As noted in the introduction, Tversky (1972b) showed that EBA is a random utility model, but did not obtain a utility function for the alternatives or a statistical distribution over the utilities. The equivalence between EBA and probabilistic lexicographic rules allows us to do so.

On a choice occasion, a person using a probabilistic lexicographic rule selects some sequence $s \in S$ with probability $p(s)$. Any such sequence specifies a lexicographic preference ordering over the set of alternatives that are available to the person. In turn, each lexicographic preference ordering can be represented by a utility function over the aspects. We provide the details below.

Let $\mathscr{C}$ denote the set of all alternatives. Let $\mathscr{A}$ denote the set of aspects associated with the alternatives in $\mathscr{C}$. Let $\mathscr{S}$ denote the set of all possible aspect orderings of these $|\mathscr{A}|$ aspects.

Let $C \subseteq \mathscr{C}$ denote a subset of the $|\mathscr{C}|$ alternatives, and let $A \subseteq \mathscr{A}$ denote the set of aspects associated with the alternatives in C, where $|A| = n$. As in the preceding sections, let S denote the set of $n!$ orderings of the aspects in A. Observe that unless $A = \mathscr{A}$, S is not a subset of the sequences in $\mathscr{S}$ because each $s \in S$ is defined over n aspects, whereas each sequence in $\mathscr{S}$ is defined over $|\mathscr{A}|$ aspects.

Each sequence $s = (k_1, \ldots, k_n) \in S$ specifies a lexicographic preference ordering over the alternatives in $C \subseteq \mathscr{C}$. Given any sequence $s \in S$, we wish to specify (i) a function that assigns a utility to each alternative in C, and (ii) a probability with which the alternatives in C simultaneously obtain these utility values.

Equation (6) gives the probability $p(s)$ that the probabilistic lexicographic rule uses the sequence $s \in S$. Kohli and Jedidi (2007) showed that the ordering of the alternatives in C, obtained by a lexicographic rule using sequence $s = (k_1, \ldots, k_n)$, can be represented by the utility function

$$u(x_i) = \frac{x_{ik_1}}{2^1} + \frac{x_{ik_2}}{2^2} + \cdots + \frac{x_{ik_n}}{2^n},$$

where $x_i = (x_{i1}, \ldots, x_{in})$ is the profile of alternative $i \in C$, and $x_{ik} = 1(0)$ if alternative i has (does not have) aspect k. Note that each sequence $s \in S$ simultaneously specifies the utilities for all alternatives $i \in C$.

Since EBA does not use aspects that are common to all alternatives in a choice set, any aspect $k \in \mathscr{A} \backslash A$ appears in either all or none of the alternatives in a choice set C.

However, it makes no difference to the ordering of the alternatives in C if we compute their utility values by using the subset of aspects in A or by using all aspects in $\mathscr{A}$. The reason is that if aspect k appears in none of the alternatives in C (that is, if $x_{ik}=0$ for all $i\in C$), then including aspect k in the utility function does not change the utility of any alternative $i\in C$. And if aspect k appears in all of the alternatives in C (that is, if $x_{ik}=1$ for all $i\in C$), then including aspect k in the utility function changes the utility of each alternative $i\in C$ by the same value.

We now discuss the link between the utilities of alternatives and their EBA choice probabilities. Given a set $C\subseteq\mathscr{C}$, let $p(x_i, C)$ denote the probability that a person selects alternative $i\in C$ on a choice occasion. Then

$$p(x_i, C)=\Pr\Big[u(x_i)=\max_{l\in C} u(x_l)\Big].$$

Let

$$K(i)=\Big\{s\in S \mid i=\arg\max_{l\in C} u(x_l)\Big\}$$

denote the subset of sequences in S for which alternative $i\in C$ has the highest utility, $u(x_i)=\max_{l\in C} u(x_l)$. Since an alternative with a higher utility has a higher lexicographic ranking, alternative $i\in C$ is ranked first by each sequence $s\in K(i)$ when the sequence is used by a lexicographic rule to arrange the alternatives in C in decreasing preference order. Let

$$K(i)=K_1(i)\cup K_2(i)\cup\cdots\cup K_h(i),\quad h\geqslant 1,$$

where each $K_t(i)\subseteq K(i)$, $t=1,\ldots,h$, is a subset of sequences that, when used in a lexicographic rule, satisfy the same precedence conditions; that is, $\sigma(k_{j_r})<\sigma(k_j)$, for all $j_r+1<j\leqslant n$ and $r=1,\ldots,m$. Then, by definition, the choice probability for alternative $i\in C$ is

$$p(x_i, C)=\sum_{t=1}^{h} p(K_t(i)),$$

where $p(K_t(i))$ is given by Theorem 1, and the summation runs over the h subsets of sequences that choose the same alternative $i\in C$.

Example 6. We reconsider Example 1. Let $\mathscr{A}=\{1,2,3,4\}$ denote a set of four aspects. Let $\mathscr{S}=\{(1,2,3,4),\ldots,(4,3,2,1)\}$ denote the set of $4!=24$ sequences, which are listed in Table 1. Let $\mathscr{C}=\{1,\ldots,15\}$ denote the set of $2^4-1=15$ alternatives, each of which has one or more aspects. Then $A=\mathscr{A}$, $S=\mathscr{S}$, and $C=\{1,2,3\}\subset\mathscr{C}$, where $x_1=(1\ 0\ 1\ 0)$ and $x_2=(1\ 0\ 0\ 1)$ are the two Beethoven albums, and $x_3=(0\ 1\ 0\ 0)$ is the Debussy album. Consider the sequence $s=(k_1,k_2,k_3)$, which is used in a probabilistic lexicographic rule with the probability

$$p(s)=\frac{\alpha_{k_1}}{\alpha_{k_1}+\alpha_{k_2}+\alpha_{k_3}+\alpha_{k_4}}\cdot\frac{\alpha_{k_2}}{\alpha_{k_2}+\alpha_{k_3}+\alpha_{k_4}}\cdot\frac{\alpha_{k_3}}{\alpha_{k_3}+\alpha_{k_4}}.$$

Let

$$u(x_i)=\frac{x_{ik_1}}{2}+\frac{x_{ik_2}}{2^2}+\frac{x_{ik_3}}{2^3}+\frac{x_{ik_4}}{2^4}$$

denote the utility of alternative $i\in C$. Then the vector $u=(u(x_1),u(x_2),u(x_3))$ is obtained with probability $p(s)$, for all $s\in S$. For example, if $s=(1,2,3,4)$, then

$$u(x_1)=\frac{1}{2}+\frac{0}{2^2}+\frac{1}{2^3}+\frac{0}{2^4}=\frac{10}{16},$$
$$u(x_2)=\frac{1}{2}+\frac{0}{2^2}+\frac{0}{2^3}+\frac{1}{2^4}=\frac{9}{16},$$
$$u(x_3)=\frac{0}{2}+\frac{1}{2^2}+\frac{0}{2^3}+\frac{0}{2^4}=\frac{4}{16}.$$

We note that $u(x_1)>u(x_2)>u(x_3)$, which is consistent with the preference ordering $1,2,3$ associated with the alternatives by a lexicographic rule using the aspect ordering $s=(1,2,3,4)$. We say that the alternatives in C have the utilities $(u(x_1),u(x_2),u(x_3))=(10/16,9/16,4/16)$ with probability

$$p(s)=\frac{\alpha_1}{\alpha_1+\alpha_2+\alpha_3+\alpha_4}\cdot\frac{\alpha_2}{\alpha_2+\alpha_3+\alpha_4}\cdot\frac{\alpha_3}{\alpha_3+\alpha_4},\quad\text{for } s=(1,2,3,4).$$

Similarly, if $s=(3,2,4,1)$, then

$$u(x_1)=\frac{1}{2}+\frac{0}{2^2}+\frac{0}{2^3}+\frac{1}{2^4}=\frac{9}{16},$$
$$u(x_2)=\frac{0}{2}+\frac{0}{2^2}+\frac{1}{2^3}+\frac{1}{2^4}=\frac{3}{16},$$
$$u(x_3)=\frac{0}{2}+\frac{1}{2^2}+\frac{0}{2^3}+\frac{0}{2^4}=\frac{4}{16}.$$

We note that $u(x_1)>u(x_3)>u(x_2)$, which is consistent with the preference ordering $1,3,2$ of the alternatives obtained by a lexicographic rule using the sequence $s=(3,2,4,1)$. We say that the alternatives in C have the utilities $(u(x_1),u(x_2),u(x_3))=(9/16,3/16,4/16)$ with probability

$$p(s)=\frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3+\alpha_4}\cdot\frac{\alpha_2}{\alpha_1+\alpha_2+\alpha_4}\cdot\frac{\alpha_4}{\alpha_1+\alpha_4},\quad\text{for } s=(3,2,4,1).$$

Similar probabilities can be associated with the utility vectors for alternatives in any choice set $C\subseteq\mathscr{C}$.

Table 1 shows the partitions (subsets of sequences) that choose the same alternative from the choice set $C=\{1,2,3\}$. For example, alternative 1 is selected by the first and the fourth partitions shown in the table. Example 5 gives the choice probabilities for each alternative.

5. Implications for Model Estimation

The error theory we developed in this paper implies that EBA is a member of the logit family of models. Suppose we choose to represent each alternative by a single, unique aspect. Then each EBA instance stops after $m = 1$ step. The EBA model becomes equivalent to the multinomial logit model. Aspect utilities become indistinguishable from the utilities of alternatives, and the probability of choosing an aspect is equal to the probability of choosing an alternative. Now suppose an EBA instance eliminates a single alternative at each step. Then it terminates in $m = n$ steps, and its occurrence probability is given by the rank-ordered logit model. If $1 < m < n$, the probability associated with an EBA instance is a product of m terms, the jth term corresponding to the multinomial logit probability of choosing aspect $k_j \in A_j$ at step j.

The choice probability for an alternative is obtained by adding the probabilities of EBA instances that result in its choice (see Equations (2)–(4) in Example 1). Since the probability of an EBA instance is a function of the aspect choice probabilities, the latter can be used to formulate a likelihood function, which can be maximized to estimate the parameters of an EBA model. As errors are associated with aspects, heterogeneity and covariates can be introduced at the aspect level.

Heterogeneity in EBA rules can be specified by allowing the deterministic aspect utilities, v_k, to vary across individuals or segments. For example, a random effects model can be formulated in which each individual i is considered to associate utility $v_{ik} = \bar{v}_k + \gamma' z_i + e_i$ with aspect k, where $\bar{v}_k$ is an intercept term, z_i is a vector of individual-level covariates (e.g., demographic variables), γ' is the associated vector of parameters, and e_i has a Normal distribution across individuals, with mean zero and variance τ_k^2. This formulation allows both observed and unobserved heterogeneity in the aspect utilities. A latent class formulation can also be obtained by allowing each v_{ik} value to be a draw from a mixture of Normal distributions $N(v_{gk}, \tau_{gk}^2)$, where $g = 1, \ldots, G$ is an index denoting segments.

Since aspects have utilities, their deterministic components can be further decomposed in the same way as the utilities of alternatives are decomposed in a logit model. For example, such a decomposition may be useful for an aspect like the affordability of a durable good, which may be a composite of its price, maintenance cost, and expected life; or for an aspect like the convenience of a travel mode, which may be a composite of ease of access, waiting time and travel time. This reparametrization allows the modeling of aspect utilities as functions of other variables that are common to subsets of aspects. Consider an aspect with utility $u_k = \bar{v}_k + \epsilon_k$, where ϵ_k has the extreme value distribution. Suppose that aspect k is composed of other features; that is,

$$\bar{v}_k = \beta_{k0} + \beta_1 y_{k1} + \cdots + \beta_t y_{kt},$$

where y_{kj} are the values of the variables $y_1, \ldots, y_t$ associated with aspect k. The standard EBA model corresponds to $\bar{v}_k = \beta_{k0}$ for all $k \in A$.

6. Conclusion

Elimination by aspects is a model of bounded rationality that is considered to represent consumer choice more accurately than models assuming utility maximizing consumers. One limitation of the model has been the lack of an error theory. We obtain an error theory and show that EBA can be derived by assuming that aspects have random utilities with independent, extreme value distributions. The result implies that EBA is consistent with the assumption that consumers maximize random utility when ordering the aspects they use to eliminate alternatives from a choice set. EBA generalizes the multinomial logit and rank-ordered logit models. Maximum likelihood methods currently available for logit models can be modified to estimate EBA parameters. The model can be extended to allow aspect utilities to be functions of covariates and to capture consumer heterogeneity using latent-class or random-effects approaches.

The proposed error structure can also be associated with aspects in a special case of EBA called preference trees (Tversky and Sattah 1979). Since the latter have the same hierarchical structure as nested logit models, it can be useful to compare the empirical performance of the two models. Conceptually, the proposed error structure implies that preference trees do not require the assumption of a generalized extreme value distribution, which is needed for nested logit models. Batley and Daly (2006) showed that there is a mapping between the parameters of an EBA model and a nested logit model when there are three alternatives and two aspects. The problem they considered is equivalent to the problem of choosing between Beethoven and Debussy record albums discussed in the present paper. It may be useful to examine if the proposed error theory can be useful for establishing a more general correspondence between nested logit models and preference trees.

EBA uses the Luce choice rule for selecting aspects at each stage of elimination. Since the Luce choice rule assumes IIA, EBA assumes that aspect selection is independent of irrelevant aspects. It may be useful to examine how the EBA model changes if the Luce choice rule is replaced by another model for aspect selection. This can be achieved by considering alternative statistical distributions for aspect utilities. These distributions can be used to derive the choice probabilities of alternatives using a probabilistic lexicographic rule, and thus for an equivalent EBA model.

Acknowledgments

The authors are grateful to Omar Besbes and Khaled Boughanmi for their help in proving Theorem 1.

Appendix. Proofs of Lemmas and Theorems

Lemma 1. *Let sequence $s=(k_1,\dots,k_n)$ select alternative $i\in C$ using m steps, where $1\leqslant m<n$. Let $t_j=1$ for $j=j_1,\dots,j_m$ and $t_j=0$ otherwise, where $0<j_1\cdots<j_m$. Then sequence s', which assigns position $\sigma(k_j)$ to aspect k_j, also selects alternative $i\in C$ using m steps, eliminating alternatives in the same order as sequence s, if $\sigma(k_{j_r})<\sigma(k_j)$, for all $j_r+1\leqslant j\leqslant n$ and $1\leqslant r\leqslant m$.*

Proof. Consider sequence s. Let $t_{j_1}=\cdots=t_{j_m}=1$, where $1\leqslant j_1<\cdots<j_m<n$; and $t_j=0$, for all $j_r<j<j_{r+1}$ and $r=1,\dots,m$ (define $j_{m+1}=n+1$). Then $C_j=C_{j_r}$, for all $j_r<j<j_{r+1}$, because $t_j=0$ means that aspect k_j is not used to eliminate alternatives. Since $C_j\subseteq C_{j-1}$ for all $j\geqslant 1$, aspect k_j cannot appear in alternatives in the subsets $C_{j_{r+1}},\dots,C_{j_m}$ that are associated with sequence s.

Now consider sequence s'. If $j_1>1$, then aspects $k_1,\dots,k_{j_1-1}$ appear in none or all of the alternatives in C and therefore are not used to eliminate alternatives. Thus, these aspects can appear at any position in s' without affecting the order in which s' eliminates alternatives from C. Similarly, if $j_2>j_1+1$, then aspects $k_{j_1+1},\dots,k_{j_2-1}$ appear in none or all of the alternatives in C_1 and therefore are not used to eliminate alternatives. Thus, these aspects can appear at any position in s' after aspect k_{j_1} without affecting the order in which s' eliminates alternatives from C. More generally, since $\sigma(k_{j_1})<\cdots<\sigma(k_{j_m})$, assigning position $\sigma(k_j)>\sigma(k_{j_r})$ to aspect k_j when $j_r<j<j_{r+1}$, for all $r=1,\dots,m$, has no effect on the sequence in which s' eliminates alternatives from C. It follows that sequence s' also eliminates alternatives from C in the same order as sequence s. □

Lemma 2.

$$I_r=\int_{u_{k_{j_r}}=-\infty}^{u_{k_{j_{r-1}}}} G_r\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_{r+1}}^{n}e^{v_{k_j}}\Big)\,du_{k_{j_r}}
=\frac{e^{v_{k_{j_r}}}}{\sum_{j=j_r}^{n}e^{v_{k_j}}}\exp\Big(-e^{-u_{k_{j_{r-1}}}}\sum_{j=j_r}^{n}e^{v_{k_j}}\Big),$$

for all $r=1,\dots,m-1$.

Proof. By definition,

$$I_r=\int_{u_{k_{j_r}}=-\infty}^{u_{k_{j_{r-1}}}} f(u_{k_{j_r}})\prod_{j=j_r+1}^{j_{r+1}-1}F_{k_j}(u_{k_{j_r}})\cdot\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_{r+1}}^{n}e^{v_{k_j}}\Big)\,du_{k_{j_r}}.$$

Substituting the expressions for $f_k(\cdot)$ and $F_k(\cdot)$ gives

$$\begin{aligned}I_r&=\int_{u_{k_{j_r}}=-\infty}^{u_{k_{j_{r-1}}}}\exp(-e^{-u_{k_{j_r}}}e^{v_{k_{j_r}}})e^{-u_{k_{j_r}}}e^{v_{k_{j_r}}}\\&\quad\times\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_r+1}^{j_{r+1}-1}e^{v_{k_j}}\Big)\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_{r+1}}^{n}e^{v_{k_j}}\Big)\,du_{k_{j_r}}\\&=e^{v_{k_{j_r}}}\int_{u_{k_{j_r}}=-\infty}^{u_{k_{j_{r-1}}}}\exp(-e^{-u_{k_{j_r}}}e^{v_{k_{j_r}}})e^{-u_{k_{j_r}}}\\&\quad\times\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_r+1}^{j_{r+1}-1}e^{v_{k_j}}\Big)\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_{r+1}}^{n}e^{v_{k_j}}\Big)\,du_{k_{j_r}}.\end{aligned}$$

Assembling terms and multiplying and dividing by $\sum_{j=j_r}^{n}e^{v_{k_j}}$ gives

$$\begin{aligned}I_r&=\frac{e^{v_{k_{j_r}}}}{\sum_{j=j_r}^{n}e^{v_{k_j}}}\int_{u_{k_{j_r}}=-\infty}^{u_{k_{j_{r-1}}}}\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_r}^{n}e^{v_{k_j}}\Big)e^{-u_{k_{j_r}}}\cdot\sum_{j=j_r}^{n}e^{v_{k_j}}\,du_{k_{j_r}}\\&=\frac{e^{v_{k_{j_r}}}}{\sum_{j=j_r}^{n}e^{v_{k_j}}}\Big[\exp\Big(-e^{-u_{k_{j_r}}}\sum_{j=j_r}^{n}e^{v_{k_j}}\Big)\Big]_{-\infty}^{u_{k_{j_r}}=u_{k_{j_{r-1}}}}\\&=\frac{e^{v_{k_{j_r}}}}{\sum_{j=j_r}^{n}e^{v_{k_j}}}\exp\Big(-e^{-u_{k_{j_{r-1}}}}\sum_{j=j_r}^{n}e^{v_{k_j}}\Big).\quad\square\end{aligned}$$

Theorem 1.

$$p(K)=\prod_{r=1}^{m}\frac{w_{k_{j_r}}}{w_{k_{j_r}}+\cdots+w_{k_{j_n}}},$$

where $w_k=e^{v_k}$, for all $k=1,\dots,n$.

Proof. We can write the expression for $p(K)$ as

$$\begin{aligned}p(K)&=p(u_{k_{j_r}}\geqslant u_{k_j},\ \text{for all } j_r+1\leqslant j\leqslant n \text{ and } r=1,\dots,m)\\&=p(u_{k_{j_r}}\geqslant u_{k_j},\ \text{for all } j_r+1\leqslant j\leqslant j_{r+1} \text{ and } r=1,\dots,m)\\&=\int_{u_{k_{j_1}}=-\infty}^{u_{k_{j_0}}}G_1\Big(\int_{u_{k_{j_2}}=-\infty}^{u_{k_{j_1}}}G_2\Big(\int_{u_{k_{j_3}}=-\infty}^{u_{k_{j_2}}}G_3\\&\qquad\cdots\Big(\int_{u_{k_{j_{m-1}}}=-\infty}^{u_{k_{j_{m-2}}}}G_{m-1}\Big(\int_{u_{k_{j_m}}=-\infty}^{u_{k_{j_{m-1}}}}G_m du_{k_{j_m}}\Big)du_{k_{j_{m-1}}}\Big)\\&\qquad\cdots\Big)du_{k_{j_2}}\Big)\,du_{k_{j_1}}\end{aligned}$$

First substitute[4]

$$\int_{u_{k_{j_m}}=-\infty}^{u_{k_{j_{m-1}}}}G_m du_{k_{j_m}}=\frac{e^{v_{k_{j_m}}}}{\sum_{j=j_m}^{n}e^{v_{k_j}}}\exp\Big(-e^{-u_{k_{j_{m-1}}}}\sum_{j=j_m}^{n}e^{v_{k_j}}\Big)$$

in the preceding expression for $p(K)$ to obtain

$$\begin{aligned}p(K)&=\frac{e^{v_{k_{j_m}}}}{\sum_{j=j_m}^{n}e^{v_{k_j}}}\\&\quad\cdot\int_{u_{k_{j_1}}=-\infty}^{u_{k_{j_0}}}G_1\Big(\int_{u_{k_{j_2}}=-\infty}^{u_{k_{j_1}}}G_2\Big(\int_{u_{k_{j_3}}=-\infty}^{u_{k_{j_2}}}G_3\\&\quad\cdots\Big(\int_{u_{k_{j_{m-1}}}=-\infty}^{u_{k_{j_{m-2}}}}G_{m-1}\exp\Big(-e^{-u_{k_{j_{m-1}}}}\sum_{j=j_m}^{n}e^{v_{k_j}}\Big)\Big)du_{k_{j_{m-1}}}\\&\quad\cdots\Big)du_{k_{j_2}}\Big)\,du_{k_{j_1}}.\end{aligned}$$

Next, use Lemma 2 to make a sequence of $m-1$ substitutions in the expression for $p(K)$. The substitutions begin with $r=m-1$ and end with $r=1$. Thus,

$$\begin{aligned}I_{m-1}&=\int_{u_{k_{j_{m-1}}}=-\infty}^{u_{k_{j_{m-2}}}}G_{m-1}\exp\Big(-e^{-u_{k_{j_{m-1}}}}\sum_{j=m}^{n}e^{v_{k_j}}\Big)\,du_{k_{j_{m-1}}}\\&=\frac{e^{v_{k_{j_{m-1}}}}}{\sum_{j=j_{m-1}}^{n}e^{v_{k_j}}}\cdot\exp\Big(-e^{-u_{k_{j_{m-2}}}}\sum_{j=j_{m-1}}^{n}e^{v_{k_j}}\Big).\end{aligned}$$

Making this substitution in the preceding expression for $p(K)$ gives

$$p(K) = \frac{e^{v_{k_{j_{m-1}}}}}{\sum_{j=j_{m-1}}^{n} e^{v_{k_j}}} \cdot \frac{e^{v_{k_{j_m}}}}{\sum_{j=j_m}^{n} e^{v_{k_j}}} \cdot \int_{u_{k_{j_1}}=-\infty}^{u_{k_{j_0}}} G_1 \left(\int_{u_{k_{j_2}}=-\infty}^{u_{k_{j_1}}} G_2 \left(\int_{u_{k_{j_3}}=-\infty}^{u_{k_{j_2}}} G_3 \cdots \left(\int_{u_{k_{j_{m-2}}}=-\infty}^{u_{k_{j_{m-3}}}} G_{m-2} \cdot \exp\left(-e^{-u_{k_{j_{m-2}}}} \sum_{j=j_{m-1}}^{n} e^{v_{k_j}} \right) du_{k_{j_{m-2}}} \right) \cdots \right) du_{k_{j_2}} \right) du_{k_{j_1}}.$$

Recursive substitutions using Lemma 2 for $r = m-2, \ldots, 1$, in the expression for $p(K)$ yield

$$p(K) = \prod_{r=1}^{m} \frac{e^{v_{k_{j_r}}}}{\sum_{j=j_r}^{n} e^{v_{k_j}}} = \prod_{r=1}^{m} \frac{w_{k_{j_r}}}{w_{k_{j_r}} + \cdots + w_{k_{j_n}}},$$

where $w_k = e^{v_k}$, for all $k = 1, \ldots, n$. □

THEOREM 2. *Elimination by aspects is equivalent to a probabilistic lexicographic rule.*

PROOF. Consider an instance of EBA that selects aspect $k_j \in A_j$. By definition,

$$\pi(k_j) < \pi(k), \quad \text{for all } k \in A_j \backslash \{k_j\} \text{ and } j = 1, \ldots, m;$$

and

$$q = \prod_{j=1}^{m} q(k_j, A_j) = \prod_{j=1}^{m} \frac{\alpha_{k_j}}{\sum_{k \in A_j} \alpha_k}.$$

Let K denote the set of sequences that satisfy

$$\sigma(k_j) < \sigma(k), \quad \text{if } \pi(k_j) < \pi(k), \text{ for all } k \in A_k \backslash \{k_j\}.$$

From Lemma 1, each sequence $s \in K$ eliminates alternatives in the same order and chooses the same alternative $i \in C$, in m steps, where $1 \leqslant m \leqslant n$. Theorem 1 implies that

$$p(K) = \prod_{j=1}^{m} \frac{w_{k_j}}{\sum_{k \in A_j} w_k}.$$

It follows that $q = p(K)$ if $\alpha_k = w_k$, for all $k \in A$.

Conversely, suppose $K \subseteq S$ is a subset of sequences that satisfies

$$\sigma(k_{j_r}) < \sigma(k_j) \quad \text{for all } j_r + 1 \leqslant j \leqslant n \text{ and } r = 1, \ldots m.$$

From Theorem 1,

$$p(K) = \prod_{j=1}^{m} \frac{w_{k_j}}{w_{k_r} + \cdots + w_{k_n}}.$$

Consider

$$A_r = \{k_{j_r}, k_{j_r+1}, \ldots, k_n\}, \quad \text{for all } r = 1, \ldots, m.$$

Then each sequence $s \in K$ eliminates alternatives in the same order, and select the same alternative $i \in C$, as an instance of EBA that selects aspect $k_{j_r} \in A_r$ at step r, for all $r = 1, \ldots, m$. Thus, there is an instance of EBA that selects alternative $i \in A$ with probability

$$q = \prod_{j=1}^{m} \frac{\alpha_{k_j}}{\sum_{k \in A_j} \alpha_k}.$$

It follows that $p(K) = q$ if $w_k = \alpha_k$, for all $k \in A$. □

Endnotes

1. Order independence is equivalent to simple scalability, which requires the choice probabilities of alternatives to be monotone functions of their scale values.
2. Note that a fixed aspect ordering does not mean that the probability of selecting an aspect is constant across stages. Instead, it depends upon the relative importances of the aspects that are available at any given stage of EBA.
3. The same example is sometimes described as the red-bus/blue-bus problem, in which a person chooses between a car (Debussy suite) and two equally good buses (Beethoven symphony), one with red color (album 1) and the other with blue color (album 2).
4. The evaluation of this integral follows the same steps as the evaluation of I_r in Lemma 2.

References

Batley R, Daly A (2006) On the equivalence between elimination-by-aspects and generalized extreme value models of choice behavior. *J. Math. Psych.* 50(5):456–467.

Batsell RR, Polking JC, Cramer R, Miller C (2003) Useful mathematical relationships embedded in Tversky's elimination by aspects model. *J. Math. Psych.* 47(5–6):538–544.

Becker GM, DeGroot MH, Marschak J (1963) Stochastic models of choice behavior. *Behavioral Sci.* 8(1):41–55.

Beggs S, Cardell S, Hausman J (1981) Assessing the potential demand for electric cars. *J. Econometrics* 17(1):1–19.

Block HD, Marschak J (1960) Random orderings and stochastic theories of responses. Olkin I, Ghurye S, Hoeffding W, Madow W, Mann H, eds. *Contributions to Probability and Statistics* (Stanford University Press, Stanford, CA), 97–132.

Chapman RG, Staelin R (1982) Exploiting rank ordered choice set data within the stochastic utility model. *J. Marketing Res.* 19(3):288–301.

Chipman JS (1960) Stochastic choice and subjective probability. Willner D, ed. *Decisions, Values, and Groups*, Vol. 1 (Pergamon Press, New York), 70–95.

Coombs CH (1958) On the use of inconsistency of preferences in psychological measurement. *J. Experiment. Psych.* 55(1):1–7.

Corbin R, Marley AAJ (1974) Random utility models with equality: An apparent, but not actual, generalization of random utility models. *J. Math. Psych.* 11(3):274–293.

Debreu G (1960) Review of R. D. Luce, individual choice behavior: A theoretical analysis. *Amer. Econom. Rev.* 50:186–188.

Fader P, McAlister L (1990) An elimination by aspects model of consumer response to promotion calibrated on UPC scanner data. *J. Marketing Res.* 27(3):322–332.

Gilbride T, Allenby G (2006) Estimating heterogeneous EBA and economic screening rule choice models. *Marketing Sci.* 25(5):494–509.

Görür D, Jäkel F, Rasmussen CE (2006) A choice model with infinitely many latent features. *Proc. 23rd Internat. Conf. Machine Learn. Pittsburgh, PA* (ACM, New York), 361–368.

Kohli R, Jedidi K (2007) Representation and inference of lexicographic preference models and their variants. *Marketing Sci.* 26(3):380–399.

Krantz DH (1967) Rational distance function for multidimensional scaling. *J. Math. Psych.* 4(2):226–245.

Manrai AK, Sinha PK (1989) Elimination-by-cutoffs. *Marketing Sci.* 8(2):133–152.

McFadden D (1981) Econometric models of probabilistic choice. Manski C, McFadden D, eds. *Structural Analysis of Discrete Data with Econometric Applications* (MIT Press, Cambridge, MA), 198–272.

Park SJ, Choi S (2013) A theoretical note on the number of free parameters in the elimination-by-aspects model. *J. Math. Psych.* 57(5): 255–259.

Pihlens DA (2008) The multi-attribute eliminations by aspects (MEBA) model. Doctoral dissertation, University of Technology Sydney, Sydney, Australia.

Tversky A (1972a) Elimination by aspects: A theory of choice. *Psych. Rev.* 79(4):281–299.
Tversky A (1972b) Choice by elimination. *J. Math. Psych.* 9(4):341–367.
Tversky A, Russo JE (1969) Similarity and substitutability in binary choices. *J. Math. Psych.* 6:1–12.
Tversky A, Sattah S (1979) Preference trees. *Psych. Rev.* 86(6):542–573.

Rajeev Kohli is the Ira Leon Rennert Professor of Business, Columbia University. His research interests include choice models, new product development, discrete optimization, and analysis of algorithms. His previous research in the *Journal of Marketing Research* was the finalist for the Paul Green Award in 2008 and the O'Dell Award in 2000.

Kamel Jedidi is the John A. Howard Professor of Business, Columbia University. His research interests include pricing, diffusion of innovations, and the long-term impact of advertising and promotions. He received Best Article Awards from *International Journal of Research in Marketing* in 1998 and the Marketing Science Institute in 2000 and was finalist for 2009 Paul Green Award for the *Journal of Marketing Research* and for the 2009 Long-term Impact Paper Award for *Marketing/Management Science.*