Article:

Batley, R.P. and Daly, A.J. (2006) On the equivalence between elimination-by-aspects and generalised extreme value models of choice behaviour. Journal of Mathematical Psychology, 50 (5). pp. 456-467. ISSN 0022-2496

https://doi.org/10.1016/j.jmp.2006.05.003

Published paper
Batley, R.P.; Daly, A.J. (2006) On the equivalence between elimination-by-aspects and generalised extreme value models of choice behaviour. Journal of Mathematical Psychology, 50(5), pp.456-467.

ON THE EQUIVALENCE BETWEEN ELIMINATION-BY-ASPECTS AND GENERALISED EXTREME VALUE MODELS OF CHOICE BEHAVIOUR

Richard Batley
Institute for Transport Studies, University of Leeds

Andrew Daly
Institute for Transport Studies, University of Leeds and RAND Europe

Author for correspondence: Dr Richard Batley
Institute for Transport Studies
University of Leeds
Leeds
LS2 9JT
UK

Telephone: +44 (0) 113 343 6607
Fax: +44 (0) 113 343 5334
E-mail: rbatley@its.leeds.ac.uk

ABSTRACT

Elimination-by-aspects and generalised extreme value offer competing paradigms for the representation of a common behaviour, that of individual discrete choice. Observing certain consistencies in their mathematical structure, several eminent authors have commented on the degree of equivalence between the two paradigms. Most contributions to this debate have, however, been less than definitive. More fundamentally, the contributions lack consensus. We advance the debate considerably by establishing formal mathematical conditions under which three-alternative tree models from the two paradigms are exactly equivalent. We then extend our analysis to consider more general models, showing that equivalence can be established for general tree models, but not for cross-nested models.

Keywords: elimination-by-aspects, hierarchical elimination-by-aspects, random utility model, generalised extreme value, nested logit, recursive nested extreme value

1. INTRODUCTION

This paper considers the relationship - in particular, the existence of a formal mathematical equivalence - between two alternative paradigms for the representation of individual discrete choice. An understanding of individual discrete choice is pertinent to many of the behavioural sciences, although we restrict attention here to the disciplinary interests of microeconomics and mathematical psychology. Despite sharing common origins in the literature of psychophysics, the paradigms that now prevail in the two disciplines are well-developed and distinct. That they differ, in fundamental respects, should perhaps come as no surprise. Microeconomics and mathematical psychology may well have interest in the same behavioural observations, but they do so with substantive differences in their scientific philosophy and analytical focus.

In conducting economic analysis of individual discrete choice, convention is to apply the Random Utility Model (RUM), first proposed by Marschak (1960) and Block and Marschak (1960). RUM offers an apparatus for attending to one (or both) of the principal needs of the working economist. First, to elicit the monetary value of factors that influence individual choice. Second, to forecast behaviour - in terms of the aggregation of discrete choices across a population of individuals - in scenarios that differ from those observed. In this paper we emphasise the latter. Finally, for purposes of implementation, it is useful to equip RUM with additional theory. A popular such implementation is the Generalised Extreme Value (GEV) model of McFadden (1978, 1981); this will be adopted in what follows.

The focus of mathematical psychology, in contrast, is more on understanding the cognitive processes underlying individual behaviour. A consequence, perhaps inevitably, is that consensus around a generic paradigm is less evident than in economics; rather there exists a diverse range of paradigms specific to context. This paper focuses on one such context, that of cognitive processing in 'complex' discrete choice tasks. Complexity has attracted significant attention in the behavioural science literature, wherein Herbert Simon's contributions (e.g. 1955, 1959, 1989, 1990) have achieved seminal status. For our chosen context, we focus further on one (albeit well known) model of discrete choice from mathematical psychology, namely Elimination-By-Aspects (EBA). EBA, which was proposed by Tversky (1972a, 1972b), offers a simplifying heuristic for complex tasks, conceptualising discrete choice as a process of sequentially eliminating alternatives according to criteria of acceptability, until a single 'chosen' alternative remains.

Although they differ fundamentally in their paradigmatic basis, both EBA and GEV adopt a probabilistic notion of choice, and represent the choice set as a 'preference tree', with subsets of similar alternatives nested together. In this way, both models relax the property of independence from irrelevant alternatives (IIA), which would require that for any two alternatives, the ratio of their choice probabilities is unaffected by the presence or absence of any other alternatives in the choice set (Debreu, 1960). This has prompted several eminent authors - both in economics and in mathematical psychology - to make observations regarding the degree of 'equivalence' between EBA and GEV (Tversky and Sattath, 1979; McFadden, 1981; Small, 1987; Vovsha, 1997; Train, 2003). Since these observations show significant inconsistencies, however, it would appear that some, if not all, must be incorrect. That notwithstanding, a characteristic shared by most of these works is an absence of clear reasoning behind their respective assertions.

Recent work by Daly and Bierlaire (2005) has introduced a more general GEV model, known as recursive nested EV (RNEV) or network EV. This gives the potential for greater equivalence between EBA and the GEV family. The contribution of our paper is to establish the conditions under which the probability statements of EBA and RNEV are equivalent, thereby offering a more definitive result than previous researchers. The importance of this work is that our interpretation of behaviour will differ under each of the two paradigms. It is necessary, therefore, to design experiments that reveal which, if either, of the paradigms offers the correct representation of behaviour. If, however, the models developed from the two paradigms yield indistinguishable predictions of behaviour under specific circumstances, then we cannot expect experiments conducted under those circumstances to inform their discrimination. Discrimination must, in this case, be left to other means of experiment; in particular, means that are not reliant on the probability statement.

2. STOCHASTIC MODEL OF DISCRETE CHOICE

Our interest in choice is restricted to the particular context of discrete choice by an individual, whereby the number of alternatives in the choice set is finite, alternatives are mutually exclusive, and the choice set is exhaustive. Further, we restrict attention to choice under certain outcomes. While this is a considerable restriction, choice with certainty is at least the first step to developing a model of choice with uncertainty.

This paper does not seek to offer a comprehensive account of theories of discrete choice; such needs are attended to elsewhere in the literature (e.g. Marschak *et al*., 1963; Luce *et al*., 1965; Restle and Greeno, 1970; McFadden, 1981; Suppes *et al*., 1989). Rather we employ as our starting point a general, and uncontroversial, form of discrete choice model defined by Marschak *et al.* (1963) as the 'stochastic model'. Formally:

Let T be the universal set of discrete choice alternatives, and let $S = \{1,...,J\}$ be a subset $S \subseteq T$ that constitutes the offered or 'choice' set of J alternatives. A stochastic model specifies, for each choice set S and each alternative $i \in S$, the probability $P(i|S)$ that an individual will choose i from S. Under any such model, it must be true that $0 \leq P(i|S) \leq 1$ and $\sum_{i \in S} P(i|S) = 1$ for all i and all S.

We make three observations in relation to the stochastic model. First, it is based on the concept of a single individual making independent repeated choices, i.e. choice probability can be conceptually based on the relative frequency of choices by an individual in repeated experiments. Second, the source of randomness is unspecified; it may be associated with incomplete information on the part of the analyst about the preferences of the consumer, or with intrinsically random behaviour, or with both. Third, randomness permits apparent 'inconsistencies' in choice, which might imply some degree of variation in the underlying preferences of the individual.

The following sections consider two model forms - GEV and EBA - that can be considered special cases of the stochastic model.

3. GENERALISED EXTREME VALUE

RUM was first proposed in Marschak (1960) and Block and Marschak (1960) as a development of Thurstone's (1927) Law of Comparative Judgement. Perhaps unsurprisingly, given these origins, RUM takes a nod towards psychophysical theory, in the sense that the model accommodates the possibility that behaviour may not always be 'consistent'. In economic parlance, this means that the repeated choices of an individual may imply variability in the underlying preference ordering. Moreover, RUM can be considered a stochastic analogue of the more conventional microeconomic analysis of individual choice (e.g. Debreu, 1954). Marschak *et al.* (1963) provided the following definition of RUM:

A stochastic model is said to be a random utility model if there exists a random vector $(U_1,...,U_J)$ such that, for each choice set S and all $i \in S$:

$$P(i|S) = \Pr\{U_i \geq U_j \quad \text{for all} \quad j \in S\}.$$

A consequence of this definition [they also note] is that under any random utility model:

$$\Pr\{U_i = U_j\} = 0 \quad \text{for all} \quad j \neq i.$$

To interpret, for any choice set, an individual can rank alternatives in terms of their utility; this ranking is however random and may change between replications.

In implementing RUM, microeconomics has shown heavy reliance on a subset of RUM derived from McFadden's (1978, 1981) GEV model. Prompted by the interest of economists in aggregate behaviour, McFadden re-interpreted RUM as representative, not of an individual engaged in repeated independent discrete choice tasks, but of a population of decision-makers with explicitly varying tastes, each facing a single discrete choice task. McFadden (1981) argued that the two interpretations are formally equivalent, i.e. that the probability of a choice made by a specific individual in (conceptually) repeated experiments also applies to the probability of that choice made by a randomly chosen individual in an extensive population.

Let $U_i = V_i + \varepsilon_i$, where V_i is deterministic utility, and ε_i is a random variable. Following McFadden (1978), GEV is based on a function $G = G(Y_1,...,Y_J)$, where $Y_i = \exp(V_i)$. Let $G_i = \partial G/\partial Y_i$. If G satisfies a series of properties then a GEV model is given by:

$$P(i) = \frac{Y_i G_i}{G} \tag{1}$$

This model has a mean utility given by $V = \log G + \gamma$ where γ is Euler's constant. The properties are as follows:

(a) $G \geq 0$ for all positive values of Y_i, for all $i \in S$.

(b) G is homogenous of degree one.

(c) $G \to \infty$ as $Y_i \to \infty$ for any i.

(d) The cross partial derivatives of G switch sign in a particular way, specifically: $G_i \geq 0$ for all i, $G_{ij} = \partial G_i / \partial Y_j \leq 0$ for all $j \neq i$, $G_{ijk} = \partial G_{ij} / \partial Y_k \geq 0$ for any distinct i, j and k, and so on for higher order cross partial derivatives.

A few comments are appropriate. First, an oversight is apparent in McFadden's property (a); with reference to (1) and the mean utility result, it is clear that G must be positive. Second, with reference to property (b), G may actually be homogenous of degree $\mu > 0$ (Ben-Akiva and Lerman, 1985). Our third comment concerns ease of implementation, since while properties (a), (b) and (c) are reasonably straightforward to verify, property (d) is not (see Daly and Bierlaire, 2005).

3.1 Tree models

GEV models with particular properties can be derived through appropriate specification of the G function. A popular form is tree or nested logit (NL), which groups related alternatives in mutually exclusive 'nests' across a series of 'levels'. Consider the derivation of the simplest possible form of NL, consisting of only two levels, from GEV. Specify a G function:

$$G = \sum_{m=1}^{M} \left(\sum_{h \in R_m} \exp {V_h}/{\mu_m} \right)^{\mu_m}$$

where the choice set $S = \{1,..., J\}$ is partitioned into M non-overlapping subsets or nests $R_m \subseteq S, m = 1,..., M$. The probability of choosing alternative i from S is given by:

$$P(i|S) = P(R_m) \times P(i|R_m)$$

where the marginal probability of choosing nest R_m is given by:

$$P(R_m) = \frac{\left(\sum_{h \in R_m} \exp {V_h}/{\mu_m} \right)^{\mu_m}}{\sum_{m=1}^{M} \left(\sum_{h \in R_m} \exp {V_h}/{\mu_m} \right)^{\mu_m}}$$

and the conditional probability of choosing alternative *i* from nest R_m is given by:

$$P(i|R_m) = \frac{\exp^{V_i/\mu_m}}{\sum_{h \in R_m} \exp^{V_h/\mu_m}}$$

where μ_m is the dissimilarity parameter relating to nest R_m. Consistency with RUM requires that $0 < \mu_m \leq 1$ for all μ_m. In the limiting case where $\mu_m = 1$ for all R_m, NL collapses to the multinomial logit (MNL) model, which is characterised by IIA.

Although the above relates to the simple case, it can be readily extended to more than two levels. An alternative generalisation is to relax the constraint that nests are mutually exclusive. The latter yields the cross-nested logit (CNL) model, first suggested (albeit somewhat vaguely) by McFadden (1978). CNL allows alternatives to belong to more than one nest, potentially with different 'degrees' of membership, although the model is restricted to two levels.

3.2 General cross-nested models

RNEV and its network interpretation have been derived by Daly and Bierlaire (2005). The network interpretation allows an analyst to draw an arbitrary cross-nesting structure, emanating from a single 'root' and with any number of levels but without cycles, and to calculate the conditional choice probabilities by logit functions at each node in the network. Daly and Bierlaire showed that this structure is consistent with McFadden's GEV requirements and is therefore consistent with RUM. RNEV generalises NL because it allows cross-nesting and generalises CNL because it allows multiple levels. The Daly and Bierlaire proof that RNEV is consistent with GEV can thus be used to prove that these special cases are also consistent with RUM.

Formally, the RNEV model can be defined by a recursive relationship at each node b of a single root (i.e. connected), non-cyclic network:

$$\exp V_b = \sum_{a \in \mathrm{succ}(b)} \lambda_{ba} \exp \mu_b {V_a}/{\mu_a}$$

where $\mathrm{succ}(b)$ is the set of immediate successors to b in the network; λ and μ are sets of positive parameters with $\mu_a \geq \mu_b$, if $a \in \mathrm{succ}(b)$. Daly and Bierlaire showed that $\exp V_r$ is a GEV G function, where r is the single root of the network, defined over the functions $\exp V_e$ for the elementary nodes (i.e. those that have no successors) which form the alternatives in the model. The conditional choice probabilities in this model are given by:

$$P(a|b) = \frac{\lambda_{ba} \exp \mu_b \, {}^{V_a}\!/_{\mu_a}}{\sum_{a' \in \mathrm{succ}(b)} \lambda_{ba'} \exp \mu_b \, {}^{V_{a'}}\!/_{\mu_{a'}}}$$

i.e. the standard logit probabilities with the addition of the inclusion parameters λ. The marginal probability of choice for each alternative can then be calculated as the sum of the choice probabilities for each of the paths connecting the root to that alternative, where the path choice probabilities are calculated as the products of the conditional probabilities at each node on the path (including the root).

4. ELIMINATION-BY-ASPECTS

An alternative paradigm, commonly attributed to Simon (e.g. 1959), is that individual decision-makers, instead of seeking to utility-maximise, are more inclined to use simplifying heuristics, achieving 'approximate' solution by means of 'modest' computational effort. A number of formal mathematical models based on this idea have been proposed; such models are often based on some form of attribute-based sequential elimination. See Manrai (1995) for a useful review.

4.1 General cross-nested models

One of the better-known elimination models is EBA, which was proposed by Tversky (1972a, 1972b). According to EBA, each alternative is conceptualised as a collection of aspects '*…that denote all valued attributes of the options including quantitative attributes (e.g., price, quality) and nominal attributes (e.g., automatic transmission on a car, fried rice on a menu)*' (Tversky and Sattath, 1979 p543). Choice is modelled as a sequential elimination process whereby an aspect is selected, with probability proportional to its 'value' or 'utility' to the decision-maker, and all alternatives not possessing the aspect are eliminated. (It is clear that continuous attributes must be expressed as 0/1 indicators (e.g. by formulating a series of thresholds) to incorporate them in the model). This process is repeated until only one alternative remains. The model can be formalised following Tversky and Sattath:

Again let $S = \{1,...,J\}$ be the choice set of alternatives. Let each i in S be associated with a finite non-empty set $i' = \{\alpha, \beta, ...\}$ of aspects of i. If aspect α is an element of i', alternative i is said to 'include' α. Let S' be the set of aspects that are included in at least one alternative in S, i.e. $S' = \{\alpha | \alpha \in i' \text{ for some } i \in S\}$. For any α in S', let $R_\alpha = \{i \in S | \alpha \in i'\}$ denote the subset of alternatives that include α.

A family of choice probabilities $P(i|S)$, for each choice set S and all $i \in S$, is EBA if there exists a non-negative scale u on S' such that:

$$P(i|S) = \frac{\sum_{\alpha \in i'} u(\alpha) P(i|R_\alpha)}{\sum_{\beta \in S'} u(\beta)} \qquad (2)$$

The recursive formula (2) is thus a weighted sum of the probabilities of selecting *i* from aspect-based subsets of S, where the weights are interpreted as aspect-specific utility or value. Note that - in common with RNEV - there is no requirement that the subsets be non-overlapping.

To facilitate discussion, and without loss of generality, the following simplifying properties of aspects and alternatives can be assumed to be part of the EBA definition:

I. eliminate from consideration all aspects which are not possessed by any alternative (this is just making sure the individual does not require an aspect that is not available);

II. eliminate from consideration all aspects that are possessed by all alternatives (these do not help in making a choice);

III. eliminate cases where a single alternative possesses all of the aspects (it would always be chosen);

IV. consider any two or more aspects which are possessed by exactly the same set of alternatives as effectively forming a single aspect, since they always apply together;

V. consider as indistinguishable any alternatives that possess the same set of aspects.

In case IV, when aspects α_1 and α_2 are considered together as α, we can write:

$$u(\alpha) = u(\alpha_1) + u(\alpha_2)$$

and this gives the same probabilities as (2) because $R_{\alpha_1} = R_{\alpha_2} = R_\alpha$ and so:

$$u(\alpha_1)P(i|R_{\alpha_1}) + u(\alpha_2)P(i|R_{\alpha_2}) = u(\alpha)P(i|R_\alpha)$$

Moreover, equation (2) can be seen to be an extension of the following equation, which is always true:

$$P(i|S) = \sum_{\alpha \in i'} P(R_\alpha|S) P(i|R_\alpha)$$

where $P(R_\alpha|S)$, the probability that choice will be in R_α given that it is in S, is defined by the specific conditional probability:

$$P(R_\alpha|S) = \frac{u(\alpha)}{\sum_{\beta \in S'} u(\beta)}$$

Tversky (1972a) derived the following properties of EBA probabilities.

- *Regularity* asserts that the choice probability of specific alternatives cannot be increased through enlargement of the choice set.
- *Multiplicative inequality* asserts that the probability of choosing *i* from $\{i, j, k\}$ cannot be less than the probability of choosing *i* from both $\{i, j\}$ and $\{i, k\}$ in two independent choices.
- *Moderate stochastic transitivity* is a generalisation of the transitivity axiom of microeconomic theory.

Tversky (1972a) presented the above properties as 'estimation-free' tests of EBA, meaning that they can be applied to choice outcomes without a necessity to estimate parameters. It is however important to note that, while necessary to determine consistency with EBA, the above properties are not sufficient. Indeed, it can be shown that NL is characterised by the same three properties of regularity, multiplicative inequality and moderate stochastic transitivity.

4.2 Tree models

Tversky and Sattath (1979) proposed the PRETREE model as a special case of EBA. PRETREE imposes a restriction on the choice set such that subsets of alternatives sharing aspects form a hierarchical structure. Two alternative but entirely equivalent interpretations of PRETREE were proposed, Elimination-By-Tree (EBT) and Hierarchical Elimination-By-Aspects (HEBA).

Extending the above analysis of EBA, we now consider PRETREE more formally, focussing on the HEBA interpretation. Let R_α be the set of alternatives in S that include the link α, i.e. $R_\alpha = \{i \in S | \alpha \in i'\}$. Define $\alpha | \beta$ if β follows directly from α. Let $u(\alpha)$ be the length (or utility) of α, and let $w(\alpha)$ be the total length of all links that follow from α, including α. If T^* is a tree and $S \subset T$, the set $S^* = \{i' | i \in S\}$ is referred to as a 'sub-tree'. Finally, for $R_\beta \subset R_\alpha$, let $P(R_\beta, R_\alpha)$ be the probability that an alternative chosen from R_α is also an element of R_β, i.e. $P(R_\beta, R_\alpha) = \sum_{i \in R_\beta} P(i | R_\alpha)$.

A family of choice probabilities $P(i|S)$, for each choice set S and all $i \in S$, is HEBA if there exists a tree T^* with measure u such that the following three conditions hold:

1. If $\gamma|\beta$ and $\beta|\alpha$ then $P(R_\alpha, R_\gamma) = P(R_\alpha, R_\beta)P(R_\beta, R_\gamma)$.
2. If $\gamma|\beta$ and $\gamma|\alpha$ then: $\dfrac{P(R_\alpha, R_\gamma)}{P(R_\beta, R_\gamma)} = \dfrac{w(\alpha)}{w(\beta)}$, provided $P(R_\beta, R_\gamma) \neq 0$.
3. The above conditions also hold for any sub-tree S^* of T^*, with the induced structure on S^*.

Since it is a special case of EBA, the properties of regularity, moderate stochastic transitivity and multiplicative inequality apply also to HEBA. Tversky and Sattath (1979) presented two further properties, namely the trinary and quarternary conditions, as specific to HEBA. Indeed, they asserted that: '…*the trinary and the quarternary conditions are not only necessary but are also sufficient to ensure the representation of binary choice probabilities as a* [PRETREE]' (p551). Since we will refer to the trinary condition in subsequent discussion, we define it formally:

The *trinary condition* states that any three alternatives that form a sub-tree $\{i, j\}k$, e.g. $R_\alpha = \{i, j\}$, $R_\beta = \{k\}$ in HEBA, satisfy the condition:

$$\text{If } \frac{P(i\{i,j\})}{P(j\{i,j\})} \geq 1, \text{ then } \frac{P(i\{i,j\})}{P(j\{i,j\})} \geq \left[\frac{P(i\{i,k\})}{P(k\{i,k\})} \bigg/ \frac{P(j\{j,k\})}{P(k\{j,k\})} \right] \geq 1$$

where strict inequality (equality) in the hypothesis implies strict inequality (equality) in the result. The quarternary condition applies, in analogous fashion, to the four-alternative case; for reasons of brevity, we refer readers to the definition in Tversky and Sattath (1979).

5. EQUIVALENCE BETWEEN EBA AND GEV

Although they differ fundamentally in their paradigmatic basis, HEBA and NL show significant commonality in the following respects. First, both are special cases of the common generic framework offered by the stochastic model. Second, both represent the choice problem as a 'preference tree', with mutually exclusive subsets of alternatives defined on the basis of covariance or similarity. Third, the two models share common properties; for example, both HEBA and NL are characterised by regularity, moderate stochastic transitivity and multiplicative inequality. These characteristics - in particular it would seem the second - have led several eminent authors to make comment on the existence of equivalence between HEBA and NL.

Tversky and Sattath (1979) asserted; *'Although the nested logit model does not coincide with pretree, the two models are sufficiently close that the former may be regarded as a random utility counterpart of the latter'* (p567); but offered no further insight. McFadden (1981) presented a more detailed analysis. For a preference tree consisting of three alternatives, multinomial choice probabilities were derived analytically from binary choice probabilities for three models: NL, HEBA and multinomial probit (MNP). McFadden compared the multinomial choice probabilities forecast by the three models for given binary choice probabilities, finding that the multinomial choice probabilities forecast by NL, HEBA and MNP were, not only intuitively plausible, but extremely close to each other. He concluded that *'...at least for simple preference trees...these models are for all practical purposes indistinguishable'* (p236). Researchers now understand the relationship of MNP to GEV models better and extensive analytical and simulation testing has illuminated the comparatively small but sometimes significant differences between MNP and NL (Whelan *et al.*, 2002).

Extending the analysis, it is insightful to also consider the degree of equivalence between EBA and CNL, which generalise HEBA and NL respectively to accommodate a 'cross-nested' structure. Again, a number of authors have made observations. Small (1987) asserted that ordered GEV (OGEV), which is a special case of CNL, is: *'…an example of A. Tversky's* *"elimination by aspect" class of choice models'* (p414). Vovsha (1997) noted: *'…within the cross-nested framework…choice can be conditionally made from several marginal nests reflecting different attraction focuses…From this point of view the cross-nested structure comes close to the elimination-by-aspects theory'* (p9). Unfortunately, neither Tversky nor Vovsha elaborated on these comments, but perhaps even less transparent is the contribution of Train (2003), who asserted: *'With positive* [log sum parameter], *the nested logit approaches the "elimination by aspects" model of Tversky (1972) as* [log sum parameter] $\rightarrow$ *0'* (p93).

In the subsequent analysis we establish formal conditions under which EBA and GEV are equivalent. Our initial investigation is restricted to HEBA and NL, and requires two stages of analysis. First, we establish equivalence, in a mathematical sense, between the full set of choice probabilities of HEBA and NL. While necessary to establish formal equivalence, this is not however sufficient. For this, a second stage is required of verifying that the equivalence, as established, does not violate the internal conditions of either HEBA or NL (i.e. both models remain valid according to their own rules).

Implicit in the above is consistency in the level of aggregation at which the two models are represented. HEBA in its most basic form is a model of repeated choices by a single individual; this contrasts with NL, which typically relies on a single choice from each of a sample of individuals. In the context of our investigation of equivalence, consistency of aggregation level is achieved through deference to Tversky and Sattath's (1979) re-interpretation of HEBA as an aggregate model. Having investigated equivalence between HEBA and NL, we extend the analysis to EBA and RNEV.

Before proceeding, it is worth noting that Tversky (1972a) presented a much-neglected proof that EBA is RUM. This result is based on a re-interpretation of EBA as a ranking model. 'Standard EBA' establishes, at each stage of the choice process, a partition between acceptable and unacceptable alternatives. 'Ranking EBA' extends this partitioning to the unacceptable alternatives as well as to the acceptable alternatives. Having established a complete ranking of alternatives, a utility vector is assigned to this ranking, and consistency with RUM established.

Our own investigation differs from Tversky's in the following respects. First, we focus both on the GEV subset of RUM as opposed to RUM generally, and on Standard EBA as opposed to Ranking EBA. Second, we consider equivalence, which is considerably stronger than consistency. Moreover, Tversky's proof that EBA is RUM could, in principle, be extended to any choice model that is able to derive a complete preference ordering. Whilst demonstrating the generality of RUM, the proof does not necessarily inform our own investigation.

5.1 Equivalence between HEBA and NL

Following McFadden (1981), consider the simplest possible form of HEBA; that is, a choice between three alternatives, two of which show some degree of similarity. More formally, consider a choice from the set $S = \{i, j, k\}$, where i, j and k are characterised by unique aspects α_i, α_j and α_k, respectively, and i and j share a common aspect α_{ij} which is not possessed by k.

Create two subsets of S, the first containing the two alternatives i and j that share α_{ij}, and the second containing the remaining alternative k:

$$R_{\alpha_{ij}} = \{i, j\}$$

$$R_{\alpha_k} = \{k\}$$

In this two-level model, under both HEBA and NL specifications of this problem, choice probability can be written as follows:

$$P(i|S) = P\left(R_{\alpha_{ij}}\right) \times P\left(i|R_{\alpha_{ij}}\right)$$

$$P(j|S) = P(R_{\alpha_{ij}}) \times P(j|R_{\alpha_{ij}})$$

$$P(k|S) = P(R_{\alpha_k})$$

Moreover, the choice problem under consideration, whether specified as HEBA or as NL, can be defined completely by either of the two marginal probabilities and either of the two conditional probabilities. That is, a choice model for this situation has just two effective degrees of freedom.

Consider first HEBA. Following Tversky and Sattath (1979), the marginal probability of choosing the subset $R_{\alpha_{ij}}$ is given by:

$$P(R_{\alpha_{ij}}) = \frac{u(\alpha_i) + u(\alpha_j) + u(\alpha_{ij})}{u(\alpha_i) + u(\alpha_j) + u(\alpha_{ij}) + u(\alpha_k)} \quad (3)$$

and the conditional probability of choosing *i* from the set $R_{\alpha_{ij}}$ is given by:

$$P(i|R_{\alpha_{ij}}) = \frac{u(\alpha_i)}{u(\alpha_i) + u(\alpha_j)} \quad (4)$$

For HEBA to hold, distinct from the Luce model (Luce, 1959), it is necessary that $u(\alpha_{ij}) > 0$. Similarly, for alternatives i, j and k to exist as distinct alternatives, it must be the case that $u(\alpha_i), u(\alpha_j), u(\alpha_k) > 0$.

Now consider identification, where two transformations are relevant:

H1. Define $u'(\alpha_n) = \gamma u(\alpha_n)$ for all n. The factor γ cancels out in (3) and (4), so that the entire model is unchanged.

H2. Define $u''(\alpha_n) = \phi u(\alpha_n)$ for $n = i, j$. With $u(\alpha_k)$ held constant, the model remains unchanged if we define:

$$u''(\alpha_{ij}) = u(\alpha_{ij}) + c$$

provided that:

$$u(\alpha_i) + u(\alpha_j) + u(\alpha_{ij}) = u''(\alpha_i) + u''(\alpha_j) + u''(\alpha_{ij}) = \phi u(\alpha_i) + \phi u(\alpha_j) + u(\alpha_{ij}) + c$$

i.e.

$$c = (1 - \phi)[u(\alpha_i) + u(\alpha_j)]$$

We must have $u''(\alpha_{ij}) > 0$, but if $\phi < 1$ it is always positive and no problems arise, so that the model is once again unchanged.

This discussion reveals that HEBA is inherently over-specified in two distinct ways. Given that the four variables $u(\alpha_i)$, $u(\alpha_j)$, $u(\alpha_k)$ and $u(\alpha_{ij})$ are defined, there are, as already noted, two effective degrees of freedom. HEBA cannot, therefore, be estimated without imposing two appropriate constraints.

An important observation is that (3) and (4) also yield the probabilities of choice in reduced choice sets, by considering these as limiting cases when individual aspect weights becomes zero. For example, if $u(\alpha_k)=0$ in (3), then $P(R_{\alpha_{ij}})=1$ and (4) gives the probability of choosing i from the reduced choice set $\{i, j\}$. Similarly, if $u(\alpha_j)=0$ in (4), then $P(i|R_{\alpha_{ij}})=1$ and (3) gives the probability of choosing i from the reduced choice set $\{i,k\}$. Thus the probability equations for the three-alternative case yield the relevant probability equations for each of the three possible binary choice subsets.

The properties established for the three-alternative case, such as the excess degrees of freedom H1 and H2, then extend to these binary cases. Note however that if $u(\alpha_{ij})=0$ then $u(\alpha_{ij})$ becomes irrelevant by property I; similarly, if $u(\alpha_i)=0$ then $u(\alpha_j)$ and $u(\alpha_{ij})$ can be combined by property IV. In binary choice subsets, therefore, there is always one redundant degree of freedom.

For the same choice problem, we now write the analogous NL probabilities of choice:

$$P\left(R_{\alpha_{ij}}\right)=\frac{\exp\left\{V\left(\alpha_{ij}\right)+\mu\log\left[\exp^{V(\alpha_i)}\!/\!_{\mu}+\exp^{V(\alpha_j)}\!/\!_{\mu}\right]\right\}}{\exp\left\{V\left(\alpha_{ij}\right)+\mu\log\left[\exp^{V(\alpha_i)}\!/\!_{\mu}+\exp^{V(\alpha_j)}\!/\!_{\mu}\right]\right\}+\exp V\left(\alpha_k\right)} \tag{5}$$

$$P\left(i\middle|R_{\alpha_{ij}}\right)=\frac{\exp^{V(\alpha_i)}\!/\!_{\mu}}{\exp^{V(\alpha_i)}\!/\!_{\mu}+\exp^{V(\alpha_j)}\!/\!_{\mu}} \tag{6}$$

If we again consider identification, it is apparent that NL is over-specified in three distinct ways, since:

N1. The addition of a constant to all the elementary utilities has no effect on probability.

N2. The addition of a constant to $V\left(\alpha_k\right)$ and $V\left(\alpha_{ij}\right)$ has no effect on probability.

N3. The multiplying of μ by a factor has no effect on probability, provided that the lower level utilities are adjusted appropriately.

It is important to note that properties N1 and N2 are analogous to HEBA, whereas property N3 is specific to NL. Obviously, the introduction of the above brings three additional degrees of freedom to NL. Here, we see that the five variables used in defining the model ($V(\alpha_i)$, $V(\alpha_j)$, $V(\alpha_k)$, $V(\alpha_{ij})$ and μ) are again reduced to two degrees of freedom by the constraint imposed.

In a similar manner to before, the NL probabilities of choice for reduced choice sets can be derived from (5) and (6) by considering the form these equations take when individual utility weights tend to minus infinity. For example, if $V(\alpha_k) \to -\infty$ in (5), then $P(R_{\alpha_{ij}}) \to 1$ and (6) gives the probability of choosing i from the reduced choice set $\{i, j\}$. Similarly, if $V(\alpha_j) \to -\infty$ in (6), then $P(i|R_{\alpha_{ij}}) \to 1$ and (5) gives the probability of choosing i from the reduced choice set $\{i, k\}$.

In order to establish equivalence between HEBA and NL, a first and necessary requirement is that the choice probabilities of HEBA and NL are equivalent not only for the complete choice set, but for each and every choice subset thereof. For the three-alternative case, equivalence must therefore hold for the complete choice set $\{i, j, k\}$, as well as for the binary subsets $\{i, j\}$, $\{i, k\}$ and $\{j, k\}$. As the preceding discussion has demonstrated, however, the probability statements applying to any of the reduced choice sets exist as limiting cases of the probability statements applying to the complete choice set; hence our earlier comment that HEBA and NL can be defined completely by either of the two marginal probabilities and either of the two conditional probabilities. Moreover, our interest in establishing equivalence between HEBA and NL for all choice subsets of the three-alternative case amounts to a requirement that there exists simultaneous equivalence between both (3) and (5) and (4) and (6).

This requirement can be established in a number of ways, because of the excess degrees of freedom in the two models. A simple equivalence is obtained through the solution of a simultaneous equation problem consisting of four equations, as follows:

$$\exp^{V(\alpha_i)/\mu} = u(\alpha_i) \quad (7)$$

$$\exp^{V(\alpha_j)/\mu} = u(\alpha_j) \quad (8)$$

$$\exp V(\alpha_k) = u(\alpha_k) \quad (9)$$

$$\exp\{V(\alpha_{ij}) + \mu W_{ij}\} = \exp W_{ij} + u(\alpha_{ij}) \quad (10)$$

$$\text{where } W_{ij} = \log\left[\exp^{V(\alpha_i)/\mu} + \exp^{V(\alpha_j)/\mu}\right] = \log[u(\alpha_i) + u(\alpha_j)]$$

Solving the equation system (7) to (10) in terms of $V(\alpha_i)$, $V(\alpha_j)$, $V(\alpha_k)$ and $V(\alpha_{ij})$, we arrive at the following:

$$V(\alpha_i) = \mu \log[u(\alpha_i)]$$

$$V(\alpha_j) = \mu \log[u(\alpha_j)]$$

$$V(\alpha_k) = \log[u(\alpha_k)]$$

$$V(\alpha_{ij}) = \log[\exp W_{ij} + u(\alpha_{ij})] - \mu W_{ij} \quad (11)$$

Substituting these expressions for $u(\alpha_i)$, $u(\alpha_j)$, $u(\alpha_k)$, $u(\alpha_{ij})$ into (3) and (4), and $V(\alpha_i)$, $V(\alpha_j)$, $V(\alpha_k)$, $V(\alpha_{ij})$ into (5) and (6), or indeed into any limiting case of these probabilities, one can confirm the equivalence between NL and HEBA, ***whatever the value of*** μ[1]. This flexibility with respect to μ corresponds to the additional degree of over-specification of NL indicated above. However, note that $V(\alpha_{ij})$ depends on both $u(\alpha_i)$ and $u(\alpha_j)$ unless $u(\alpha_{ij}) = 0$ and $\mu = 1$, when we get $V(\alpha_{ij}) = 0$ (i.e. the Luce model).

The equivalence can also be set up in the opposite direction, deriving $u(\alpha_i)$, $u(\alpha_j)$ and $u(\alpha_k)$ straightforwardly from the corresponding V's as in (7) to (9) and setting:

$$u(\alpha_{ij}) = \exp\{V(\alpha_{ij}) + \mu W_{ij}\} - \exp W_{ij} \quad (12)$$

[1] Adding the caveat that NL must remain internally consistent - discussion of this follows.

This is not, however, the end of the story, since we must consider whether the above equivalence violates the internal conditions of either HEBA or NL. For the equivalence to work within the terms of HEBA, all aspect weights must be positive. It is clear from (7) to (9) that this always holds for $u(\alpha_i)$, $u(\alpha_j)$ and $u(\alpha_k)$. The case of $u(\alpha_{ij})$ is less straightforward, since it requires that:

$$\exp V(\alpha_{ij}) > [\exp W_{ij}]^{1-\mu} \tag{13}$$

With reference to over-specification N2, however, it is clear that the addition of a sufficiently large constant to $V(\alpha_{ij})$ and $V(\alpha_k)$ will ensure that (13) is always satisfied.

Turning to NL, the issue of internal validity rests on consistency with RUM. Perusal of (11), which is expanded below as (14), reveals that although $V(\alpha_{ij})$ is a component of the utility of alternative *i*, it is a function of $u(\alpha_j)$, which itself depends on attributes of alternative *j*. Similarly, $V(\alpha_{ij})$ forms part of the utility of alternative *j*, but depends partly on attributes of alternative *i*.

$$V(\alpha_{ij}) = \log[u(\alpha_i) + u(\alpha_j) + u(\alpha_{ij})] - \mu \log[u(\alpha_i) + u(\alpha_j)] \tag{14}$$

Since $u(\alpha_i)$ and $u(\alpha_j)$ will necessarily be significantly different from zero for compliance with HEBA, the implication is that the utility of one nested alternative will be a function of the attributes of other alternatives in the same nest. Moreover, even if $0 < \mu \leq 1$, consistency with RUM is not ensured.

Consider, however, the limiting case where:

$$V(\alpha_i) \to V(\alpha_i) + V(\alpha_{ij})$$

$$V(\alpha_j) \to V(\alpha_j) + V(\alpha_{ij})$$

$$V(\alpha_{ij}) = 0$$

such that (12) now simplifies to:

$$0 < u(\alpha_{ij}) = \exp\{\mu W_{ij}\} - \exp W_{ij}$$

which holds provided $0 < \mu < 1$ (i.e. NL does not collapse to MNL) and $W_{ij} < 0$, which again can be ensured by subtracting a sufficiently large constant from $V(\alpha_i)$, $V(\alpha_j)$ and $V(\alpha_k)$.

All that now remains is to verify that HEBA can always be adjusted to ensure that $V(\alpha_{ij}) = 0$. With reference to identification issue H2, the relevant transformation is:

$$u(\alpha_{ij}) \rightarrow u(\alpha_{ij}) + c\,,$$

$u(\alpha_i)$ and $u(\alpha_j)$ multiplied by factor ϕ with $c = (1-\phi)\exp W_{ij}$. With this transformation, (11) changes to:

$$V(\alpha_{ij}) = \log[u(\alpha_{ij}) + c + \phi \exp W_{ij}] - \mu(W_{ij})^{\phi}$$

$$= \log[u(\alpha_{ij}) + \exp W_{ij}] - \mu(W_{ij} + \log\phi)$$

which can be zero for an appropriate choice of ϕ. Thus by exploiting the redundant degrees of freedom of HEBA, it is possible to ensure that the RUM-violating term is always zero.

We conclude that it is always possible, in the three-alternative case, to specify a valid NL that is equivalent to a HEBA, and vice versa.

Before turning to the more general case with larger numbers of alternatives, it is useful to set our apparently simple finding in the context of the fundamental papers on GEV and HEBA. McFadden (1981), a key paper in establishing the GEV model, also considered the three-alternative comparison between HEBA and GEV but failed to arrive at any conclusive result on their equivalence. Tversky and Sattath (1979), again an influential contribution to the literature, acknowledged the similarity between HEBA and GEV, but appeared to rule out the possibility of equivalence: *'…the nested logit model does not coincide with pretree…'* (p567). We have thus improved considerably on the clarity of McFadden's result, and shown that Tversky and Sattath's assertion was incorrect.

5.2 Equivalence between EBA and RNEV

The equivalence established between HEBA and NL suggests that there may also be equivalence between EBA and RNEV, an equivalence which can be investigated now that the RNEV model is available as a GEV counterpart to EBA. In each case cross-nested network structures with an indefinite number of levels can be constructed to represent the models, with conditional choice probabilities of transition between successive nodes until a single alternative is reached.

However, the situation is complicated and there are limits to the equivalence that can be achieved. To keep the discussion clear, we assume that the simplifying properties I to V in section 4.1 above apply to the aspects and alternatives in an EBA model and consider how an RNEV model can be constructed which might be equivalent to it.

If there are n aspects, define the 'index set' N over the aspects, i.e. the set of all possible combinations of aspects. Note that $|N| = 2^n$, providing we allow the set Φ which contains no aspects. Define $X \subset N$ to be the set of combinations of aspects which are possessed by some alternatives. Note that X contains more than Φ, because of simplifying property I, and X is less than N because of simplifying property V, but that X certainly contains Φ, because there is no aspect that is possessed by all the alternatives (simplifying property II).

Every subset of alternatives has a well-defined set of aspects which all members of the set possess (this set of aspects may be Φ). Every set of aspects in X has at least one corresponding set of alternatives, because of the definition of X, but there may be several sets of alternatives that possess a given set of aspects. We can associate with each element of X the **largest** set of alternatives that possess the relevant set of aspects – this largest set contains **all** the alternatives that possess these aspects – thus obtaining a one-to-one correspondence between X and a subset of the index set of alternatives.

A network can be defined connecting the elements of X. A link is defined from element x to element y if y represents the addition of a single aspect to x. This network is not circular because each successive link adds another element to the set of aspects. It is connected, because every element x (other than Φ) has at least one predecessor formed by removing an aspect from the set of aspects at x and this predecessor is a member of X because all of the alternatives possessing the set at x also possess the set at x minus one aspect. Thus the network fits the specification for RNEV required by Daly and Bierlaire (2005) and Φ is its root.

Moreover, choice in the network represents exactly the EBA model. Each selection of an aspect will move to a new element in X and may (or may not) reduce the number of alternatives that are still open for consideration. Selection of an aspect at any stage that would mean that no alternatives possess the required set (i.e. would take us outside X) is prohibited by the rules of EBA. Eventually, because alternatives each possess a unique set of aspects, we arrive at a point where only one alternative is acceptable and the process is complete.

It appears that there is considerable similarity between the EBA and RNEV models in terms of their structure and because of the simple form of the conditional choice probability at each stage in the process. It is therefore reasonable to enquire as to the full extent of that equivalence.

Because RNEV is at each stage a GEV model, the conditional choice probability at each stage is given by McFadden's GEV formula:

$$P(z|x) = G_z . \frac{\partial G_x / \partial G_z}{\mu^* G_x}$$

where G is the utility function at each node and is itself given by the RNEV formula (Daly and Bierlaire, 2005):

$$G_x = \sum_{z'} \lambda_{xz'} G_z^{\mu_x / \mu_{z'}}$$

for some positive constants λ and μ, and μ^* is the constant in the homogeneity property of the GEV model.

To complete RNEV, for each node x in the network defined over the set X we must define the constants $\mu_x \geq \mu_z > 0$, where z is any successor of x, and for each link xz in the network we must define the constants $\lambda_{xz} > 0$. With these constants we then obtain:

$$P(z|x) = G_z . \frac{\partial G_x / \partial G_z}{(\mu^* G_x)} = \frac{\lambda_{xz} G_z^{\mu_x / \mu_z}}{\sum_{z'} \lambda_{xz'} G_z^{\mu_x / \mu_{z'}}}$$

since μ^* is given in this case by:

$$\mu^* = \mu_z / \mu_x$$

The formula for $P(z|x)$ is clearly the basic logit form. To achieve equivalence with EBA, these conditional choice probabilities must be exactly proportional to the importance of the attribute that is included in z and not in x, which can be achieved if (and only if):

$$\lambda_{xz} G_z^{\mu_x / \mu_z} \propto \prod_{i \in z} u(\alpha_i) \qquad (15)$$

where the product is over all of the aspects possessed by the alternative. If this proportionality applied, the conditional choice probabilities would then be given by:

$$P(z|x) = \frac{\prod_{i \in z} u(\alpha_i)}{\sum_{z'} \prod_{i \in z'} u(\alpha_i)} = \frac{u(\alpha_{xz})}{\sum_{z'} u(\alpha_{xz'})}$$

where α_{xz} is the aspect possessed by z and not by x, because each successor z' of x possesses exactly one such aspect. These conditional probabilities are exactly what is required to give the unconditional probabilities of equation (2).

Thus equivalence between EBA and RNEV depends on specifying G, μ and λ in an RNEV model to achieve the proportionality (15). For example, if we could arrange that:

$$\lambda_{xz} = \frac{1}{\left(\sum_{z} u(\alpha_{xz})^{\mu_x}\right)}$$

and that at each node z in the network, we had a function:

$$G_z = \prod_{i \in z} \left(u(\alpha_i) \right)^{\mu_z}$$

where x is the predecessor node of the alternative, then we would satisfy the proportionality requirement for the choice probabilities and also obtain at each node:

$$G_x = \sum_z \lambda_{xz} G_z^{\mu_x / \mu_z} = \frac{\sum_z \left(\prod_{i \in z} \left(u(\alpha_i) \right)^{\mu_z} \right)^{\mu_x / \mu_z}}{\sum_z u(\alpha_{xz})^{\mu_x}}$$

$$= \frac{\sum_z \prod_{i \in z} \left(u(\alpha_i) \right)^{\mu_x}}{\sum_z u(\alpha_{xz})^{\mu_x}} = \prod_{i \in x} u(\alpha_i)^{\mu_x}$$

the final simplification being possible because the aspects possessed by each z are different but each just one more than those possessed by x. Thus the form of the function could be maintained throughout the network.

The propagation of the G functions can be proved formally by induction, once the elementary alternatives have been defined to have a function of the required form. However, the form specified for λ is obviously **not** a constant as is required in the RNEV specification. The issue is then whether and how this problem can be solved.

One approach is to view the λ's as being defined in the base situation and not subject to change if the weights of the aspects should change in a future situation. Forecasting and interpretation in EBA is in any case not clearly defined. However this approach is not entirely satisfactory.

A second approach is to consider the λ's as defining an additional utility component at each node, along the lines of $V(\alpha_{ij})$ introduced in equation (5) above. The problem with this approach is that an additional utility component of this type does not form part of the RNEV specification and, as in the NL case, prevents the model from being truly utility-maximising, since the utility of one alternative then depends on characteristics of other alternatives.

A final approach, as applied in the simple NL case of section 5.1, is to 'demote' the utility component corresponding to the λ's to the lowest level of the choice structure and incorporate it in the utility of the elementary alternatives. This approach exploits redundant degrees of freedom in EBA. To implement this approach, we have to incorporate the λ value at x in the utility of all its successors z. The problem here is then that, because of the potential cross-nesting, the utility of z must incorporate λ values from all its predecessors, while a different successor z' of x may be incorporating λ

values from a different set of predecessors. Taking in values from different sets of predecessors would in general then destroy the proportionality which is required for full consistency with EBA.

Thus a valid RNEV model can be constructed that is fully consistent with an EBA model if and only if there is no cross-nesting, i.e. when there is a unique set of predecessors for each node – a tree model. Note that the values of the μ's are not relevant to this discussion, as was the case in the discussion of the simple NL case above. The similarity between alternatives represented by μ is of a different type to the similarity represented in EBA. We conclude that complete equivalence between EBA and RNEV can be established in a fully satisfactory way only for tree models. For cross-nested models, equivalence can be achieved but only at the expense of unsatisfactory features in the models, such as a failure to adhere fully to RUM.

6. SUMMARY AND CONCLUSIONS

The discussion above has demonstrated that tree models (i.e. HEBA and NL) are formally equivalent for the three-alternative case. Equivalence exists in the sense that the mathematical formulae for choice probability are identical in the two models, whilst each model remains valid in respect of its own internal conditions. For larger tree models equivalence can also be established.

For general cross-nested models (i.e. EBA and RNEV models), however, equivalence can be established, but only at the cost of breaching the internal conditions of the models. For example, the RNEV equivalent of EBA can be specified, but this may require the utility functions of the RNEV to include terms that breach RUM. In principle, given a data set, the significance of these terms in influencing behaviour can be tested, so that the validity of RNEV can be established. A failure of this test - i.e. the finding that a utility-breaching term was significant - would indicate that EBA was more appropriate (or that a different RNEV specification was required). In many cases, however, identification of the most appropriate model specification for representing a given set of observed choices may be ambiguous, i.e. both paradigms give fits to the data which are closely comparable.

Our analysis of EBA and GEV equivalence represents a considerable advance in the clarity with which conditions for equivalence can be given. We finish by drawing some implications from our work, focussing on the respective interests of microeconomics and mathematical psychology, as identified in section 1.

In seeking to understand the cognitive process underlying individual discrete choice, mathematical psychology has interest in the discrimination between EBA and GEV. Tversky and Sattath (1979) devoted considerable attention to the application of HEBA, similarly interested in discriminating between HEBA and the Luce model. Their approach consisted of two complementary activities, as follows. First, maximum likelihood estimation of both the HEBA and Luce models, and their comparison via a likelihood ratio test. Second, estimation-free comparison of the models via the trinary condition. Since both activities rely on the probability statement, however, it is necessary to issue qualification in the light of our own work. Hence, either activity carries validity as a means of identifying the Luce model, but neither activity is capable of discriminating between HEBA and NL. Moreover, our work places significant doubt on the validity of using choice probability (or any derivative thereof, such as log-likelihood or 'testable' properties including regularity, multiplicative inequality, moderate stochastic transitivity, or the trinary and quarternary conditions) as the basis for discriminating between EBA and GEV. Since none of the properties are sufficient, discrimination is left to other means, which will have to be established by further research.

In simple tree models, and more general models by approximation, the description of EBA and GEV behaviour will lead to the same probability statement and hence to an equally good explanation of observed behaviour. Not only does this mean that observed behaviour can be equally well explained by either paradigm, but also it means that *both* types of behaviour can be indistinguishably mixed in the same observations. Some people may be choosing by elimination procedures, others may be carefully maximising utility: they cannot be distinguished. Most interestingly, people may be using mixed strategies, for example, reducing the choice set to reasonable dimensions by elimination strategies, then choosing to maximise utility in the reduced choice set. This possibility has been suggested, in a simple way requiring strong assumptions, by Swait (2001).

Turning to microeconomics, we consider the practical need to value and forecast. Economic valuations are typically derived from RUM as the ratio of utility coefficients (or more correctly, the ratio of marginal utilities); this ratio has behavioural interpretation in economics as 'marginal rate of substitution'. The result that EBA and GEV are substantially equivalent (in particular the possibility that GEV could be an equivalent representation of 'true' underlying EBA behaviour) perhaps confuses this interpretation. In short, where the true behaviour is EBA, do the economic concepts of marginal rate of substitution and value remain valid? A similar point applies to forecasting which, in RUM, typically involves the extrapolation of utility to attribute combinations not covered by existing data. Without understanding of the underlying decision process, however, it would seem impossible to specify, in any meaningful way, which parts of the models will or will not change in the forecast scenario.

It thus becomes clear that, whether the disciplinary interest is in microeconomics or mathematical psychology, equivalence between EBA and GEV brings a potential for serious error in model interpretation and application. Discrimination between EBA and GEV is essential, but recourse to the probability statement offers little insight in this regard.

ACKNOWLEDGEMENTS

This work was partly funded by the UK Department for Transport under New Horizons research project B36. The first-named author acknowledges the support of a Departmental Research Fellowship at the Institute for Transport Studies, University of Leeds.

REFERENCES

Ben-Akiva, M.E. and Lerman, S.R. (1985) *Discrete Choice Analysis*. The MIT Press, Cambridge, Massachusetts.

Block, H.D. and Marschak, J. (1960) Random orderings and stochastic theories of responses. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1). D. Reidel, Dordrecht.

Daly, A. and Bierlaire, M. (2005) A general and operational representation of Generalised Extreme Value models. In press, Transportation Research B (corrected proof available at http://www.elsevier.com as of 16 September 2005).

Debreu, G. (1954) Representation of a preference ordering by a numerical function. In Thrall, R.M., Coombs, C.H. and Davis, R.L. (eds) *Decision Processes*. John Wiley & Sons, New York.

Debreu, G. (1960) Review of R.D. Luce, Individual choice behaviour: a theoretical analysis. *American Economic Review*, 50 (1), pp186-188.

Luce, R.D. (1959) *Individual Choice Behaviour: A Theoretical Analysis*. John Wiley & Sons, New York.

Luce , R.D. and Suppes, P. (1965) Preference, utility and subjective probability. In Luce, R.D., Bush, R.R., and Galanter, E. (eds) *Handbook of Mathematical Psychology*, Volume III. John Wiley & Sons, New York.

McFadden, D. (1978) Modelling the choice of residential location. In Karlqvist, A., Lundqvist, L., Snickars, F. and Weibull, J. (eds) *Spatial Interaction Theory and Residential Location*. North-Holland, Amsterdam.

McFadden, D. (1981) Econometric models of probabilistic choice. In Manski, C. and McFadden, D. (eds) *Structural Analysis of Discrete Data: With Econometric Applications*. The MIT Press, Cambridge, Massachusetts.

Manrai, A.K. (1995) Mathematical models of brand choice behaviour. *European Journal of Operational Research*, 82 (1), pp1-17.

Marschak, J. (1960) Binary choice constraints and random utility indicators. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1). D. Reidel, Dordrecht.

Marschak, J., Becker, G.M. and DeGroot, M.H. (1963) Stochastic models of choice behavior. In Marschak, J. (1974) *Economic Information, Decision and Prediction: Selected Essays* (Volume 1). D. Reidel, Dordrecht.

Restle, F. and Greeno, J.G. (1970) *Introduction to Mathematical Psychology*. Addison-Wesley, Reading, Massachusetts.

Simon, H.A. (1955) A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, pp99-118.

Simon, H.A. (1959) Theories of decision-making in economics and behavioral science. *American Economic Review*, 49 (1), pp253-283.

Simon, H.A. (1989) Cognitive architectures and rational analysis: comments. *21st Annual Symposium on Cognition*. Department of Psychology, Carnegie-Mellon University.

Simon, H.A. (1990) Invariants of human behaviour. *Annual Review of Psychology*, 41, pp1-19.

Small, K.A. (1987) A discrete choice model for ordered alternatives. *Econometrica*, 55 (2), pp409-424.

Suppes, P., Krantz, D.M., Luce, R.D., and Tversky, A. (1989) *Foundations of Measurement*, Volume II, Geometerical, threshold and probabilistic representations. Academic Press, San Diego, California.

Swait, J. (2001) Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research B*, 35, pp643-666.

Thurstone, L. (1927) A law of comparative judgement. *Psychological Review*, 34, pp273-286.

Train, K. (2003) *Discrete Choice Methods With Simulation*. Cambridge University Press, Cambridge.

Tversky, A. (1972a) Choice by elimination. *Journal of Mathematical Psychology*, 9, pp341-367.

Tversky, A. (1972b) Elimination by aspects: a theory of choice. *Psychological Review*, 79 (4), pp281-299.

Tversky, A. and Sattath, S. (1979) Preference trees. *Psychological Review*, 86 (6), pp542-573.

Vovsha, P. (1997) Application of cross-nested logit model to mode choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record,* 1607, pp6-15.

Whelan, G., Batley, R., Fowkes, T. and Daly, A. (2002) Flexible models for analysing route and departure time choice. Seminar on Methodological Innovations, Proceedings of the European Transport Conference, 10-12 September 2002, Homerton College, Cambridge.