Published paper
Batley, R., Toner J. (2003) *Hierarchical Elimination-by-Aspects and Nested Logit Models of Stated Preferences for Alternative Fuel Vehicles* - Association of European Transport Conference, Strasbourg, 8-10 October 2003.

HIERARCHICAL ELIMINATION-BY-ASPECTS AND NESTED LOGIT MODELS OF STATED PREFERENCES FOR ALTERNATIVE-FUEL VEHICLES

Richard Batley and Jeremy Toner
Institute for Transport Studies, University of Leeds

1. INTRODUCTION

Since the late 1960s, transport demand analysis has been the context for significant developments in model forms for the representation of discrete choice behaviour. Such developments have adhered almost exclusively to the behavioural paradigm of Random Utility Maximisation (RUM), first proposed by Marschak (1960) and Block and Marschak (1960). A common argument for the allegiance to RUM is that it ensures consistency with the fundamental axioms of microeconomic consumer theory and, it follows, permits interface between the demand model and the concepts of welfare economics (e.g. Koppelman and Wen, 2001).

The desire to better represent observed choice, which has driven developments in RUM models, has been somewhat at odds, however, with the frequent assault on the utility maximisation paradigm, and by implication RUM, from a range of literatures. This critique has challenged the empirical validity of the fundamental axioms (e.g. Kahneman and Tversky, 2000; McIntosh and Ryan, 2002; Saelensmide, 1999) and, more generally, the realism of the notion of instrumental rationality inherent in utility maximisation (e.g. Hargreaves-Heap, 1992; McFadden, 1999; Camerer, 1998).

Emanating from these literatures has been an alternative family of so-called non-RUM models, which seek to offer greater realism in the representation of how individuals actually process choice tasks. The workshop on Methodological Developments at the 2000 Conference of the International Association for Travel Behaviour Research concluded: *'Non-RUM models deserve to be evaluated side-by-side with RUM models to determine their practicality, ability to describe behaviour, and usefulness for transportation policy. The research agenda should include tests of these models'* (Bolduc and McFadden, 2001 p326). The present paper, together with a companion paper, Batley and Daly (2003), offer a timely contribution to this research priority.

Batley and Daly (2003) present a detailed account of the theoretical derivation of RUM, and consider the relationships of two specific RUM forms; nested logit [NL] (Ben-Akiva, 1974; Williams, 1977; Daly and Zachary, 1976; McFadden, 1978) and recursive nested extreme value [RNEV] (Daly, 2001; Bierlaire, 2002; Daly and Bierlaire, 2003); to two specific non-RUM forms; elimination-by-aspects [EBA] (Tversky, 1972a, 1972b) and hierarchical EBA [HEBA] (Tversky and Sattath, 1979). In particular, Batley and Daly (2003) establish conditions under which NL and RNEV derive equivalent choice probabilities to HEBA and EBA, respectively. These findings would seem to ameliorate the concern that the application of RUM models to data generated by non-RUM choice processes could introduce significant biases. That aside, substantive issues remain as to how non-RUM models can best be specified so as to yield useful and robust information in both estimation and forecasting contexts, and how their empirical performance compares with RUM models. Such issues are the focus of the present paper, which applies non-RUM models to a real empirical context.

2. STATED PREFERENCES FOR ALTERNATIVE-FUEL VEHICLES

The context for the investigation is an analysis of stated preference [SP] choices between conventional and alternative-fuel vehicles [AFVs][1] in hypothetical car purchase scenarios.

The principal survey instrument was a mail-back SP experiment. This was supplemented by a focus group analysis, the aim of which was to explore the decision processes involved in vehicle ownership, purchase and use and the interaction between the ownership and use decisions. SP questionnaires were also administered to the attendees of the focus groups. In focussing this paper on conceptual issues of model specification, we make only cursory reference to the focus group analysis and previous SP studies of AFV demand. For greater detail on each, interested readers are referred to Hodgson (2002) and Batley *et al.* (2003), respectively.

Conventional SP design practice in the UK is to avoid 'overloading' the respondent, whether in terms of the number of alternatives or attributes presented. If, from the analyst's perspective, there is an interest in a large number of attributes, conventional practice is to develop a series of small designs featuring one common attribute, and to merge the designs at the modelling stage. Moreover, a typical design in UK applications might feature, say, two alternatives and four attributes. The experiment developed in this paper was much larger in dimension, featuring three alternatives and eight attributes.

The SP experiment involved a purchase choice between three cars, as follows:

- Car A was broadly consistent with a conventional petrol or diesel car;
- Car C was broadly consistent with a near-term AFV;
- Car B was a compromise option, which might represent some form of future 'clean' petrol or diesel vehicle or a future AFV with performance features more comparable with a petrol or diesel vehicle.

Following review of previous research and canvassing of expert opinion, it was decided to describe each car in terms of the following attributes:

1. on-the-road price (abbreviated in what follows by *orp*);
2. running costs (*rc*);
3. range on a full refuel or recharge (*rfr*);
4. time for a full refuel or recharge (*tfr*);

5. top speed (*ts*);
6. time taken to accelerate from 0-60 mph (*ac*);
7. retained value after 3 years or 36,000 miles (*rv*);
8. emissions as a percentage of a year 2000 petrol car (*em*).

An important point to note is that, given the current state of technology, AFVs typically offer inferior performance over at least some characteristics - for example *rfr* - as compared with conventional vehicles. This performance gap was represented as realistically as possible in the experiment.

A common justification for the application of non-RUM choice models is the proposition that individual decision-makers, particularly when faced with complex choice tasks, do not seek to utility maximise, but instead resort to simplifying heuristics in the achievement of a 'satisficing' solution (e.g. Simon, 1955; 1959; 1989; 1990). *A priori*, we hypothesised that, in the context of our SP investigation, such heuristics might arise in two (mutually consistent) ways:

- First, the inferior performance of AFVs relative to conventional vehicles might prompt some respondents to employ threshold-based heuristics, such as a minimum range threshold.
- Second, the complexity of the SP task arising from the relatively large quantity of alternatives and attributes might encourage the use of filtering heuristics. Again, the likes of a minimum range threshold might become relevant in this regard.

3. DESIGN AND IMPLEMENTATION OF THE STATED PREFERENCE EXPERIMENT

3.1 Design

Although we were able to draw on recent SP studies on AFV demand from the US and Australia (Brownstone and Train, 1999; Brownstone *et al.* 2000; Hensher and Greene, 2001), together with UK public policy documents (e.g. Cleaner Vehicles Task Force, 2000; Powershift, 2003), the obtaining of reliable data on the design attributes and their likely valuations proved difficult. Given this dearth of information, the design process steered away from concepts such as boundary values (Fowkes, 1991). Rather, the strategy was to maintain a high degree of orthogonality, whilst presenting seemingly sensible values for the attributes, and including a variety of interesting trade-offs. Moreover, each of the attributes was specified at four levels as detailed in Table 1.

An important design consideration was that, for any given attribute and pair of alternatives, none of the attribute levels should overlap. This simplified the identification of threshold-based choice behaviour in the SP responses. For example, were a respondent to adopt a rule of always choosing the alternative with the highest *rfr*, one would expect Car B to be chosen in each and every replication faced by that respondent.

The large numbers of attributes and levels not only precluded the use of a full factorial design, since this would have required 4^{24} replications, but were beyond the scope of conventional fractional factorial design plans such as (Kocur *et al.*, 1982). Design was instead carried out using the OPTEX and FACTEX procedures in the SAS software (SAS Institute Inc., 1999). The FACTEX procedure was used to construct an orthogonal fractional factorial design, which was supplied as a 'candidate' design to the OPTEX procedure. The OPTEX procedure was used to generate a saturated second-order design, that is, a design as small as possible but permitting estimation of all main effects and two-factor interactions. The search method employed was the modified Fedorov procedure described by Cook and Nachtsheim (1980) and the optimal design identified as the one maximising D-efficiency. The final SP design consisted of 25 replications.

Table 1: Attributes and levels represented in SP experiment

Attributes	Car A	Car B	Car C
orp (£ '000)	10, 10.5, 11, 11.5	12, 12.5, 13, 13.5	14, 15.5, 20, 30
rc (p/mile)	48, 50, 55, 75	40, 42, 44, 46	15, 20, 25, 35
rfr (miles)	350, 400, 450, 500	555, 600, 700, 800	50, 100, 150, 200
tfr (minutes)	8, 9, 10, 15	4, 5, 6, 7	30, 60, 180, 360
ts (mph)	110, 115, 120, 125	85, 90, 95, 100	50, 60, 70, 80
ac (seconds)	10, 10.5, 11, 11.5	12, 12.5, 13, 13.5	14, 16, 18, 20
rv (%)	30, 36, 38, 40	42, 44, 46, 48	50, 52, 55, 60
em (as % of year 2000 petrol car)	70, 75, 80, 95	30, 40, 50, 65	5, 10, 15, 20

3.2 Implementation

The SP experiment was implemented in the form of a two-part self-completion written questionnaire. The first part of the questionnaire posed questions principally about the respondent's household and its ownership and use of cars, although some socio-economic and demographic data were also elicited. The second part of the questionnaire presented the SP experiment. The rubric to the SP element of the questionnaire asked respondents to imagine that they were considering the purchase of a new car. Each replication of the design offered a choice between Cars A, B and C, each car being described in terms of the eight attributes.

Since many car buyers pay by credit, *orp* was presented with an equivalent credit price based on an interest rate of 10% (10.47% APR). For clarification, respondents were informed that *rc* included depreciation and maintenance as well as fuel, that *rv* was expressed as a percentage of the original on-the-road price, and that *em* was expressed as a percentage of those from an average petrol car in the year 2000. The questionnaire advised that, apart from the eight attributes explicitly specified, the three cars were identical in all other respects. For each question, respondents were given the option of choosing Car A, Car B, Car C or 'none of the three cars'. The SP design was split into five blocks of five replications, and a separate version of the questionnaire was developed for each block. Each of the five versions presented the attributes in a different order, thereby allowing the testing of order effects.

The design was tested in a small pilot study, which yielded encouraging results. This gave us the confidence to proceed to the main field study. During the Summer of 2002, 500 questionnaires (100 of each ordering) were delivered to homes in each of two locations: the Roundhay area of Leeds and the Pannal area of Harrogate (a commuter village 13.5 miles north of Leeds). Both areas are relatively affluent, with higher-than-average incomes and larger-than-average residential properties. Moreover, one would expect households residing in these areas to be pre-adapted to AFV ownership, with available funds to purchase and operate the vehicles, and available parking facilities to allow overnight home refuelling or recharging. Since many householders in Pannal commute to Leeds, the selection of Pannal as a survey location was motivated by an interest in how commuting behaviour, when coupled with the limited range of AFVs, might influence car choices. Along with the questionnaire, a pre-paid return envelope and covering letter were delivered. The covering letter described the background to the research project and advised that households without access to a car should ignore the questionnaire.

Finally, a small number of questionnaires were distributed to attendees of the focus group experiments. Since the focus groups involved more detailed investigation into the barriers to, and implications of, AFV ownership and use, distribution of SP questionnaires in this context allowed investigation into the effect of information on choice behaviour.

Of the 500 questionnaires distributed in Roundhay, 157 were returned, giving a response rate of 31.4%. Of the 500 questionnaires distributed in Pannal, 174 were returned, giving a response rate of 34.8%. All of the 29 questionnaires distributed in the focus groups were returned. Before any analysis was conducted, the data were cleaned to remove any non-usable choice observations i.e. either non-responses to the choice task or 'none of the three cars' responses. Following cleaning, there remained 595 observations from Roundhay, 639 observations from Pannal, 275 observations from the focus groups, and 1509 observations in total, for application to model estimation.

4. ESTIMATION

4.1 Multinomial logit, and analysis of order and error variances effects

Initial analysis involved the estimation of a single MNL model for the entire data without segmentation (Table 2)[2]. Perusal of Table 2 reveals that *orp*, *rc*, *rfr* and *tfr* are significantly different from zero at 1%, while *em* is significantly different from zero at 5%. The *t*-statistic of the *rfr* coefficient is particularly large. All of these parameters have the expected signs. For the remaining three attributes, *ts*, *ac* and *rv*, the null hypothesis of non-significance is not rejected at 5%. The above model was developed further to test for evidence of systematic biases arising from (i) the ordering of the attributes in the SP experiment and, (ii) differential error variance across sub-samples. No evidence of either effect was detected.

Table 2: MNL based on full data set following cleaning

Coefficient	Estimate (*t*-statistic)
orp	-0.0435 (-2.595)
rc	-0.0451 (-6.315)
rfr	0.0045 (20.596)
tfr	-0.0033 (-3.742)
ts	0.0112 (1.865)
ac	0.0067 (0.170)
rv	0.0192 (1.502)
em	-0.0094 (-2.411)
Observations	1509
Mean log-likelihood	-0.620713

4.2 Tree models, and analysis of decision process

The main focus of the modelling effort was to investigate the choice processes employed by SP respondents: in particular, the performance of HEBA as compared with MNL and NL. For a detailed theoretical analysis of HEBA, and its relationship with MNL and NL, the reader is referred to Batley and Daly (2003). Here we concentrate on issues of implementation and empirical performance.

Both NL and HEBA conceptualise the choice problem as a 'preference tree', with subsets of similar alternatives nested together. For the choice problem under consideration, each of the three permutations of the preference tree was analysed i.e.

Tree 1: Cars A and B nested together, with Car C represented in a single-alternative nest

Tree 2: Cars A and C nested together, with Car B represented in a single-alternative nest

Tree 3: Cars B and C nested together, with Car A represented in a single-alternative nest

A priori, we expected Tree 1 or Tree 3 to be better specifications that Tree 2. Tree 2 is the specification which has Car B on its own and Cars A and C nested together; yet in all respects, bar *rfr* and *tfr*, Car B performs at a level between Car A and Car C, as can be confirmed by perusal of Table 1. Moreover, Car A is adjacent to Car B throughout the design; Car C never comes between Car A and Car B. Thus, in terms of similarities, Car A is always more similar to Car B than it is to Car C. It is possible that Cars B and C could be nested together - this is an empirical issue. For Cars A and C to be nested together, either *rfr* is the dominating characteristic (the only one where Car A lies between Cars B and C <u>and</u> Car A is better than Car B) or there are kinks, holes and discontinuities in peoples' preferences. We would rather discount the latter possibility at this stage in our research.

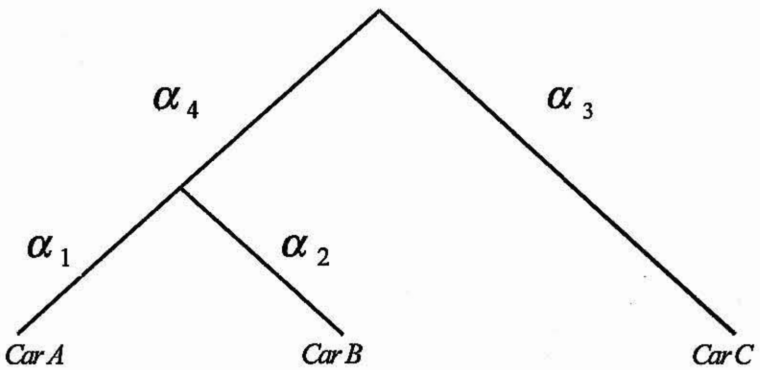

Figure 1: HEBA model for Tree 1

For illustration, we show the derivation of HEBA for Tree 1. With reference to Figure 1, it can be observed that aspect α_4 is common and unique to Cars A and B, and aspects α_1, α_2 and α_3 are unique to Cars A, B and C, respectively. The probabilities of choosing Cars A, B and C are, according to HEBA, given by the following:

$$P_A = \frac{u(\alpha_1)+u(\alpha_2)+u(\alpha_4)}{u(\alpha_1)+u(\alpha_2)+u(\alpha_3)+u(\alpha_4)} \times \frac{u(\alpha_1)}{u(\alpha_1)+u(\alpha_2)}$$

$$P_B = \frac{u(\alpha_1)+u(\alpha_2)+u(\alpha_4)}{u(\alpha_1)+u(\alpha_2)+u(\alpha_3)+u(\alpha_4)} \times \frac{u(\alpha_2)}{u(\alpha_1)+u(\alpha_2)}$$

$$P_C = \frac{u(\alpha_3)}{u(\alpha_1)+u(\alpha_2)+u(\alpha_3)+u(\alpha_4)}$$

where $u(\alpha_1)$, $u(\alpha_2)$, $u(\alpha_3)$ and $u(\alpha_4)$ are non-negative parameters relating to aspects α_1, α_2, α_3 and α_4, respectively.

In implementing the above specification, several issues arise. First, it is important to note Batley and Daly's (2003) finding that the model is over-specified by two degrees of freedom, such that the four parameters $u(\alpha_1)$, $u(\alpha_2)$, $u(\alpha_3)$ and $u(\alpha_4)$ cannot be estimated without appropriate restrictions. A second, more general, observation is that the above specification has limited usefulness for policy analysis. Tversky and Sattath (1979) asserted that HEBA: '*...represents choice alternatives as collections 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)'* (p543). Since the specification does not, however, represent the u's as an explicit function of policy variables such as orp, it is difficult to draw any clear inferences regarding the effects of changes in such variables on choice behaviour. McFadden (1981) addressed this limitation by suggesting that, with little loss in generality, the u's could be specified as log-linear functions of relevant policy variables.

Motivated by McFadden's suggestion, we developed three alternative specifications of the u's that overcome the identification problems referred to above. Before introducing these specifications, however, it is important to note a third and final implementation problem, which is the key finding of Batley and Daly (2003). Specifically, no matter how the u's are specified in HEBA, there always exists a valid NL model that is equivalent in terms of choice probability. Furthermore, equivalence can also, in principle, exist in the opposite direction, such that HEBA can be specified to be equivalent to NL. This latter equivalence does not, however, ensure that the 'equivalent HEBA' is a valid HEBA model, since this requires that all the u's are non-negative. In short, a valid NL equivalent of HEBA always exists, but whether the reverse exists is an empirical issue.

Specification 1

$$\begin{aligned} u(\alpha_1) &= \exp\left[V(\alpha_1) - \beta^* X_A^*\right] \\ u(\alpha_2) &= \exp\left[V(\alpha_2) - \beta^* X_B^*\right] \\ u(\alpha_3) &= \exp V(\alpha_3) \\ u(\alpha_4) &= \exp\left[\beta^* \max or \min\left(X_A^*, X_B^*\right)\right] \end{aligned} \tag{1}$$

Specification 2

$$\begin{aligned} u(\alpha_1) &= \exp V(\alpha_1) \\ u(\alpha_2) &= \exp V(\alpha_2) \\ u(\alpha_3) &= \exp V(\alpha_3) \\ u(\alpha_4) &= \exp V(\alpha_4) \end{aligned}$$

Specification 3

$$\begin{aligned} u(\alpha_1) &= \exp\left\{V(\alpha_1)/\mu\right\} \\ u(\alpha_2) &= \exp\left\{V(\alpha_2)/\mu\right\} \\ u(\alpha_3) &= \exp V(\alpha_3) \\ u(\alpha_4) &= \exp\left\{V(\alpha_4) + \mu \ln\left[\exp\left\{V(\alpha_1)/\mu\right\} + \exp\left\{V(\alpha_2)/\mu\right\}\right]\right\} - \left[u(\alpha_1) + u(\alpha_2)\right] \end{aligned}$$

where:

$$V(\alpha_1) = \beta_{orp} orp_A + \beta_{rc} rc_A + \beta_{rfr} rfr_A + \beta_{tfr} tfr_A + \beta_{ts} ts_A + \beta_{ac} ac_A + \beta_{rv} rv_A + \beta_{em} em_A$$
$$V(\alpha_2) = \beta_{orp} orp_B + \beta_{rc} rc_B + \beta_{rfr} rfr_B + \beta_{tfr} tfr_B + \beta_{ts} ts_B + \beta_{ac} ac_B + \beta_{rv} rv_B + \beta_{em} em_B$$
$$V(\alpha_3) = \beta_{orp} orp_C + \beta_{rc} rc_C + \beta_{rfr} rfr_C + \beta_{tfr} tfr_C + \beta_{ts} ts_C + \beta_{ac} ac_C + \beta_{rv} rv_C + \beta_{em} em_C$$
$$V(\alpha_4) = \beta_{nsc}$$

, and the β 's are parameters to be estimated, μ is the log sum parameter in NL, and nsc denotes 'nest-specific constant'.

Following McFadden (1981), the three specifications express the u's as log-linear functions of the SP design variables. The three specifications differ principally in their representation of $u(\alpha_4)$, which relates to the aspects common to the nested alternatives. Specification 1 is motivated by an interest in the prevalence of threshold-based choice heuristics, representing $u(\alpha_4)$ as a function of the maximum or minimum value of a given attribute in the nest (Equation 1). If there were an interest, for example, in the propensity of respondents to filter alternatives according to a maximum orp threshold, $u(\alpha_4)$ would be specified as a function of the maximum orp in the nest, β^* would be estimated as the coefficient of this threshold, and orp would be removed from the functions $u(\alpha_1)$ and $u(\alpha_2)$. Specification 2 stays closer to Tversky and Sattath's (1979) original specification, representing $u(\alpha_4)$ simply as a nest-specific constant, and all design variables remaining specific to $u(\alpha_1)$, $u(\alpha_2)$ and $u(\alpha_3)$. With reference to Batley and Daly (2003), Specification 3 expresses $u(\alpha_4)$ in the form that achieves equivalence with a 'Classic' McFadden (1978) NL specification. It should be noted, however, that equivalent (perfectly valid) NL forms also exist for each of Specifications 1 and 2. With reference to earlier discussion on NL-HEBA equivalence, Specifications 1 and 2 are always HEBA-consistent. As regards Specification 3, $u(\alpha_1)$, $u(\alpha_2)$ and $u(\alpha_3)$ always satisfy the requirement of non-negativity, but overall HEBA consistency rests on the sign of $u(\alpha_4)$, which is non-negative provided the following inequality holds:

$$\exp V(\alpha_4) > [u(\alpha_1) + u(\alpha_2)]^{1-\mu} \qquad (2)$$

Table 3: HEBA models for Specification 1

	Tree 1	Tree 2	Tree 3
Coefficient	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)
orp	-0.0485 (-2.896)	-0.0477 (-2.735)	**-0.1774 (-6.290)**
rc	-0.0485 (-6.587)	-0.0397 (-4.637)	-0.0333 (-4.538)
rfr	0.0037 (6.102)	**0.0029 (11.944)**	0.0023 (5.324)
tfr	-0.0024 (-2.807)	-0.0030 (-3.287)	-0.0030 (-3.594)
ts	**0.0315 (3.379)**	0.0325 (6.303)	0.0350 (5.144)
ac	-0.0125 (-0.371)	-0.0088 (-0.219)	-0.0843 (-2.025)
rv	0.0160 (1.550)	0.0291 (2.582)	-0.0005 (-0.041)
em	-0.0109(-2.554)	-0.0068 (-1.680)	-0.0017 (-0.419)
Observations	1509	1509	1509
Mean log-likelihood	-0.610981	-0.615731	-0.609700

For each tree structure, Specification 1 was estimated eight times, each time specifying a different design variable in $u(\alpha_4)$. The above tables show only the preferred Specification 1 models for each tree. The bold type indicates which design attribute was specified in $u(\alpha_4)$ of the preferred model; evidently this differs across the three models, with *ts* represented in Tree 1, *rfr* in Tree 2 and *orp* in Tree 3. The selection of these preferred models was influenced by several criteria including overall explanatory power, the significance of the parameter estimates, whether the signs of the estimated parameters accorded with intuition, and whether the $u(\alpha_4)$ variable was consistent with the relevant tree structure. To elaborate on the final criteria, reference to Table 1 reveals that the *ts* of Car C was always higher that that of either Car A or Car B. Thus a *ts* threshold would create a dichotomy consistent with Tree 1. Note that Tree 2 is the case where we expected a good model only if *rfr* were a dominant attribute. Not surprisingly, therefore, the best performing Tree 2 model is that with *rfr* in the upper part of the tree.

All three of the Specification 1 models presented demonstrate similar fit. In all three cases, six of the eight estimated parameters are significantly different from zero at 5%; indeed the pattern of significance is similar across the three models, with the first five coefficients always significant; and the signs of the significant parameters are as expected. The significant parameters are, for most design variables, of a similar order of magnitude across the three specifications, although an obvious exception to this is *orp* in Tree 3. Tree 3 demonstrates the best fit of the three models in Table 3. Finally, it should be noted that, with the same number of parameters, each of the Specification 1 models shows superior fit to the MNL reported in Table 2.

Table 4: HEBA models for Specification 2

	Tree 1	Tree 2	Tree 3
Coefficient	Estimate (*t*-value)	Estimate (*t*-value)	Estimate (*t*-value)
orp	-0.0492 (-2.930)		-0.0469 (-2.731)
rc	-0.0493 (-6.592)		-0.0340 (-4.413)
rfr	0.0032 (5.942)		0.0025 (5.009)
tfr	-0.0026 (-3.038)		-0.0025 (-2.780)
ts	0.0103 (1.670)		0.0302 (4.022)
ac	0.0358 (0.893)		-0.0603 (-1.393)
rv	0.0344 (2.509)		0.0054 (0.391)
em	-0.0125 (-2.613)		-0.0045 (-0.955)
nsc	2.8812 (2.447)		3.1958 (2.513)
Observations	1509	1509	1509
Mean log-likelihood	-0.613430	FTC[3]	-0.606218

Unlike Specification 1, implementation of Specifications 2 and 3 required the estimation of only a single model for each tree structure. Turning to Specification 2, a converged model was estimated only for Trees 1 and 3; our assessment is therefore restricted to these two models (Table 4). Tree 3 demonstrates superior fit to Tree 1, but the latter yields seven significant parameters at 5% - one more than Tree 3. The *nsc* is significantly different from zero at 5% in both models. The signs of all significant parameters are intuitive, and the magnitudes of the significant parameters show a degree of consistency across the two models.

Comparing the three models based on Specification 3 (Table 5), it is apparent that only Tree 3 yields a log sum parameter that is significantly less than one at 5%. Thus, in both Trees 1 and 2, NL collapses to MNL, implying that no significant correlation between the nested alternatives exists. Tree 2 offers the best fit of the three models, although Tree 1, with eight, yields the greatest number of significant parameters at 5%. The signs of all significant parameters accord with intuition.

Table 5: HEBA models for Specification 3

	Tree 1	Tree 2	Tree 3
Coefficient	Estimate (t-value)	Estimate (t-value)	Estimate (t-value)
orp	-0.0464 (-2.737)	-0.0462 (-2.741)	-0.0198 (-1.687)
rc	-0.0490 (-3.783)	-0.0378 (-4.771)	-0.0237 (-3.276)
rfr	0.0026 (3.529)	0.0007 (1.169)	0.0014 (3.086)
tfr	-0.0027 (-3.022)	-0.0017 (-2.164)	-0.0011 (-1.828)
ts	0.0068 (1.096)	0.0227 (3.555)	0.0173 (2.437)
ac	0.0338 (0.852)	-0.0332 (-0.922)	-0.0440 (-1.689)
rv	0.0293 (2.103)	0.0175 (1.413)	-0.0007 (-0.076)
em	-0.0138 (-2.798)	-0.0068 (-1.780)	-0.0031 (-1.085)
nsc	1.7122 (3.569)	-1.5348 (-5.769)	1.7772 (8.520)
log sum	0.9576 (4.125)	0.8537 (4.014)	0.5038 (2.910)
Observations	1509	1509	1509
Mean log-likelihood	-0.616188	-0.605362	-0.606310

Finally, a comparison of each specification across tree structures provokes the following comments. As regards explanatory power, Specification 1 demonstrates the best fit for Tree 1 despite having the fewest parameters; similarly, Specification 2 demonstrates superior fit to Specification 3 for Tree 3 despite having one less parameter. Thus, Specifications 1 and 2 achieve, in some cases, better fit than Classic NL. The magnitudes of the significant estimated parameters show a reasonable degree of consistency across the specifications for any given tree structure, with some exceptions; in particular *nsc* in Trees 1 and 3 and *orp* in Tree 3.

A further observation is that the different specifications show some differences in their patterns of parameter significance for a given tree. In particular, the high degree of significance of the $u(\alpha_4)$ parameters in each of the Specification 1 models (i.e. *ts, rfr* and *orp* in Trees 1, 2 and 3, respectively) contrasts with the insignificance (at 5%) of the same variables in their corresponding Specification 3 models. Thus, the application of Specifications 1 and 3 could yield quite different inferences regarding the influence of policy variables on demand.

With reference to earlier discussion of NL-HEBA equivalence, it is insightful to consider the compliance of each of the Specification 3 (i.e. Classic NL) models with the requirements for HEBA. This can be investigated by means of applying the estimated parameters to the experimental design data and assessing the non-negativity (or otherwise) of $u(\alpha_4)$ for each replication (equation 2). Such an investigation revealed that:

- For Tree 1, HEBA holds on all 25 replications of the design
- For Tree 2, HEBA holds on none of the 25 replications
- For Tree 3, HEBA holds on 9 out of the 25 replications

Moreover, in the case of Tree 1, NL collapses to MNL, but HEBA always holds; in the case of Tree 2, NL collapses to MNL, but HEBA never holds; and in the case of Tree 3, NL holds, while HEBA holds on a proportion of the design replications.

Before turning to the forecasting capabilities of the different tree structures and model specifications, we consider why HEBA sometimes holds and sometimes does not. It will be recalled that equation 2 is the crucial determinant of whether HEBA holds under Specification 3. It can clearly be seen from equation 2 that in the special case where $\mu = 1$, the critical condition is $V(\alpha_4) > 0$. $V(\alpha_4)$ is the constant nsc. So for Trees 1 and 2, where $\mu \approx 1$, the validity of the HEBA model turns on the sign of nsc. In Tree 1, it is positive and thus the HEBA model is valid for all replications (as stated above). In Tree 2, it is negative and thus the HEBA model is invalid for all replications. Re-estimations of all three specifications constraining $\mu = 1$ are reported in Table 6, where it can be clearly seen that nsc remains positive (negative) for Tree 1 (Tree 2).

In the cases where $\mu \neq 1$, the story is a little more complicated. Whether or not a model is HEBA-compliant on any given replication depends on:

(i) the relative shares of the two alternatives nested together;

(ii) μ;

(iii) the actual value of the utility which drives the within-nest relative shares i.e. excluding any utility common to the nested alternatives.

Consider Tree 3. Table 5 reports a perfectly respectable model using Specification 3, albeit for application purposes we would want to do something with the insignificant parameters. The question is in what circumstances this valid NL model is also HEBA-compliant. The necessary and sufficient condition, equation 2, can be restated as:

$$V(\alpha_4) > \left\{(1-\mu)^{V(\alpha_2)}\big/_{\mu}\right\} + \left\{(1-\mu)\ln\left(1+{}^{P_C}\big/_{P_B}\right)\right\}$$

or

$$V(\alpha_4) > \left\{(1-\mu)^{V(\alpha_3)}\big/_{\mu}\right\} + \left\{(1-\mu)\ln\left(1+{}^{P_B}\big/_{P_C}\right)\right\}$$

If equation 2 holds, both of these also hold since they are directly equivalent statements. For interpretation purposes it is helpful to consider which is the greater of $V(\alpha_2)$ and $V(\alpha_3)$, and hence which is greater of P_B and P_C, and use the first equation if $P_B > P_C$ and the second otherwise. When we do that, the ratio of the probabilities has a maximum value of one and a sufficient but not necessary condition for the model to be HEBA-compliant is:

$$V(\alpha_4) > \left\{(1-\mu)^{V(\alpha_m)}\!\Big/_{\mu}\right\} + \{(1-\mu)\ln(2)\}$$

where $V(\alpha_m) = \max\{V(\alpha_2), V(\alpha_3)\}$

If $V(\alpha_m)$ is sufficiently negative, and in the extreme case for $P_B = P_C$ the requirement is $V(\alpha_m) < -\mu\ln(2)$, then $V(\alpha_4) > 0$ is sufficient for HEBA-compliance. As $V(\alpha_m)$ becomes more negative, so the possibility of even negative $V(\alpha_4)$ being HEBA-compliant increases. Of course, for specific applications where the choice probabilities of the nested alternatives are known, we can refine the sufficient condition for HEBA-compliance so that it becomes necessary and sufficient.

This has interesting implications for our understanding of the relationship between NL and HEBA. Whereas HEBA requires the common aspect of nested alternatives to be positive, it is perfectly feasible for this to be reflected in negative utility of the common aspect when the model is framed as the equivalent NL model. In other words, when nested alternatives share some utility relative to the non-nested alternative, that commonality can be either a bad or a good *ceteris paribus*, although in practice it seems that the commonality being a good increases the chances of the NL model being HEBA-compliant.

In our particular case, 9 of the 25 replications were found to be HEBA-compliant; in other words the common part of utility, *nsc* in the Tree 3 model in Table 5, was sufficiently positive to outweigh the specific utility of the preferred choice out of Cars B and C plus the ratio of probabilities term. Note that the parameters on *rfr* and *ts* are positive. Thus better performance of the preferred alternative on these two aspects reduces the chance of the condition being met. In practice, it also turned out that the non-HEBA replications were those where Car B had higher levels of range and top speed.

Table 6: HEBA models for ('forced' MNL) Specification 3

	Tree 1	Tree 2	Tree 3
Coefficient	Estimate (*t*-value)	Estimate (*t*-value)	Estimate (*t*-value)
orp	-0.0461 (-2.703)	-0.0503 (-2.968)	-0.0498 (-2.963)
rc	-0.0511 (-6.824)	-0.0410 (-5.655)	-0.0325 (-4.376)
rfr	0.0027 (4.738)	0.0006 (1.072)	0.0020 (4.547)
tfr	-0.0026 (-2.998)	-0.0019 (-2.158)	-0.0024 (-2.745)
ts	0.0068 (1.107)	0.0239 (3.724)	0.0307 (4.372)
ac	0.0318 (0.794)	-0.0373 (-0.938)	-0.0700 (-1.687)
rv	0.0299 (2.280)	0.0188 (1.450)	0.0044 (0.334)
em	-0.0146 (-3.448)	-0.0064 (-1.618)	-0.0008 (-0.203)
nsc	1.7283 (3.582)	-1.6324 (-6.729)	1.9225 (6.135)

log sum	1.0	1.0	1.0
Observations	1509	1509	1509
Mean log-likelihood	-0.616199	-0.605476	-0.607775

When we estimated a MNL for Specification 3 (Tree 3 in Table 6), we found that *nsc* was positive. In other words, falsely estimating MNL when in fact the data support NL could lead to the erroneous conclusion that the data were consistent with HEBA when that is only partly true. Thus, whereas only a minority of replications were HEBA-compliant, forcing an all-or-nothing estimation can overstate the case for compliance. The question remains as to how many HEBA-compliant replications could be construed as being overall broadly HEBA-consistent.

5. FORECASTING

Following estimation, the models reported in section 4.2 were compared in a forecasting context. Here we investigated the ability of the three HEBA specifications to accurately forecast observed market shares for 'extreme' attribute data i.e. attribute levels, such as a relatively high *orp*, that might be particularly susceptible to threshold-based heuristics. In this regard, it was decided that Trees 1 and 3 represented the most interesting cases, and the analysis was therefore focused on these structures.

With reference to Tree 1, thresholds relating to all eight of the design variables appeared consistent with the chosen tree structure. Similarly, with reference to Tree 3, thresholds relating to six of the design variables - all except *rfr* and *tfr* - appeared consistent with the chosen tree structure. For each of these attributes, a threshold was applied to the data set, dichotomising the observations into those passing or failing the threshold (Tables 7 and 8 for Trees 1 and 3, respectively). The modelling reported above was then repeated for the larger of the two-sub-samples in each case, retaining the smaller sub-sample as a holdback sample. Each re-estimated model was then applied to forecasting, using the holdback sample and sample enumeration techniques. In the Specification 1 models, the appropriate 'threshold' attribute from Tables 7 and 8 was specified in $u(\alpha_4)$.

Table 7: Thresholds applied to Tree 1, and numbers of passes and fails

Attribute	Threshold	No. Passes	No. Fails
orp (£ '000)	< 13.5	1088 (Model)	421 (Holdback)
rc (p/mile)	< 35	941 (Model)	568 (Holdback)
rfr (miles)	> 150	359 (Holdback)	1150 (Model)
tfr (mins)	< 15	1102 (Model)	407 (Holdback)
ts (mph)	> 85	1268 (Model)	241 (Holdback)
ac (seconds)	< 13.5	1081 (Model)	428 (Holdback)
rv (%)	> 50	1212 (Model)	297 (Holdback)
em (as % of year 2000 petrol car)	< 20	1021 (Model)	488 (Holdback)

Table 8: Thresholds applied to Tree 3, and numbers of passes and fails

Attribute	Threshold	No. Passes	No. Fails

orp (£ '000)	< 11.5	1026 (Model)	483 (Holdback)
rc (p/mile)	< 50	436 (Holdback)	1073 (Model)
ts (mph)	> 120	368 (Holdback)	1141 (Model)
ac (seconds)	< 11.5	1206 (Model)	303 (Holdback)
rv (%)	> 38	482 (Holdback)	1027 (Model)
em (as % of year 2000 petrol car)	< 75	363 (Holdback)	1146 (Model)

The observed and forecast market shares for Trees 1 and 3 are reported in Tables 9 and 10. To assess the relative performance of the different trees and specifications, we conducted a series of χ^2 tests. For Tree 1, 8 out of 24 sets of forecasts varied significantly from the observed data and for Tree 3, 5 out of 18. In all six cases when *rc* was the variable used to create the holdback sample, the forecasts were significantly different. This is despite segmentation by *rc* never featuring as the preferred Specification 1 model in Table 3 and *rc* not being implicated in the failure of a Specification 3 model to achieve HEBA-compliance in Table 5. We have no explanation for this outcome. Other than this, there is no obvious pattern to the ability of the three HEBA specifications to forecast observed market shares accurately. It would therefore seem that it is possible to apply HEBA models to forecast beyond the range on which they were calibrated without serious deleterious implications and that the precise specification of the HEBA model has no effect on this. Even in the *rc* case, where to do so has proven deleterious, we have no reason *a priori* or *ex post* why this is so. We thus acknowledge that further work remains necessary in this area.

Table 9: Forecast market shares for holdback samples in Tree 1

	Observed			Specification 1			Specification 2			Specification 3		
	A	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
orp	0.11	0.79	0.10	0.12	0.79	0.09	0.12	0.79	0.09	0.13	0.78	0.09
rc	0.12	0.84	0.05	0.14	0.75	0.11	0.16	0.78	0.06	0.15	0.77	0.08
rfr	0.11	0.76	0.13	0.13	0.77	0.10	0.13	0.75	0.11	0.13	0.74	0.14
tfr	0.09	0.84	0.07	0.09	0.83	0.08	0.09	0.83	0.08	0.09	0.83	0.08
ts	0.12	0.83	0.05	0.09	0.84	0.07	0.09	0.83	0.09	0.09	0.82	0.09
ac	0.14	0.75	0.10	0.16	0.76	0.08	0.16	0.75	0.09	0.16	0.76	0.08
rv	0.09	0.78	0.13	0.09	0.81	0.10	0.10	0.84	0.06	0.10	0.83	0.06
em	0.11	0.84	0.05	0.12	0.74	0.14	0.13	0.79	0.07	0.13	0.78	0.08

Table 10: Forecast market shares for holdback samples in Tree 3

	Observed			Specification 1			Specification 2			Specification 3		
	A	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
orp	0.09	0.83	0.07	0.12	0.79	0.09	0.12	0.79	0.09	0.13	0.78	0.09
rc	0.15	0.75	0.10	0.10	0.82	0.08	0.10	0.82	0.08	0.10	0.82	0.07
ts	0.11	0.77	0.12	0.11	0.81	0.08	0.12	0.80	0.08	0.12	0.80	0.08
ac	0.14	0.78	0.08	0.11	0.81	0.08	0.11	0.82	0.08	0.11	0.81	0.08
rv	0.13	0.79	0.08	0.10	0.82	0.08	0.10	0.82	0.08	0.11	0.81	0.08
em	0.11	0.83	0.06	0.13	0.78	0.09	0.12	0.79	0.10	0.12	0.79	0.09

6. SUMMARY AND CONCLUSION

In this paper, we have sought to estimate and apply a series of models which were, at least potentially, theoretically consistent with both NL and HEBA formulations as outlined in Batley and Daly (2003). This has involved, in a simple three-alternative case, testing each of the three possible tree structures using three different HEBA-type specifications. The first two specifications were guaranteed to generate valid HEBA models, from which it is always possible to derive an equivalent NL model, while the third specification was an NL model which was then tested for HEBA-compliance. We have developed and presented a rule for testing HEBA-compliance and also given interpretation to enable greater understanding of the relationship between HEBA and NL models.

We found no substantial difference between the performance of HEBA models which allowed for the possibility of respondents to filter alternatives according to a maxmimum or minimum performance threshold on any given attribute and a simpler formulation. However, all our preferred models of these two types performed better than a standard MNL using a criterion of maximum mean log-likelihood.

When estimating NL models, we found that in two of the three tree structures NL collapsed to MNL. This simplifies the assessment of HEBA-compliance. Interestingly, in one of these two cases, even though NL suggested that a tree structure was not appropriate since the log sum parameter was insignificantly different from one, the same model was a valid HEBA model which does imply that a tree structure is appropriate. A precise understanding of 'treeness' thus remains elusive, although Batley and Daly (2003) offer some insight on this issue.

Where the question of HEBA-compliance was dependent on a log sum parameter as well as on some common utility, we found that it is perfectly possible to have common bads defining a NL tree structure which translate into a shared positive aspect under a HEBA formulation. The more two nested alternatives are different from the third in NL terms (i.e. log sum closer to zero), the greater the chance of shared disutility still being consistent with HEBA, *ceteris paribus*.

As regards the forecasting performance of different model specifications and tree structures, we found it difficult to elucidate any substantive difference between rival forms, at least not in ways we might have expected, nor in ways which we could subsequently explain. While this finding requires further exploration and testing, it does perhaps suggest that it is at least possible in some circumstances to apply HEBA models beyond the range on which they were calibrated. The circumstances where this does or does not hold remain, as yet, the subject of speculation.

NOTES

[1] We define an AFV as any vehicle that can be used to partially or fully replace conventional petrol or diesel fuel.

[2] All models presented in the paper were estimated by maximum likelihood methods using self-written GAUSS code (Aptech Systems Inc. 1996a, 1996b, 1996c).
[3] FTC denotes 'failed-to-converge'.

ACKNOWLEDGEMENTS

Different stages of the work reported above were funded by the UK Economic and Social Research Council under research grant R000223244, and by the UK Department for Transport under New Horizons research project B36. The first-named author would also like to acknowledge the personal support of a Departmental Research Fellowship at the Institute for Transport Studies, University of Leeds.

REFERENCES

Aptech Systems Inc. (1996a) GAUSS Mathematical and Statistical System, Volume I, System and Graphics Manual.

Aptech Systems Inc. (1996b) GAUSS Mathematical and Statistical System, Volume II, Command Reference.

Aptech Systems Inc. (1996c) GAUSS Maximum Likelihood Application Module.

Batley, R. and Daly, A. (2003) Establishing equivalence between nested logit and hierarchical elimination-by-aspects choice models. Paper to be presented at the European Transport Conference, Strasbourg, October 2003.

Batley, R., Knight, M. and Toner, J. (2003) A mixed logit model of the demand for alternative-fuel vehicles. Paper presented at the 35th Annual Conference of the Universities Transport Study Group, Loughborough University, January 2003.

Ben-Akiva, M.E. (1974) Structure of passenger travel demand models. *Transportation Research Record*, 526, pp26-42.

Bierlaire, M. (2002) The Network GEV model. Paper presented at the Swiss Transport Research Conference 2002, Monte Verita.

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.

Bolduc, D. and McFadden, D. (2001) Methodological developments in travel behaviour modelling. In Hensher, D. (ed) *Travel Behaviour Research: The Leading Edge*. Pergamon, Amsterdam.

Brownstone, D. and Train, K. (1999) Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics*, 89, pp109-129.

Brownstone, D., Bunch, D. and Train, K. (2000) Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles. *Transportation Research B*, 34, pp315-338.

Camerer, C. (1998) Bounded rationality in individual decision-making, *Experimental Economics*, 1 (2), pp163-183.

Cleaner Vehicles Task Force (CVTF) (2000) The report of the Alternative Fuels Group of the Cleaner Vehicles Task Force: An assessment of the emissions performance of alternative and conventional fuels. Available at: http://www.autoindustry.co.uk/library/books_reports/74290.pdf [31/03/03].

Cook, R.D. and Nachtsheim, C.J. (1980) A comparison of algorithms for constructing exact D-optimal designs. *Technometrics*, 22, p315-324.

Daly, A. (2001) Recursive Nested Extreme Value Model. Working Paper 559, Institute for Transport Studies, University of Leeds.

Daly, A. and Bierlaire, M. (2003) A general and operational representation of GEV models. Technical Report RO-030502, Institute of Mathematics, Operations Research Group ROSO, EPFL, Lausanne, Switzerland.

Daly, A.J. and Zachary, S. (1976) Improved multiple choice models. In *Proceedings of the Fourth PTRC Summer Annual Meeting*. PTRC, London.

Fowkes, A.S. (1991) Recent developments in stated preference techniques in transport research. In *PTRC Transportation Planning Methods*, Proceedings of Seminar G, PTRC 19th Summer Annual Meeting, pp251-263. PTRC, London.

Hargreaves Heap, S. (1992) Rationality. In Hargreaves Heap, S., Hollis, M., Lyons, B., Sugden, R. and Weale, A. (eds) *The Theory of Choice: A Critical Guide*. Blackwell, Oxford.

Hensher, D.A. and Greene, W.H. (2001) Choosing between conventional, electric and LPG/CNG vehicles in single-vehicle households. Working Paper ITS-WP-01-05, The University of Sydney.

Hodgson, F.C. (2002) Estimating household demand for alternative-fuel vehicles: report of focus group analysis. Internal project report, Institute for Transport Studies, University of Leeds.

Kahneman, D. and Tversky, A. (2000) *Choices, Values and Frames*. Cambridge University Press, Cambridge.

Kocur, G., Adler, T., Hyman, W. and Audet, E. (1982) Guide To Forecasting Travel Demand With Direct Utility Measurement. UMTA, USA Department of Transportation, Washington D.C.

Koppelman, F. and Wen, C.H. (2001) Alternative nested logit models: a response to comments by Andrew Daly on an earlier paper of Frank Koppelman and Chieh-hua Wen. *Transportation Research B*, 35, pp725-729.

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.

McFadden, D. (1999) Rationality for economists? *Journal of Risk and Uncertainty*, 19 (1-3), pp73-105.

McIntosh, E. and Ryan, M. (2002) Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: implications of discontinuous preferences. *Journal of Economic Psychology*, 23, pp367-382.

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.

Powershift (2003) Homepage. Available at: http://www.powershift.org.uk/ [01/04/03].

Saelensmide, K. (1999) The impact of choice inconsistencies in stated choice studies, in Valuation of non-market goods for use in cost-benefit analyses: methodological issues. Dissertation for the degree of Doctor Scientiarum, Department of Economics and Social Sciences, Agricultural University of Norway.

SAS Institute Inc. (1999) *SAS/STAT User's Guide, Version 8*. SAS Institute, Cary, NC.

Simon, H.A. (1955) Behavioural models 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.

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.

Williams, H.C.W.L. (1977) On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9 (3), pp285-344.