



ITLS

WORKING PAPER
ITLS-WP-05-16

**The implications on willingness
to pay of a stochastic treatment
of attribute processing in stated
choice studies**

By

David A Hensher, John Rose &
Tony Bertoia

June 2005

Design of Designs Paper #3
Forthcoming Transportation Research D
Research funded under the Australian Research
Council Large Grants Scheme, Grant A00103962

ISSN 1832-570X

**INSTITUTE of TRANSPORT and
LOGISTICS STUDIES**

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER: Working Paper ITLS-WP-05-16

TITLE: The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies

ABSTRACT: Individuals processing the information in a stated choice experiment are asked to evaluate a set of attributes offered and to choose their most preferred alternative. It has always been thought that some attributes are not attended to in this process for many reasons, including a coping strategy to handle their perception of the complexity of the choice task. However analysts proceed, with rare exception, to estimate discrete choice models as if all attributes have influenced the outcome *to some degree*. In this paper we investigate the implications of bounding the attribute processing task by attribute elimination through not attending to one or more attributes. Using a sample of car non-commuters in Sydney we estimate a mixed logit model in which all attributes are assumed to be attended to, and models which assume that certain attribute(s) are not attended to, based on supplementary information provided by respondents. The supplementary information is accounted for in a deterministic and a stochastic way; the latter in recognition of the analyst's lack of full information on why a specific attribute processing (AP) strategy was adopted by each sampled individual. We compare the value of travel time savings distribution under alternative attribute processing regimes. As expected, there are noticeable variations in the mean and standard deviation willingness to pay (WTP) across the three AP strategies.

KEY WORDS: *Stated choice designs, attribute processing, willingness to pay*

AUTHOR/S: David Hensher, John Rose & Tony Bertoia

CONTACT: Institute of Transport and Logistics Studies (C37)
An Australian Key Centre
The University of Sydney NSW 2006 Australia

Telephone: +61 9351 0071
Facsimile: +61 9351 0088
E-mail: itlsinfo@itls.usyd.edu.au
Internet: <http://www.itls.usyd.edu.au>

DATE: June 2005

1. Introduction

In the majority of SC studies it is assumed that all attributes are deemed relevant (i.e., are attended to in varying degrees) in the assessment of the alternatives. To what extent might individuals adopt differing attribute processing (AP) strategies either to cope with the ‘complexity’ of an SC experiment and/or because specific attributes are not relevant in their choice? It is reasonable to propose that individuals do have a variety of AP styles, including the *simplifying* strategy of ignoring certain attributes (for whatever reason). Heterogeneity in AP strategies is widely reported in consumer research (see for example Hensher, 2004; DeShazo and Fermo, 2004) and its existence in choice experiments is supported by observation of lexicographic choice behaviour in segments of respondents completing SP surveys (see for example, Saelendsminde, 2002)¹. Failure to account for such an AP strategy is tantamount to the imposition of the assumption that all designs are comprehensible, all design attributes are relevant (to some degree) and the design has accommodated the relevant amount of ‘complexity’ necessary to make the choice experiment meaningful (Hensher, et al., 2005).

The (implicit) assumption in SC studies that all attributes are processed by all respondents has been challenged by a number of researchers (e.g., DeShazo and Fermo, 2004, Hensher, 2004, *in press*, Hensher, et al., 2005) who argue that it is more likely that individuals react to increasingly ‘complex’ choice situations by adopting one of two AP strategies, broadly defined by the rival *passive bounded rationality* and *rationality-adaptive* behavioural models. Under the passive bounded rationality model, individuals are thought to continue assessing all available attributes, however, do so with increasing levels of error in the processing of that information as choice complexity increases (de Palma, et al. 1994). The rationally-adaptive model assumes that individuals recognise that their limited cognition may have positive opportunity costs and react accordingly. As DeShazo and Fermo (2004) state: “Individuals will therefore allocate their attention across alternative-attribute information within a choice set in a rationally-adaptive manner by seeking to minimise the cost and maximise the benefit of information evaluation” (page 3).

It is important to recognise that simplistic designs may be ‘complex’ in a perceptual sense, since an individual expects more information which they know is relevant in making such a choice in a real market setting. The development of a stated choice experiment, supplemented with questions on how an individual processed the information, enables the researcher to explore sources of systematic influences on choice that if not attended to, can lead to biases in key outputs such as willingness to pay.

In Hensher, et al. (2005), we treated the exogenous information of attribute inclusion/exclusion deterministically. We assumed that the analyst knows for certain which attributes are used by which respondents. More realistically, the exogenous

¹ Significant research effort has been expended on how to optimise the outputs derived from respondents completing choice tasks derived from these single design plans generated using statistical design theory (e.g., Bunch, et al., 1994; Huber and Zwerina, 1996; Kanninen, 2002; Kuhfeld, et al., 1994; Lazari and Anderson, 1994; Sandor and Wedel, 2001), whilst minimising the amount of cognitive effort required of respondents (e.g., Louviere and Timmermans, 1990; Oppewal, et al., 1994; Wang, et al., 2001).

information points to the correct likelihood specification for a respondent with error, so that the likelihood for a respondent is a probabilistic mixture of likelihoods. In this paper we set out two models to investigate the empirical implications of assuming knowledge of the respondent-level likelihood of attribute processing with certainty, and to contrast this with a stochastic specification where the deterministic assumption is relaxed. In the stochastic model the exogenous covariate is probabilistically related to the structural heterogeneity specification, through an expected maximum utility index derived from a choice of attribute processing strategy model, conditioning the preference heterogeneity distribution for each random parameter associated with the attributes of the SC model.

Using a 2004 sample of car non-commuters in Sydney we estimate mixed logit models which assume that all attributes are candidate contributors, and models which assume that certain attributes are not attended to (based on supplementary information provided by respondents), taking into account the treatments that the analyst might select to represent their knowledge of the sampled population's true AP rule. We derive individual-respondent parameters (strictly parameter estimates drawn randomly from a common-choice distribution), using mixed logit, and compare the value of travel time savings distribution under alternative information processing regimes.

The paper is organised as follows. The next section draws on the extant literature to support treatments of attribute processing. We then set out the particular specification of the mixed logit that incorporates the various specifications of AP. A brief overview of the empirical data is then given followed by the set of model results comparing the traditional full attention to all attributes SC model with the models that account for the way that attributes are attended to by each individual. The substantive implications of the analysis, especially the variation in values of travel time savings, are set out followed by some conclusions and directions for ongoing research.

2. Behavioural Advantages of Using a Stochastic Specification of Attribute Processing

There is widespread evidence in the psychology literature concerning the behavioural variability, unpredictability and inconsistency regularly demonstrated in decision making and choices (e.g., Gonzales-Vallejo, 2002; Slovic, 1995), reflecting an assumption that goes back at least to Thurstone's law of comparative judgment (1927). One of the particularly important advantages of using a stochastic representation of decision strategies, as promoted herein, is that it enables a more behaviourally realistic analysis of variation in decision strategies.

There is a substantial extant literature in the psychology domain in regards to the influence of various factors on the amount of information processed in decision tasks. Recent evidence demonstrates the importance of such factors as time pressure (e.g., Diederich, 2003), cognitive load (e.g., Drolet and Luce, 2004), and task complexity (Swait and Adamowicz, 2001) in influencing the decision strategy employed during complex decision tasks. There is also a great deal of variability in decision strategies employed in different contexts, and this variability adds to the complexity in understanding the behavioural mechanisms involved in decision making and choice. A

recent attempt to define a typology of decision strategies (e.g., Payne, et al., 1992) has been particularly useful in providing a framework within which to understand decision strategies.

Payne, et al. (1992) characterised decision strategies along three dimensions: basis of processing, amount of processing, and consistency of processing. Decision strategies are said to differ in regards to whether many attributes within an alternative are considered before another alternative is considered (*alternative-based* processing) or whether values across alternatives on a single attribute are processed before another attribute is processed (*attribute-based* processing). Strategies are also said to differ in terms of the amount of information processed (i.e. in terms of whether any information is ignored or not processed before a decision may be made). Finally, decision strategies can also be grouped in terms of whether the same amount of information for each alternative is examined (*consistent* processing) or whether the amount of processing varies depending on the alternative (*selective* processing).

On the basis of this typology, Payne, et al. (1992) identified six specific decision strategies, three of which are attribute-based and three alternative-based approaches. The attribute-based approaches included the elimination-by-aspects (EBA), lexicographic choice (LEX), and majority of confirming dimensions (MCD) strategies. The alternative-based approaches included the weighted additive (WADD), satisficing (SAT), and equal-weight (EQW) strategies. These strategies are further described in Table 1 below. See Payne, et al. (1992) for a full description of these strategies. The main argument posited by Payne, et al. (1992) was that individuals construct strategies depending on the task demands and the information they are faced with.

Table 1: Typology of Decision Strategies (Payne, et al., 1992)

Strategy	Attribute or Alternative-based	Amount of Information	Consistency
EBA	Attribute-based	Depends on values of alternatives and cut-offs	Selective
LEX	Attribute-based	Depends on values of alternatives and cut-offs	Selective
MCD	Attribute-based	Ignores probability or weight information	Consistent
WADD	Alternative-based	All information processed	Consistent
SAT	Alternative-based	Depends on values of alternatives and cut-offs	Selective
EQW	Alternative-based	Ignores probability or weight information	Consistent

Given our position, as described above, that rationally-adaptive behavioural models are more likely to be behaviourally valid descriptions of choice behaviour, we can clearly discount the WADD strategy, since it assumes that all information is processed (this remains a testable assumption, however). Furthermore, given that we are focusing on stochastic representations of *attribute-based* processes which may not be consistent across different decision tasks, it is clear that we are left with two potentially useful strategies that can help to explain choice behaviour. Table 1 above demonstrates that the only two attribute-based strategies capable of explaining inconsistent and variable decision strategies are Elimination-by-aspects (EBA) and Lexicographic Choice (LEX),

described by Payne, et al. (1988), which satisfy the criteria in the current study on stochastic specifications of attribute processing strategies.

EBA (See Starmer, 2000) involves a determination of the most important attribute (usually defined by the attribute with the highest weight/probability) and the cut-off value for that attribute (i.e., a threshold). An alternative is eliminated if the value of its most important attribute falls below this cut-off value. This process of elimination continues for the second most important attribute, and so on, until a final alternative remains. Thus, the EBA strategy is best characterized as a '*threshold*' attribute processing strategy. The LEX strategy, in its strictest sense, involves a direct comparison between alternatives on the *most important* attribute. In the event of a tie, the second most important attribute is used as a comparison, and so on until an alternative is chosen. The LEX strategy is thus best characterized as a '*relative comparison*' strategy. Thus, we can clearly differentiate two classes of attribute processing strategies: threshold and relative comparison.

A major deficit in these strategies is that although they assume selectivity in attribute processing across different decision task contexts, they assume consistency in attribute strategy within the same decision context. In other words, once a strategy is selected for a given task (or choice), it does not change within the task.

This issue is further complicated by an influential psychological theory which identifies two main stages in the decision process. Differentiation and Consolidation (Diff Con) Theory, developed by Svenson (1992), assumes that decision-making is a goal-oriented task which incorporates the pre-decision process of differentiation and the post-decision process of consolidation. This theory is crucial in encouraging a disaggregation of the entire decision process.

The two issues discussed above, regarding the *adaptive* nature of strategies and the *disaggregation* of the decision process, are issues that can only be assessed realistically within a paradigm that relaxes the deterministic assumption of most rational and normative models of decision-making.

In other words, a stochastic specification of attribute processing capable of accommodating the widespread consensus in the decision-making literature that decision-making is an active process which may require different decision making strategies in different contexts and at different stages of the decision process (e.g., Stewart, et al., 2003). As the relevance of attributes in a decision task changes, so too must our approach to modelling the strategies individuals employ when adapting to such changes. In the next section, we outline the particular specification of the mixed logit that incorporates the stochastic specification of AP.

3. Revealing Preference Heterogeneity in Mixed Logit

Mixed logit is increasingly used to estimate choice models. There are a number of useful summaries of the method (such as Train, 2003 and Hensher and Greene, 2003, Hensher, et al., 2005) and so we will not detail it here. What we do want to do is to highlight the way preference heterogeneity is handled in the model, since it provides the mechanism

for conditioning the parameter estimates of each attribute defining the SC alternatives on the attribute processing strategy adopted by each sampled respondent, known up to a probability by the analyst.

We assume that a sampled individual q ($q=1, \dots, Q$) faces a choice among J alternatives in each of T choice situations. Individual q is assumed to consider the full set of offered alternatives in choice situation t and to choose the alternatives with the highest utility. The utility associated with each alternative j as evaluated by each individual q in choice situation t , is represented in a discrete choice model by a utility expression of the well known general form in (1).

$$U_{jtq} = \beta'_q \mathbf{x}_{jtq} + \varepsilon_{jtq}, \quad (1)$$

where \mathbf{x}_{jtq} is the vector of explanatory variables, including attributes of the alternatives, characteristics of the individual and descriptors of the decision context and choice task itself in choice situation t . The components β_q and ε_{jtq} are not observed by the analyst and are treated as stochastic influences.

In assessing each alternative, we assume within a sampled population, that there are a range of strategies in respect to how each individual attends to each of the attributes. These attribute processing (AP) strategies take into account the level of each attribute in the context of the levels offered of the other attributes present. Prior to making a choice, it is assumed that a suite of bounded rationality decision heuristics are invoked across the sample to include or exclude each attribute in arriving at a choice amongst the offered choice set of alternatives. The analyst does not have full knowledge of what these heuristics are, but knows only which attributes were stated by each sampled individual as attended to (i.e., included) or not (i.e., ignored). For any sampled individual we are only able to establish the attribute processing rule up to a probability, based on observing the frequency distribution of each AP rule over the sampled population. By imposing an extreme value type I distribution on the unobserved influences, and parameterising the role of the attribute levels offered, as well as any contextual effects such as socio economic characteristics, we are able to engender an index of expected maximum utility that can be used to condition the preference heterogeneity with a mixed logit model for each attribute associated with each alternative. The way in which we link this conditioning of attributes is through the decomposition of the structural parameters for each attribute as indicated in equation (2).

Specifically, we do this by introducing alternative-specific heterogeneity into the utility function through β_q . Thus,

$$\beta_q = \beta + \Delta \mathbf{z}_q + \Gamma_q \mathbf{v}_q = \beta + \Delta \mathbf{z}_q + \boldsymbol{\eta}_q, \quad (2)$$

or $\beta_{qk} = \beta_k + \boldsymbol{\delta}'_k \mathbf{z}_q + \eta_{qk}$, where β_{qk} is the random parameter whose distribution over individuals depends in general on underlying parameters, \mathbf{z}_q is data representing the expected maximum utility (EMU) derived from the choice amongst attribute processing strategies, and the random vector $\boldsymbol{\eta}_q$ endows the random parameter with its stochastic properties (including the random error in the estimated EMU). The EMU index is the classical measure of expected maximum utility from a multinomial logit model (i.e. the

log sum formulation). In our application context there are nine attribute processing strategies invoked by the sample (see Table 4 below). It is treated like any attribute used to decompose the mean of a random parameter, except that EMU is estimated data. We set up a hierarchical model with mixing over the attribute processing rules chosen by each respondent, fixed, in this application, over multiple choice tasks by the same respondent.²

For convenience in isolating the model components, we define \mathbf{v}_q to be a vector of uncorrelated random variables with known variances and denote the matrix of known variances of the random draws as \mathbf{W} .

Since β_q may contain individual specific constants, η_{qk} may also vary across choices and, in addition, may thus induce correlation across choices. Note that β_q and its components are structural parameters $(\beta, \Delta, \Gamma_q)$ and choice situation invariant characteristics of the individual, \mathbf{z}_q . They do not vary across choice situations or across choices, save for the extent that components of \mathbf{x}_{jitq} are choice specific. The terms $\beta + \Delta \mathbf{z}_q$ accommodate heterogeneity in the mean of the distribution of the random parameters, accounting for the impact of the attribute processing strategy, up to probability, on the distribution of preference heterogeneity. Thus we condition the variance of the random parameter for an attribute on the (probability weighted) attribute processing indicator, given in (3).

$$\text{Var}[\beta_q | \mathbf{z}_q] = \Gamma \Sigma^{1/2} \mathbf{W} \Sigma^{1/2} \Gamma' \quad (3)$$

We will assume that Γ is an unrestricted lower triangular matrix. Thus, with no loss of generality, we assume that \mathbf{W} is diagonal and contains no unknown parameters. Variance heterogeneity can also be introduced into the model. Let $\Sigma_q = \text{Diag}[\sigma_{1q}, \sigma_{2q}, \dots, \sigma_{Kq}]$ where

$$\sigma_{qk} = \sigma_k \times \exp(\boldsymbol{\theta}_k' \mathbf{h}_q), \quad (4)$$

and \mathbf{h}_q is a vector of M variables that enters the variances, which can be the attribute processing index. Thus the attribute processing strategy can impact on preference heterogeneity via one or both of $\mathbf{z}_q, \mathbf{h}_q$. The full model for the variances is given as equation (5).

$$\text{Var}[\beta_q | \boldsymbol{\Omega}, \mathbf{z}_q, \mathbf{h}_q] = \Phi_q = \Gamma \Sigma_q^{1/2} \mathbf{W} \Sigma_q^{1/2} \Gamma' \quad (5)$$

where $\boldsymbol{\Omega} = (\beta, \Delta, \Gamma, \Sigma, \mathbf{W})$, the component structural parameters of β_q . We now have a functional form for an attribute's marginal (dis)utility in which its preference profile across a sample is represented by a mean and a standard deviation expression of the general form (Hensher, 2005) given in (6).

$$\beta_{qk} = \pm \exp[\beta_k + \boldsymbol{\delta}_k' \mathbf{z}_q + \sigma_k \exp(\boldsymbol{\theta}_k' \mathbf{h}_q) v_q] \quad (6)$$

The sign for the entire expression (i.e., positive or negative) is imposed by the analyst to represent the behaviourally required sign, v_q is an analytical distribution selected by the analyst, and all other terms are defined above.

² In ongoing research, we are assessing the behavioural implications of allowing for a different attribute processing strategy after each choice task, primarily to establish if the levels of attributes influence the processing strategy.

The *mixed logit* class of models assumes a general distribution for β_{qk} and an IID extreme value type 1 distribution for ϵ_{jtq} . That is, β_{qk} can take on different distributional forms. For a given value of β_q , the *conditional* (on \mathbf{z}_q , \mathbf{h}_q and \mathbf{v}_q) probability for choice j in choice situation t is multinomial logit, since the remaining random term, ϵ_{jtq} , is IID extreme value:

$$P_{jtq}(\text{choice } j \mid \Omega, \mathbf{X}_{tq}, \mathbf{z}_q, \mathbf{h}_q, \mathbf{v}_q) = \exp(\beta_q' \mathbf{x}_{jtq}) / \sum_j \exp(\beta_q' \mathbf{x}_{jtq}) \quad (7)$$

where the full set of attributes and characteristics is gathered in $\mathbf{X}_{tq} = [\mathbf{x}_{1tq}, \mathbf{x}_{2tq}, \dots, \mathbf{x}_{Jtq}]$.

4. The Design Plan and Sample Context

The literature on the design of SC experiments is extensive and growing (Louviere, et al., 2000; Hensher, et al., 2005), with substantial developments in the methods used to construct designs that are optimal in both a statistical and a behavioural sense (Bliemer and Rose, 2005). The statistical state of the art of designing SC experiments has moved away from orthogonal designs to D-optimal designs; and the behavioural state of the art has moved to promoting designs that are pivotted around the knowledge base of travellers in recognition of a number of supporting theories in behavioural and cognitive psychology and economics such as prospect theory, case-based decisions theory and minimum-regret theory³. Starmer (2000, p 353) makes a very strong plea in support of the use of reference points (i.e., a current trip):

“While some economists might be tempted to think that questions about how reference points are determined sound more like psychological than economic issues, recent research is showing that understanding the role of reference points may be an important step in explaining real economic behaviour in the field”

A total of 344 face to face computer aided personal interview (CAPI) surveys were undertaken in the Sydney metropolitan area in late 2004 for car non-commuters. Full details of the sampling and response rates are given in Hensher and Rose (2004). The choice set assessed by each respondent involved a current trip and two SC alternatives, all defined as unlabelled routes. The trip attributes associated with each route are summarised in Table 2. These were identified from reviews of the literature and through the effectiveness of previous valuation of travel time savings studies.

Table 2: Trip Attributes in Stated Choice Design

Routes A and B
Free flow travel time
Slowed down travel time
Trip travel time variability
Running cost
Toll Cost

³ See Starmer, 2000; Hensher, 2004; Kahnemann and Tversky, 1979; Gilboa, et al., 2002.

All attributes of the two SC alternatives are based on the values of the current trip. Variability in travel time for the current alternative was calculated as the difference between the longest and shortest trip time provided in non-SC questions. The SC alternative values for this attribute are variations around the total trip time. For all other attributes, the values for the SC alternatives are variations around the values for the current trip. The variations used for each attribute are given in Table 3.

Table 3: Profile of the Attribute range in the SC design

	Free-flow time	Slowed down time	Variability	Running costs	Toll costs
Level 1	- 50%	- 50%	+ 5%	- 50%	- 100%
Level 2	- 20%	- 20%	+ 10%	- 20%	+ 20%
Level 3	+ 10%	+ 10%	+ 15%	+ 10%	+ 40%
Level 4	+ 40%	+ 40%	+ 20%	+ 40%	+ 60%

The experimental design has one version of 16 choice sets. It allows the estimation of all linear and quadratic main effects. The design has no dominance given the assumptions that less of all attributes is better.

Some specific cases had to be accounted for in the CAPI program. If someone has a current trip with no slowed down time, it is set to 10 percent of the current free-flow time in order to construct the SC alternatives. If someone has a current trip without free-flow time, it is set to 10 percent of the slowed down time and slowed down time is decreased to 90 percent of its current value.

For tolls in-between the above numbers, the values of the levels are also in-between those indicated. For the toll attribute, these amounts are assumed for a 90-minute trip. However, we want shorter trips to have adjusted tolls. For this reason, the toll set by the design is adjusted by a coefficient calculated according to the formula $(0.005 \times TotalTime + 0.55)^4$ where TotalTime is the sum of Free-flow and Slowed down times. This formula returns more realistic SC alternatives. An example of a stated choice screen is shown as Figure 1.

⁴ This formula is based on extensive simulations to ensure that the attribute levels are sensible.

5. Mixed Logit Results

Three specifications of mixed logit were estimated in which we (i) did not account for the presence or absence of one or more attributes in information processing (M1); (ii) where we removed an attribute if the individual stated that they ignored it in the assessment of the alternatives (M2). We refer to this as a deterministic model since it uses the exogenous information deterministically; and (iii) a stochastic specification (M3) which assumes that the analyst does not know for certain which attributes are used by which respondents. For case (iii), we can only establish, up to probability, what attribute mix a sampled individual attends to, drawing inferences from the distribution across the sampled population.

The incidence of not attending to one or more attributes is summarised in Table 4. This is the attribute processing choice set for the sample. We focus the analysis on four attributes – free flow time, slowed down time, running costs and toll cost. Just over half (i.e., 52 percent) of the sample attended to every attribute and not one respondent attended to none of the attributes. Running cost was the least attended to attribute when one attribute was ignored (i.e., 17.9 percent of the sample); in contrast the toll cost was attended to nearly for 96 percent of the sample. Free flow time was not attended to by 13 percent of the sample, with 8.5 percentage point of this being when both components of travel time were ignored and the focus was totally on cost. The message from this information is that 78 percent of the sample attended to the components of travel time and 69 percent attended to the components of cost.

Table 4: Incidence of Mixtures of Attributes Processed

<i>Attribute Processing Profile</i>	<i>Sample no. of observations=3568</i>
All attributes attended to (v1)	1856
<i>Attributes not attended to:</i>	
Running cost (v2)	640
Running and toll cost (v3)	192
Toll Cost (v4)	96
Slowed down time (v5)	192
Free flow and slowed down time (v6)	304
Free flow time (v7)	112
Slowed down time and running cost (v8)	64
Free flow and slowed down time and toll cost (v9)	48

To account for the assumption that the analyst is not able to identify the attribute processing strategy of a specific individual (just like the analyst does not have full information of what drives an individual's choice) we can only infer the attention to attributes up to a probability. We have estimated a separate model to establish the probability of a sampled individual drawn from a population choosing a specific attribute processing strategy in terms of the portfolio of attributes that are attended and not-attended to. Table 4 defines the choice set of nine alternatives for estimating an attribute inclusion/exclusion processing model. The estimated parameters are used to derive, for each individual, an index of the expected maximum utility associated with the portfolio of attending to strategies for an individual drawn from the sampled population, calculated as the usual logsum formula in a nested logit model (i.e., $\ln \sum_{i=1}^9 \exp V_i$). The distribution of this index for the sample is given in Figure 3. This

index empirically is a function of the attribute levels for free flow time, slowed down time, running cost and toll cost as well as the respondent's age and household income. The utility expressions for each of the nine attribute processing rules in Table 4 are given in Table 5. Importantly the attribute processing rule recognises the role of the *level* of each attribute in influencing an individual's AP rule. An EMU for each sampled individual is introduced (sequentially) into the mixed logit model M3 (see Table 4) as a way of conditioning the marginal utility of the attributes of each alternative in the stated choice experiment where the conditioning is found to be statistically significant. The presence of estimated parameters in EMU is accounted for through an assumption of additive (common) error with the η_q , although we might reasonably assume that the difference between the true and estimated parameters in EMU is small relative to the preference heterogeneity captured in η_q attributable to other influences.

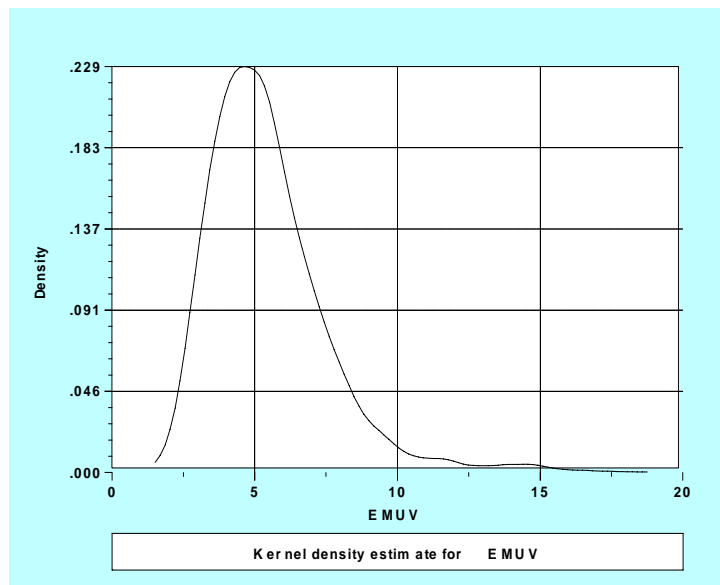


Figure 3: Distribution of expected maximum utility of attending to an attribute mix

Table 5: Utility expressions for attribute attention profiles, estimated as multinomial logit

$v_1 = 2.0909 + .02872 * \text{age} - .01088 * \text{income} - .03606 * \text{ff} + .11071 * \text{sdt} + .1969 * \text{cost} + .06767 * \text{toll}$
 $v_2 = 1.7487 + .019159 * \text{age} - .011466 * \text{income} - .03545 * \text{ff} + .10151 * \text{sdt} + .17557 * \text{cost} + .06932 * \text{toll}$
 $v_3 = -1.49000 + .01978 * \text{age} - .001379 * \text{income} - .00194 * \text{ff} + .13364 * \text{sdt} + .07899 * \text{cost} + .01865 * \text{toll}$
 $V_4 = -3.055 + .01147 * \text{age} + .01349 * \text{income} - .020047 * \text{ff} + .1175 * \text{sdt} + .20619 * \text{cost} + .07678 * \text{toll}$
 $V_5 = 0.82309 + .03845 * \text{age} - .01994 * \text{income} - .01032 * \text{ff} - .05525 * \text{sdt} + .33109 * \text{cost} + .00305 * \text{toll}$
 $V_6 = 1.68608 + .01397 * \text{age} - .02204 * \text{income} - .061966 * \text{ff} + .126399 * \text{sdt} + .2674 * \text{cost} + .0999 * \text{toll}$
 $V_7 = 1.5842 - .02523 * \text{age} - .003078 * \text{income} - .017136 * \text{ff} + .07665 * \text{sdt} + .14232 * \text{cost} - .016056 * \text{toll}$
 $V_8 = -4.10832 + 0.07469 * \text{age} - .0112178 * \text{income} - .03349 * \text{ff} + .12575 * \text{sdt} + .23752 * \text{cost} - .00806 * \text{toll}$
 $V_9 = 0$
 Pseudo- $R^2 = 0.179$, bolded= statistically non-significant at 95 percent confidence level

The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies

Hensher, Rose & Bertoia

The random parameters in the mixed logit models in Table 6 have a triangular distribution which is constrained (as per equation 6) to ensure that the willingness to pay for travel time savings was non-negative. For the triangular distribution, the density function looks like a tent: a peak in the centre and dropping off linearly on both sides of the centre⁵. The overall goodness of fit of all models is impressive, as is often the experience with stated choice studies. All parameters are statistically significant and of the expected sign. The toll-route quality bonus, defined as a dummy variable to account for the relative benefits of a tolled route (compared to a free route) after accounting for the levels of service engendered in the measured attributes, is statistically significant in M1 (as a random parameter) and M3 (as a fixed parameter), but is not significant in M2. We investigated a number of specifications for this attribute but were unable to establish any improvement over the reported specifications. The M2 result suggests that the exclusion of specific attributes (setting their marginal utility to zero) has impacted in a significant way on the discriminating contribution of the toll route quality effect relative to a free route.

Table 6: Mixed Logit Choice Models with alternative information processing conditions
(3,568 observations). Time is in minutes, cost is in dollars. (500 Halton draws)

Attribute	All attributes assumed to be attended to	Deterministic attribute exclusion	Stochastic attribute exclusion
	M1	M2	M3
<i>Random Parameters</i>			
<i>Mean of random parameters</i>			
Free flow time	-0.0755 (-16.3)	-0.0758 (-15.1)	-0.1676 (-10.1)
Slowed down time	-0.0928 (-16.8)	-0.1034 (-15.9)	-0.1249 (-9.78)
Toll-route quality bonus	0.6624 (4.52)	0.0998 (0.74)	0.6849 (4.94)
<i>Standard deviations of random parameters:</i>			
Free flow time	0.0755 (16.3)	0.0758 (15.1)	0.1676 (10.1)
Slowed down time	0.0928 (16.8)	0.1034 (15.9)	0.1249 (9.78)
Toll-route quality bonus	2.397 (3.45)	3.6910 (5.49)	
<i>Heterogeneity around mean:</i>			
Free flow time x expected maximum utility from attending to specific attributes			0.01550 (5.99)
Slowed down time x expected maximum utility from attending to specific attributes			0.00396 (2.75)
<i>Non Random Parameters</i>			
Running cost	-0.3321 (-12.7)	-0.3619 (-12.12)	-0.3444 (-13.4)
Toll cost	-0.6282 (-14.0)	-0.5824 (-12.03)	-0.62501 (-17.9)
<i>Model Fits</i>			
Pseudo-R ²	0.300	0.292	0.307
Log-Likelihood	-2739.65	-2772.62	-2714.5
<i>Number of respondents who ignored this attribute</i>			
Free flow excluded		496	
Slowed down time excluded		624	
Running cost excluded		976	
Toll cost excluded		304	

⁵ Let c be the centre and s the spread. The density starts at $c-s$, rises linearly to c , and then drops linearly to $c+s$. It is zero below $c-s$ and above $c+s$. The mean and mode are c . The standard deviation is the spread divided by $\sqrt{6}$; hence the spread is the standard deviation times $\sqrt{6}$. The height of the tent at c is $1/s$ (such that each side of the tent has area $s \times (1/s) \times (1/2) = 1/2$, and both sides have area $1/2 + 1/2 = 1$, as required for a density). The slope is $1/s^2$.

The values of travel time savings (VTTS) are reported in Table 7. In our example, where we treat the travel time parameter as random and the cost parameter as fixed, we see some similarities and some differences in the distributions of VTTS under alternative attribute processing assumptions. Most notably, the mean VTTS varies from \$7.21 to \$7.95 for free flow time and from \$8.86 to \$10.65 for slowed down time. Although the free flow values may appear very similar on average, looking at the full distribution, we find great similarity in free flow time between the specifications that assume all attributes are attended to and the deterministic inclusion/exclusion rule. The stochastic specification displays greater preference heterogeneity across the sample. For slowed down time, there are greater differences in the mean and standard deviations for VTTS. Specifically the model that assumes all attributes are attended to delivers a lower mean VTTS and a lower standard deviation, except for free flow time where the standard deviation is virtually the same as the deterministic AP rule.

*Table 7: Values of travel time savings
(\$ per person hour car non-commuter driver)*

(i) time = random parameter, cost = fixed parameter

Attribute	All attributes assumed to be attended to		Deterministic attribute exclusion		Stochastic attribute exclusion	
	Sample mean	Sample Std dev	Sample mean	Sample Std dev	Sample mean	Sample Std dev
Free flow time	7.21	0.44	7.81	0.46	7.95	3.59
Slowed down time	8.86	0.54	10.65	0.67	9.91	1.22
Ratio slowed to free flow time	1.23	1.22	1.36	1.46	1.25	0.69
Sample Size	3568		3071/2944*		3568	

* 3,071 relates to free flow and 2,944 to slowed down time.

(ii) ratio of non-ignored (Model 1) to deterministic and stochastic specifications

Attribute	Deterministic attribute exclusion		Stochastic attribute exclusion	
	mean	Std dev	mean	Std dev
Free flow time	0.923	0.956	0.907	0.123
Slowed down time	0.832	0.806	0.894	0.443

In order to test for differences in the variances of the free flow and slowed down time VTTS distributions over the three models (i.e., M1, M2 and M3), Brown and Forsyth (1974) tests for homogeneity of variances were conducted. The Brown and Forsyth test tends to produce quite accurate error rates when the underlying distributions of the deviations from the group medians deviate significantly from the normal distribution. Test statistics for the Brown and Forsyth tests of homogeneity of variances of the free flow and slowed down time VTTS behavioural distributions were 124.695 and 3562.21, respectively. The test is asymptotically F distributed with two and 3801.3 degrees of freedom for the free-flow VTTS and two and 7115.561 degrees of freedom for the slowed down time VTTS distribution. In both cases we reject the null hypothesis of homogeneity of variances and conclude that the variances for the VTTS distributions are significantly different from one another over the three models (i.e., M1, M2 and M3).

Given differences in variances of the VTTS distributions over the three models, it is inappropriate to perform an ANOVA test, to test differences in the means of the distributions. We therefore conduct a Kruskal-Wallis test, which is the non-parametric equivalent to the ANOVA test (for more details of the test, see Siegel and Castellan, 1988). For the VTTS distributions obtained from the three models, chi-square statistics of 1809.071 and 5066.843 were obtained for the free flow and slowed down time VTTS distributions respectively, which we compare with a critical value of 5.99 (i.e., χ^2_2 at the 95 percent confidence level). We conclude that the both the means and variances of the VTTS distributions for both attributes, as derived from the M1, M2 and M3 model specifications are statistically different.

This evidence suggests a deflating effect on VTTS when one ignores the attribute processing strategy and assumes that all attributes are attended to. While the differences do not appear to be large at the mean for free flow, they are sizeable for slowed down time, and when converted to time savings benefits in road projects would make a substantial difference to the user benefits, given the dominance of travel time savings. The direction of impact shown in this study cannot be assumed to be transferable to other studies since this is a single study. In contrast, Hensher, et al. (2005) found that when we do not condition parameter estimation on the attribute processing strategy of the respondent, we get a significantly higher estimate of the mean value of travel time savings, on average of the order of 18-62 percent depending on the specific attribute.

6. Conclusions

The evidence in this paper, that recognition of varying attribute processing strategies in respect to how specific attributes are processed, in terms of exclusion and inclusion, is compelling. Although we cannot suggest whether the exclusion of an attribute is due to some underlying behavioural rationale for the attribute's role, or simply a coping strategy in processing the amount of information presented in the stated choice experiment, we have empirical evidence which supports the position that imposing a condition of unlimited human capacity to process information of varying degrees of magnitude and quality is not a reflection of how individuals actually make choices. It also artificially produces a willingness to pay distribution for a specific attribute (in our case valuation of travel time savings) that is only true when we assume that all presented attributes matter and individuals are capable of processing the information content of all attributes, as well as wishing to process it.

Indeed in real markets the choice process is simplified for both behavioural and process coping reasons, and as such both sources of potential influence are at play, interacting to produce a specific choice outcome and implied trade-off, and hence valuation of attributes influencing the choice outcome. Accounting for the inclusion versus exclusion of an attribute in an individual's decision calculus does appear to impact on the behavioural outputs of a discrete choice model; in our example the behavioural value of travel time savings distribution and its associated moments appear to be influenced by the assumption made on how attributes are processed. Given that the majority of practitioners continue to use mean estimates of VTTS, even if such estimates are

obtained from models capable of producing distributions, the differences in means will deliver sizeable differences in travel time benefits for projects⁶.

Acknowledgement

The comments of a referee have materially improved this paper.

References

- Bliemer, M.C. and Rose J.M. (2005) Efficient Designs for Alternative Specific Choice Experiments, Institute of Transport and Logistics Studies, University of Sydney, February.
- Brown, M.B. and Forsythe, A.B. (1974) Robust tests for equality of variances, *Journal of American Statistical Association*, 69, 264-267.
- Bunch, David S., Jordan J. Louviere, and Donald A. Anderson (1996) A Comparison of Experimental Design Strategies for Choice-Based Conjoint Analysis with Generic-Attribute Multinomial Logit Models, Working Paper, Graduate School of Management, University of California, Davis.
- DePalma, Andre, Meyers, Gordon M. and Papageorgiou, Yorgos Y. (1994) Rational choice under an imperfect ability to choose, *American Economic Review*, 84, 419-440.
- DeShazo, J.R. and Fermo, G. (2004) Implications of rationally-adaptive pre-choice behaviour for the design and estimation of choice models, working paper, School of Public Policy and Social Research, University of California at Los Angeles, August.
- Diederich, A. (2003). MDFT account of decision making under time pressure. *Psychonomic Bulletin and Review*, 10, 157-166.
- Drolet, A. and Luce, M.F. (2004). The rationalizing effects of cognitive load on emotion-based trade-off avoidance. *Journal of Consumer Research*, 31, 63-77.
- Gilboa, I., Schmeidler, D. and Wakker, P. (2002) Utility in case-based decision theory, *Journal of Economic Theory*, 105, 483-502.
- González-Vallejo, C. (2002). Making trade-offs: a probabilistic and context-sensitive model of choice behavior. *Psychological Review*, 109, 137-155.
- Heiner, R.A. (1983) The origin of predictable behaviour, *American Economic Review*, 73, 560-595.
- Hensher, D.A. (2004) Accounting for Stated Choice Design Dimensionality in Willingness to Pay for Travel Time Savings, *Journal of Transport Economics and Policy*, 38 (2), 425-446.
- Hensher, D.A. (2005) Imposing a Globally Signed Condition on a Generalised Parameterisation of Attributes in Choice Models to Produce Sign Compliant Willingness to Pay Distributions, Institute of Transport and Logistics Studies, University of Sydney, February.

⁶ Our experience in advising banks on toll road infrastructure projects, suggests that differences in means that are displayed herein do matter and influence bankability.

Hensher, D.A. (*in press*) Revealing differences in behavioural response due to the dimensionality of stated choice designs: an initial assessment, *Environmental and Resource Economics*, Invited Paper.

Hensher, D.A., Rose, J. and Greene, W. (2005) The Implications on Willingness to Pay of Respondents Ignoring Specific Attributes *Transportation*, 32(2).

Hensher, D.A. and Greene, W.H. (2003) Mixed logit models: state of practice, *Transportation*, 30(2), 133-176.

Hensher, David A., John M. Rose, and William H. Greene (2005) *Applied Choice Analysis: A Primer* Cambridge University Press, Cambridge.

Huber, Joel. and Klaus Zwerina (1996) The Importance of utility Balance and Efficient Choice Designs, *Journal of Marketing Research*, 33 (3), 307-317.

Kahnemann, D. and Tversky, A. (1979) Prospect theory: an analysis of decisions under risk, *Econometrica*, 47 (2), 263-91.

Kanninen, Barbara J. (2002) Optimal Design for Multinomial Choice Experiments, *Journal of Marketing Research*, 39 (2), 214-217.

Kuhfeld, Warren F., Randal D. Tobias and Mark Garratt (1994) Efficient Experimental Design with Marketing Research Applications, *Journal of Marketing Research*, 21 (4), 545-557.

Lazari, Andreas G. and Donald A. Anderson (1994) Designs of Discrete Choice Experiments for Estimating Both Attribute and Availability Cross Effects, *Journal of Marketing Research*, 31 (3), 375-383.

Louviere, Jordan J. and Harry J.P. Timmermans (1990) Hierarchical information integration applied to residential choice behaviour, *Geographical Analysis*, 22, 127-145.

Oppewal, Harman, Jordan J. Louviere, Harry J.P Timmermans (1994) Modeling hierarchical information integration processes with integrated conjoint choice experiments, *Journal of Marketing Research*, 31, 92-105.

Payne, J.W. Bettman, J.R. and Johnson, E.J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534-552.

Payne, J.W., Bettman, J.R., Coupey, E. and Johnson, E.J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, 80, 107-141.

Sælensminde, Kjartan (1994) The Impact of Choice Inconsistencies in Stated Choice Studies, *Environmental and Resource Economics*, 23 (4), 403-420.

Sandor, Zsolt and Michel Wedel (2001) Designing Conjoint Choice Experiments Using Managers' Prior Beliefs, *Journal of Marketing Research*, 38 (4), 430-444.

Siegel S. and Castellan, N. (1988) *Nonparametric Statistics for the Behavioral Sciences*, McGraw Hill, New York.

Slovic, P. (1995). The Construction of Preference. *American Psychologist*, 50, 364-371.

Starmer, C. (2000) Developments in non-expected utility theory: the hunt for a descriptive theory of choice under risk, *Journal of Economic Literature*, XXXVIII, 332-382.

Stewart, N., Chater, N., Stott, H.P. and Reimers, S. (2003). Prospect relativity: how choice options influence decision under risk. *Journal of Experimental Psychology: General*, 132, 23-46.

Svenson, O. (1992). Differentiation and Consolidation Theory of human decision making: A frame of reference for the study of pre-and post-decision processes. *Acta Psychologica*, 80, 143-168.

Swait, J. and Adamowicz, W. (2001). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, 28, 135-148.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 23, 273-286.

Tversky, A. (1972), Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281-299.

Train, Ken. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge.

Wang, Donggen, Jiuqun, Li and Harry J.P. Timmermans (2001) Reducing respondent burden, information processing and incomprehensibility in stated preference surveys: principles and properties of paired conjoint analysis, *Transportation Research Record*, 1768, 71-78.