Attribute Processing, Heuristics, and Preference Construction in Choice Analysis

David A. Hensher
Institute of Transport and Logistics Studies (ITLS)
Faculty of Economics and Business
The University of Sydney, NSW 2006 Australia
D.Hensher@itls.usyd.edu.au

March 29 2009 (Version 5.1)

Invited plenary paper for the First International Conference on Choice Analysis, Harrogate, UK March 29 - April 3 2009

Abstract

It has long been recognized that humans draw from a large pool of processing aids to help manage the everyday challenges of life. It is not uncommon to observe individual's adopting simplifying strategies when faced with ever increasing amounts of information to process, and especially for decisions where the chosen outcome will have a very marginal impact on their well being. The transactions costs associated with processing all new information often exceed the benefits from such a comprehensive review. The accumulating life experiences of individuals are also often brought to bear as reference points to assist in selectively evaluating information placed in front of them. These features of human processing and cognition are not new to the broad literature on judgment and decision making, where heuristics are offered up as deliberative analytic procedures intentionally designed to simplify choice. What is surprising is the limited recognition of heuristics that individuals use to process the attributes in stated choice experiments. In this paper we present a case for a utility-based framework within which some appealing processing strategies are embedded (without the aid of supplementary self-stated intentions), as well as models conditioned on self-stated intentions represented as single items of process advice, and illustrate the implications on willingness to pay for travel time savings of embedding each heuristic in the choice process. Given the controversy surrounding the reliability of self-stated intentions, we introduce a framework in which mixtures of process advice embedded within a belief function might be used in future empirical studies to condition choice, as a way of increasingly judging the strength of the evidence.

Keywords: heuristics and rules, common-metric attribute processing, parameter transfer, choice experiments, thresholds, asymmetry, willingness to pay, referencing, self-stated intentions, belief functions, plausibility

Acknowledgements. The ideas presented herein are an accumulation of research activity undertaken with a number of colleagues. I especially acknowledge the contributions made by William Greene, John Rose, David Layton, Sean Puckett, Ric Scarpa, Stephane Hess, and Joffre Swait. Discussions with Stewart Jones on belief functions were especially useful. This research is partially funded by the Australian Research Council Discovery Project Grant DP0770618.

1. Introduction

Any economic decision or judgment has an associated, often subconscious, psychological process prodding it along, in ways that makes the 'neoclassical ambition of avoiding [this] necessity ….unrealizable' (Simon 1978, 507). The translation of this fundamental statement on human behaviour has become associated with the identification of heuristics that individuals use to simplify preference construction and hence make choices, or to make the representation of what matters relevant, regardless of the degree of complexity as perceived by the decision maker and/or the analyst.

Despite the recognition in behaviour research as far back as the 1950s (see Svenson 1998, Gilovich et al. 2002), that cognitive processes have a key role in preference revelation, and the reminders throughout the literature (see McFadden 1998, Yoon and Simonson 2008) about rule-driven behaviour (e.g., Swait and Ben-Akiva 1987, Gilbride and Allenby 2004, Martinez et al. 2009, Arana et al. 2008, Gabaix and Laibon 2000), we still see relatively little of the belief incorporated into stated choice modelling which is, increasingly, becoming the mainstream empirical context for preference measurement and willingness to pay derivatives[1].

There is an extensive literature on what might broadly be described as heuristics and biases, and which is crystallized in the notion of *process*, in contrast to *outcome*. Choice has both elements of process and outcome, which in combination represent the endogeneity of choice in stated choice studies. The failure to recognise process and the maintenance of a linear additive utility expression under full attribute and parameter preservation is an admission, by default, that individuals when faced with a stated choice experiment, deem all attributes relevant, and that a compensatory decision rule is used to arrive at a choice.

Although there should be no suggestion that such compensatory rules are always invalid, indeed they may be, in aggregate, an acceptable representation or approximation of many process circumstances, there is a strong belief that process heterogeneity exists as a consequence of mixtures of genuine cognitive processing strategies that simplify decision making in real markets, for all manner of reason, and the presence of new states that are introduced through the design of choice experiments that are no more than new circumstances to process. Whether the processing rules adopted are natural to real choices, or are artifacts of the design of an experiment or some other survey instrument (including revealed preference surveys) in front of an individual, is in some senses irrelevant; what is relevant is the manner in which such choice assessments are processed in respect of the role that each design attribute and the mixture of attributes, plays in the outcome. Yoon and Simonson (2008) and Park et al. (2008)[2] provide some interesting perspectives from marketing research on preference revelation.

[1] Consultants still adopt, almost without exception, a full compensatory approach in which all attributes are 'relevant'.

[2] Park et al. (2008) promotes the idea of starting with a basic product profile and upgrading it one attribute at a time, identifying the willingness to pay for that additional attribute given the budgets available.

Recent research by Hensher (2006, 2008), Greene and Hensher (2008), Layton and Hensher (2008), Hensher and Rose (in press), Hensher and Layton (2008), Hess and Hensher (2008), Puckett and Hensher (2008), Swait (2001), Cantillo et al. (2006), Cameron (2008), Scarpa et al. (2009,2009a), Beharry and Scarpa (2008), Cantillo and Ortúzar (2005), Carlsson et al. (2008), Caussade et al. (2005) and Hensher et al. (2005, 2009), amongst others, are examples of a growing interest in the way that individuals evaluate a package of attributes associated with ordered or unordered alternatives in real or hypothetical markets, and make choices[3]. The accumulating empirical evidence suggests that individuals use a number of strategies derived from heuristics, to represent the way that information embedded within attributes defining alternatives is used to process the context and arrive at a choice outcome. These include cancellation or attribute exclusion, degrees of attention paid to attributes in a package of attributes, referencing of new or hypothetical attribute packages around a recent or past experience, and attribute aggregation where attributes are in common units (see Gilovich et al. 2002 for a series of papers that synthesise the evidence under the theme of heuristics and biases). Importantly, as shown herein, the heuristics are likely to be context specific, such that the nature of the information shown in stated choice experiments, for example, conditions the choice of rules adopted.

Hensher (2006b, 2008) argues that individuals appear to adopt a range of 'coping' or editing strategies in hypothetical choice settings that are consistent with how they normally process information in real markets. Choice experiments have varying amounts of information to process, but importantly, aligning 'choice complexity' with the amount of information to process is potentially misleading. *Relevancy* is what matters (Hensher 2006b)[4], and the heuristics adopted by individuals to evaluate a circumstance is what needs to be captured through frameworks that can empirically identify rules adopted by individuals.

There are at least two ways in which information on processing might be identified. One involves direct questioning of respondents after each choice scenario (what we refer to as self-stated intentions); the other involves probabilistic conditions imposed on the model form through specification of the utility expressions associated with each alternative, that enables inference on the way that specific attributes are processed. Both may be complementary. The focus of this paper draws on a stream of active research by Hensher, Rose, Puckett, Layton, Greene, Scarpa, and Hess, in which we are systematically investigating process rules to establish the behavioural implications on marginal willingness to pay[5]. The functional forms presented herein, as well as responses to self-stated intention questions, enable the analyst to infer, up to a probability, the presence of

[3] This paper does not consider other aspects of process in choice experiments such as uncertainty in the choice response. See Lundhede et al. (2009).

[4] The emphasis on cognitive load may well be misplaced. Arana et al. (2008) suggest that individual's cognitive abilities can be interrelated with affective and emotional states, i.e., every aspect of human decision making could be seen as simultaneously influencing the final outcome of the choice process. They show that complexity effects could be non-significant for particular emotional states.

[5] The methods used to identify MWTP also have relevance in the calculation of total WTP.

some very specific attribute processing strategies such as: (i) common metric attribute aggregation, (ii) common-metric parameter transfer, and (iii) attribute non-attendance.

This paper is organised as follows. In the next section a utility-based framework is set out in which special cases of attribute processing can be specified, including a latent class model that can accommodate all heuristics of current interest. The empirical context is then summarised, followed by the empirical analysis, with a focus on marginal willingness to pay (WTP). A further section considers the influence that self-stated intention responses play in WTP when we accept the full validity of such evidence. Given concerns about the face validity of self-stated intentions, we consider the role that belief functions might play in adjusting the self-stated intentions, so that the evidence is more plausible. We conclude with suggestions for ongoing research.

2. Incorporating Attribute Processing Heuristics through Non-linear Processing

2.1 Process I: Common Metric Attribute Aggregation

In this section we present a utility specification that captures two heuristics[6] in non-linear attribute processing of common-metric attributes over a continuum that accommodates preservation of attribute partitioning and attribute aggregation[7]. Importantly, the approach allows for mixtures of heuristics within a single model form associated with a sampled population, in contrast to the many studies that impose a single heuristic on the entire sample, and compare separate models in which each is defined by a single heuristic. A recent study by Arana et al. (2008) also considers multiple heuristics. With more than one heuristic within a model form, we are able to capture an individual choosing heuristic h with expected payoff $E[V_h]$, and define the decision rule h^* as optimal if $E[V_{h^*}] = \max_h E[V_h]$. This specification is similar to the way in which a nested logit model is defined, with the new twist that the upper level of a two-level tree defines a choice set of heuristics.

Consider a utility function defined in terms of two attributes labelled x_1 and x_2 (in the empirical setting below these might be route-specific free flow time and congestion time, both in common units) and other attributes such as running cost and toll cost x_3 and x_4:

$$U = f\left(x_1, x_2, x_3, x_4\right) + \varepsilon \qquad (1)$$

where

[6] Generalisation to more than two heuristics is feasible.
[7] The functional form selected is one of many possible forms, but is useful in illustrating the way in which the utility expression can be defined to test for specific heuristics applications across a sampled population.

$$f(x_1, x_2, x_3, x_4) = \begin{cases} \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 & if\ (x_1 - x_2)^2 > \alpha \\ \beta_{12}(x_1 + x_2) + \beta_3 x_3 + \beta_4 x_4 & if\ (x_1 - x_2)^2 < \alpha \end{cases} \tag{2}$$

β_1, β_2, β_3, β_4, β_{12}, are estimated parameters. β_{12} does not necessarily equal a combination of β_1 and β_2. We assume that the standard random utility alternative-specific error ε is not dependent on which form of $f(x_1, x_2)$ is operative. The term $(x_1 - x_2)^2$ is introduced as the basis of a heuristic, and represents the 'distance' between x_1 and x_2. A squared form is computationally convenient, but another form could be used. Intuitively, the heuristic associated with this functional specification is as follows (Layton and Hensher 2008): when the difference between the common metric attributes x_1 and x_2 is great enough, the agent's process preserves attribute partitioning, and thus treats each attribute as separate entities, and evaluates their contribution to utility in the standard random utility model manner with parameters β_1 and β_2. On the other hand, when the difference between the common metric attributes x_1 and x_2 is relatively small, the agent's process aggregates the attributes and thus treats the sum of x_1 and x_2 as a single attribute with utility weight β_{12}.

We can enrich the model by allowing the α_i for person i to be randomly distributed (with $\alpha_i > 0$). A useful candidate distribution is that α_i is exponential with mean $1/\lambda$ and density $f(\alpha) = \lambda e^{-\lambda \alpha}$ (Layton and Hensher 2008). This density generally has a large mass near zero, and so allows for some fraction of the population to behave essentially as standard optimisers. Still others behave as standard optimisers when attributes are dis-similar, but aggregate when attributes are similar. Importantly, this density also allows for a tail of others who more frequently are aggregating the two attributes. The probability conditions are given in (3). In this model, we assume that there is an exponentially distributed threshold parameter, IID across alternatives and respondents, which indicates how the respondent views the attribute components.[8]

$$P\left((x_1 - x_2)^2 > \alpha\right) = 1 - \exp^{-\lambda (x_1 - x_2)^2} \tag{3a}$$

and

$$P\left((x_1 - x_2)^2 < \alpha\right) = \exp^{-\lambda (x_1 - x_2)^2} \tag{3b}$$

Integrating over the α_i we write U in conditional form (Layton and Hensher 2008):

$$\begin{aligned} U = {} & f\left(x_1, x_2 \mid \left[(x_1 - x_2)^2 > \alpha\right]\right) P\left(\left[(x_1 - x_2)^2 > \alpha\right]\right) + \\ & f\left(x_1, x_2 \mid \left[(x_1 - x_2)^2 < \alpha\right]\right) P\left(\left[(x_1 - x_2)^2 < \alpha\right]\right) + \varepsilon \end{aligned} \tag{4}$$

[8] At much greater computational cost one might allow for the α_i's to be constant across alternatives for a given respondent. We leave refinements like this for future work.

Equation (4) implies that:

$$U = (\beta_1 x_1 + \beta_2 x_2)\left(1 - \exp^{-\lambda(x_1 - x_2)^2}\right) + \beta_{12}(x_1 + x_2)\left(\exp^{-\lambda(x_1 - x_2)^2}\right) + \varepsilon \tag{5}$$

Equation (5) together with the equivalent treatment of x_3 and x_4 implies that:

$$U = (\beta_1 x_1 + \beta_2 x_2)\left(1 - \exp^{-\lambda 1(x_1 - x_2)^2}\right) + \beta_{12}(x_1 + x_2)\left(\exp^{-\lambda 1(x_1 - x_2)^2}\right) + (\beta_3 x_3 + \beta_4 x_4)\left(1 - \exp^{-\lambda 2(x_3 - x_4)^2}\right) + \beta_{34}(x_3 + x_4)\left(\exp^{-\lambda 2(x_3 - x_4)^2}\right) + \varepsilon \tag{5a}$$

Equation (5a) is a non-linear form in x_1, x_2, x_3, x_4. As λ_i, i=1,2, tends toward ∞ the distribution becomes degenerate at zero. In this case, all individuals are always standard optimisers who partition the common metric attributes and we obtain the linear additive form (6).

$$U = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \tag{6}$$

If λ tends toward 0, every individual becomes a common metric aggregator, as they perceive no difference between the two attributes[9]. Equation (5a) is the estimable utility expression for each alternative in a stated or revealed choice model.

The willingness to pay (WTP) function is nonlinear. The derivative of the utility expression with respect to a specific attribute is given in equation (7), using free flow time (defined as x_1), and in equation (8) using slowed down time (x_2) as examples of the common form. The exact same functional form for equations (7) and (8) applies to running cost and toll cost, respectively.

[9] As an example, imagine an experimental design with x_1 and x_2 being dummy variables, and the only combinations considered are (1,0) and (0,1). In both cases $(x_1 - x_2)^2 = 1$, and so we have condition:

$$U = (\beta_1 x_1 + \beta_2 x_2)\left(1 - \exp^{-\lambda}\right) + \beta_{12}(x_1 + x_2)\left(\exp^{-\lambda}\right) + \varepsilon$$

If $x_1 = 1$ and $x_2 = 0$, we have condition (a), equivalent to (b).

$$U = (\beta_1 x_1)\left(1 - \exp^{-\lambda}\right) + \beta_{12}(x_1)\left(\exp^{-\lambda}\right) + \varepsilon \tag{a}$$

$$U = (\beta_1 x_1) + (\beta_{12} - \beta_1) x_1 \left(\exp^{-\lambda}\right) + \varepsilon = \left\{\beta_1 + (\beta_{12} - \beta_1)\left(\exp^{-\lambda}\right)\right\} x_1 + \varepsilon \tag{b}$$

The same functional expression applies for x_2. In both cases we have a co-mingling of parameters. If we include the combinations of (1,1) and (0,0), then we have equation (c).

$$U = \beta_{12}(x_1 + x_2) + \varepsilon \tag{c}$$

$$\partial V / \partial x_1 = \beta_1 \left(1 - \exp^{-\lambda(x_1 - x_2)^2}\right) + 2\left(\beta_1 x_1 + \beta_2 x_2\right)\lambda\left(x_1 - x_2\right)\exp^{-\lambda(x_1 - x_2)^2}$$
$$+\beta_{12}\exp^{-\lambda(x_1 - x_2)^2} - 2\beta_{12}\left(x_1 + x_2\right)\lambda\left(x_1 - x_2\right)\exp^{-\lambda(x_1 - x_2)^2} \quad (7)$$

$$\partial V / \partial x_2 = \beta_2 \left(1 - \exp^{-\lambda(x_1 - x_2)^2}\right) - 2\left(\beta_1 x_1 + \beta_2 x_2\right)\lambda\left(x_1 - x_2\right)\exp^{-\lambda(x_1 - x_2)^2}$$
$$+\beta_{12}\exp^{-\lambda(x_1 - x_2)^2} + 2\beta_{12}\left(x_1 + x_2\right)\lambda\left(x_1 - x_2\right)\exp^{-\lambda(x_1 - x_2)^2} \quad (8)$$

2.2 Process II: Common Metric Attribute Parameter Assignment

We now introduce a new heuristic on top of this general non-linear specification, to account for parameter transfer. Essentially, *we replace the aggregation of the two attributes with a parameter transfer rule.* The attribute process model proposed assumes that if a common-metric attribute (i.e., time or cost components) is greater in magnitude to the other attribute, then individuals transfer the parameter assigned initially to the former attribute to the latter attribute. We call this process 'attribute marginal disutility referencing' (Hensher and Layton 2008).

In this new model, the processing sets determined by α and the *x*'s are more complicated. First we note that for each pair of common metric attributes, say x_1 and x_2, there are three regimes, whether $x_1 > x_2$, $x_1 = x_2$, or $x_1 < x_2$. Next, in the language of our model, when x_1 and x_2 are cognitively close, neither attribute is expected to dominate, and hence the *α* conditions are reversed from the previous heuristic. Allowing α to follow an exponential distribution, as above, results in the following utilities for a situation in which there will be two sets of common metric attributes for two time measured attributes (x_{t1} and x_{t2} associated with λ_t), and two cost measured attributes (x_{c1} and x_{c2} associated with λ_c). We can write the overall utility in terms of the sub-utilities for time, V_t, and cost, V_c.

If $x_{t1} > x_{t2}$:

$$V_t = \left(\beta_{t1}x_{t1} + \beta_{t2}x_{t2}\right)\left(\exp^{-\lambda_t(x_{t1} - x_{t2})^2}\right) + \beta_{t1}\left(x_{t1} + x_{t2}\right)\left(1 - \exp^{-\lambda_t(x_{t1} - x_{t2})^2}\right) \quad (9)$$

If $x_{t1} < x_{t2}$:

$$V_t = \left(\beta_{t1}x_{t1} + \beta_{t2}x_{t2}\right)\left(\exp^{-\lambda_t(x_{t1} - x_{t2})^2}\right) + \beta_{t2}\left(x_{t1} + x_{t2}\right)\left(1 - \exp^{-\lambda_t(x_{t1} - x_{t2})^2}\right) \quad (10)$$

In the case of $x_{t1} = x_{t2}$, evaluating either (9) or (10) at $x_{t1} = x_{t2}$ yields:

$$V_t = \left(\beta_{t1} x_{t1} + \beta_{t2} x_{t2}\right) \tag{11}$$

If $x_{c1} > x_{c2}$:

$$V_c = \left(\beta_{c1} x_{c1} + \beta_{c2} x_{c2}\right)\left(\exp^{-\lambda_c (x_{c1} - x_{c2})^2}\right) + \beta_{c1}\left(x_{c1} + x_{c2}\right)\left(1 - \exp^{-\lambda_c (x_{c1} - x_{c2})^2}\right) \tag{12}$$

If $x_{c1} < x_{c2}$:

$$V_c = \left(\beta_{c1} x_{c1} + \beta_{c2} x_{c2}\right)\left(\exp^{-\lambda_c (x_{c1} - x_{c2})^2}\right) + \beta_{c2}\left(x_{c1} + x_{c2}\right)\left(1 - \exp^{-\lambda_c (x_{c1} - x_{c2})^2}\right) \tag{13}$$

In the case of $x_{c1} = x_{c2}$, evaluating either (12) or (13) at $x_{c1} = x_{c2}$ yields:

$$V_c = \left(\beta_{c1} x_{c1} + \beta_{c2} x_{c2}\right) \tag{14}$$

The overall utility for alternative k is given in equation (15).

$$U_k = V_{time,k} + V_{\cos t,k} + \varepsilon_k \tag{15}$$

In any single utility expression setting, the rules will identify pairs of time and cost situations given above. There are three time conditions and three cost conditions, giving nine possible combinations for each alternative in a choice set. If λ tends toward ∞, every individual becomes a common metric re-packager using the parameter transfer rule for the marginal (dis)utility of each attribute. For all cases, as the common metric attributes become equal the standard compensatory model results.

The willingness to pay expressions for Process II are summarised below. The exact same functional form for equations (16) and (17) applies to running cost and toll cost respectively.

If $x_{t1} > x_{t2}$:

$$\partial V_t \Big/ \partial X_{t1} = (\beta_{t1})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right) - 2(\beta_{t1}x_{t1} + \beta_{t2}x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2})) +$$
$$\beta_{t1}\left(1-\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right) + 2\beta_{t1}(x_{t1}+x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2})) =$$

$$(\beta_{t1}) \ - 2(\beta_{t1}x_{t1} + \beta_{t2}x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2})) +$$
$$2\beta_{t1}(x_{t1}+x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2}))$$

(16)

If $x_{t1} < x_{t2}$:

$$\partial V_t \Big/ \partial X_{t1} = (\beta_{t1})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right) - 2(\beta_{t1}x_{t1} + \beta_{t2}x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2})) +$$
$$\beta_{t1}\left(1-\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right) - 2\beta_{t1}(x_{t1}+x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2})) =$$

(17)

$$(\beta_{t1}) \ - 2(\beta_{t1}x_{t1} + \beta_{t2}x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2})) -$$
$$2\beta_{t1}(x_{t1}+x_{t2})\left(\exp^{-\lambda_t (x_{t1}-x_{t2})^2}\right)(\lambda_t(x_{t1}-x_{t2}))$$

If $x_1 > x_2$:

$$\partial V / \partial x_1 = \beta_1; \partial V / \partial x_2 = \beta_2 \qquad (18)$$

2.3 Process III: Attribute Non-Attendance

The general form above is not suitable for the attribute non-attendance heuristic, since it collapses down to a simple linear model. Given four attributes, the proposed utility for alternative k is given in equations (19) and (20).

$$U_k = V_{freeflowtime,k} + V_{sloweddowntime,k} + V_{toll\,cost,k} + V_{runnung\,cost,k} + \varepsilon_k \qquad (19)$$

or (suppressing the subscript k):

$$U = \beta_{ff}\, x_{ff}\left(1-\exp^{-\lambda_{ff} x_{ff}}\right) + \beta_{sd}\, x_{sd}\left(1-\exp^{-\lambda_{sd} x_{sd}}\right) + \beta_{rc}\, x_{rc}\left(1-\exp^{-\lambda_{rc} x_{rc}}\right) + \beta_{toll}\, x_{toll}\left(1-\exp^{-\lambda_{toll} x_{toll}}\right) + \varepsilon \qquad (20)$$

As λ tends toward 0, the probability of an attribute being non-attended increases; as λ tends toward ∞, the probability of full preservation increases. The WTP as the derivative of the utility expression with respect to a specific attribute is given in equation (21), using free flow time (defined as x_1) as an example of the common form.

$$\partial V / \partial x_{ff} = \beta_{ff} - \beta_{ff} \left(\exp^{-\lambda x_{ff}} \right) + \beta_{ff} \, x_{ff} \lambda \left(\exp^{-\lambda x_{ff}} \right) \qquad (21)$$

The focus above is a (potentially) behaviourally richer specification of the utility expression in a simple MNL model that embeds a number of process heuristics adopted by choice makers, which offers new opportunities to extract greater behavioural richness from simpler econometric specifications, in contrast to preserving the full attribute and parameter preservation assumption and introducing random parameters through mixed logit models. In time, we see the research evidence herein being extended to more advanced econometric specifications, but a reappraisal in the context of attribute processing under a simple MNL framework has merit in gaining a better understanding of the role of processing strategies in conditioning the parameters of specific attributes, and hence willingness to pay for such attributes.

2.4 Process IV: Latent Class Specification: Non-Attendance and Dual Processing of Common-Metric Attributes in Choice Analysis

The three classes of processing heuristics presented above can also be evaluated within a latent class model framework (see Hensher and Greene 2008)[10]. The underlying theory of the latent class model posits that individual behaviour depends on observable attributes and on latent heterogeneity that varies with factors that are unobserved by the analyst. It is assumed that individuals are implicitly sorted into a set of Q processing classes, but which class contains any particular individual, whether known or not to that individual, is unknown to the analyst. The behavioural model is a logit model for discrete choice among J_i alternatives, by individual i observed in T_i choice situations, given in (22).

$$\text{Prob[choice } j \text{ by individual } i \text{ in choice situation } t \mid \text{class } q] = \frac{\exp(\mathbf{x}'_{it,j}\boldsymbol{\beta}_q)}{\sum_{j=1}^{J_i} \exp(\mathbf{x}'_{it,j}\boldsymbol{\beta}_q)} \qquad (22)$$

The number of observations and the size of the choice set may vary by individual. In principle, the choice set could vary by choice situation as well. For convenience, we allow y_{it} to denote the specific choice made, so that the model provides:

$$P_{it \mid q}(j) = \text{Prob}(y_{it} = j \mid \text{class} = q). \qquad (23)$$

[10] In 2007 Stephane Hess gave a presentation in Norway where he considered a latent class model to accommodate attribute non-attendance and aggregation (see Hess and Rose 2007). This was brought to my attention after completing the paper by Hensher and Greene (2008). Swait and Adamowicz (2001a) have also used latent class modelling to accommodate decision complexity.

For convenience, we also simplify this further to $P_{it\,|\,q}$. For the given class assignment, the contribution of individual i to the likelihood is the joint probability of the sequence $\mathbf{y}_i = [y_{i1}, y_{i2}, \ldots y_{iT}]$, given in (24).

$$P_{i|q} = \prod_{t=1}^{T_i} P_{it|q} \tag{24}$$

The class assignment is unknown. Let H_{iq} denote the prior probability for class q for individual i. A convenient form is the multinomial logit (MNL) (equation 25).

$$H_{iq} = \frac{\exp\left(\mathbf{z}_i'\boldsymbol{\theta}_q\right)}{\sum_{q=1}^{Q}\exp\left(\mathbf{z}_i'\boldsymbol{\theta}_q\right)},\ q = 1,\ldots,Q,\ \boldsymbol{\theta}_Q = \mathbf{0}, \tag{25}$$

where $\mathbf{z}_i$ denotes a set of observable characteristics which enter the model for class membership. To account for possible heuristics defined in the domains of attribute non-attendance, aggregation, and common-metric parameter transfer, we impose restrictions on parameters within each latent class, each class representing a particular process heuristic[11]. For example, to impose the condition of non-attendance of a specific attribute, we set its parameter to zero; to impose common-metric aggregation, we constrain two parameters to be equal; and to allow for parameter transfer, we define a single parameter based on the parameter associated with a specific attribute[12].

3. Empirical Illustration

To illustrate the implications of each attribute processing strategy, we use a data set drawn from a study undertaken in Sydney in 2004, in the context of car driving commuters making choices from a range of level of service packages defined in terms of travel times and costs, including a toll where applicable. The stated choice questionnaire presented respondents with sixteen choice situations, each giving a choice between their current (reference) route and two alternative routes with varying trip attributes. The sample of 243 effective interviews, each responding to 16 choice sets, resulted in 3,888 observations for model estimation.

To ensure that we captured a large number of travel circumstances and potential attribute processing rules, we sampled individuals who had recently undertaken trips of various travel times, in locations where tollroads currently exist.[13] To ensure some variety in trip length, an individual was assigned to one of the three trip length segments based on a

[11] Importantly, the number of classes selected is determined by the number of heuristics to investigate, and not by the usual BIC and AIC comparisons across a varying number of classes.
[12] Unlike process rule II, which is defined on the absolute levels of each attribute, the latent class model transfers the parameter to the entire sample *within* the class rule. Hence the parameter transfer rule under process IV is strictly different to that under process rule II.
[13] Sydney has a growing number of operating tollroads; hence drivers have had a lot of exposure to paying tolls.

recent commuting trip: no more than 30 minutes, 31 to 60 minutes, and more than 61 minutes (capped at two hours). A telephone call was used to establish eligible participants from households stratified geographically, and a time and location agreed to for a face-to-face computer aided personal interview (CAPI).

A statistically efficient design (see Rose and Bliemer 2007, Sandor and Wedel 2002) that is pivotted around the knowledge base of travellers, is used to establish the attribute packages in each choice scenario, in recognition of supporting theories in behavioural and cognitive psychology and economics, such as prospect theory. A pivot design recognises the useful information contained in a revealed preference alternative, capturing the accumulated exposure to the studied context. Further details of the design of the choice experiment and merits of pivot or referenced designs are provided in Hensher and Layton (2008), Hensher (2008a) and Rose et al. (2008).

The two stated choice alternatives are unlabelled routes. The trip attributes associated with each route are free flow time, slowed down time, trip time variability, running cost and toll cost. All attributes of the stated choice (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 1.

Table 1: 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 (games). The design has no dominance.[14] The distinction between free flow and slowed down time is designed to promote the differences in the quality of travel time between various routes – especially a tolled route and a non-tolled route, and is separate to the influence of total time. Free flow time is interpreted with reference to a trip at 3 am in the morning when there are no delays due to traffic.[15] An example of a stated choice screen is shown as Figure 1.

[14] The survey designs are available from the author on request.
[15] This distinction does not imply that there is a specific minute of a trip that is free flow per se, but it does tell respondents that there is a certain amount of the total time that is slowed down due to traffic etc., and hence a balance is not slowed down (i.e., is free flow like one observes typically at 3am in the morning).

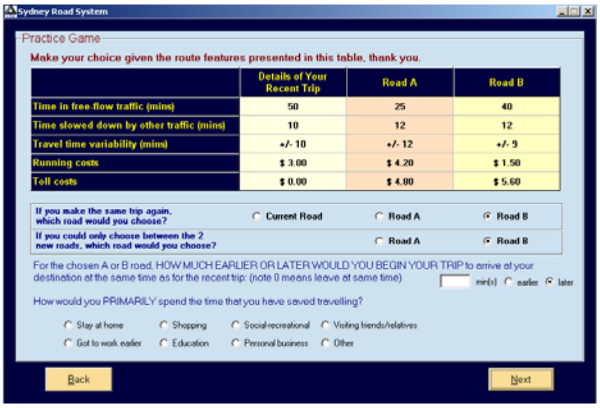

Figure 1: An example of a stated choice screen

4. Evidence on Marginal Willingness to Pay: Value of Travel Time Savings

4.1 Evidence from Processing Models I-IV

In this section, we bring together the evidence on value of travel time savings (VTTS) when one or more processing strategies are accounted for in modeling choice outcomes. The estimated models are not presented herein since they are given in Layton and Hensher (2008), Hensher and Layton (2008), Hensher and Rose (in press), and Hensher and Greene (2008). In all cases, we have accounted for the panel structure of the data. Our interest in this paper is on establishing the extent of under or over estimates of mean VTTS, in contrast to full relevancy and compensatory rules, when account is taken of the various process rules set out above.

To obtain a VTTS distribution for each of free flow and slowed down time, we have to either simulate the distribution across values for the attribute(s) of interest, or apply the formula to a sample of observations. We chose the latter, using the same data used to estimate the models. Given that the denominator in the WTP expression is a weighted average of the role of running cost and toll cost, where the weights reflect the incidence of running and toll cost, and the numerator includes both attributes with a common metric, the WTP for a specific trip time component (i.e., free flow or slowed down time) is dependent on a mix of levels of all four attributes.

We summarise the evidence in Table 2, including the reference source. The major finding is that all mean estimates of VTTS are higher when one or more processing rules are accounted for, in contrast to the traditional MNL model that assumes full attribute and parameter preservation. There is a clear trend here that, if reinforced by other data sets, sends a warning about the under-estimation of VTTS when processing heuristics are not accounted for. The extent of under-estimation appears significant; for the overall weighted average travel time it ranges from a high of 34.7 percent for the full set of process rules in the latent class model to a low of 7.3 percent for attribute aggregation for both time and cost[16].

Table 2: Summary of Willingness to Pay Estimates *(2004$/person hour)*

Process Rule	VTTS: Free flow time	VTTS: Slowed down time	VTTS: Weighted average time	Reference
Full preservation of attributes and parameters MNL	11.76	15.72	14.07*	Hensher and Greene (2008)
Full preservation of attributes and parameters Mixed Logit	14.11	16.78	15.67*	Hensher and Greene (2008)
Process I: Attribute aggregation	12.87	16.78	15.10	Layton and Hensher (2008)[17]
Process II: Parameter transfer	13.37	19.44	16.91	Hensher and Layton (2008)
Process III: Attribute non-attendance	15.28 (1.91)	22.05 (2.74)	19.23	Hensher and Rose (in press)
Process IV: Latent class mixture of all rules	-	-	19.62*	Hensher and Greene (2008)

Notes: * The standard errors have been obtained by bootstrapping. The mean standard deviations for MNL, Mixed Logit and Latent Class are respectively $1.42, $3.71 and $5.10. We can reject the null of no difference between LC and MNL and between LC and mixed logit but not between MNL and mixed logit.

We take a closer look at the findings from the latent class model, summarised in Table 3. There is a range of mean estimates of the value of travel time savings across the latent classes. The range is $1.35 to $42.19, after dividing the marginal disutility of each time component by the weighted average cost parameter, where the weights are the levels of running and toll cost. To obtain an overall sample average, we have to weight each mean estimate by the probability of class membership.

The overall sample weighted average for total time is $19.62, which contrasts with $14.07 for the traditional MNL specification in Table 1 (Hensher and Greene 2008, Table 3). The mean estimate of VTTS is 39.4 percent higher when process heterogeneity is accounted for across three classes of heuristics. A closer look at the contribution of each

[16] It is worth noting that the attribute aggregation model (Process I) allowed for aggregation of both the time and the cost components. By contrast, the latent class model (Process IV) only found time aggregation statistically significant, but did identify a significant effect from the heuristic that transferred the toll cost parameter to the running cost attribute. What this latter evidence suggests is that individuals do not tend to add up the cost components, but tend to re-weight their influence by the parameter transfer rule.

[17] In order to estimate the model as a panel, Layton and Hensher (2008) used a combination of many start values and simulated annealing (code written by Tsionas 9/4/95, available at the American university Gauss Archive: http://www.american.edu/academic.depts/cas/econ/gaussres/GAUSSIDX.HTM). Using the maximum from the simulated annealing approach, we then computed a Newton-Raphson iteration using 500 replications of the simulator, and computed the covariance from all terms except for λ_t and λ_c.

heuristic suggests that attribute addition for the two time components produces the highest mean estimate contribution to VTTS, after controlling for class membership. Ignoring free flow time is the next contributor, followed by full attendance to all attributes. Ignoring running cost and slowed down time is the next contribution.

Table 3: Values of travel time savings from a Latent Class Model *(2004$ per person hour)*

NAT = not attended to ParT = parameter transfer	Class membership probability	Free flow time	Slowed down time	Total time
All attributes attended to	0.2817	5.87	9.89	8.22
Free flow NAT	0.1119		23.02	23.02
Toll cost NAT	0.0359	3.95	8.93	6.85
Slowed down time NAT	0.0643	1.35		1.35
Running cost and slowed down time NAT	0.0497	42.19		42.19
Free flow and slowed down time added	0.2978	37.57		37.57
Free flow to slowed down and vice versa ParT	0.0758	4.57		4.57
Free flow to slowed down ParT and running cost to toll cost and vice versa ParT	0.0829	9.26		9.26
Class membership weighted VTTS				**19.62 (5.10)**

Source: Hensher and Greene (2008). Note: standard error for each component VTTS are available on request.

4.2 Evidence from Self-stated Processing Response for Common-Metric Addition

The focus of the previous sections was on exploring a way in which we are able to allow for the possibility of heterogeneity in the way that individuals process common-metric attributes in making choices, focusing on a number of potential heuristics, without having to ask supplementary (deterministic) elicitation questions. In addition to the stated choice experiment, in the survey we did however ask supplementary elicitation questions shown in Figure 2[18]. In this section we investigate the possible implications of conditioning the preference function use to derive WTP estimates, using the response to question 2 to illustrate the empirical implications. A large percentage of the respondents stated, in

[18] This question was asked after completion of all 16 choice tasks. An alternative approach is to ask these questions after each choice task as was the case in Puckett and Hensher (2008, 2009), and Scarpa, Thiene and Hensher (2008). Our preference is for choice-task-specific self-stated processing questions, especially where attribute level matters; however this comes at the risk of cognitive burden and the possibility that the number of choice tasks might have to be reduced. We also recognise the potential limitation of such questions, and devote a later section to future ways of investigating question structure, and the believability/plausibility of the evidence.

supplementary questions (see Hensher 2008), that they added the components: 88.06 percent and 76.5 percent respectively for time and cost.

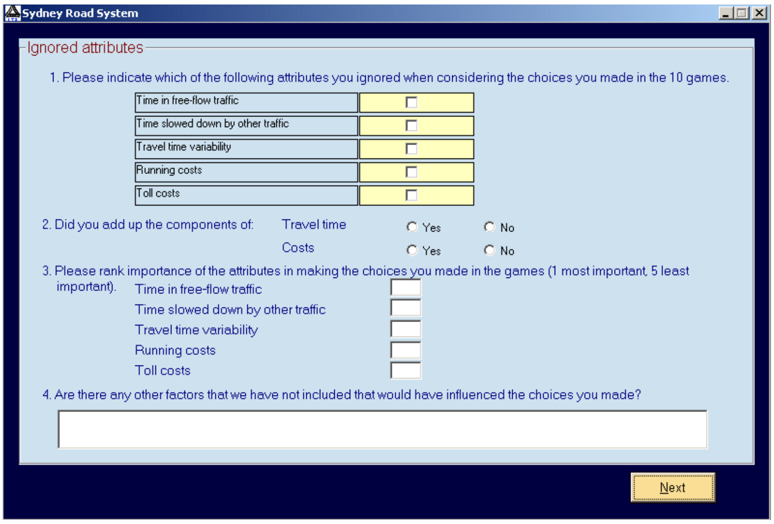

Figure 2: CAPI questions on attribute relevance

We estimated five panel-specification models - two mixed logit (with and without error components), and three latent class models. One mixed logit model ignored the attribute processing rule, and the other accommodated it through the specification of separate parameters to capture the following conditions: (i) added up times but not costs, (ii) added up costs but not times, (iii) added up both times and costs, and (iv) preserved all four attributes as separate components. One latent class model defined four class memberships as per (i)-(iv) above without recourse to information from the supplementary questions, whereas another latent class model conditioned class membership on conditions (i)-(iv). A base latent class model assumed all attributes are treated separately, but three classes were identified with statistically significant latent class probabilities. The findings are summarised in Table 4. Mixed logit and latent class models are well documented in the literature.

Table 4: Influence of Self-Stated Attribute Processing Strategy (APS) on VTTS

(i) Mixed Logit Models (panel specification)

Attributes	Mixed Logit model (constrained triangular for random parameters), t-ratios in brackets except for VTTS, which is the standard deviation	
	No allowance for self-stated APS	Allowance for self-stated APS
Random parameters:		
Free flow time (FF)	-0.10023 (-17.33)	-0.0497 (-3.64)
Slowed down time (SDT)	-0.1147 (-21.94)	-0.687 (-5.98)
Aggregated FF and SDT	-	-0.1236 (-22.5)
Running cost (RC)	-0.4167 (-14.58)	-0.1945 (-4.11)
Toll cost (TC)	-0.188 (-22.99)	-0.2905 (-9.70)
Aggregated RC and TC	-	-0.6103 (-21.62)
Fixed parameter:		

Non-reference alternative dummy	-0.1344 (-2.88)	-0.1669 (-3.61)
Log-likelihood at convergence	-2762.80	-2711.88
Log-likelihood at zero	4271.41	
Weighted average VTTS ($Aus2004 per person hour)	**$15.87 ($10.14)**	**$20.12 ($16.01)**

(ii) Mixed Logit Models (panel specification) with Error Component

Attributes	Mixed Logit model (constrained triangular for random parameters), t-ratios in brackets except for VTTS which is standard deviation	
	No allowance for self-stated APS	Allowance for self-stated APS
Random parameters:		
Free flow time (FF)	-0.11190 (-31.45)	-0.08113 (-5.50)
Slowed down time (SDT)	-0.12746 (-34.25)	-0.07514(-7.06)
Aggregated FF and SDT	-	-0.13076 (-19.37)
Running cost (RC)	-0.49740 (-19.74)	-0.23583(-3.96)
Toll cost (TC)	-0.55193(-32.95)	-0.26234(-7.489)
Aggregated RC and TC	-	-0.65814(-17.19)
Fixed parameter:		
Non-reference alternative dummy	0.18195 (1.95)	-0.27233 (-2.13)
Standard deviation of latent random effect	*2.43423 (24.5)*	*2.3357 (28.21)*
Log-likelihood at convergence	-2485.03	-2447.43
Log-likelihood at zero	4271.41	
Weighted average VTTS ($Aus2004 per person hour)	**$16.11($10.87)**	**$22.63 ($23.26)**

(iii) Latent Class Models (panel specification)

Base Model:

	Class 1	Class 2	Class3
Free flow time	-0.04006 (-4.7)	-0.2022 (-28.9)	-0.0338 (-7.5)
Slowed down time	-0.0603 (-9.6)	-0.2009 (-31.6)	-0.0749 (-22.0)
Running cost	-0.3323 (-8.9)	-0.3399 (-10.7)	-0.4739 (-15.3)
Toll cost	-0.2883 (-10.7)	-0.3417 (-24.2)	-0.6115 (-33.6)
Non-reference alternative	2.5043 (12.3)	0.3947 (-7.2)	-1.0281 (-23.3)
Class membership probability	0.263 (6.92)	0.361 (10.45)	0.376 (11.14)
Log-likelihood at convergence	-2542.74		
Log-likelihood at zero	-4271.41		
Weighted average VTTS ($Aus2004 per person hour)	**$17.89**		

Models Allowing for Attribute Processing:

	No allowance for self-stated APS		Allowance for self-stated APS	
Latent Class attributes:	Class membership probability	Parameter estimates for FF,SDT,RC,TC, NONSQ (t-ratios in brackets)	Class membership probability	Parameter estimates for FF,SDT,RC,TC,NONSQ (t-ratios in brackets)
All attributes treated separately	0.379	-.049,-.090,-.638,-.743, -.622 (-5.5,-13.0,-11.3,-19.1,-6.9)	0.381	-.055,-.092,-.648, -.748,-.637 (-5.0,-12.1,-10.1,-16.3,-6.7)
Time components aggregated	0.050	-.057,-.057,-0.29,-0.38,-3.9 (-3.3,-3.3,-1.9,-9.2,-11.1)	0.052	-.054,-.054,-.332, -.370,-3.82 (-3.2,-3.2,-2.0,-8.4,-10.4)
Cost components aggregated	0.318	-.217,-.212,-.319,-.319,-.428 (-26.9,-29.2,-19.1,-19.1,-6.8)	0.310	-.221,-.215,-.317,-.317,-.410 (-25.1,-27.8,-17.5,-17.5,-6.3)
Time and Cost components aggregated	0.253	-.052 ,-.052,-.282,-.282,2.58 (-17.4,-17.4,-25.4,-25.4,22.2)	0.257	-.050,-.050,-.277,-.277,2.49 (-16.1,-16.1,-23.2,-23.2,21.9)
Theta in Class Probability:				
			Constant, FF,SDT,FFSDT,RC,TC,RCTC Note: all covariates are in minutes or dollars, except the constant. Statistically significant: *= 5% , **= 10 % level	
All attributes treated separately			1.35**,-.006,.003,-.005,-.33,-.079,-.093 (2.4,-.17,.14,-.61,-1.1,-.45,-1.4)	
Time components			-1.59,.18*,-.45,.009,.52,-.61,-.13 (-1.2,1.9,-1.4,.44,1.6,-	

aggregated		1.1,-.7)
Cost components aggregated		1.16*,-.02,-.03,-.009,.35*,-.15,-.13* (1.9,-.7,-1.1,-.9,1.7,-.9,-1.7)
Log-likelihood at convergence	-2427.57	-2399.64
Log-likelihood at zero	-4271.41	
Weighted average VTTS ($Aus2004 per person hour)	**$18.02 ($15.02)**	**$18.05 ($15.28)**

For mixed logit, we have selected a quasi-constrained triangular distribution for each random parameter, in which the spread[19] estimate is constrained to equal the mean estimate for the random parameters. If the scale equals 1.0, the range is 0 to 2 β_1. This is an appealing way of capturing the random taste heterogeneity, avoiding the search for such heterogeneity at the extremes of unconstrained distributions[20]. The triangular distribution was first used for random coefficients by Train and Revelt (2000) later incorporated into Train (2003), and it is increasingly being used in empirical studies.

The overall goodness-of-fit for the models, with allowance for self-stated APS, are statistically better than when self-stated APS is not accounted for. The mixed logit models differ in the way that the time and cost attributes are included in the utility expressions, but in both models all parameters have the expected negative signs, and are statistically significant at the one percent level. Given the different ways that free flow and slowed down time are handled, the most sensible representation of the value of travel time savings is as a weighted average estimate, with weights associated with the contribution of each of the three specifications of cost and of time. The VTTS in Table 4 are based on conditional distributions (that is, conditional on the alternative chosen). The VTTS in the mixed logit model is significantly higher when the self-stated APS is accounted for, i.e., $20.12 (22.63 with error components) per person hour, compared to $15.87 ($16.11 with error components) per person hour.

The latent class model is based on four attribute addition rules (i)-(iv), and all time and cost parameters are statistically significant at the one percent level, and of the expected sign when class membership is conditioned on the self-stated APS; however when the self-stated APS are not included, all but one parameter is statistically significant at the one percent level, the exception being running cost in the second class, which has a 10 percent significance level. The overall log-likelihood at convergence is greatly improved over the mixed logit model for both latent class models, suggesting that the discrete nature of heterogeneity captured through latent class is a statistical improvement over the continuous representation of heterogeneity in the mixed logit model. The weighted average VTTS are derived first across classes for each attribute, based on conditional distributions associated with the probability of class membership of each respondent within each class, and then a further weighting is undertaken using weights that reflect the magnitudes of the components of time and cost.

[19] The spread is the standard deviation times $\sqrt{6}$.

[20] We acknowledge that this restriction is controversial; although we prefer to adopt it in contrast to unconstrained distributions where sign changes are common, or eliminating all negative VTTS as some analysts do.

The weighted average VTTS in the two latent class models that account for attribute processing are virtually identical. What this suggests is that once we have captured the alternative processing rules, though the definition of latent classes, the inclusion of the self-stated APS rules as conditions on class membership do not contribute additional statistically useful evidence to revise the findings, in the aggregate. This is consistent with the statistical non-significance of most of the self-stated APS variables; with only three parameters having a 10 percent significance level (excluding the constants). There were no parameters with one or five percent significance levels. However, when we contrast this evidence to the base latent class model which makes no allowance for attribute processing, the mean VTTS is only slightly lower (i.e., $17.89 per person hour compared to $18.02, and $14.07 for the MNL model). What this may suggest is that the latent class specification may have done a good job is approximating the way in which attributes are processed.

These findings support the hypothesis that allowance for attribute processing rules tends to result, on average, in a higher mean estimate of WTP for travel time savings. This is consistent, directionally, with other studies undertaken by Rose et al. (2005) and Hensher and Layton (2008).

5. Other Perspectives - Belief and Support Functions to Establish Judgment of Evidence Strength

A growing number of studies ask supplementary questions, such as those illustrated in the previous section, to elicit how respondents processed specific attributes (see e.g., Hensher 2008, Hess and Hensher 2008). The reliability of responses to such questions (e.g., 'which attributes did you ignore?' or 'which attributes did you add up?') is not without controversy (see Bertrand and Mullainathan 2001), with preliminary evidence suggesting that the marginal WTP, when the responses to supplementary intention questions are used to condition the treatment of an attribute in model estimation, are sometimes higher and sometimes lower than when processing is excluded. In contrast, the (limited but growing) evidence appears to be consistently in the upward direction when heuristics are tested through the functional specification of non-linear utility expressions. So which tendency is 'correct'? The answer is far from clear. Furthermore, some studies have shown that the expectation of a parameter approaching zero, when a respondent claims that they ignored an attribute, is not proven (Hess and Hensher 2008); in contrast a recent study by Rose et al. (2008a) using Bayesian methods, found encouraging evidence that self-stated responses on attribute non-preservation were indeed consistent with a statistically non-significant difference around a zero parameter estimate.

One potentially fruitful way forward is to transform the self-stated processing responses to recognise the potential for error in response[21]. While there are a number of ways that

[21] Another interesting approach is to recognise the role of heterogeneity and to identify what are the strongest parametric assumptions and see how they can be relaxed through modelling heterogeneity. King

this might be undertaken, one way that is worthy of investigation is known as the belief-function perspective. There is a large literature on believability, emanating from the works of Dempster and Shafer in the 1960s and 1970s (Shafer 1976, Dubois and Prade 1988). Although not focused on attribute processing in choice analysis per se, the sentiment is aligned[22].

The focus is on the uncertainty that arises because of the lack of knowledge of the true state of nature, where we not only lack the knowledge of a stable frequency (how can we be sure that the heuristic adopted is stable over time? (see also Yoon and Simonson 2008), as implied by the process heuristics in previous sections), but also we lack the means to specify fully the fixed conditions under which repetitions can be performed (Shafer and Srivastava 1990). The Dempster-Shafer theory of belief functions is used to assess reliability of evidence, which provides support for the presence or absence of such a variable in situations where the event cannot be treated as a random variable. Dempster (1967) introduces belief functions from a statistical perspective in terms of carrying a (frequentist) probability measure from a 'space of observations' to a 'space of interpretations of these observations' by a 'point-to-set mapping' (Dubois and Prade 1988).

Many individuals are influenced by the views of others, suggesting that additional information on the believability of an individual's response may be aided by this extra evidence[23]. Thus we need to find ways in which we can triangulate evidence from various sources, in order to establish a measure of belief of the evidence offered by an individual on how they processed specific attributes. The level of belief, on whether the person in question processed an attribute using a specific rule, or not, depends on the items of evidence, and their credibility. A belief function treatment of such problems provides a possible framework. It involves three constructs - belief functions (BF), plausibility functions (PF), and a measure of ambiguity. When combined, especially BF and PF, we obtain Dempster's rule of what I term 'rule reliability'. We now explain this rule in more detail, and suggest the nature of data required in future studies to embed the rule-reliability measure into the estimation of choice models.

The Dempster-Shafer theory of belief functions is similar to probability theory, however, with one difference. Under probability theory, uncertainty is assigned to the state of nature based on the knowledge of frequency of occurrence. However, under belief functions, uncertainty is assigned to the state of nature or assertion of interest in an indirect way, *based on the probability knowledge in another frame*, by mapping that knowledge onto the frame of interest. This mapping may not necessarily be one-to-one.

---

and Wand (2007) offer some interesting ideas in the context of ordered choices and anchoring Vignettes identified from supplementary information designed to account for differences in perceptions across respondents. See Greene and Hensher (2009, chapter 7).

[22] A linked literature in social psychology is focused on the meaning and moderators of attitude strength, where strong attitudes are characterized by high levels of confidence and stability. Attitude confidence is defined as the degree to which an individual is certain that his attitude is correct. See Krosnick and Schuman (1988).

[23] Extra evidence may be obtained from other questions asked to the same individual, as well as questions to other persons.

To illustrate, suppose we have a variable, say A, with n possible mutually exclusive and exhaustive set of values: $a_1, a_2, a_3, \ldots, a_n$. These values could be alternative ways that an attribute is processed (in isolation or in conjunction with other attributes), and/or processing responses to different question structures, including a simple binary statement of 'ignored or did not ignore'[24] the attribute, or 'added up or did not add up two attributes of a common metric'. Define the frame, $\Theta = \{a_1, a_2, a_3, \ldots, a_n\}$ of discernment[25] for the variable A. Under probability theory, for such a set, we assign a probability mass, $P(a_i)$, to each state a_i such that the sum of all these probabilities equals one, i.e., $\sum_{i=1}^{n} P(a_i) = 1$. However, under the Dempster-Shafer theory of belief functions, uncertainties are assigned in terms of belief masses to not only singletons, but also to all the sub-sets of the frame, and to the entire frame Θ. These belief masses add to one, similar to probability masses.

The entire frame Θ in our example might be a binary setting of 'ignored' (a_1) and 'not ignored' (a_2) for a specific attribute associated with an alternative and/or a choice task[26]. These belief masses define a function called the *basic belief mass function* (Shafer, 1976). We can write a belief mass assigned to a subset B as m(B), where B could be a single element, or a subset of two, a sub-set of three (e.g., degrees of attribute preservation), and so on, or the entire frame, Θ. The sum of such belief masses equals one, i.e., $\sum_{B \subseteq \Theta} m(B) = 1$. When the non-zero belief masses are only defined on the singletons, the belief function reduces to probability theory. Thus, one can argue that probability theory is a special case of Dempster-Shafer theory of belief functions.

To crystallise this distinction in an example, suppose we were able to determine, from a number of sources, that m(IG) =0.3, m(NIG)=0, and m(IG,NIG)=0.7[27]. IG stands for 'the ignore response being a reasonable representation of reality', and NIG stands for 'the ignored response being *either* materially misstated *or* not reflecting acceptable views of others'.[28] The belief function interpretation of these belief masses is that the analyst has 0.3 level of support for 'IG', no support for 'NIG, and 0.7 level of support remains uncommitted which represents ignorance[29] (Dubois and Prade 1988)[30]. However, if we

[24] Including different wording, such as 'attending to or not attending to'.
[25] That is the quality of being able to grasp and comprehend what is obscure.
[26] It could also be degrees of attribute relevance ($a_1, a_2, \ldots, a_n$), from totally relevant (not ignored) to totally irrelevant (ignored).
[27] Establishing these probabilities is the great challenge.
[28] Information to gauge the reliability of stated self-intentions could be sought from the very same person along similar lines to supplementary questions used in reducing the hypothetical bias (HB) gap in willingness to pay. An example in the HB context is a supplementary *certainty scale* question after each choice scenario, along lines suggested by Johannesson et al. (1999), on a scale 0 (very unsure) to 10 (very sure), to indicate how sure or certain the respondent is that they would actually chose that route (or not at all) at the indicated price and travel time.
[29] A "*complete ignorance heuristic*" (CI) reflects the case in which the individual is not aware of the influence of the attributes in their utility function. Arana et al. (2008, 757) suggest an interpretation as follows: "It collects individuals who do not care about the consequences of their responses, or who do not

had to express the above judgment in terms of probabilities, we get into problems, because we will assign P(IG) = 0.3 and P(NIG) = 0.7, which implies that there is a 70 percent chance that the response to the question is '*materially misstated or does not reflect acceptable views of others'*. However this is not what the analyst's judgment is; he has no information or evidence that ignoring an attribute is materially misstated. Simply knowing the fact that the response appears to be reasonable, compared to the predicted values based on the average views of others, including additional information obtained from the specific individual, provide no evidence that the response to the question on whether an attribute is ignored is materially misstated. It only provides *some level of support* that the subjective response is accurately stated.

The *Belief Function* is defined as follows: The belief in B, Bel(B), for a subset B of elements of a frame, Θ, represents the total belief in B, and is equal to the belief mass, m(B), assigned to B *plus* the sum of all the belief masses assigned to the set of elements (C) that are contained in B. In terms of symbols: $\text{Bel(B)} = \sum_{C \subseteq B} m(C)$.[31]

The *Plausibility Function* is defined as follows: Intuitively, the plausibility of B is the degree to which B is plausible given the evidence. In other words, Pl(B) represents the maximum belief that could be assigned to B, given that all the evidence collected in the future support B. In mathematical terms, one can define plausibility of B as $\text{Pl(B)} = \sum_{B \cap C = \varnothing} m(C)$, which can also be expressed as: Pl(B) = 1 - Bel(~B), which is the degree to which we do not assign belief to its negation (~B). The belief-function measure of ambiguity in an assertion, say B, is the difference between the plausibility of B, and the belief in B (Wong and Wang 1993).

Dempster's rule (Shafer 1976) combines *more than one independent items of evidence,* similar to Bayes' rule in probability theory. Dempster's rule reduces to Bayes' rule under the condition when all the belief masses defined on the frame are zero, except the ones for the singletons. For example, for two independent items of evidence[32] pertaining to a frame of discernment, Θ, we can write the combined belief mass for a sub-set B in Θ using Dempster's rule of combination as:

$$m(B) = \sum_{C1 \cap C2 = B} m_1(C1) m_2(C2)/K, \tag{26}$$

where

---

pay attention to the experiment. In other words, CI is utilized by those individuals who make choices using a completely random process."

[30] Dubois and Prade (1988, 55) state "Probability theory is not very good at modeling weak states of knowledge where the uncertainty about some event is but loosely related to the uncertainty about the contrary event. Especially, total ignorance … cannot be expressed by a single probability measure. Another way of putting it is that probability cannot distinguish between the absence of belief in not-A and the belief in A".

[31] By definition, the belief mass assigned to an empty set is always zero, i.e., $m(\varnothing) = 0$.

[32] Such as the supplementary self-stated intention questions and the reliability question (see footnote 28).

$$K = 1 - \sum_{C1 \cap C2 = \varnothing} m_1(C1) m_2(C2) \qquad (27)$$

The symbols $m_1(C1)$ and $m_2(C2)$ determine the belief masses of C1 and C2, respectively, from the two independent items of evidence represented by the subscripts. K is a re-normalisation constant. The second term in K represents the conflict between the two items of evidence[33]; the two items of evidence are not combinable if the conflict term is 1.

Let us consider an example to illustrate the details of Dempster's rule. Suppose we have the following sets of belief masses obtained from two independent items of evidence[34] related to the accurate representation of whether an attribute is ignored (IG) or not ignored (NIG):

Evidence 1: $m_1(IG) = 0.3$, $m_1(NIG) = 0.0$, $m_1(\{IG, NIG\}) = 0.7$,
Evidence 2: $m_2(IG) = 0.6$, $m_2(NIG) = 0.1$, $m_2(\{IG, NIG\}) = 0.3$.

The re-normalisation constant for the above case is:

$$K = 1 - [m_1(IG)m_2(NIG) + m_1(NIG)m_2(IG)] = 1 - [0.3*0.1 + 0.0*0.6] = 0.97.$$

Using Dempster's rule (26), the combined belief masses for 'IG', 'NIG', and {IG, NIG} are given by[35]:

---

[33] A challenging problem in combining uncertain information is to decide what to do with conflicts. Generally, combining information issued from conflicting sources leads to un-normalized uncertainty measures. Shafer (1976), in advocating Dempster's rule, suggests that the resulting uncertainty measure should be re-normalised. He motivates his choice by the Sherlock Holmes principle saying that 'having discarded the impossible, what remains, however improbable, is the truth'. But this principle assumes that sources of information are both totally reliable, an over-optimistic assumption in some situations. The weight $m(\varnothing)$, and more generally, the amount of sub-normalisation, assesses the extent to which both sources are indeed reliable. Moreover, the normalisation operation introduces discontinuities in the combination rule. See Dubois and Prade (1987) for a discussion of combination rules for belief functions and possibility measures, including the case of unequally reliable sources.

[34] It is straightforward to generalise to any number of evidence sources.

[35] The term "$m_2(\{IG, NIG\})m_1(\{IG, NIG)\}$" represents the "unknown" or ambiguity factor. This is assigned to m({IG, NIG}). Dempster's rule for combining two items of evidence on a frame {x, ~x} of a binary variable X dictates that the combined belief masses on the frame should be determined as follows: $m(x) = [m_1(x)m_2(x) + m_1(x)m_2(\{x, \sim x\}) + m_1(\{x, \sim x\})m_2(x)]/K$ where $K = 1 - [m_1(x)m_2(\sim x) + m_1(\sim x)m_2(x)]$. The above equation comes from the simple logic that state 'x' is true if both items of evidence suggest that 'x' is true, i.e., $m_1(x)m_2(x)$, or one item of evidence suggests that 'x' is true but the other one is not sure whether it is 'x' or '~x', i.e., $m_1(x)m_2(\{x, \sim x\})$ and $m_1(\{x, \sim x\})m_2(x)$. K is a re-normalization constant to make sure that the combined m-values add to one. As you can see above, the second term in K, i.e., $[m_1(x)m_2(\sim x) + m_1(\sim x)m_2(x)]$, represents conflict between the two items of evidence; one item of evidence is suggesting that 'x' is true and the other one is suggesting that '~x' is true. $m(\sim x) = [m_1(\sim x)m_2(\sim x) + m_1(\sim x)m_2(\{x, \sim x\}) + m_1(\{x, \sim x\})m_2(\sim x)]/K$. The above equation again comes from the logic that state '~x' is true if both items of evidence suggest that '~x' is true, i.e., $m_1(\sim x)m_2(\sim x)$, or one item of evidence suggests that '~x' is true but other one is not sure whether it is 'x' or '~x', i.e., $m1(\sim x)m_2(\{x, \sim x\})$ and $m_1(\{x, \sim x\})m2(\sim x)$. $m(\{x, \sim x\}) = [m_1(\{x, \sim x\})m_2(x, \sim x\})]/K$. The above equation suggests that if both items of evidence are not sure whether it is 'x' or '~x', then the combined evidence also is not sure

$$m(IG) = [m_1(IG)m_2(IG) + m_1(IG)m_2(\{IG, NIG\}) + m_1(\{IG, NIG\})m_2(IG)]/K$$
$$= [0.3*0.6 + 0.3*0.3 + 0.7*0.6]/0.97 = 0.69/0.97 = 0.711,$$
$$m(NIG)=[m_1(NIG)m_2(NIG)+m_1(NIG)m_2(\{IG,NIG\})+m_1(\{IG,NIG\})m_2(NIG)]/K$$
$$= [0.0*0.1 + 0.0*0.3 + 0.7*0.1]/0.97 = 0.07/0.97 = 0.0721,$$
$$m(\{IG, NIG\}) = m_1(\{IG, NIG\})m_2(\{IG, NIG\})/K = 0.7*0.3/0.97 = 0.21/0.97$$
$$= 0.2165. \qquad (28)$$

The combined beliefs and plausibilities that attribute processing is not misstated are:

$$Bel(IG) = m(IG) = 0.711, \text{ and } Bel(NIG) = m(NIG) = 0.0721, \qquad (29)$$
$$Pl(IG) = 1 – Bel(NIG) = 0.928, \text{ and } Pl(NIG) = 1 – Bel(IG) = 0.289. \qquad (30)$$

The choice model, for each individual observation, can have each attribute discounted by the 'plausibility factors': Pl(IG) (=0.928) and Pl(NIG) (=0.289). This might be a decomposition of a random parameter in a mixed logit model, or interaction terms in MNL and latent class models, or through conditioning the scale parameters. These plausibility factors would be applied to all observations, based on evidence obtained from supplementary questions. The challenge for ongoing research is to identify a relevant set of questions posed to the respondent and other agents that can be used to quantify evidence, suitable for deriving the belief and plausibility functions for each respondent.

The Dempster-Shafer theory of belief has links to support theory (Tversky and Koehler 1994, Fox and Tversky 1998, Idson et al. 2001, Hensher 2009), a psychological model of degree of belief, which argues that different descriptions of the same event often give rise to systematically different responses, and hence the judged probability of the union of disjoint events is generally smaller than the sum of judged probabilities of these events. Support theory assumes that subjective probability is attached to *descriptions of events* (e.g., 'which attribute(s) did you ignore?' or 'did you ignore attribute x?'), and not events per se, and hence different descriptions of the same event may be assigned different probabilities.

There is a key distinction, however, between the Dempster-Shafer theory of belief and support theory, linked to the extensionality principle, which states that events with the same extension are assigned the same probability. The extensionality principle is problematic in that alternative descriptions of the same event can produce systematically different judgments. For example, in the context of two states of the world, called events, such as 'two attributes were added up' and 'two attributes were treated separately through unpacking retention', we might ask an individual to consider two routes to take for a given trip. In the first experiment we might offer total times of 30 and 40 minutes (and associated costs of $4 and $2.50); in the second experiment we might offer a free flow time of 20 minutes and a congested time of 10 minutes versus 25 minutes for free flow

---

whether it is 'x' or '~x'. Thus, $[m_1 1(\{x, \sim x\})m_2 (x, \sim x\})]$ maps to $m(\{x, \sim x\})$. We thank Stewart Jones for this clarification.

and 15 minutes for congested time, keeping costs the same as the first experiment. The responses are almost always different across a sample (see Layton and Hensher 2008), since some individuals will initially add up the travel time components and undertake the comparison, essentially equivalencing experiments 1 and 2; whereas other individuals will evaluate the unpacked time components and make a judgment (i.e., the choice) which implicitly has weighted the components differently to the weight obtained for the aggregation of time.

The previous paragraph is essentially re-iterating a point made by Krantz (1991) that Dempster-Shafer's model is more suitable for *judgments of evidence strength* than for *judgments of probability*, the latter being what we focus on in the estimation of the choice model to explain the choice amongst alternatives, in contrast to how we enter a specific attribute into the choice model. The *judgments of evidence strength* is the very role that the plausibility function plays in the context of specifying the way that a specific attribute is processed in the context of stated choice experiments. We are not using the belief (or indeed the support) theory to establish probabilities of outcomes, since that is accommodated though the choice model. The specific feature of the belief paradigm is the idea of superadditivity; namely that multiple sources of evidence (obtained through more detail, or what is commonly referred to as unpacking in the psychology literature) results in a belief in the disjunction of disjoint events being greater than or equal to the sum of the beliefs in each of the components. For example, if we have four attributes (x_1,x_2,x_3,x_4), of which the first two have a common metric (e.g., travel time) and the last two have a common metric (e.g., cost), we might have a number of ways in which we can structure questions suitable for establishing how each specific attribute is processed (in the context of how the package of attributes is processed).

There are a number of possible ways of evaluating an attribute in arriving at a decision on how it will be processed in the context of a choice task. These might, for example, be based on five items of evidence (or heuristics) in relation to the processing of x_1 (the responses could include: (i) ignored or not, (ii) added up with another common-metric attributes, and (iii) transferred the parameter to another common-metric attribute):

E_α = E(x_1): 'I evaluated only x_1 in deciding what role x_1 plays'[36]
E_β = E(x_1,x_2); E(x_1,x_3); E(x_1,x_4);…; E(E(x_1,x_2,x_3,x_4): 'I evaluated x_1 in the context of a subset of the attributes offered'
E_γ = E(x_1,x_2): 'I evaluated x_1 in the context of attributes that have a common metric with x_1'
E_ε = E(x_1+x_2, x_3+x_4): 'I evaluated x_1 by adding up attributes that have a common metric (e.g., times and costs)'
E_η = E(x_1,x_2,x_3,x_4): 'I evaluated every attribute in deciding what role x_1 plays.'

For *each* of these candidate heuristics, the analysts might ask, in the context of whether an attribute x_1 was ignored or not: *'Please allocate 100 points between the three possible ways you might (or did) respond to reflect your assessment of how you believe you used each of the processing rules in determining the role of attribute x_1 :*

[36] The approach commonly adopted in supplementary questions. The precise wording would vary according to the nature of the empirical study.

	E_α	E_β	E_γ	E_ε	E_η
• I definitely ignored (IG) x_1	____	____	____	____	____
• I did not fully ignore, or fully not ignore {IG,NIG}, x_1	____	____	____	____	____
• I definitely did not ignore (NIG) x_1	____	____	____	____	____
	100	100	100	100	100

These heuristics may be randomly assigned to each respondent, or all might be assigned to each respondent (in a randomised order). There are some clear (cognitive) disadvantages of assigning all heuristics to each respondent, yet this may be necessary in order to obtain the required data to calculate a plausibility expression. It might also be of interest to have each respondent rank the heuristics in order of applicability (in the example above, this is a rank from 1 to 5, where 1 = most applicable).

If the focus is on whether an attribute x_1 was ignored or not, then we might identify the following evidence:

$$
\begin{array}{ll}
E\_\alpha: E\_\alpha(IG) = 0.4,\ E\_\alpha(NIG) = 0.2,\ E\_\alpha(\{IG,NIG\}) = 0.4;\ \text{rank} = 4 & \\
E\_\beta: E\_\beta\,(IG) = 0.4,\ E\_\beta\,(NIG) = 0.3,\ E\_\beta\,(\{IG,NIG\}) = 0.3;\ \text{rank} = 3 & \\
E\_\gamma: E\_\gamma\,(IG) = 0.5,\ E\_\gamma\,(NIG) = 0.3,\ E\_\gamma\,(\{IG,NIG\}) = 0.2;\ \text{rank} = 2 & (31) \\
E\_\varepsilon: E\_\varepsilon\,(IG) = 0.3,\ E\_\varepsilon\,(NIG) = 0.3,\ E\_\varepsilon\,(\{IG,NIG\}) = 0.6;\ \text{rank} = 5 & \\
E\_\eta: E\_\eta\,(IG) = 0.5,\ E\_\eta\,(NIG) = 0.2,\ E\_\eta\,(\{IG,NIG\}) = 0.3;\ \text{rank} = 1 &
\end{array}
$$

The responses to (31) can be fed into equation (28) to obtain the belief and plausibility values in (29) and (30), which can then be interacted in a choice model (for ordered or unordered alternatives), with each attribute and/or the scale parameters to account for the attribute processing strategy of each respondent at an alternative and at a choice set level[37].

6. Conclusions

This paper brings together an accumulating set of interesting processing rules that are hypothesised to be applied, to varying degrees, by respondents in assessing choice scenarios in choice experiments. The rules are ways of cognitively rationalising the information on offer in order to make a choice. The paper synthesises the empirical evidence presented in Layton and Hensher (2008), Hensher and Layton (2008), Hensher and Rose (in press), and Hensher and Greene (2008), and offers new evidence to support the view that failure to identify and account for process heterogeneity tends to result in potentially significant differences in the marginal willingness to pay for travel time savings. If this evidence accumulates, and is shown to be applicable to a wider set of marginal willingness to pay attributes and contexts, then we should be concerned about the standard evidence, especially in an economic appraisal and demand forecasting context.

[37] A way of accounting for the rank order requires ongoing research.

The paper also draws on a literature not connected to discrete choice analysis that recognizes the errors of response in qualitative questions typically used to establish the presence of a specific processing rule. Although we have no new empirical evidence to quantify the notion of believability and plausibility associated with the Dempster-Shafer belief function, the approach to discounting self-stated explication on how attributes are processed is appealing, and worthy of investigation.

In ongoing research, we are investigating additional heuristics, including alternative functional forms for the heuristics herein, and ways of combining more than two heuristics into a single choice model that enable each heuristic to evolve up to a probability as continuum across a sampled population and /or across observations obtained from the same individuals (as is a panel such as a set of choice scenarios in a stated choice experiment). Another topic of especial interest is the relationship between WTP findings from self-stated intentions and specific functional forms of utility expressions. The evidence to date, from both sources, is not supportive empirically of the other in terms of the magnitude of marginal WTP relative to the simple fully compensatory linear model. The roles of both approaches are yet to be clarified.

Tangential to the current study is the growing literature on hypothetical bias in stated choice studies, which suggests that the marginal WTP (MWTP) is under-estimated for VTTS in stated choice studies, compared to actual market-based evidence, possibly by as much as 50 percent (see Brownstone and Small 2005, and Hensher 2008a). Isacsson (2007), in the context of trading time with money, found that the MWTP based on a hypothetical experiment was almost 50 percent lower at the mean than the real experiment MWTP, supporting the conclusions by Brownstone and Small (2005), in a transport context, that "…the value of time saved on the morning commute is quite high (between $20 and $40 per hour) when based on revealed behavior, and less than half that amount when based on hypothetical behavior" (page 279). It may be that the failure to accommodate process heterogeneity is a significant contributing influence.

References

Arana, J.E., Leon, C.J. and Hanemann, M.W. (200j): 'Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly', *Journal of Health Economics*, 27, 753-769.

Beharry, N. and Scarpa, R. (2008): 'Who should select the attributes in choice-experiments for non-market valuation? An application to coastal water quality in Tobago', Sustainability Research Institute, The University of Leeds, Leeds, UK.

Bertrand, M. and Mullainathan, S. (2001): 'Do people mean what they say? Implications for subjective survey data,' *American Economic Review Papers and Proceedings*, 91(2), May 2001: 67-72

Brownstone, D. and Small, K. (2005): 'Valuing time and reliability: assessing the evidence from road pricing demonstrations', *Transportation Research* Part A, 39, 279-293.

Cameron, T. (2008): ‘Differential attention to attributes in utility-theoretic choice models’, Department of Economics, University of Oregon.

Cantillo, V. and Ortúzar, J. de D. (2005): ‘A semi-compensatory discrete choice model with explicit attribute thresholds of perception’, *Transportation Research* 39B, 641–657.

Cantillo, V., Heydecker, B. and Ortuzar, J. de Dios (2006): ‘A discrete choice model incorporating thresholds for perception in attribute values’, *Transportation Research* B, 40 (9), 807-825.

Carlsson, F., Kataria, M. and Lampi, E. (2008): ‘Ignoring attributes in choice experiments’, *Proceedings EAERE Conference,* 25-28 June 2008, Gothenburg, Sweden.

Caussade, S., Ortúzar, Juan de Dios Ortúzar, J., Rizzi L. and Hensher, D.A. (2005): ‘Assessing the influence of design dimensions on stated choice experiment estimates’, *Transportation Research B*, 39 (7), 621-640.

Dempster, A.P. (1967): ‘Upper and lower probabilities induced by a multiple-valued mapping’, *Ann.Math. Stat.* 38, 325-339.

Dubois, D. and Prade, H. (1988): ‘Modelling uncertainty and inductive inference: A survey of recent non-additive probability systems’, *Acta Psychologica*, 68, 53-78.

Dubois, D. and Prade, H. (1987): ‘Representation and combination of uncertainty with belief functions and possibility measures’, *Computational Intelligence*, 170 (11), 909-924.

Fox, C. and Tversky. A. (1998): ‘A belief-based account of decision under uncertainty’, *Management Science*, 44 (7), July 870-895.

Gabaix, X. and Laibon, D. (2000): ‘A boundedly rational decision algorithm’, *American Economic Review Papers and Proceeding*s, 90 (2), May, 433-438.

Gilbride, T. and Allenby, G. (2004): ‘A choice model with conjunctive, disjunctive, and compensatory screening rules’, *Marketing Science*, 23 93), 391-406.

Gilovich, T., Griffin, D. and Kahneman, D. (Eds.) (2002): *Heuristics and Biases – The Psychology of Intuitive Judgment*, Cambridge University Press, Cambridge.

Greene, W.H. and Hensher, D.A. (2008 in press): ‘Ordered choice, heterogeneity, and attribute processing’, *Journal of Transport Economics and Policy*.

Greene, W.H. and Hensher, D.A. (2009, in press): *Modeling Ordered Choices*, Cambridge University Press, Cambridge.

Hensher, D.A. (2006): ‘How do respondents handle stated choice experiments? - Attribute processing strategies under varying information load’, *Journal of Applied Econometrics,* 21, 861-878

Hensher, D.A. (2008): ‘Joint estimation of process and outcome in choice experiments and implications for willingness to pay’, *Journal of Transport Economics and Policy*, 42 (2), May, 297-322.

Hensher, D.A. (2008a): ‘Hypothetical bias, stated choice studies and willingness to pay’, forthcoming in *Transportation Research* B.

Hensher, D.A. (2009): ‘Event descripton and support theory as a framework for representing process in choice analysis’, Institute of Transport and Logistics Studies, University of Sydney, February.

Hensher, D.A. and Greene, W.H. (2008): 'Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification', revised for *Empirical Economics*, December.

Hensher, D.A. and Layton, D. (2008): 'common-metric attribute parameter transfer and cognitive rationalisation: implications for willingness to pay', Institute of Transport and Logistics Studies, University of Sydney, July.

Hensher, D.A. and Rose, J. (in press): 'Simplifying choice through attribute preservation or non-attendance: implications for willingness to pay', *Transportation Research E*

Hensher, D.A., Rose, J. and Greene, W. (2005): 'The implications on willingness to pay of respondents ignoring specific attributes', *Transportation*, 32 (3), 203-222.

Hensher, D., Scarpa, R. and Campbell, D. (2009): Non-attendance to attributes in environmental choice analysis: a latent class specification, submitted to *Australian Journal of Agriculture and Resource Economics*, revised January 2009.

Hess, S. and Rose, J.M. (2007): 'A latent class approach to modelling heterogeneous information processing strategies in SP studies', paper presented at the *Oslo Workshop on Valuation Methods in Transport Planning*, Oslo.

Hess, S. and Hensher, D. (2008): 'Using conditioning on observed choices to retrieve individual-specific attribute processing strategies', submitted to a Special Issue of *Transportation Research* B, revised December.

Idson, L.C., Krantz, D.H., Osherson, D. and Bonini, N. (2001): 'The relation between probability and evidence judgment: an extension of support theory', *The Journal of Risk and Uncertainty*, 22 (3), 227-249.

Isacsson, G. (2007): 'The trade off between time and money: Is there a difference between real and hypothetical choices?', Swedish National Road and Transport Research Institute, Borlange, Sweden.

Johannesson, M., Blomquist, G., Blumenschien, K., Johansson, P., Liljas, B. and O'Connor, R. (1999): 'Calibrating hypothetical willingness to pay responses', *Journal of Risk and Uncertainty*, 8, 21-32.

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

King, G. and Wand, J. (2007): 'Comparing incomparable survey responses: new tools for anchoring vignettes,' *Political Analysis*, 15, 46–66.

Krantz, D.H. (1991): 'From indices to mappings: the representational approach to measurement', in *Frontiers of Mathematical Psychology: Essays in Honour of Clyde Coombs*, Brown, D, and Smith, E. (eds.), Springer Verlag, New York, 1-52.

Krosnick, J.A. and Schuman, H. (1988): 'Attitude intensity importance, and certainty and susceptibility to response effects', *Journal of Personality and Social Psychology*, 54 (6), 940-52.

Layton, D. and Hensher, D.A. (2008): 'Aggregation of common-metric attributes in preference revelation in choice experiments and implications for willingness to pay', revised for *Transportation Research D* Special Issue, November.

Lundhede, T.H., Olsen, S.B., Jacobsen, J.B. and B.J. Thorsen (2009): 'Handling respondent uncertainty in choice experiments: evaluating recoding approaches against explicit modelling of uncertainty', Faculty of Life Sciences, University of Copenhagen.

Martinez, F., Aguila, F. and Hurtubia, R. (2009): ‘The constrained multinomial logit: A semi-compensatory choice model’, *Transportation Research B*, 43, 365-377.
McFadden, D. (1998): ‘Measuring willingness-to-pay for transportation improvements’ in Garling, T., Laitila, T. and Westin, K. (eds.) *Theoretical Foundations of Travel Choice Modeling*, Elsevier, Oxford, 339-264.
Park, Y-H., Ding, M and Rao, V. (2008): ‘Eliciting preference for complex products: a web-based upgrading method’, *Journal of Marketing Research*, XLV (October), 562-574.
Puckett, S.M. and Hensher, D.A. (2008): ‘The role of attribute processing strategies in estimating the preferences of road freight stakeholders under variable road user charges’, *Transportation Research* E, 44, 379-395.
Puckett, S. and Hensher, D.A. (2009): ‘Revealing the extent of process heterogeneity in choice analysis: An empirical assessment’, *Transportation Research A*, 43 (1), 117-126.
Rose, J.M. and M.C.J. Bliemer (2007): ‘Stated preference experimental design strategies’ in Hensher, D.A. and K. Button, (eds.) *Transport Modelling*, Second Edition, Handbooks in Transport, Vol. 1, Elsevier Science, Oxford, Chapter 8.
Rose, J., Hensher, D.A., and Greene, W. (2005): ‘Recovering costs through price and service differentiation: accounting for exogenous information on attribute processing strategies in airline choice’ *Journal of Air Transport Management*, 11, 400-407.
Rose, J.M., Bliemer, M.C., Hensher and Collins, A. T. (2008): ‘Designing efficient stated choice experiments in the presence of reference alternatives’, *Transportation Research B* 42 (4), 395-406
Rose, J.M., Hensher, D.A., Greene, W.H., Washington, S.P. and Black, I. (2008a): ‘Accounting for exogenous information on decision maker processing strategies in models of discrete choice: attribute exclusion strategies in airline choice’, Institute of Transport and Logistics Studies, University of Sydney, July.
Sandor, Z. and Wedel, M. (2002): ‘Profile construction in experimental choice designs for mixed logit models’, *Marketing Science*, 21(4), 455–475.
Scarpa, R., Gilbride, T., Campbell, D. and Hensher, D.A. (2009): ‘Modeling attribute non-attendence in choice experiments: does it matter?’ revised for *European Journal of Agricultural Economics*, February.
Scarpa, R., Thiene, M. and Hensher, D.A. (2009a): ‘Monitoring choice task attribute attendance in non-market valuation of multiple park management services: does it matter?’ revised for *Land Economics*, February.
Shafer, G. (1976): *A Mathematical Theory of Evidence,* Princeton, N.J.: Princeton University Press.
Shafer, G. and Srivastava, R. P. (1990): ‘The Bayesian And Belief-Function Formalisms: A General Perspective for Auditing’, *Auditing: A Journal of Practice and Theory*, (Supplement), 110-148.
Simon, H. (1978): ‘Rational decision making in organisations’, *The American Economic Review*, 69 (4), 493-513.
Svenson, O. (1998): ‘The perspective from behavioral decision theory on modeling travel choice’, in Garling, T., Laitila, T. and Westin, K. (eds.) *Theoretical Foundations of Travel Choice Modeling*, Elsevier, Oxford, 141-172.

Swait, J. (2001): ‘A non-compensatory choice model incorporating attribute cut-offs’, *Transportation Research* B, 35(10), 903-928.

Swait, J.D. and Ben-Akiva, M. (1987): ‘Empirical test of a constrained discrete choice model: mode choice in Sao Paula, Brazil’ *Transportation Research* B, 21 (2), 103-115.

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

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

Train, K. and Revelt, D. (2000): ‘Customer-specific taste parameters and mixed logit, working paper’, Department of Economics, University of California, Berkeley, http://elsa.berkeley.edu/wp/train0999.pdf

Tversky, A. and Koehler, D. (1994): ‘Support theory: a nonextensional representation of subjective probability’, *Psychological Review*, 1010, 547-567.

Wong, S. K. M., and Wang, Z. W. (1993): “Qualitative measures of ambiguity”, in Hackerman D. and Mamdani A., eds., *Proceedings of The Ninth Conference on Uncertainty in Artificial Intelligence*, San Mateo, California: Morgan Kaufmann, 443-450.

Yoon, S-O. and Simonson, I. (2008): ‘Choice set configuration as a determinant of preference attribution and strength’ *Journal of Consumer Research*, 35 (August), 324-336.