Attribute Processing as a Behavioural Strategy in Choice Making

Version: 22 November 2012

Forthcoming as: Hensher, D.A. Attribute processing as a behavioural strategy in choice making, in Hess, S. and Daly, A.J. (eds.), Handbook of Discrete Choice Modelling, Edward Elgar, UK, Chapter 4.1.

1. Introduction

Choosing is a complex process that is typically simplified by human beings in many ways in order to ensure that the expected benefits outweigh the assumed costs of an outcome. Regardless of whether the context entails habitual or variety seeking behaviour, individuals draw on decision rules, often referred to as heuristics, to provide guidance on making choices. Such rules might be associated with an accumulation of overt experience; but whatever the basis of rule selection, there are many forces at play, often called cognitive processes, conscious or unconscious, that dictate responses in settings that researchers use to study choice making.

Despite the recognition in behavioural research, as long ago as the 1950s (see Simon 1955, Svenson 1992, also **Busemeyer and Rieskamp Ch 2.2**), that cognitive processes have a key role in preference revelation, and the reminders throughout the choice literature (see McFadden 2001) about rule-driven behaviour, we still see relatively little of the decision processing literature incorporated into mainstream discrete choice modelling which is, increasingly, becoming the preferred empirical context for individual preference measurement and willingness to pay derivatives.
There is an extensive literature outside of discrete choice modelling focussing on these matters, broadly 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 choice studies. The failure to recognise process, and the maintenance of a linear in parameters and additive in attributes (including allowance for attribute interactions) utility expression under full attribute and parameter preservation, is an admission, by default, that individuals when faced with a choice situation deem all attributes (and alternatives) relevant, and that a fully compensatory decision rule is used by all agents to arrive at a choice. In recent years we have started to see a growing interest in alternative processing strategies at the attribute, alternative and choice set levels, with empirical evidence suggesting that inclusion of process matters in a non-marginal way, in the determination of important behavioural outputs such as estimates of willingness to pay, elasticities, and predicted choice outcomes.

Research contributions such as Hensher (2006, 2008), Layton and Hensher (2010), Hensher and Rose (2009), Hensher and Layton (2010), Hess and Hensher (2010), Puckett and Hensher (2008), Swait (2001), Cantillo et al. (2006), Cameron (2008), Scarpa *et al.* (2008), Beharry and Scarpa (2008), Cantillo *et al.*(2006) , Scarpa *et al.* (2012) and Hensher et al. (2009), amongst others, are examples of a growing interest in the way that individuals evaluate a package of attributes associated mutually exclusive alternatives in real or hypothetical markets, and make choices[1]. The accumulating empirical evidence, in part represented in the references above, 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 under assessment 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, imposing thresholds on attribute levels to represent acceptable levels (e.g., Swait 2001, Hensher and Rose 2012), and attribute aggregation where they are in common units (e.g., Layton and Hensher 2010). Gilovich *et al.* (2002) synthesise the evidence under the theme of heuristics and biases. Importantly, the heuristics are likely to be context specific, such that the design and hence the nature of the information shown in stated choice experiments, for example, conditions in part the choice of rules adopted.

The broad multidisciplinary literature on behavioural decision making (see Gilovich *et al.* 2002) 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 2006, 2006a, 2008), 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, which may or may not be conditioned by the instrument being used to capture evidence.

There are at least two ways in which information used in processing might be empirically identified. One involves direct questioning of respondents after each choice scenario (what is increasingly referred 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 as recently investigated by Scarpa et al. (2012).

The purpose of this chapter is to review some of the findings and models that have emerged from the literature that might be used to gain an understanding of choice making and hence improve the choice modelling process. This chapter focuses on the role of attribute processing in stated choice experiments, the dominant discrete choice setting within which attribute processing has been studied, but we note that the heuristics also apply in the context of revealed preference data[2]. The chapter draws on both direct questioning and inferential methods to synthesise what is known about the role of mixtures of processing rules in order to establish the behavioural implications on key outputs such as marginal willingness to pay. 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 some very specific attribute

[1] This chapter does not consider other aspects of process in choice experiments such as uncertainty in the choice response. See Lundhede et al. (2010).

[2] This chapter is focused on stated choice surveys. We recognise that at some level, one might expect these attribute processing effects to be more prominent in revealed preference data given that, for example, advertising/branding is designed to encourage not paying attention to attributes, while in other instances such as putting high sugar cereals on low shelves in grocery stores or putting important detail in fine print there are intentional efforts to obscure details. But it is also possible to make the case that survey respondents may pay less attention to details.

processing strategies such as attribute non-attendance in the presence or otherwise of attribute thresholds and referencing.

We restrict the scope of this chapter, given the extensive literature on heuristics and biases (covered in part by **Chorus *et al.* in Ch 4.2**), to attribute processing strategies that researchers have found to be behaviourally appealing, to date, in the context of discrete choice analysis studies. We focus on the following five themes: attribute non-attendance, attribute thresholds, the majority of confirming dimensions, reference point revision, and value learning. We suggest however that it is too early to offer a view as to whether one decision making process is gaining greater empirical support to 'explain' real choice making. What we can say is that there is a need for much more research on *process* to complement *outcome* in choice modelling. The current disproportionate interest in attribute non-attendance may in large measure reflect the relative ease of studying this phenomenon in contrast to the much more complex non-linear propositions associated with the other attribute processing rules presented in this review.

2. Attribute non-attendance

A behavioural rule which is attracting particular attention in stated choice studies is the extent to which respondents attend to, or ignore, one or more attributes in processing the information on offer, resulting in a (stated) choice outcome. Some agents do not appear to put any weight on some attributes. The question then is whether the heterogeneity with respect to placing a zero weight on some attributes is effectively exogenous, that is simply preference heterogeneity, whether it is a function of the characteristics of the choice sets agents faced, or more likely both factors play a role. One can probably never rule out that there is exogenous preference heterogeneity among agents with respect to placing a zero weight on one or more attributes, but it is here that by running choice experiments that one can show that the nature of the choice sets that agents see influences the pattern and extent of particular attributes given no weight. Given a continuum of relevance, distinguishing a zero weight from a very low level of relevance (approximating but not equal to zero) creates a research challenge (see below). We present below evidence on the contextual influence of the design of the choice experiment.

In the popular stated choice approach it is assumed, in the main, that all attributes are processed in what DeShazo and Fermo (2004) describe as the *passive bounded rationality* model. This model assumes that individuals attend to all information in the choice set, but increasingly make mistakes in processing that information, as the volume of information increases. Contrasting this is the *rationally-adaptive* model which assumes that individuals recognise that their limited cognition has positive opportunity costs. Whether *rationally-adaptive* behaviour is a product of the survey instrument and/or the nature of an individual's processing of any information, is an empirical matter.

In stated choice (SC) studies, respondents are typically asked to choose their preferred alternative among several hypothetical alternatives in a sequence of experimentally designed choice tasks (see **Rose and Bliemer Ch 3.4**). The standard behavioural assumption underlying most SC studies is that respondents make trade-offs between all attributes describing each of the alternatives, and are expected to choose their most preferred alternative in a choice set. This rules out the possibility that respondents focus solely on a subset of offered attributes, ignoring all other differences between the alternatives (see Hensher

2006a). Ignoring attributes in the choice task implies some form of non (or semi-)-compensatory behaviour, because no matter how much the level of a given attribute is improved, the improvement will fail to compensate for worsening in the levels of other attributes if the attribute itself is ignored by the respondent (Spash 2000, Rekola 2003, Sælensminde 2002, Lockwood 1996), or what Rigby and Burton (2006) describe as ‘disinterest’. There may be one exception where choosing to ignore an attribute may be influenced by the levels of the other attributes, and hence a switch between compensatory and non-compensatory behaviour may be legitimate as the attribute levels change within a choice experiment. This can be tested at a choice set level (see Puckett and Hensher 2008), but is problematic if the test relates to the entire set of choice sets.

There is potential for attribute non-attendance (henceforth AN-A) to have serious consequences on the derivation of prediction and welfare estimates, especially when the object of neglect is the monetary attribute, such as the cost of an alternative, although it applies equally to the numerator in any calculation of willingness to pay. The detection and statistical handling of AN-A raises technical issues for the practice of discrete choice modelling, especially where specific processing rules are observed or predicted for a sample, which then have to be applied to a population.

2.1 Two emerging approaches

Two choice methods are emerging to investigate the role of specific heuristics – one involving supplementary questions on whether specific attributes are ignored, referred to as self-stated intentions (see Hensher *et al.* 2005 for an initial contribution), and the other involving a specification of a model that can reveal the extent to which each attribute is preserved across a sample without the need for supplementary data (e.g., Hess and Hensher 2010). Although it is not possible to suggest which method is closer to the ‘truth’ in capturing process strategy, there is ongoing research designed to understand the behavioural implications of each method, and in time to establish a mapping between the two methods (see Hess and Hensher 2012, Scarpa *et al.* 2012).

2.2 Attribute non-attendance on supplementary questions

Hensher and co-authors have initiated several explorations of attribute non-attendance within a standard multivariate discrete choice setting. For example, the attribute ranges in Hensher (2006a) were varied simultaneously across all attributes but the self-stated ANA response is available only at the level of the individual, not the choice set. In contrast, Cameron and DeShazo (2011) consider differences in the ranges of attributes within a single choice set as additional potential determinants of attention, and therefore of apparent marginal utilities, and ultimately estimates of willingness to pay. Very relevant to Cameron and DeShazo (2011) is Hensher’s finding that individuals’ processing strategies depend on the *nature* of the attribute information in the choice set, not just the *quantity* of such information (i.e., the number of attributes).

Hensher *et al.* (2005) use a specific follow-up question about which attributes the respondent did not use in making their choices. Hensher *et al.* (2007) also uses the same follow-up question to identify nine distinct attribute processing rules. Respondent adherence to these rules is modelled as stochastic. The authors then use a modified mixed logit model which conditions each parameter on whether a respondent included or excluded an attribute in their attribute processing strategy. In their conclusions, the authors acknowledge that there may be differences "between what people say they think and what they really think" (p. 216), and they question whether the "simply conscious statements" made by survey respondents, no matter how much detail is obtained, represent an adequate measure of information processing. They emphasise that regardless of the source of information on attribute processing, individuals' information processing strategies "should be built into the estimation of choice data from stated choice studies" (p. 214).

A related study is Puckett and Hensher (2008) which builds on Hensher *et al.* (2006a) in that it considers the effects of APSs utilised by respondents for every alternative in every choice set, including across choice tasks faced by a given respondent. This approach can accommodate cases where attribute level mixes are outside of the acceptable choice bounds for the individual. The wording of their debriefing question for each choice was: "Is any of the information shown not relevant when you make your choice? If an attribute did not matter to your decision, please click on the label of the attribute below. If any particular attributes for a given alternative did not matter to your decision, please click on the specific attribute." Subjective all-or-nothing attention to different attributes is thus elicited directly from each respondent, rather than being inferred from choice behaviour.

2.3 The Role of the design of the choice experiment in ANA defined by self-stated intentions

The actual design of a choice experiment may itself be a source of heterogeneity in induced attribute processing, in recognition that the design of a choice experiment can itself induce particular processing strategies. Hensher (2006a) was a first study to investigate the dimensionality of a stated choice experiment with 16 different choice experiment designs, each varying according to the number of attributes (3,4,5,6), the number of alternatives (2,3,4), the number of choice sets (6,9,12,15), and the range of each attribute (narrower, base, wider) the role of (i) the dimensionality of a stated choice experiment. See Table 1 for a summary of the designs. Attribute non-attendance was defined by self-stated intentions.

Table 1: The sub-designs of the overall design (Hensher 2006a)

Choice set of size	Number of alternatives	Number of attributes	Number of levels of attributes	Range of attribute levels
15	3	4	3	Base
12	3	4	4	Wider than base
15	2	5	2	Wider than base
9	2	5	4	Base
6	2	3	3	Wider than base
15	2	3	4	Narrower than base
6	3	6	2	Narrower than base
9	4	3	4	Wider than base
15	4	6	4	Base
6	4	6	3	Wider than base
6	3	5	4	Narrower than base
9	4	4	2	Narrower than base
12	3	6	2	Base
12	2	3	3	Narrower than base

9	2	4	2	Base
12	4	5	3	Narrower than base

Note: Column 1 refers to the number of choice sets. The 16 rows represent the set of designs

The key findings from this study in terms of number of attributes attended to are:

1. The probability of considering more attributes increases *dramatically* as the number of levels per attribute decreases, *ceteris paribus*.
2. The probability of considering more attributes from the offered set decreases as an attribute's range narrows, *ceteris paribus*. That is, respondents ignore more attributes when the difference between attribute levels is small. This result is perhaps due to the fact that evaluation of small differences is more difficult than evaluation of large differences. An important implication is that if an analyst continues to include, in model estimation, an attribute across the entire sample that is *ignored by a respondent*, then there is a much greater likelihood of mis-specified parameter estimates in circumstances where the attribute range is narrower than wider[3].
3. These two results can be combined and stated in the converse as follows: respondents tend to consider more attributes (i.e., ignore fewer attributes) when the attributes have only a few levels that differ greatly, such that evaluation of each attribute is easier. Overall the respondent seems to trade-off effort spent on each attribute against the number of attributes considered.
4. As we increase the 'number of alternatives' to evaluate, *ceteris paribus*, the importance of considering more attributes increases, as a way of making it easier to differentiate between the alternatives. This is an important finding that runs counter to some views, for example, that individuals will tend to ignore increasing amounts of attribute information as the number of alternatives increases. Our evidence suggests that the processing strategy is dependent on the nature of the attribute information, and not strictly on the quantity.
5. Overall, we see a picture emerging that design dimensionality seems to have less of an influence on the attribute processing strategy when we have fewer items to process. This makes good sense but should not be taken to imply that designs with fewer items are preferred; but that preference heterogeneity in invoking an attribute processing strategy, appears to decline substantially as the information content declines, for real or spurious reasons. Contrariwise, individuals appear to increasingly invoke a relevancy strategy as the amount of information to process increases. The need to capture this growing heterogeneity in IP strategies is clear and should be accounted for in behavioural choice models.
6. The evidence on sources of influence on how many attributes are considered, relative to the full set offered, is important in revealing candidate influences on attribute processing, and the extent to which the empirical policy outputs, such as willingness to pay, vary as a consequence of the SC design and its context. Where we might find evidence of attribute reduction (through exclusion and/or aggregation), we might reasonably speculate that the selected attribute processing strategy has elements of coping *and* relevancy. This should not necessarily be interpreted as a response to complexity, but part of the natural process of decision making.

[3] This finding has interesting implications for the growing evidence that mean willingness to pay (WTP) for an attribute tends to be higher under a wider range for the numerator attribute (Louviere and Hensher 2001). Simply put, the greater relevance in preserving the attribute content under a wider range will mean that such an attribute is relatively more important to the outcome, than it is under a narrow range specification, and hence a higher mean WTP is inferred.

2.4 The role of attribute non-attendance through model inference

Attribute non-attendance on supplementary questions is designed to establish whether a respondent had ignored an attribute or not: they could be asked either after each choice set or after completing all choice scenario assessments. However, as argued in a number of papers, such as Hensher and Rose (2009), Hess and Hensher (2010), and Hensher (2010), there is concern about the reliability of responses to such supplementary questions. Although the jury is still out on this issue, there is growing interest in identifying the role of attribute non-attendance through model inference, rather than directly asking each respondent. The most recent examples are Scarpa *et al.* (2010), Hess and Hensher (2010), Hensher and Greene (2010), and Hole (2011).

A growing research theme is the issue of how to incorporate this phenomenon in statistical models when data on self-reported AN-A are not available or are deemed problematic. There are some intuitive ways of addressing this issue, building on basic models that are commonly employed by practitioners. In particular, panel mixed logit models are an appealing setting within which to account for repeated attribute exclusion in the evaluation of proposed alternatives by a given respondent. What is intended here is that the identification of AN-A behaviour is achieved by analysing the observed response pattern using a statistical model with degenerate distributions of taste intensities at zero, which implies non-attendance. This contrasts with the approaches that rely on the self-stated intentions (Hensher 2008; 2008; Carlsson *et al*., 2008) that ask respondents which attributes they paid attention to or were important. Methods are now available that do not require self-reported information on attendance (see Hess and Hensher 2010, 2012; Hoyos *et al.* 2010; Hole 2011).

Hess and Hensher (2010) infer AN-A through the analysis of respondent-specific parameter distributions, obtained through conditioning on stated choices. Their results suggest that some respondents do indeed ignore a subset of explanatory variables. There is also some evidence that these *inferred* attribute processing strategies are not necessarily consistent with the responses given to supplementary questions about attribute attendance, when mapping is available. This raises questions about how both types of data can be used to assist in improving behavioural relevance.

The results in Hess and Hensher (2010) for example, show that respondents who indicate that they ignored a given attribute often still show non-zero sensitivity to that attribute, albeit one that is (potentially substantially) lower than that for the remainder of the population. A possible interpretation of these results is that respondents who indicate that they did not attend to a given attribute simply assigned it a lower importance, and that the probability of indicating that they ignored a given attribute increases as the perceived importance of that attribute is reduced, an argument put forward recently by Hess (2011). In a similar manner, Scarpa *et al.* (2009) implement two ways of modelling AN-A; the first involves constraining coefficients to zero in a latent class framework, while the second is based on stochastic attribute selection, and grounded in Bayesian estimation. In all studies, the results indicate that accounting for non-attendance significantly improves model fit in comparison to models that assume full attribute attendance, and yields estimates of willingness to pay for specific attributes that are typically different.

2.4.1 The growing popularity of the latent class framework to accommodate

probabilistic decision processes

Much marketing research is designed to determine what people do not care about. What does cause a problem is heterogeneity in the sense that there is a substantive fraction of the population that places a zero weight on some attribute. The latent class models with zero for some attribute for some classes are an ideal statistical modelling solution for this phenomenon. There is a large literature in economics (e.g., Pudney, 1989) that looks at how to model corner solutions. Not allowing for corner solutions gets one trapped into absurd propositions as suggested above whereby for example, increasing the contrast in the colour of the paint under the carpets in an automobile (which the agent does not see and does not care about) is a way to compensate for lower fuel economy. It is common to see Tobit models, hurdle/spike models, and various types of count data models. Much of the research output reported here can be seen as the analogue of how to do this in choice models with respect to individual attributes.

Hess and Rose (2007), Hensher and Greene (2010) and Campbell *et al*. (2010) use a latent class framework (without drawing on evidence from stated non attendance questions) as a way of capturing a probabilistic decision rule process, in which specific restrictions are imposed on the utility expressions for each class, to represent hypotheses of pre-defined attribute processing strategies. A growing number of authors are showing that the constrained latent class model appears to outperform its competitors that are based on continuous mixing of taste such as mixed logit. However, while a number of the classes relate to attribute non-attendance (and other candidate attribute processing rules), these studies excluded the possibility of combinations of more than one attribute non-attendance rule *in a class*. Investigating all combinations, while appealing, becomes increasingly complex and infeasible as the number of attributes (K) increases, given a 2^K rule for the combination of attendance or non-attendance. With four attributes, for example, we have 16 possible combinations, and with eight attributes we have 256. Nevertheless the approach has appeal and is presented below as one way of recognising attribute processing heterogeneity.

Formally, assume respondent i = 1,2, …, I is asked to select from amongst J alternatives, j = 1,2 ,… ,J. Assuming that the basic analytical framework is a standard MNL choice model, the probability that respondent i chooses alternative j is given as

$$\text{Prob}(i,j) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{i,j})}{\Sigma_{j=1}^{J}\exp(\boldsymbol{\beta}'\mathbf{x}_{i,j})}. \qquad (1)$$

where $x_{i,j}$ represents the attributes associated with alternative j as observed by respondent i and $\boldsymbol{\beta}'$ is a vector of parameter weights related to the attributes.

Non-attendance is accommodated by supposing that individuals sort themselves into one of 2^K (or $q=1,\ldots,Q$) classes, distinguished by which of the attributes were considered in their choice process (see Hole 2011). If the configuration chosen by the individual is not directly observed (as, for example, in a supplementary question), then in the model, this sorting can only be done probabilistically. In the context of (1), we can model this by writing equation (2).

$$\text{Prob}(i,j|q) = \frac{\exp(\boldsymbol{\beta}_q'\mathbf{x}_{i,j})}{\Sigma_{j=1}^{J}\exp(\boldsymbol{\beta}_q'\mathbf{x}_{i,j})}. \qquad (2)$$

$\boldsymbol{\beta}_q$ is one of the 2^K possible vectors $\boldsymbol{\beta}$ in which m of the elements are zero and K-m are nonzero. Specifically, q can be thought of as a masking vector of the form (δ_1, δ_2, δ_3, $\delta_{4,\ldots}$), where each δ takes the possible values 0,1. β_q is then the "element for element product" of this masking vector, with the standard coefficient vector β, indicating that the masking vector interacts with the coefficient vector. For example, for two attributes (classes), the parameter vectors would appear $\boldsymbol{\beta}_1$=(0,0), $\boldsymbol{\beta}_2$=($\boldsymbol{\beta}_A$,0), $\boldsymbol{\beta}_3$=(0,$\boldsymbol{\beta}_B$), $\boldsymbol{\beta}_4$=($\boldsymbol{\beta}_A$,$\boldsymbol{\beta}_B$)[4]. An important part of the underlying theory is that the class q is not defined by the attribute taking value zero within the class, but by the corresponding coefficient taking the value zero. Thus the "random parameters" aspect of the model is a discrete distribution of preference structures across individuals who are distinguished by whether they pay attention to the particular attribute or not.

Since the sorting is not observable, we cannot directly construct the likelihood function for estimation of the parameters. In keeping with the latent class approach, we need to estimate a set of probabilities (π_q) that each individual i falls into class q. While this could be conditioned on individual characteristics, or indeed any exogenous information such as respondent stated reasons as to why they did not attend to an attribute (e.g., it is not important, it simplified choosing, or the attribute levels were out of an acceptable range), in this case we have assumed that the same set applies equally to all respondents, so that the probabilities reflect the class proportions.

The marginal probability that individual i will choose alternative j is found by averaging over the classes, as in (3).

$$\text{Prob}(i,j) = \sum_{q=1}^{2^K} \pi_q \frac{\exp(\boldsymbol{\beta}_q' \mathbf{x}_{i,j})}{\Sigma_{j=1}^{J} \exp(\boldsymbol{\beta}_q' \mathbf{x}_{i,j})} \text{ where } \sum_{q=1}^{2^K} \pi_q = 1. \tag{3}$$

As formulated, this is a type of finite mixture, or latent class model. It differs from more familiar formulations in that the nonzero elements in $\boldsymbol{\beta}_q$ are the same across the classes and the classes have specific behavioural meaning, as opposed to merely being groupings defined on the basis of responses as in the strict latent class formulation, hence the reference to a probabilistic decision process model. Estimation of the probabilistic decision process model is straightforward as a latent class MNL model with linear constraints on the coefficients, as suggested above and can allow for random as well as fixed parameters within each class (as in Hensher et al. 2012, Hess et al. 2012 and Collins et al. 2012).

In the presence of AN-A, the model differs from more familiar formulations of latent class models in that the nonzero elements in $\boldsymbol{\beta}_q$ can be allowed to be the same or free across the classes, and the classes have specific behavioural meaning, as opposed to merely being groupings defined on the basis of responses, as in the strict latent class formulation, hence the reference to a probabilistic decision process model.

As an example of this approach, Hensher *et al.* (2012), using data collected in Australia in the context of car commuters choosing between tolled and untolled roads, estimated a

[4] In this example, there is one unrestricted parameter vector in the model, shown as β_4 = (β_A, β_B). The other parameter vectors are constructed from the same two parameters either by setting one or both elements to zero or by equating elements to those in β_4. Thus, β_3 = (0,β_B) is obtained as a linear restriction on β_4, namely that one element equal zero and a second element equal the corresponding element in β_4.

multinomial logit (MNL) model in which all attributes are assumed to be attended to, and then a probabilistic decision process model with 2^K possible attribute attendance "rules". The model that accounts for attribute non-attendance was a significant improvement on the model that assumes all attributes are attended to, in terms of log-likelihood and Bayes information criterion (BIC). Although the probabilistic decision process model has additional parameters, namely the class probabilities π_q, the choice probability part of the model has the same number of parameters as MNL. For this application, mean values of travel time savings (VTTS) are obtained for the attribute non-attendance model.

To calculate the overall VTTS across the attribute attendance rule classes, we have to weight each class by the membership probability. Under allowance for AN-A, a mean VTTS of $12.77 per person hour is obtained, which is lower that the MNL model mean VTTS of $12.81 per person hour; however if we were to exclude the two classes where there is no time-cost trade off, we would obtain $17.96 per person hour. This suggests an under-estimate from the MNL model of the mean VTTS by 36 percent. However, this implies that for some respondents (28.9% in the sample) a VTTS does not exist, which is doubtful. This is a major concern for applications of VTTS, and indeed any WTP study (see Scarpa *et al.* 2009), since we can reasonably assume that everyone does in reality value travel time savings, despite the inability to measure this under certain AN-A rules.

We believe that this situation has arisen as a result of the design of the stated choice experiment. In particular, the range and levels of specific attributes might be such that some respondents do not see merit in some of the levels of times and costs being traded, with one or both attributes having levels that do not matter[5]. In real markets, it is not unreasonable to suggest that there exist levels of time and cost that do matter, implying that the empirical instrument might not be adequate to pick up the real behavioural response at work. However, there might be some individuals, who would deem a specific attribute not relevant, no matter what a sufficiently wide attribute range was considered (e.g., a very wealthy person who does not care about the running cost), and hence never trade-off time with running cost. Furthermore, the situation of a very low level of an attribute might be processed in such a way that relevance only applies when a specific threshold level is reached (see the following section). This suggests that a more careful assessment of *respondent-specific attribute ranges* is called for in future choice experiment designs.

We suspect this finding is not uncommon in choice experiments, but is never known until an analyst undertakes the modelling exercise. Scarpa *et al.* 2009 for example, find that over 90 percent of the sample ignore the cost attribute in the context of a stated preference survey designed to value landscapes in Ireland, where the cost attribute was specified as the value in Euros that the respondent would personally have to pay per year through their income tax and value added tax contributions.

Greene and Hensher (2012), Bujosa et al. (2010) and Hess et al. (2011) introduce a natural extension of the fixed parameter latent class model as a random parameter latent class model which allows for another layer of preference heterogeneity within each class; however to date only Hess et al. (2012), Hensher et al. (2012a) and Collins et al. (2012) have developed this model form in the context of AN-A. What we then have is a latent class model that allows for heterogeneity both within and across groups. To accommodate the two layers of

[5] Puckett and Hensher (2008) suggest that the range and relative equivalence of the price attribute levels among alternatives in a particular choice task may lead respondents to ignore the price attribute in some choice tasks and not in others.

heterogeneity, we allow for continuous variation of the parameters within classes. The latent class aspect of the model is given as (4) and (5).

$$f(y_i|\mathbf{x}_i,\text{class} = q) = g(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}_{i|q}) \tag{4}$$

$$\text{Prob}(\text{class} = q) = \pi_q(\boldsymbol{\theta}),\ q = 1,...,Q. \tag{5}$$

The within-class heterogeneity is structured as

$$\boldsymbol{\beta}_{i|q} = \boldsymbol{\beta}_q + \mathbf{w}_{i|q} \tag{6}$$

$$\mathbf{w}_{i|q} \sim E[\mathbf{w}_{i|q}|\mathbf{X}] = \mathbf{0},\ \text{Var}[\mathbf{w}_{i|q} \mid \mathbf{X}] = \boldsymbol{\Sigma}_\text{q} \tag{7}$$

where the $\mathbf{X}$ indicates that $\text{w}_{i|q}$ is uncorrelated with all exogenous data in the sample. We typically assume that the underlying distribution for the within-class heterogeneity is normal, with mean **0** and covariance matrix $\boldsymbol{\Sigma}$. In a given application, it may be appropriate to further assume that certain rows and corresponding columns of $\boldsymbol{\Sigma}_q$ equal zero, indicating that the variation of the corresponding parameter is entirely across classes.

The contribution of individual i to the log likelihood for the model is obtained for each individual in the sample by integrating out the within-class heterogeneity and then the class heterogeneity. We can allow for a panel data setting, hence the observed vector of outcomes is denoted $\mathbf{y}_i$ and the observed data on exogenous variables are collected in $\mathbf{X}_i = [\mathbf{X}_{i1},..,\mathbf{X}_{iT_i}]$. An individual is assumed to engage in T_i choice situations, where $\text{T}_\text{i} \geq 1$. The generic model is given in (8).

$$f(\mathbf{y}_i|\mathbf{X}_i,\boldsymbol{\beta}_1,...,\boldsymbol{\beta}_Q,\boldsymbol{\theta},\boldsymbol{\Sigma}_1,...,\boldsymbol{\Sigma}_Q) = \sum\nolimits_{q=1}^{Q} \pi_q(\boldsymbol{\theta})\int_{w_i} \prod\nolimits_{t=1}^{T_i} f[\mathbf{y}_{it} \mid (\boldsymbol{\beta}_q + \mathbf{w}_i), \mathbf{X}_{it}]h(\mathbf{w}_i \mid \boldsymbol{\Sigma}_q)d\mathbf{w}_i \tag{8}$$

The model is called a latent class, mixed multinomial logit (LC_MMNL) model. Individual i chooses among J alternatives with conditional probabilities given as (11).

$$f[\mathbf{y}_{it} \mid (\boldsymbol{\beta}_q + \mathbf{w}_i), \mathbf{X}_{it}] = \frac{\exp[\Sigma_{j=1}^{J} y_{it,j}(\boldsymbol{\beta}_q + \mathbf{w}_i)'\mathbf{x}_{it,j}]}{\sum_{j=1}^{J} \exp[\Sigma_{j=1}^{J} y_{it,j}(\boldsymbol{\beta}_q + \mathbf{w}_i)'\mathbf{x}_{it,j}]},\ j = 1,...,J, \tag{9}$$

$y_{it,j} = 1$ for the j corresponding to the alternative chosen and 0 for all others, and $\mathbf{x}_{it,j}$ is the vector of attributes of alternative j for individual i in choice situation t. Applications are given in Greene and Hensher (2012), Bujosa et al. (2010),Hess et al. (2011) and Hensher et al. (2012a).

Hess *et al.* (2011) show forcefully that particular modelling assumptions influence the appearance of how much attribute non-attendance there is versus simply a substantial fraction of the sample only placing weak weight (in a mixed logit sense) on the attribute. The more general point is that any estimate of the fraction of the sample found to be doing something like placing a zero weight on an attribute is conditional on other maintained assumptions. One could, for instance, introduce different types of non-linearities and also shift this fraction.

Hensher *et al.* (2012a) conclude that, despite the marginal influence of preference heterogeneity in the overall fit of the models, they find potentially important behavioural evidence to suggest that inclusion of random parameters may be a way of accommodating small marginal disutilities (in contrast to AN-A set equal to zero marginal disutility), and small differences in marginal disutilities (in contrast to equal marginal disutilities under aggregated common metric attributes), as observed by a 'move back' to full attribute attendance when fixed parameter become random parameters under attribute processing. If this argument has merit and can be confirmed using other data sets, they suggest that they may have identified one way of recognising what the broader literature (e.g., Hess *et al.* 2011) refers to as low sensitivity in contrast to zero sensitivity.

An important output is willingness to pay (WTP) estimates, computed using the familiar result, WTP = $-\beta_x/\beta_{cost}$. For the most general model with random parameters within class, since there is heterogeneity of the parameters within the classes as well as across classes, the result is best averaged to produce an overall estimate. The averaging can be undertaken for the random parameters within each class and then again across classes using the posterior probabilities as weights. Collecting the results, the procedure is given as (10) (from Hensher et al. 2012a).

$$WTP = \frac{1}{N}\sum_{i=1}^{N} \sum_{qAPR=1}^{QAPR} \left\{\pi_{qAPR}\left(\hat{\theta}\right)|\,i\right\} \left[\frac{\frac{1}{R}\sum_{r=1}^{R} L_{ir|qAPR} \frac{-\hat{\beta}_{x,ir|qAPR}}{\hat{\beta}_{\cos t,ir|qAPR}}}{\frac{1}{R}\sum_{r=1}^{R} L_{ir|q}}\right] \tag{10}$$

$$= \frac{1}{N}\sum_{i=1}^{N} \sum_{qAPR=1}^{QAPR} \left\{\pi_{qAPR}\left(\hat{\theta}\right)|\,i\right\} \frac{1}{R}\sum_{r=1}^{R} W_{ir|qAPR} WTP_{x,ir|qAPR},$$

where R is the number of draws in the simulation and r indexes the draws,

$\hat{\beta}_{x,ir|qAPR} = \hat{\beta}_{x|qAPR} + \hat{\sigma}_{x|qAPR} w_{x,ir|qAPR}$

and likewise for $\hat{\beta}_{\cos t,ir|qAPR}$, $L_{ir|qAPR}$ is the contribution of individual i to the class specific likelihood, and $\pi_{qAPR}\left(\hat{\theta}\right)|\,i$ is the estimated posterior class probability for individual i;

$$\pi_{qAPR}\left(\hat{\theta}\right)|\,i = \frac{\pi_{qAPR}(\hat{\boldsymbol{\theta}})\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}\,|\,(\hat{\boldsymbol{\beta}}_{q_{APR}} + \mathbf{w}_{i,r}), \mathbf{X}_{it}]}{\sum_{q_{APR}=1}^{Q_{APR}} \pi_{qAPR}(\hat{\boldsymbol{\theta}})\frac{1}{R}\sum_{r=1}^{R} \prod_{t=1}^{T_i} f[\mathbf{y}_{it}\,|\,(\hat{\boldsymbol{\beta}}_{q_{APR}} + \mathbf{w}_{i,r}), \mathbf{X}_{it}]}. \tag{11}$$

$\boldsymbol{\theta}$ is the vector of latent class parameters attached to candidate sources of systematic influence on class membership; and within-class heterogeneity, $\mathbf{w}_i$, is structured as

$$\boldsymbol{\beta}_{i|q_{APR}} = \boldsymbol{\beta}_{q_{APR}} + \mathbf{w}_{i|q_{APR}} \tag{12}$$

$$\mathbf{w}_{i|q_{APR}} \sim E[\mathbf{w}_{i|q_{APR}}|\mathbf{X}] = \mathbf{0},\ \mathrm{Var}[\mathbf{w}_{i|q_{APR}}\,|\,\mathbf{X}] = \boldsymbol{\Sigma}_{q_{APR}} \tag{13}$$

where the **X** indicates that $w_{i|q_{APR}}$ is uncorrelated with all exogenous data in the sample. See Hensher *et al.* (2012a) for more details.

2.5 Recent new approaches to accommodating attribute non-attendance

Hess and Hensher (2012) building on the contribution of Hensher (2008) note that AN-A is treated as an exogenous rule in the majority of studies, when in fact it may be endogenous[6], just like the choice outcome. They explicitly recognise the endogeneity induced by attribute non-attendance, and condition attribute parameters on underlying unobserved attribute importance ratings (or indeed any appropriate supplementary information that may be available).

They develop a hybrid model system involving attribute processing and outcome choice models in which latent variables are introduced as explanatory variables in both parts of the model, explaining the answers to supplementary attribute processing questions and explaining heterogeneity in marginal sensitivities in the choice model. The resulting empirical model explains how lower latent attribute importance leads to a higher probability of indicating that an attribute was ignored or that it was ranked as less important, as well as increasing the probability of a reduced value for the associated marginal utility coefficient in the choice model. The model does so by treating the answers to supplementary information processing questions as dependent rather than explanatory variables (which may themselves be conditioned on a series of supplementary questions focussed on knowing why an attribute was not attended to), hence avoiding potential risk of endogeneity bias and measurement error. We refer the reader to this paper for further details.

2.6 Illustrative empirical implications of AN-A

To complete the discussion of AN-A, we provide some examples of mean estimates of willingness to pay (WTP) under full attribute attendance and AN-A. The indicative examples are drawn from transport, agricultural and environmental case studies. The directional change in mean WTP estimates varies across the studies as well as within a discipline application. The important point to make is that there are noticeable differences in mean estimates, which is enough evidence to raise questions about the implications of not considering AN-A. When translated into an aggregate measure of welfare benefits, these differences are sufficiently large to impact on the net benefits of specific projects or policy initiatives.

Table 2 Illustrative Impacts on Willingness to Pay of AN-A

Context	Full attribute attendance	Attribute non-attendance	Reference
Value of travel time savings (VTTS) car ($AUD)	10.17 (2.77)	12.42 (1.63)	Hensher et al. (2012a)
Colour of beef ($US)	4.93	3.86	Scarpa et al. (2012)
Animal welfare ($US)	15.04	11.69	
Breed origin	5.98	4.46	
VTTS air travel ($AUD)	25.14 (7.1)	38.4 (11.8)	Rose et al. (in press)
VTTS free flow (car) ($AUD)	25.87 (60.29)	22.45 (41.43)	Hess and Hensher (2010)
Stonewalls (Euros per annum)	53.22	101.27	Campbell et al. (2011)
Farmyard tidiness (Euros per annum)	198.88	60.59	

Note: the numbers in brackets are standard deviation if reported.

[6] Assuming endogeneity bias as present requires a definition and a test. Endogeneity bias can arise from a number of sources such as measurement error, missing attributes and simultaneity and is observed when a specific variable included in the observed effects is correlated with the error term associated with the utility expression containing the explanatory variable of interest. To test for endogeneity bias (that is, the part that is correlated with the random error component), analysts should undertook two tasks: first testing for the extent to which the variable has systematic influence on the standard deviation of the error component, and secondly identifying other exogenous variables that are correlated with the variable under consideration, but not with the error component that could be used as instrumental variables, or simply as evidence of no endogeneity bias.

3. Attribute thresholds

Unlike attribute non-attendance, which assumes an attributes is ignored or not in a strict binary format, given the attribute levels, research suggests that people consider the particular level of an attribute and make judgments and choices based on specific thresholds, possibly only attending to an attribute if it satisfies some threshold condition. In this sense, attribute thresholds are inherently linked to a possible explanation for attribute non-attendance.

This section looks at the way that perceived attribute thresholds (or lower and upper cutoffs), are used by respondents to condition the role of an attribute in its contribution to the acceptability and hence choice of an alternative. There is a growing literature on attribute thresholds, with some studies imposing analytical distributions on cutoffs (including just noticeable differences such as Cantillo *et al.* 2006), and other studies asking supplementary questions (e.g., Swait 2001) prior to the stated choice questions, to establish lower and upper bounds on acceptable attribute levels. Studies in transportation in the 1970s (e.g., Hensher 1976) highlighted the presence of asymmetric thresholds, but did not incorporate them into choice models.

Individuals are thought to adopt attribute thresholding in the way they process offered attribute levels associated with each alternative. Attribute thresholds have lower and upper bounds, which may be subject to measurement error, and also may be revised depending on the levels offered by other attributes. That is, there is 'softness' (in the language of Swait 2001) in the binding nature of perceived threshold levels reported by the q^{th} individual.

To capture the notion of threshold, we can define a lower cutoff and an upper cutoff. Accounting for attribute thresholds is equivalent to introducing functions that are incremental effects on the linear attribute effect throughout an attribute's entire range, and only get activated if the corresponding cutoff is in use. These cutoff penalties are typically defined as a linear function of the amount of constraint violation and defined as: $\{0{:}max(0, X_{ljq}\text{-}X_{lmin})\}$, the lower cutoff effect and deviation of the attribute level from the minimum cutoff attribute threshold where the attribute level is below the minimum cutoff (i.e., the cutoff exists), and zero otherwise (if the cutoff does not exist); and $\{0{:}max(0, X_{mmin} - X_{mjq})\}$, the upper cutoff effect and deviation of the attribute level from the maximum cutoff attribute threshold where the attribute level is above the maximum cutoff (i.e., the cutoff exists), and zero otherwise (if the cutoff does not exist). Defining X_{kjq} as the k^{th} attribute associated with the j^{th} alternative and q^{th} individual, with $l=K+1,\ldots,L$ attribute lower cut offs; $m=L+1,\ldots,M$ attribute upper cutoffs; $q=1,\ldots,Q$ respondents, and β_l and β_m are estimated penalty parameters, we can write the threshold penalty expression as equation (14).

$$\sum_{l=K+1}^{L} \beta_l \{0:\max(0, X_{ljq} - X_{lq\,\min})\} + \sum_{m=L+1}^{M} \beta_m \{0:\max(0, X_{mq\,\max} - X_{mjq})\} \tag{14}$$

Both upper and lower bounds can be behaviourally meaningful. For example, some individuals might only be interested in six cylinder cars and would not consider four and eight cylinder cars. Likewise low prices and very high prices might be rejected for different reasons, with purchasers often looking within a specific price range given their preferences.

Attribute thresholds can be introduced into a utility expression through the functional specification of an attribute, but also as a conditioning agent on an entire utility expression. To illustrate this, beginning with the standard utility expression associated with the j^{th} alternative contained in a choice set of j=1,…,J alternatives, we define R_{hq} as a dummy variable indicating whether the h^{th} attribute level is in a perceived attribute threshold rejection region or not for the q^{th} individual. This conditioning is a form of heteroscedasticity. An example of heteroscedastic conditioning, implemented in Hensher and Rose (2012) is A_{jq} = $(1+\sum_{h=1}^{H} \gamma_h R_{hq})$, where R_{hq} is defined above and γ_h are estimated parameters.

The model form for the utility expression that encapsulates the elements presented above is given in equation (15).

$$U_{jq} = (1+\sum_{h=1}^{H} \gamma_h R_{hq}))[\alpha_j + \sum_{k=1}^{K} \beta_{kj} X_{kjq} + \sum_{l=K+1}^{L} \beta_l \{0:\max(0, X_{ljq} - X_{lq\,\min})\} + \sum_{m=L+1}^{M} \beta_m \{0:\max(0, X_{mq\,\max} - X_{mjq})\}] + \varepsilon_j \quad (15)$$

All terms are defined above except α_j which are alternative-specific constants.

Equation (15) is a non-linear utility function, with utility functions defined over J_{qt} choices available to individual q in choice situation t, given in equation (16).

$$U_{jqt} = V_{jqt} + \varepsilon_{jqt},\ j = 1,...,J_{qt};\ t = 1,...,T_q;\ q = 1,...,Q \quad (16)$$

The IID, type I extreme value distribution is assumed for the random terms ε_{jqt}. Conditioned on V_{jqt}, the choice probabilities take the familiar multinomial logit (MNL) form (17).

$$Prob_{jqt} = \frac{\exp V_{jqt}}{\Sigma_{j=1}^{J_{qt}} \exp V_{jqt}} \quad (17)$$

When we allow for heteroscedasticty , equation (17) becomes equation (18).

$$Prob_{jqt} = \frac{\exp[(1+\sum_{h=1}^{H} \gamma_h R_{hq}))V_{jqt}]}{\Sigma_{j=1}^{J_{qt}} \exp[(1+\sum_{h=1}^{H} \gamma_h R_{hq}))V_{jqt}]} \quad (18)$$

Hensher and Rose (2011) have implemented this model form in the context of automobile purchases. They found a significant improvement in predictive power as well as different mean direct elasticities for heteroscedastic Gumbel scale MNL (HG-SMNL) compared to simple MNL models of the form in (18), due in large measure to the 'scaling' of the standard utility expression by a function that accounts for acceptability of each alternative and perceived attribute thresholds, as well as accounting for scale heterogeneity.

Hensher and Rose (2012) found that the relative disutility of the j^{th} alternative decreases[7] when this alternative (1) is perceived to be acceptable in contrast to not acceptable (0); and when the price attribute is in the rejection range (given the attached parameter estimate is -0.1848); this disutility is further tempered and increases. A negative parameter for the lower and upper cutoff penalties recognised that a price level outside of the lower and upper

[7] It should be noted that the overall utility expression is negative, and hence the heteroscedastic effect reduces the disutility when the alternative is acceptable, compared to not acceptable, as might be expected.

perceived thresholds of preference will add disutility, increasing the overall relative disutility. What we then have in this formulation is a way of recognising and adjusting the marginal disutility of an attribute associated with an alternative in a particular choice set.

4. The majority of confirming dimensions: dimensional vs. holistic processing strategies

The ‘majority of confirming dimensions’ (MCD) rule (Russo and Dosher 1983), is another form of attribute processing strategy that is concerned with the total count of superior attributes in each alternative. Under this test, pairs of attributes are compared in turn, with an alternative winning if it has a greater number of better attribute levels. The paired test continues until there is an overall winner.

Hensher and Collins (2011) used a choice experiment dataset to investigate the possibility of MCD. A total count of best attributes was generated for each alternative, and then entered into the utility expressions for all three alternatives. To contribute to the count for an alternative, an attribute had to be *strictly better* than that attribute in all other alternatives in the choice set. The distribution of the number of best attributes was calculated, both for the full relevance sample, and accounting for attributes being ignored, with separate reporting for all alternatives and the chosen alternative only. The distribution for the chosen alternative was found to be skewed towards a higher number of best attributes in both cases, with higher means observed, which is plausible. This alone does not suggest that MCD is being employed, as it would be expected that alternatives with a higher number of best attributes would also tend to have higher relative utilities.

Hensher and Collins (2011) did find, however, that the percentage of alternatives with zero strictly best attributes was much higher when allowing for attributes not attended to than in the ‘full relevance’ group. This might suggest that respondents are more likely to ignore an attribute when at least one attribute is outranked. On this evidence, if found true in other data, it has important behavioural implications since the analyst may wish to remove alternatives in model estimation where the number of best attributes is zero.

A series of choice models were estimated by Hensher and Collins (2011) to explore the potential for MCD when all attributes are relevant and under stated attribute non-attendance. Under full relevance of all attributes when they included a variable defined as “the number of attributes in an alternative that are best”, it was highly significant, and positive in sign, so that as the number of best attributes increases, an alternative is more likely to be chosen, as would be expected. When only the number of best attributes and the alternative-specific constants are included, and the attribute levels are omitted, the model fit was considerably worse even though “the number of best attributes” was highly significant, suggesting that the number of best attributes cannot substitute for the attribute levels themselves.

The same tests can be performed, after accounting for attributes stated as being ignored. i.e., any ignored attributes were not included in the count of the number of best attributes. The model fit was found to improve substantially when all attributes are assumed to be not attended to, with MCD complementing the parameterisation of attributes attended to. Hensher and Collins (2011) calculated values of travel time savings which varied sufficiently between full relevance and allowing for attributes being ignored, but not between models within each of this attribute processing settings when allowance was made for the number of attributes that are best.

The evidence suggests that all respondents simultaneously consider and trade between both the attribute levels in a typical compensatory fashion (both under full relevance and after ignoring some attributes if applicable), and the number of best attributes in each alternative. However to investigate whether there maybe two classes of respondent, with heuristic application distinguishing between them, two latent class models[8] were also estimated. The first class contained the attribute levels and alternative-specific constants, as per the base model, while the second class contained only the number of best attributes. A further improvement in model fit was obtained with this model. These results suggest that some respondents are employing the MCD heuristic. Under the heuristic, trading was not occurring on the absolute attribute levels. What appeared to matter instead is which alternative has the *best* level for each attribute, where tallies of the number of best attributes appeared to act as a supplementary step when determining the best alternative. Overall, the mean probability of class membership of each class in both models was over 80 percent for processing of the constituent attributes, and between 15 and 18 percent for the number of attributes being the determining influence.

5. Reference point revision and value learning

The final attribute processing strategy reviewed was proposed by DeShazo (2002) who suggested the idea of *reference point revision* in which preferences may be well-formed, but respondents' value functions shift when a non-status-quo option is chosen (see also McNair *et al.* 2012). The shift occurs because the selection of a non-status-quo option is viewed as a transaction up to a probability, and this causes a revision of the reference point around which the asymmetric value function predicted by prospect theory is centred (Kahneman and Tversky, 1979). There is an important distinction to be made between value learning, which in its broadest meaning implies underlying preferences are changing, and reference revision which can occur when preferences are stable but the objective is to maximise the likelihood of implementation of the most preferred alternative observed *over the course of the sequence of questions*. The latter is a special case of the former. Consider a model in which we identify the chosen alternative from a previous choice set, and create a dummy variable equal to 1 associated with whatever alternative was chosen in the previous choice set, be it the initial reference alternative or one of the offered non-status quo alternatives. Hensher and Collins (2011) introduced into utility expressions a revised reference dummy variable as a way of investigating the role of value-learning. They found that when the reference alternative is revised, in the next choice scenario it increases the utility of the new 'reference' alternative. This is an important finding, supporting the hypothesis of DeShazo; it is also recognition of sequential interdependence between adjacent choice scenarios, which should be treated explicitly rather than only through a correlated error variance specification, where the latter captures many unobserved effects at the alternative level.

Another useful test relates to the relationship between the level of an attribute associated with the reference (or status quo) alternative and each of the other alternatives in a choice experiment. One might distinguish between differences where a reference alternative attribute level was better, equal and worse relative to choice experiment alternatives CE1 and CE2,

[8] See Hensher and Greene (2009) for other examples of the identification of attribute processing heuristics with the latent class model.

defined as a series of attribute specific dummy variables (e.g., $attribute_i$ better = 1 if reference $attribute_i$ minus CE1 $attribute_i$ is negative and equal to zero if reference $attribute_i$ minus CE1 $attribute_i$ is positive). The choice response variable refers to the alternative chosen. A simple logit model can be specified in which the better and worse attribute forms for all design attributes can be included. Where an attribute refers to a better level for the reference alternative (the difference for all attributes being negative on the attribute difference as illustrated above for $attribute_i$), a positive parameter estimate suggests that when the difference narrows towards zero, making the reference alternative relatively less attractive on that attribute, the probability of choosing a non-reference alternative (CE1 or CE2) increases. Hensher and Collins (2011) in their empirical inquiry found that the parameter estimate was positive for 'better'. The opposite behavioural response was found when the reference alternative is worse. Positive parameter estimates suggest that when the reference alternative becomes relatively less attractive (given it is worse), the probability of choosing CE1 or CE2 increases.

6. Conclusions

This paper has selectively reviewed the growing literature on attribute processing, as well as its intersection with a broader literature on heuristics (the latter presented in other chapters). The link between attribute processing and heuristics can loosely be described by the role that attributes, as part of a package of attributes representing an alternative, play in the way that individuals process this information in arriving at a choice outcome. The connection between this chapter and the chapter by Chorus *et al.* seems obvious (see also Chorus 2010), yet there is clear scope to focus on the topics presented herein as a subset of the heuristics literature.

What we do know is that attribute processing is part of a growing interest in returning to the study of the underlying behavioural assumptions that influence the way in which decision makers adopt coping strategies to assist in making what they believe are sensible (albeit rational) choices. The extent to which the revealed processing strategies, and subsequent choice outcomes, are truly independent of the survey context is a matter of continuing debate and research; however it is generally accepted that the world is sufficiently complex that any additional imposition from a survey instrument may not be a cause of major concern in identifying the preference functions of individuals.

It is further suggested in the growing literature on attribute processing that continued sophistication of econometric assumptions, essentially treatments of errors and parameters, cannot alone improve the behavioural fit of choice models. A number of chapters in this handbook reinforce this position.

References

Bujosa, A., Riera, A. and Hicks, R. (2010) Combining discrete and continuous representation of preference heterogeneity: a latent class approach, *Environment and Resource Economics*, 47, 477-493.

Cameron, T.A. and DeShazo J.R. (2011) Differential Attention to Attributes in Utility-Theoretic Choice Models, *Journal of Choice Modelling*, 3(3), 73-115.

Campbell, D., Hensher, D.A. and Scarpa, R. (2011) Non-attendance to Attributes in Environmental Choice Analysis: A Latent Class Specification, *Journal of Environmental Planning and Management*, 54 (8), 1061-1076.

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

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

Chorus, C.G. (2010) A new model of random regret minimization. *European Journal of Transportation and Infrastructure Research* 10(2): 181-196.

Collins, A.T., Rose, J.M., and Hensher, D.A. (2012) The random parameters attribute nonattendance model. Paper presented at The 13th International Conference on Travel Behavior Research, Toronto, July.

DeShazo J.R. (2002) Designing transactions without framing effects in iterative question formats. *Journal of Environmental Economics and Management* 43, 360-385.

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 Califormia at Los Angeles, August.

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

Greene, W.H. and Hensher, D.A. (2010) Ordered choice, heterogeneity, and attribute processing, *Journal of Transport Economics and Policy* 44(3), 331-264.

Greene, W.H. and Hensher, D.A. (2012) Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model, *Applied Economic Letters*

Hensher, D.A. (1976) The value of commuter travel time savings: empirical estimation using an alternative valuation model, Journal of Transport Economics and Policy, X (2), 167-176

Hensher, D.A. (2006) Attribute processing in choice experiments, in *Valuing Environmental Amenities using Stated Choice Studies: A Common Sense Approach to Theory and Practice*, Barbara Kanninen, ed., Springer, Dordrecht, The Netherlands, 135-158.

Hensher, D.A. (2006a) How do Respondents Process Stated Choice Experiments? – Attribute consideration under varying information load, *Journal of Applied Econometrics*, 21, 861-878.

Hensher, D.A. (2007) Reducing Sign Violation for VTTS Distributions through Recognition of an Individual's Attribute Processing Strategy, *International Journal of Transport Economics* , XXXIV (3), October, 333-349.

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. (2010) Attribute Processing, Heuristics and Preference Construction in Choice Analysis. in Hess, S. and Daly, A. (eds.) *State-of Art and State-of Practice in Choice Modelling*, Emerald Press, UK., 35-70.

Hensher, D.A. and Collins, A. (2011) Interrogation of Responses to Stated Choice Experiments: Is there sense in what respondents tell us? A Closer Look at what Respondents Choose in Stated Choice Experiments. *Journal of Choice Modelling*, 4 (1), 62-89.

Hensher, D.A. and Greene, W.H. (2010) Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification, *Empirical Economics* 39 (2), 413-426.

Hensher, D.A. and Layton, D. (2010) Parameter Transfer of Common-Metric Attributes in Choice Analysis and Cognitive Rationalisation: Implications for Willingness to Pay *Transportation* 37 (3), 473-490.

Hensher, D.A. and Rose, J.M. (2009) Simplifying Choice through Attribute Preservation or Non-Attendance: Implications for Willingness to Pay, *Transportation Research Part E,* 45, 583-590.

Hensher, D.A. and Rose, J.M. (2012) The Influence of Alternative Acceptability, Attribute Thresholds and Choice Response Certainty on Automobile Purchase Preferences, *Journal of Transport Economics and Policy*, 46 (3), 451-468.

Hensher, D.A., Greene, W.H. and Collins, A. T. (2012a) Accounting for attribute non-attendance and common-metric aggregation in a latent class mixed multinomial logit model: a warning on potential confoundment, submitted to *Transportation*, revised version October 2012.

Hensher, D.A., Rose, J. and Bertoia, T. (2007) The implications on willingness to pay of a stochastic treatment of attribute processing in stated choice studies, *Transportation Research Part E,* 43 (1), 73-89.

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.A., Rose, J.M. and Greene, W.H. (2012) Inferring attribute non-attendance from stated choice data: implications for willingness to pay estimates and a warning for stated choice experiment design, *Transportation*, 39 (2) 235-254.

Hensher, D.A., Rose, J. and Puckett, S. (2009) Selective developments in choice analysis and a reminder about the dimensionality of behavioural analysis, In R. Kitamura and T. Yoshii and T. Yamamoto (Eds.) *The Expanding Sphere of Travel Behaviour Research: Selected Papers from the 11th International Conference on Travel Behaviour Research* Emerald Press, UK., 237-276.

Hess, S. (2011) Impact of unimportant attributes in stated choice surveys. ITS working paper, Institute for Transport Studies, University of Leeds.

Hess, S. and Hensher, D. A. (2010) Using conditioning on observed choices to retrieve individual-specific attribute processing strategies, *Transportation Research Part B,* 44(6), 781-90.

Hess, S. & Hensher, D.A. (2012), Making use of respondent reported processing information to understand attribute importance: a latent variable scaling approach, *Transportation*, in press (DOI: 10.1007/s11116-012-9420-y).

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., Stathopoulos, A., Campbell, D., O'Neill, V. & Caussade, S. (2012), It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity, *Transportation*, in press (DOI: 10.1007/s11116-012-9438-1).

Hess, S., Stathopoulos, A. and Daly, A, (2012) Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies, *Transportation*, 39 (3) 565-591.

Hole, A.R. (2011). A discrete choice model with endogenous attribute attendance. *Economics Letters*, 110 (3), 203-205.

Hoyos, D., Mariel, P., and Meyerhoff, J., (2010). *Comparing the performance of different approaches to deal with attribute non-attendance in discrete choice experiments: a simulation experiment.* BILTOKI 201001, Universidad del PaVasco - Departamento de EconomAplicada III (Econometry Estadica).

Layton, D. and Hensher, D.A. (2010) Aggregation of common-metric attributes in preference revelation in choice experiments and implications for willingness to pay, *Transportation Research Part D*, 15, 394-404.

Leong, W. and Hensher, D.A. (2012) Embedding multiple heuristics into choice models: an alternative approach, *Journal of Choice Modelling*,

Lockwood, M. (1996) Non-compensatory preference structures in non-market valuation of natural area policy. *Australian Journal of Agricultural Economics* 40: 73–87.

Louviere, J.J. and Hensher, D.A. (2001) Combining Preference Data, in Hensher, D.A. (ed.) *The Leading Edge of Travel Behaviour Research*, Pergamon Press, Oxford, 125-144.

Lundhede, T., Olsen, S., Jacobsen, J. and Thorsen, B. (2010) Handling respondent uncertainty in Choice Experiments: Evaluating recoding approaches against explicit modelling of uncertainty, Journal of Choice Modelling, 2(2), 118-147.

McFadden, D. (2001) Economic choices, *American Economic Review*, 91(3), 351-378.

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 Part E,* 44, 379-395.

Pudney, S. (1989) *Modelling Individual Choice:The Econometrics of Vorners, Kinks and Holes*, New York, Basil Blackwell.

Rigby, D.,and Burton, M. (2006) Modelling disinterest and dislike:a bounded bayesian mixed logit model of the UK market for GM food. *Environmental and Resource Economics* 33, 485–509.

Rekola, M. (2003) Lexicographic preferences in contingent valuation: a theoretical framework with illustrations. *Land Economics* 79: 277–291.

Rose, J., Hensher, D., Greene, W. and Washington, S. (in press) Attribute Exclusion Strategies in Airline Choice: Accounting for Exogenous Information on Decision Maker Processing Strategies in Models of Discrete Choice, *Transportmetrica*

Russo, J. E., and Dosher, B. A. (1983) Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9(4), 676-696.

Sælensminde, K. (2002) The impact of choice inconsistencies in stated choice studies. *Environmental and Resource Economics* 23: 403–420.

Scarpa, R., Gilbride, T., Campbell, D. and Hensher, D.A. (2009) Modeling attribute non-attendance in choice experiments for rural landscape valuation *European Review of Agricultural Economics,* 36 (2), 151-174.

Scarpa, R, Zanoli, R., Bruschi, V. and Naspetti, S. (2012) Inferred and stated attribute non-attendance in food choice experiments *American Journal of Agricultural Economics*, doi: 10.1093/ajae/aas073.

Scarpa, R., Thiene, M. and Hensher, D.A., (2010) Monitoring Choice Task Attribute Attendance in Non-Market Valuation of Multiple Park Management Services: Does it Matter? *Land Economics*, 86(4), 817-839.

Simon, H.A. (1955). A behavioural model of rational choice. *Quarterly Journal of Economics* 69, 99-118.

Spash, C. L. (2000) Ecosystems, contingent valuation and ethics: the case of wetland recreation. *Ecological Economics* 34: 195–215

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. (2001) A Non-compensatory choice model incorporating attribute cut-offs, *Transportation Research Part B*, 35(10), 903-928.