AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Annual Australian Agricultural and Resource Economics Society Conference, 9-11 February 2011

Modelling heterogeneity in response behaviour towards a sequence of discrete choice questions: a probabilistic decision process model

Ben J. McNair[a,*] • **David A. Hensher**[b] • **Jeff Bennett**[a]

[a] Crawford School of Economics and Government, The Australian National University, Canberra ACT 0200, Australia.

[b] Institute of Transport and Logistics Studies, The Business School, The University of Sydney, NSW 2006, Australia. David.Hensher@sydney.edu.au

* Corresponding author. Email address: Ben.McNair@anu.edu.au

Acknowledgements We thank ActewAGL Distribution, particularly David Graham, for permission to publish the results, Andrew Collins and Yang Lan of the Institute of Transport and Logistics Studies (ITLS) for their survey programming work, and John Rose of ITLS for advice on the design of the choice experiment. Denzil Fiebig, Riccardo Scarpa, and two anonymous reviewers provided helpful comments on earlier versions of this article. Funding for this research is from ActewAGL Distribution and ARC Industry Linkage Grant LP0669754.

Abstract There is a growing body of evidence in the non-market valuation literature suggesting that responses to a sequence of discrete choice questions tend to violate the assumptions typically made by analysts regarding independence of responses and stability of preferences. Decision processes (or heuristics) such as value learning and strategic misrepresentation have been offered as explanations for these results. While a few studies have tested these heuristics as competing hypotheses, none has investigated the possibility that each explains the response behaviour of a subgroup of the population. In this paper, we make a contribution towards addressing this research gap by presenting a probabilistic decision process model designed to estimate the proportion of respondents employing defined heuristics. We demonstrate the model on binary and multinomial choice data sources and find three distinct types of response behaviour. The results suggest that accounting for heterogeneity in response behaviour may be a better way forward than attempting to identify a single heuristic to explain the behaviour of all respondents.

Keywords Choice experiment; decision process; ordering effects; strategic response; willingness to pay

JEL codes C25; L94; Q51

Introduction

Stated choice methods have become an increasingly popular approach to estimating social values for non-market goods. In particular, choice experiments, which were originally applied in the transport (Hensher and Truong 1985) and marketing (Louviere and Hensher 1983) contexts, have been adapted to estimate values for a range of environmental (Bennett and Blamey 2001) and monopoly service (Beenstock et al. 1998, Carlsson and Martinsson 2008a) attributes. Choice experiments typically involve presenting respondents with a sequence of choice tasks, where respondents indicate their preference between two or more attribute-based alternatives in each task. The presentation of multiple choice tasks per respondent is preferred, and in some cases necessary, because it greatly increases the statistical efficiency of estimation and allows estimation of the distribution of preferences for a given attribute over a population. The standard assumptions when modelling responses to these questions are that each question is answered independently and truthfully and that underlying preferences are initially well-formed and stable over the course of the sequence. Yet, several studies have found that responses violate these assumptions, in some cases causing estimates of willingness to pay (WTP) implied by various order positions in a sequence to differ (Bateman et al. 2008b, Cameron and Quiggin 1994, Day et al. 2009, Day and Pinto 2010, DeShazo 2002, Hanemann et al. 1991, McNair et al. 2011).

Several decision processes (or heuristics) have been put forward as explanations for such results. One group of heuristics predict that respondents consider alternatives accepted in previous questions when making their choices. These heuristics have generally been based on the prediction of neo-classical economic theory, recently highlighted by Carson and Groves (2007), that respondents may misrepresent their preferences in one or more questions in order to maximise the likelihood of implementation of their most preferred alternative observed in the sequence to that point. Another group of heuristics revolve around the idea that respondents have poorly-formed preferences that are influenced by the information observed in choice tasks. This phenomenon was termed anchoring (or starting-point bias) in the context of double-bounded contingent valuation surveys in which the preferences stated in the first question differed from those stated in the follow-up question (Boyle et al. 1985, Herriges and Shogren 1996). In longer sequences of questions, the phenomenon has been characterised as value learning (Plott 1996), which may be confined to the first question (Ariely et al. 2003), but could extend further into a sequence of questions (for example in the form of a ‘good deal / bad deal’ heuristic (Bateman et al. 2008b)).

A few studies have attempted to ascertain which of these heuristics best explains responses in a given data set (Day et al. 2009, Day and Pinto 2010, DeShazo 2002, McNair et al. 2011), but none has investigated the possibility of heterogeneity in response behaviour across respondents; that is, the possibility that each of the proposed heuristics explains the response behaviour of a subgroup of respondents in the survey (up to a probability). In this paper, we offer a contribution towards addressing this research gap. The objective is to demonstrate, using both binary and multinomial choice data, how a probabilistic decision process (PDP) model can be used to estimate discrete levels of heterogeneity in response behaviour towards a sequence of choice questions. In contrast to previous studies using mixture models allowing different decision processes (Araña et al. 2008), we focus on decision processes that affect WTP estimates in a full compensatory framework. Our model is similar to the equality-constrained latent class (ECLC) models previously used by Scarpa et al. (2009) and Hensher and Greene (2010) to account for simplifying heuristics, such as attribute non-attendance and dual processing of common-metric attributes. In contrast to typical latent class models, classes are defined by the analyst as separate utility functions specified by restricting certain parameters in a 'master' utility function. The model estimates the membership probability associated with each class. Parameters are restricted to be equal across classes to ensure that class membership is determined by decision process rather than by taste heterogeneity.

In the following section, we describe the heuristics that have been put forward in the literature as potential explanations for ordering anomalies. We then detail the PDP model, the data source to which it is applied, and the results from the analysis before, finally, drawing conclusions.

Background

Two of the standard assumptions when modelling responses to a sequence of stated choice questions are that:

1. all respondents truthfully answer the question being asked; and
2. true preferences are stable over the course of a sequence of questions.

The focus of this paper is on accounting for response behaviour that violates one or both of these assumptions in a way that affects estimates of WTP in a full compensatory framework. Consequently, we do not seek to estimate the effects of any simplifying heuristics such as attribute non-attendance (Hensher and Greene 2010, Scarpa et al. 2009, Zellman et al. 2010). Nor do we account for institutional learning (Braga and Starmer 2005) or respondent fatigue.[1] While these behavioural processes

[1] Two types of learning have been identified in the literature. The first, institutional learning, relates to the *process of learning* how to evaluate and complete a choice task. This process reduces random error in stated choices, increasing their predictability. The second, value learning, relates to the *discovery of preferences*. This process changes a respondent's taste intensities and these changes are related to the attribute levels presented in choice tasks. Given our focus on the effect of response behaviour on WTP, our models account for value learning, but, in the interest of simplicity, not institutional learning.

have been shown to influence 'noise' in the data, manifest as changes in the variance of the random error component (or, equivalently, scale)[2] (Bradley and Daly 1994, Caussade et al. 2005, Holmes and Boyle 2005), there is no implied relationship with WTP.

The various heuristics that do violate the standard assumptions can be grouped into two broad categories – those that involve a violation of the first standard assumption, and those that involve a violation of the second.

Strategic misrepresentation

Response behaviour that violates the first standard assumption can generally be classified as *strategic misrepresentation*. It has long been recognised in neoclassical economic theory that consumers may conceal their true preferences if it enables them to obtain a public good at a lower cost (Samuelson 1954). More recently, Carson and Groves (2007) highlighted the predictions of this theory, particularly mechanism design theory (Hurwicz 1972, Mirrlees 1971), in relation to stated choice surveys. One of the predicted patterns of response behaviour is the rejection of an alternative that is preferred to the status quo when a similar good was offered at a lower cost in a previous choice task. This rejection increases the likelihood that the respondent's most preferred option observed in the sequence of choice tasks to that point is implemented. Bateman et al. (2008b) differentiate between *strong* strategic misrepresentation, in which respondents always reject a good if it was offered at a lower cost in a previous choice task, and *weak* strategic misrepresentation, in which respondents weigh up the rejection against the perceived risk of the good not being provided at the lower cost.

DeShazo (2002) also argued that respondents do not answer questions independently, but that they evaluate choice questions in terms of deviations from references points based on previously accepted alternatives. Although DeShazo's model appeals to prospect theory (Kahneman and Tversky 1979) rather than mechanism design theory, it shares the two main predictions of the weak strategic misrepresentation heuristic; first, that respondents compare presented alternatives with alternatives accepted in previous choice tasks, and, second, that respondents consider *expected utility* based on the probability of provision. The prediction in both cases is that the WTP estimate implied by the first question in a sequence will exceed the WTP estimates implied by subsequent questions (assuming backward navigation through choice tasks is prevented).

Value learning

Response behaviour that violates the second standard assumption can generally be classified as *value learning* (Plott 1996). Value learning heuristics revolve around the idea that preferences are initially poorly formed and are discovered in the process of

[2] In the multinomial logit model, the scale parameter, λ, is an inverse function of the variance of the unobserved effects, $\sigma^2=\pi^2/6\lambda^2$.

completing choice tasks. They generally predict that discovered preferences are positively influenced by the cost levels presented in choice tasks. In dichotomous-choice contingent valuation surveys, the outcome of such response behaviour has been termed starting-point or anchoring bias (Boyle et al. 1985, Herriges and Shogren 1996). The focus in these short, one- or two-question sequences has been on the effect on preferences of the cost level observed in the first choice task. With respect to the longer sequences of questions typically employed in choice experiments, some authors have maintained this focus on the effect of the first choice task (Ariely et al. 2003, Ladenburg and Olsen 2008), while others have put forward heuristics in which the effect extends beyond the first task, potentially for the duration of the sequence of questions. For example, Bateman et al. (2008b) describe a 'good deal / bad deal' heuristic (Bateman et al. 2008b) in which an alternative is more (less) likely to be chosen if its cost level is low (high) relative to the levels presented in previous choice tasks.

If the value learning process is symmetric in terms of the effect of observed attribute levels on preferences, then choice experiments can be designed in which this response behaviour does not imply a relationship between question order and WTP. However, this behaviour does imply a relationship between WTP and the cost levels (or bid vector) used in the choice survey (Carlsson and Martinsson 2008b). As noted by Bateman et al. (2008a), this relationship "fundamentally questions the underpinnings of standard microeconomic theory, in effect suggesting that, at least to some degree, prices determine values rather than vice versa."

Empirical evidence

Turning to empirical evidence, a number of studies have found evidence of response patterns associated with a single heuristic, whether it be a strategic misrepresentation heuristic (Carson et al. 2009, Carson et al. 2006, Hensher and Collins 2010) or a value learning heuristic (Ariely et al. 2003, Carlsson and Martinsson 2008b, Herriges and Shogren 1996, Holmes and Boyle 2005, Ladenburg and Olsen 2008). However, only a few have tested the heuristics discussed above as competing hypotheses to ascertain which best explains responses in a given data set. DeShazo (2002) and Bateman et al. (2008b) found evidence that supports a strategic misrepresentation heuristic in which consideration is given to alternatives accepted in previous choice tasks and to the perceived probability of provision. The weight of evidence found by Day and Pinto (2010) supports a value learning heuristic, although the study found that no proposed heuristic unambiguously explained the 'ordering anomalies' in the data. McNair et al. (2011) found response patterns that could be explained by weak strategic misrepresentation or an asymmetric form of value learning in which lower cost levels have a greater impact on preference revision than higher cost levels.

It appears that no study has investigated the possibility of heterogeneity in response behaviour across respondents; that is, the possibility that each of the proposed heuristics explains the response behaviour of a subgroup of respondents in the survey. In this paper, we offer a contribution towards addressing this research gap.

Method

While it may not be possible to identify whether a heuristic has been employed by observing the responses of a single respondent, over a sufficiently large sample, it is possible to identify the response patterns predicted by a given heuristic in terms of relationships between responses and attribute levels observed by respondents in previous choice tasks. We use a PDP model to estimate the proportions of respondents behaving in accordance with three heuristics based on the three types of response behaviour discussed above:

1. the standard assumptions (truthful, independent response with stable preferences);
2. value learning; and
3. strategic misrepresentation.[3]

A random utility framework (McFadden 1974) is applied in which respondent utility is equal to the sum-product of observed factors, **x**, and associated taste intensities, **β**, plus the sum-product of respondent characteristics, **m**, and their marginal contributions to (dis)utility, **δ**, plus unobserved factors, ε, which are i.i.d. according to the Extreme Value Type I function. Following Hensher and Greene (2010), the resulting logit choice probability function for the discrete choice from J alternatives can be written:

$$\text{Prob[choice } j \text{ by individual } i \text{ in choice task } t \mid \text{class } q \text{]} = P_{itj|q} = \frac{\exp(V_{itjq})}{\sum_{j=1}^{J}\exp(V_{itjq})}$$

where $V_{itjq} = x'_{itj}\beta_{jq} + m_i\delta_{jq}$

The probability that individual i belongs to class q of Q is:

$$H_{iq} = \frac{\exp(\theta_q)}{\sum_{q=1}^{Q}\exp(\theta_q)} \text{ , } \theta_Q=0$$

The log-likelihood function to be maximised is the sum over individuals of the log of the expectation over classes of the joint probability of the sequence of T choices.

$$\ln L = \sum_{i=1}^{N}\ln P_i = \sum_{i=1}^{N}\ln\left[\sum_{q=1}^{Q}H_{iq}\prod_{t=1}^{T}P_{itj|q}\right]$$

In order to simplify the approach, the standard attributes, **x**, are defined so that they take the value zero in the status quo utility function. To achieve this, we simply define the attributes in terms of changes relative to the status quo. The reason for this redefinition becomes clearer in the discussion to follow.

[3] Directly asking respondents to reveal their behaviour is not a viable alternative since those employing the strategic misrepresentation heuristic would indicate that they responded to each choice task independently and truthfully.

The Q classes are defined by separate parameter vectors, $\boldsymbol{\beta}_{jq}$. Parameters are constrained to take the value zero in certain classes, but the non-zero parameters to be estimated are constrained to take the same value across classes (that is, they are assumed to be generic). These Q vectors effectively translate to Q sets of utility functions to which respondents are assigned up to a probability to maximise the log-likelihood function.

In this study, Q=3 sets of utility functions are specified to capture the response patterns associated with each of the three classes of response behaviour.[4] Given that the literature contains variants on each hypothesis, there is likely to be some argument about how the utility functions should be specified for each class. While we do not claim to have developed definitive sets of utility functions, we believe the functions described below are the most suitable for this study based on the weight of evidence in the literature and model fit testing. They are tailored to analyse responses to stated choice surveys in which similar goods are offered at very different prices over the course of a sequence. Such surveys arise in non-market valuation settings where significant heterogeneity is expected in the distribution of WTP for a public project over the population, but the set of credible project options are viewed as similar. The consequence is that value learning and strategic behaviour tend to be driven mainly by the cost attribute. Our utility functions are specified accordingly, however, the approach could potentially be expanded to incorporate the effects of other attributes.

Standard assumptions (Class 1)

The utility functions specified for the class of respondents behaving in accordance with the standard assumptions are the conventional sum-product of the k attributes as they appear in the choice task being answered and their associated taste intensities.

$$U_{it,\mathrm{SQ,class1}} = \beta_1 x_{1,it,\mathrm{SQ}} + \ldots + \beta_k x_{k,it,\mathrm{SQ}}$$

$$U_{it,\mathrm{ALT,class1}} = \beta_0 + \boldsymbol{\delta}\mathbf{m}_i + \beta_1 x_{1,it,\mathrm{ALT}} + \ldots + \beta_k x_{k,it,\mathrm{ALT}}$$

Value learning (Class 2)

The second class represents those responding in accordance with a value learning heuristic. We focus on the role of cost levels in value learning. Cost levels are generally considered to be the main influence in the value learning process, particularly in stated choice surveys in which similar goods are offered at very different prices over the course of a sequence. We specify utility functions that capture the response patterns of this group by allowing the alternative-specific preference to vary with the average of cost levels observed in the sequence up to and

[4] We limit the classes to broad types of decision process because of concerns over the stability of models dealing with richer specifications. We found that when too many very similar strategies were included in our model, some class probabilities are estimated at zero, with positive probabilities estimated for a set of sufficiently different strategies. Our concern with such models is that the assignment of zero probabilities amongst similar strategies may be unstable.

including the current choice task.[5] This equal-weight average was found to result in better model fit on our data source than a specification weighted towards more recent observations. The length of the sequence in our data source was just four choice tasks. In longer sequences, perfect recall is less likely and a weighted specification may be preferred (for example Day et al. 2009). The cost level in the current choice task is included in the average to accommodate the prediction of coherent arbitrariness (Ariely et al. 2003), anchoring, and starting-point bias (Herriges and Shogren 1996) that the cost level observed in the first choice task will influence preferences prior to response. The utility functions are as follows.

$$U_{it,\mathrm{SQ,class2}} = \beta_1 x_{1,it,\mathrm{SQ}} + \ldots + \beta_k x_{k,it,\mathrm{SQ}}$$

$$U_{it,\mathrm{ALT,class2}} = \beta_0 + \boldsymbol{\delta}\mathbf{m}_i + \beta_1 x_{1,it,\mathrm{ALT}} + \ldots + \beta_k x_{k,it,\mathrm{ALT}} + \beta_{k+1} z_{it,\mathrm{ALT}}$$

where

$z_{itj} = z^{o}_{itj} - \check{z}_j$

z^{o}_{itj} = the average of cost levels observed up to and including the current choice task

$\check{z}_j$ = the average of cost levels in the sample (across all respondents and all choice tasks)

The purpose of $\check{z}_j$ is econometric rather than behavioural. It simply 'normalises' the average observed cost variable by ensuring its sample mean is approximately zero. This prevents the model from using the coefficient, β_{k+1}, to infer heterogeneity in taste across classes, thus ensuring the model estimates only heterogeneity in decision process.

Strategic misrepresentation (Class 3)

In a third class of response behaviour, we specify utility functions that capture the response patterns predicted by a strategic misrepresentation heuristic. The heuristic has two features. The first is that respondents compare alternatives to those accepted in previous choice tasks.[6] In particular, they choose the status quo option not only when the status quo is preferred to the alternatives, but, potentially, also when a previously accepted alternative is preferred to the alternatives currently on offer. We assume that respondents effectively replace the status quo with a reference alternative once they have expressed a preference for an alternative over the status quo. We

[5] The main variation within the group of value learning heuristics lies in the length of the sequence of choice tasks over which the learning occurs. In this case, it was not possible to estimate separate classes for different lengths. We define a class in which learning occurs over the duration of the full sequence of four choice tasks, but value revision is based on changes in average observed cost, which become smaller on average over the course of a sequence.

[6] A key difference between value learning and strategic misrepresentation is that the former is driven by cost levels *observed* in previous tasks, while the latter is driven by cost levels *accepted* in previous tasks.

define the reference alternative as the highest-cost alternative previously accepted in the sequence. Over the range of cost and WTP levels that matter, this reference alternative yields the highest *expected* utility (based on the provision probabilities discussed below) of all previously accepted alternatives.[7]

The second feature of this heuristic is that respondents consider the probability of provision. When a similar good is offered at very different cost levels over the course of a sequence of choice tasks, respondents may assume, often quite rightly, that higher-cost options are more likely to be provided because the agency is more likely to proceed with the project the higher is respondents' stated WTP. The finding of McNair et al. (2011) that response patterns in this survey align more closely with *weak cost minimisation* than *strong cost minimisation* suggests that probability of provision was considered by at least some respondents. We assume the perceived probability of project provision is equal to the ratio of the maximum cost level *accepted* and the maximum cost level *observed*.[8] Consider the case where a project option priced at $4,000 is accepted in the first of a sequence of binary choice tasks. If a project option priced at $8,000 is presented in the second task, then the perceived probability of project provision is revised to 50 per cent. The respondent is faced with a trade-off. The perceived probability of provision can be increased to 100 per cent, but at the cost of accepting the more expensive ($8,000) alternative. If the alternative is accepted, it becomes the reference alternative in the next choice task. Alternatively, if a project option priced at $2,000 is presented in the second choice task, then the choice does not influence the probability of project provision (and the respondent will accept the $2,000 alternative assuming the goods are sufficiently similar). Of course, respondents may not carry out such precise calculations when making their choice. The decision process is likely to be somewhat implicit and may vary across respondents. However, the aim is to use response patterns observed over many respondents to capture the broad type of decision process.

The utility equations represent the expected utilities from the reference and current alternatives.[9]

---

[7] We showed by simulation that, if the good being offered is sufficiently similar across tasks, the highest-cost alternative previously accepted yields higher expected utility (as a reference alternative, with expectations based on the assumed provision probabilities) than all other previously accepted alternatives for all combinations of WTP and cost (in the present task) in which the present alternative yields expected utility higher than at least one previously accepted alternative. Calculations are available from the corresponding author on request.

[8] In surveys where cost is less dominant, other attributes may need to be incorporated in this proxy. We also acknowledge that the perceived probability of provision is unlikely to ever be 100 per cent due to uncertainty about others' preferences and the advisory nature of most surveys. However, it is the relative probabilities, rather than the absolute probabilities, that are important in determining the choice probabilities.

[9] The (1-p) terms are not required since utility from the status quo is zero.

$$U_{it,SQ,class3} = p_{it,SQ}(\beta_0 + \boldsymbol{\delta}\mathbf{m}_i + \beta_1 x^a_{1,it} + \ldots + \beta_k x^a_{k,it})$$

$$U_{it,ALT,class3} = p_{it,ALT}(\beta_0 + \boldsymbol{\delta}\mathbf{m}_i + \beta_1 x_{1,it,ALT} + \ldots + \beta_k x_{k,it,ALT})$$

where

$\mathbf{x}^a_{it}$ = the levels of attributes in the highest-cost alternative accepted in previous choice tasks ($x^a_{1,it}$ is the maximum cost level accepted in previous choice tasks)

$p_{it,SQ} = x^a_{1,it}/x^o_{1,it}$

$x^o_{1,it}$ = the maximum cost level observed up to and including the current choice task

$p_{it,ALT} = max[p_{it,SQ}, x_{1,it,ALT}/x^o_{1,it}]$

The importance of defining the standard attributes in terms of changes relative to the status quo now becomes clear. If a respondent has chosen the status quo in all choice tasks to a given point, then $x^a_{1,it}=0$, $p_{it,SQ}=0$ and $U_{it,SQ,class3}= U_{it,SQ,class1}= U_{it,SQ,class2}=0$. In the first question in a sequence, the class 3 utility functions are identical to those in Class 1 since $p_{it,SQ}=0$ and $p_{it,ALT}=1$. Once a respondent has chosen an alternative over the status quo, that alternative replaces the status quo as the reference point and $U_{it,SQ,class3}>0$. Alternatives presented in subsequent choice tasks are accepted if the expected utility from choosing the alternative exceeds the expected utility from choosing the reference alternative.

Class structure in the model

The three sets of utility functions are operationalised in the model by three separate sets of restrictions on a 'master' utility function. Certain parameters are restricted to be zero and certain parameters are restricted to be equal both within and across classes as shown in Table 1. The alternative-specific constants and the standard attributes, **x,** are divided into two parts – one multiplied by $p_{it,ALT}$ and another by 1- $p_{it,ALT}$. In Classes 1 and 2, coefficients on attributes multiplied by $p_{it,ALT}$ and 1- $p_{it,ALT}$ are assumed to be equal so that they represent the marginal utility of the standard attribute without consideration of the probability of provision. In Class 3, the coefficients on attributes multiplied by 1- $p_{it,ALT}$ are set to zero so that utility depends on the probability of provision. A set of reference alternative variables are restricted to hold zero value in Classes 1 and 2 (in which previously accepted alternatives are ignored), but in Class 3, they are assumed to have the same taste intensities as the equivalent variables in the non-status-quo alternatives in the current choice task. All non-zero attributes are assumed to take the same value across classes.

Data

We implement the model on data from a survey of homeowners in the Australian Capital Territory (ACT) in 2009. The main objective of the survey was to establish homeowners' willingness to pay to have overhead electricity and telecommunications

wires in their suburb replaced by new underground wires. We provide a brief overview herein and refer readers to McNair et al. (2010) for details.

Table 1: Class structure[a]

Variable	Alternative	Class 1 (Standard)	Class 2 (Learning)	Class 3 (Strategic)
$p_{it,SQ}$	SQ	0	0	β_0
$\mathbf{x}^a_{it}p_{it,SQ}$	SQ	0	0	$\boldsymbol{\beta}$
$p_{it,ALT}$	Alt	β_0	β_0	β_0
$\mathbf{x}_{it}p_{it,ALT}$	Alt	$\boldsymbol{\beta}$	$\boldsymbol{\beta}$	$\boldsymbol{\beta}$
$1-p_{it,ALT}$	Alt	β_0	β_0	0
$\mathbf{x}_{it}(1-p_{it,ALT})$	Alt	$\boldsymbol{\beta}$	$\boldsymbol{\beta}$	0
$z_{it,ALT}$	Alt	0	β_{k+1}	0

[a] $\boldsymbol{\beta}$ refers to a coefficient vector, $\beta_1, \beta_{2,\ldots,} \beta_k$, associated with $x_1,\ldots, x_k$.

Data from two elicitation formats used in the survey are analysed in this study. The first format comprised a sequence of four binary choice tasks in which respondents were presented with a description of their current (overhead) service and one undergrounding alternative (the binary choice format). The second format also comprised a sequence of four choice tasks, but each task contained the current service and *two* undergrounding alternatives (the multinomial choice format). The attributes used to describe the alternatives and the levels assigned to those attributes are presented in Table 2. The value of the alternative label embodies all of the benefits of undergrounding other than supply reliability benefits, including the amenity and safety benefits that qualitative questions showed to be the major household benefits from undergrounding. The restricted range of credible levels for supply reliability attributes meant that similar goods were offered at very different prices over the course of the choice task sequences. Consequently, opportunities for strategic misrepresentation may have been relatively obvious and, potentially, value learning may have been exacerbated.

Two blocks of four choice tasks were constructed in the multinomial choice format to maximise the Bayesian C-efficiency of the design (Scarpa and Rose 2008) and minimise the correlation between attribute levels and block assignment.[10] The binary design was created by splitting these two blocks into four blocks of four binary choice tasks. An example of a choice task from the multinomial choice format is presented in Figure 1.

[10] Bayesian priors were derived from pilot responses and from NERA and ACNielsen (2003). Default levels were assumed for supply reliability attributes in the status quo.

Table 2: Attributes and levels

Attribute	Levels	
	Status quo (overhead) alternative	Undergrounding alternatives
Your one-off undergrounding contribution (A$ 2009)	0	1,000, 1,100, 2,000, 2,100, 2,800, 3,000, 3,900, 4,000, 6,000, 6,200, 8,000, 8,200, 11,800, 12,000, 15,900, 16,000
Power cuts without warning:		
Number of power cuts each five years	Set by respondent	Proportions of status quo level: 0.25, 0.5, 0.75, 1 [a,b]
Average duration of power cuts	Set by respondent	Proportions of status quo level: 0.33, 0.66, 1.33, 1.66 [a]
Power cuts with written notice (occurring in normal business hours):		
Number of power cuts each five years	Set by respondent	Proportions of status quo level: 0.2, 0.4, 0.6, 0.8 [a,b]
Average duration of power cuts	Set by respondent	Proportions of status quo level: 0.33, 0.66, 1.33, 1.66 [a]

[a] Rounded to the nearest integer; [b] Absolute levels (0, 1 and 2) were assigned where respondents chose very low status quo levels (1 or less).

Some 292 respondents completed the web-based questionnaire in the binary choice format and 290 in the multinomial choice format.[11] Importantly, the questionnaire did not allow respondents to navigate back through the sequence of choice tasks. It was programmed to cycle through the various sample splits, blocks, and choice task orderings to ensure approximately equal representation across choice observations.

As many as 30 per cent of respondents completing the binary format and 24 per cent of respondents completing the multinomial format chose the status quo scenario in all four choice tasks.[12] The response behaviour of this group is difficult to determine because, if the value placed on undergrounding by a respondent is sufficiently low, then all three heuristics result in the same pattern of responses – selection of the status quo in every task. These respondents are omitted from the analysis in this paper to ensure that the method can be demonstrated effectively. We expect the method could be applied to full survey data sets in other studies where such responses represent a lower proportion of the sample.

[11] The data were drawn from a larger split-sample internet survey completed by 1745 of 2485 respondents agreeing to provide an email address in telephone recruitment interviews.

[12] The magnitude of these proportions could have been reduced by offering undergrounding options at a lower cost than the status quo, but such an approach was not practical in this study.

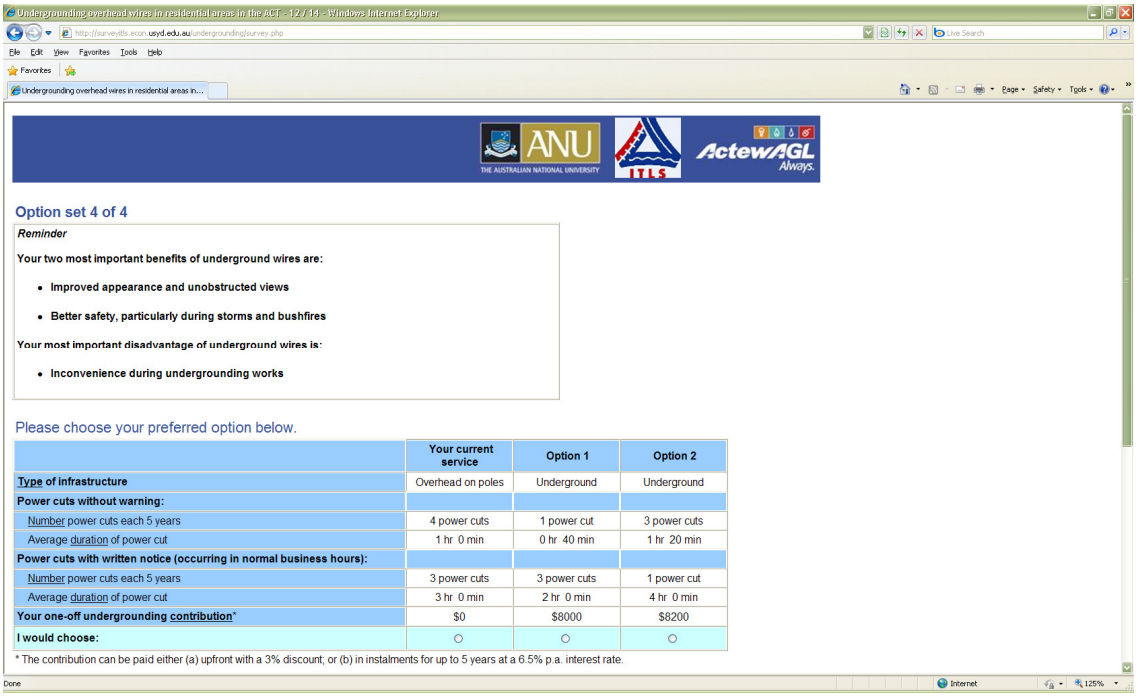

Figure 1: Example of a choice task

An important part of the method is the manipulation of variables prior to estimation. We used a spreadsheet to create the normalised average observed cost variable, $z_{it,\mathrm{ALT}}$, the provision probability proxies for the reference and current alternatives, $p_{it,\mathrm{SQ}}$ and $p_{it,\mathrm{ALT}}$, the attribute levels associated with the highest-cost alternative previously accepted, $\mathbf{x}^{\mathbf{a}}_{it}$, and the maximum cost level observed up to and including the current choice task, $x^{\mathrm{o}}{}_{1,it}$.

Results

A summary of the PDP model results for the binary (Model 1) and multinomial (Model 2) formats is presented in Table 3. The parameter estimates on the seven project attributes in each model have the expected sign where they are significant at the 0.05 level, and there is some evidence to suggest that household income and respondent age are positively related to WTP for undergrounding.[13] The positive coefficient on the normalised average observed cost variable indicates that, within Class 2, the value placed on undergrounding is influenced by the cost levels observed in previous choice tasks and the current choice task. A respondent in this class is more likely to accept an undergrounding alternative priced at $4,000 in the second choice

[13] The supply reliability attributes are more statistically significant in the multinomial choice format, since, in contrast to the binary format, respondents used these attributes to discriminate between two undergrounding options with similar cost. Age and household income variables are effects coded such that: each of the four age variables included in the model take the value -1 when age is 40-49; and, each of the six income variables included in the model take the value -1 when income is not provided. Other demographic variables, such as gender, education, and household size were found to be statistically insignificant in the estimated utility function. No demographic variable was found to be significantly related to class membership probability in this case.

task if the alternative offered in the first choice task was priced at $6,000 than if it was priced at $2,000 (all else held constant).

Table 3: Summary of estimation results

Model type	Probabilistic decision process				Standard multinomial logit			
Choice format	Binary choice		Multinomial choice		Binary choice		Multinomial choice	
	(Model 1)		(Model 2)		(Model 3)		(Model 4)	
Variable	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Undergrounding-specific constant	7.2281	6.99	5.8505	7.04	4.3267	10.57	3.4334	8.77
Log of household contribution	-3.9484	-8.34	-3.1712	-8.86	-2.4943	-13.91	-2.0670	-12.68
Change in frequency of unplanned power cuts	-0.0544	-0.87	-0.1587	-2.82	-0.0425	-0.85	-0.1751	-3.54
Change in frequency of planned power cuts	-0.1709	-1.65	0.0554	0.84	-0.1542	-1.92	0.0309	0.55
Change in average duration of unplanned power cuts	-0.0008	-0.18	-0.0062	-3.70	-0.0019	-0.57	-0.0055	-3.78
Change in average duration of planned power cuts	-0.0011	-0.75	-0.0041	-7.27	-0.0016	-1.47	-0.0038	-7.60
Normalised average observed cost (Class 2 only)	0.5009	2.21	0.4360	2.38				
Interactions with undergrounding-specific constant:								
Household income: A$18,199 or less	-2.2195	-1.80	-1.4144	-1.16	-1.7594	-1.36	-1.2642	-1.16
Household income: A$18,200 – A$51,999	0.2607	0.45	-0.0346	-0.08	0.0065	0.02	0.1686	0.50
Household income: $52,000 – A$88,399	0.2645	0.63	-0.4812	-1.27	0.1289	0.44	-0.2744	-0.98
Household income: A$88,400 – A$129,999	0.4807	1.19	0.3930	1.23	0.5332	1.85	0.3043	1.20
Household income: A$130,000 – A$181,999	0.9327	2.17	1.0071	2.33	0.7746	2.56	0.7094	2.56
Household income: A$182,000 or more	0.6511	1.39	1.1731	2.43	0.7063	2.07	0.7132	2.36
Age: 18-29	-0.9168	-1.53	-1.4320	-2.49	-0.8547	-1.97	-0.8760	-2.40
Age: 30-39	-0.1537	-0.45	0.0436	0.15	-0.0159	-0.06	0.1045	0.51
Age: 50-64	0.3265	1.32	0.7007	2.69	0.2557	1.45	0.4110	2.49
Age: 65 and over	0.5403	1.61	1.2240	2.47	0.5062	2.11	0.6664	2.48
Estimated class probabilities:								
Class 1 (standard assumptions)	0.269	2.28	0.129	0.39				
Class 2 (value learning)	0.412	4.43	0.422	2.13				
Class 3 (strategic misrepresentation)	0.319	2.81	0.450	2.35				
Model fit:								
N	800		872		800		872	
Log-likelihood	-342		-740		-362		-760	
AIC	722		1518		756		1552	

Turning to the estimated class probabilities, all except one are significant at the 0.05 level across the two models. Both models estimate that approximately 40 per cent of respondents behaved in accordance with the value learning utility specification. The

proportion behaving in line with the strategic misrepresentation specification is estimated at 32 per cent in the binary format and 45 per cent in the multinomial format. The class with the lowest membership probability in both models was that based on the standard assumptions of truthful response and stable preferences, with 27 and 13 per cent predicted by Model 1 and Model 2, respectively. No single class dominates either model, indicating significant heterogeneity in the response behaviour towards both the binary and multinomial choice formats. There is no evidence of a relationship between choice format and decision process, with the class probabilities statistically indistinguishable at the 0.05 level across the two models (p-values are 0.71, 0.95, and 0.58 based on 1000 paired differences).

The log-likelihood values associated with the PDP models indicate an improvement in model fit over the single-class, but otherwise equivalent, multinomial logit (MNL) models (also presented in Table 3). This improvement is expected given the additional parameters accommodating heterogeneity in the PDP models. Of greater interest is the improvement in the AIC value, which accounts for parameter proliferation. The improvement in this criterion suggests that accounting for heterogeneity in response behaviour using the PDP model is important even when model parsimony is considered desirable.

To confirm that the heterogeneity captured by the model is indeed related to precedent-dependent decision processes such as value learning and strategic misrepresentation, we scrambled the order of the choice tasks within each respondent in the data, re-calculated the relevant variables, and re-estimated Model 1 and Model 2. Consistent with our theoretical framework, the task-dependent behaviour all but disappeared, with statistically insignificant parameter estimates on the average observed cost variable and on the membership probabilities for Class 2 and Class 3.[14]

We turn now to implications for welfare estimates. The intention is to estimate welfare based on *prior underlying* preferences. The standard MNL model assumes that *stated* preferences are a true reflection of prior underlying preferences, and welfare is estimated accordingly. In the PDP model, prior underlying preferences are assumed to be equal across classes following Scarpa et al. (2009) and Hensher and Greene (2010), but stated preferences can differ, with divergence between stated and prior underlying preferences in Class 2 and Class 3. The estimated Class 1 utility function represents the prior underlying preferences of the representative respondent and is therefore the appropriate basis for welfare estimation.

The undergrounding choice probability (or bid acceptance) curves based on the Class 1 utility functions from the binary and multinomial choice format models are shown in Figure 2 and Figure 3 (for a 50-64 year old respondent with annual household income in the range A$88,400 to A$129,999, and all other non-cost attributes set at their sample means). Estimates of mean prior WTP, calculated as the areas under the

[14] Model results are available from the corresponding author on request.

undergrounding choice probability curves, are not significantly different at the 0.05 level across the PDP and MNL models, with p-values of 0.91 and 0.38 in the binary and multinomial formats, respectively, based on 1000 paired differences. However, this may not be the case in other data sources. The changes in the curves when moving from the MNL to the PDP model are the net effect of two separate influences – the effect of accounting for value learning (Class 2); and the effect of accounting for strategic misrepresentation (Class 3). The overall effect on WTP depends on the magnitude of each of these effects, which are determined, in part, by the associated class probabilities.

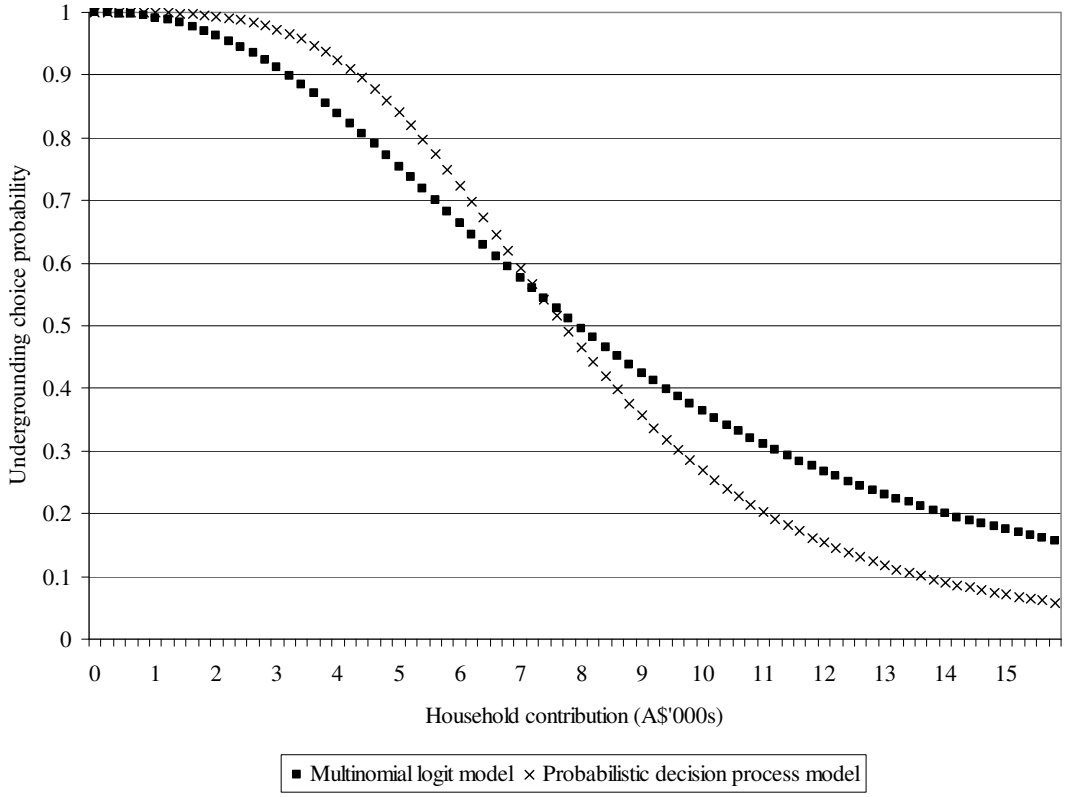

Figure 2: Undergrounding choice probability implied by models on binary choice format

The expected effect of accounting for value learning behaviour is an increase in probabilities at lower costs and a decrease in probabilities at higher costs. The reason is as follows. Average observed cost, z, is positively related to cost, x_1, for a given set of cost levels in previous choice tasks. Utility from undergrounding alternatives net of the effect of average observed cost therefore needs to be higher at lower cost levels and vice versa in order to adequately explain respondents' choices. The PDP model achieves this by altering the remaining parameters, $\boldsymbol{\beta}$. The effect is a narrowing of the distribution of total WTP with average observed cost held constant. Accounting for value learning has a relatively small effect on estimates of mean WTP in this case, because the mean cost level presented in the survey (approximately A$6,300) is similar to mean WTP.

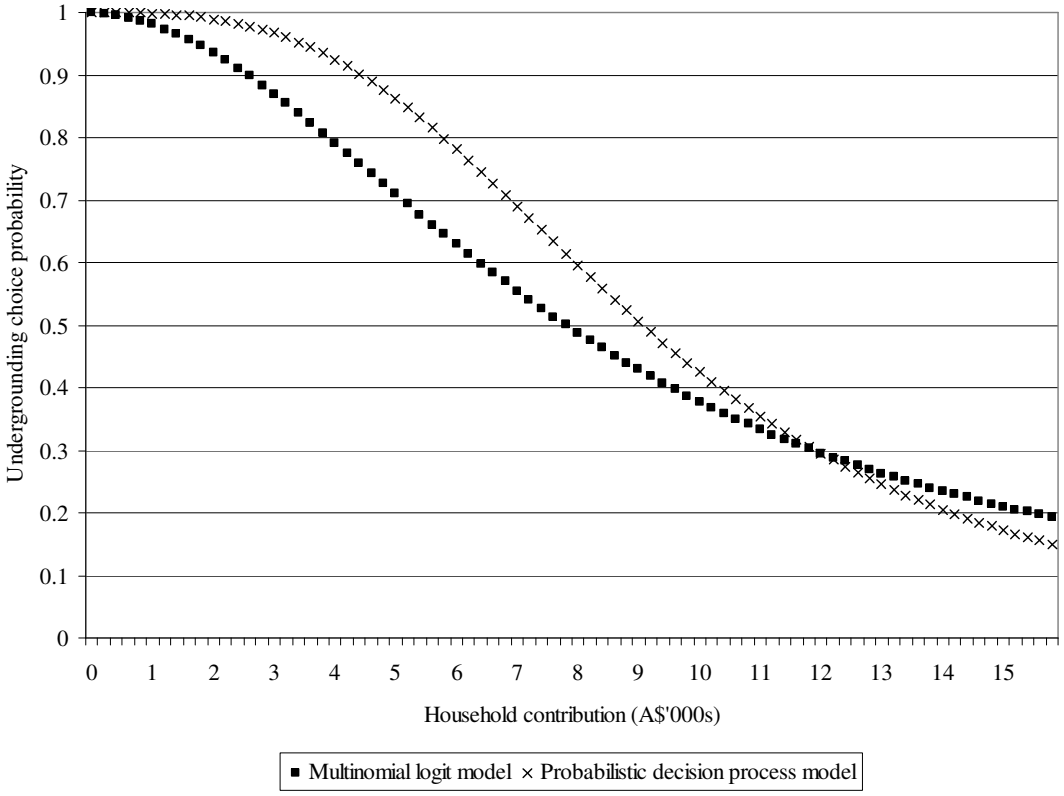

Figure 3: Undergrounding choice probability implied by models on multinomial choice format

The expected effect of accounting for strategic misrepresentation is an increase in undergrounding choice probability at all cost levels (albeit not in a linear fashion). For a given set of parameters, $\boldsymbol{\beta}$, the undergrounding choice probability for Class 3 is always less than or equal to that for Class 1 since $U_{it,SQ,class3} \geq U_{it,SQ,class1}$ and $U_{it,ALT,class3} \leq U_{it,ALT,class1}$.[15] Therefore, when switching from a Class 1 to a Class 3 utility specification, the parameters must be altered in such a way that increases the undergrounding choice probability. Accounting for strategic misrepresentation therefore leads to increased estimates of mean WTP, which is expected since strategic misrepresentation effectively involves an under stating of true WTP. Consistent with the relative mix of class probabilities, the strategic misrepresentation effect appears to be more dominant in Figure 3 (Model 2) than in Figure 2 (Model 1).

Conclusions

This paper presents a PDP model that can be used to identify heterogeneity in response behaviour towards a sequence of choice tasks. The illustrative evidence herein shows the model can be applied to choice data from both binary and multinomial choice formats where a status quo alternative is present and similar goods are offered at very different prices over the course of a sequence of questions.

[15] Note that in the first question in a sequence, the utility functions in Classes 1 and 3 are identical.

The PDP models achieved an improvement in fit over standard, single-class MNL models, even based on information criteria that account for model parsimony. Estimates of total WTP were statistically indistinguishable between the two types of model. However, this may not be the case in other data sources as it depends on several factors including the relative mix of class probabilities.

Three distinct groups were identified in both the binary and multinomial choice data. The group behaving in accordance with the standard assumptions was the smallest of the three in both models, providing further evidence that the standard assumptions do not adequately reflect the response behaviour of the majority of respondents in a survey of this type. The heterogeneity in response behaviour identified herein may explain the variation in findings across studies and the ambiguity of evidence within studies (Day and Pinto 2010, McNair et al. 2011) that have attempted to identify a single heuristic that best describes respondent behaviour towards a sequence of choice questions. It suggests that the literature may never converge to agreement on a single heuristic. The best way forward would appear to be to account for heterogeneity in response behaviour. The method presented in this paper is one approach that could be used in future studies. Further work is required to extend the approach to surveys in which cost is less dominant and to accommodate preference and scale heterogeneity, as well as information processing strategies. Clearly, other approaches are also possible and this is likely to be a fertile area for future research.

References

Araña J E, León C J and Hanemann M W (2008) Emotions and decision rules in discrete choice experiments for valuing health care programmes for the elderly. Journal of Health Economics 27: 753-769

Ariely D, Loewenstein D and Prelec D (2003) 'Coherent arbitrariness': stable demand curves without stable preferences. Quarterly Journal of Economics 118(1): 73-105

Bateman I J, Burgess D, Hutchinson W G and Matthews D I (2008a) Learning design contingent valuation: National Oceanic and Atmospheric Administration guidelines, preference learning and coherent arbitrariness. Journal of Environmental Economics and Management 55: 127-141

Bateman I J, Carson R T, Day B, Dupont D, Louviere J J, Morimoto S, Scarpa R et al (2008b) Choice set awareness and ordering effects in discrete choice experiments. CSERGE Working Paper EDM 08-01

Beenstock M, Goldin E and Haitovsky Y (1998) Response bias in conjoint analysis of power outages. Energy Economics 20: 135-156

Bennett J and Blamey R (2001) The Choice Modeling Approach to Environmental Valuation. Edward Elgar, Cheltenham, UK

Boyle K J, Bishop R C and Welsh M P (1985) Starting Point Bias in Contingent Valuation Bidding Games. Land Economics 61: 188-194

Bradley M and Daly A (1994) Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation 21(2): 167-184

Braga J and Starmer C (2005) Preference anomolies, preference elicitation and the discovered preference hypothesis. Environmental and Resource Economics 32: 55-89

Cameron T A and Quiggin J (1994) Estimation using contingent valuation data from a "dichotomous choice with follow-up" questionnaire. Journal of Environmental Economics and Management 27: 218-234

Carlsson F and Martinsson P (2008a) Does it matter when a power outage occurs? A choice experiment study on the WTP to avoid power outages. Energy Economics 30(3): 1232-1245

____ (2008b) How much is too much? An investigation of the effect of the number of choice sets, starting point and the choice of bid vectors in choice experiments. Environmental and Resource Economics 40(2): 165-176

Carson K S, Chilton S M and Hutchinson W G (2009) Necessary conditions for demand revelation in double referenda. Journal of Environmental Economics and Management 57(2): 219-225

Carson R T and Groves T (2007) Incentive and informational properties of preference questions. Environmental and Resource Economics 37: 181-210

Carson R T, Groves T and List J (2006) Probabilistic influence and supplemental benefits: a field test of the two key assumptions behind using stated preferences. Unpublished manuscript,

Caussade S, Ortuzar J d D, Rizzi L I and Hensher D A (2005) Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B 39: 621-640

Day B, Bateman I J, Carson R T, Dupont D, Louviere J J, Morimoto S, Scarpa R et al (2009) Task Independence in Stated Preference Studies: A Test of Order Effect Explanations. CSERGE Working Paper EDM 09-14,

Day B and Pinto J L (2010) Ordering anomalies in choice experiments. Journal of Environmental Economics and Management 59(3): 271-285

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

Hanemann W M, Loomis J and Kanninen B (1991) Statistical efficiency of double bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics 73: 1255-1263

Hensher D A and Collins A T (2010) Interrogation of Responses to Stated Choice Experiments: Is there sense in what respondents tell us? Submitted to Journal of Choice Modelling

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 Truong T P (1985) Valuation of Travel Time Savings: A Direct Experimental Approach. Journal of Transport Economics and Policy 19(3): 237-261

Herriges J A and Shogren J F (1996) Starting point bias in dichotomous choice valuation with follow-up questioning. Journal of Environmental Economics and Management 30: 112-131

Holmes T and Boyle K J (2005) Dynamic learning and context-dependence in sequential, attribute-based stated-preference valuation questions. Land Economics 81: 114-126

Hurwicz L (1972) On informationally decentralized systems. In: McGuire C B and Radner R (ed) Decision and organisation, North-Holland, Amsterdam

Kahneman D and Tversky A (1979) Prospect theory: An analysis of decisions under risk. Econometrica 47: 263-291

Ladenburg J and Olsen S B (2008) Gender-specific starting point bias in choice experiments: Evidence from an empirical study. Journal of Environmental Economics and Management 56: 275-285

Louviere J J and Hensher D A (1983) Using Discrete Choice Models with Experimental Design Data to Forecast Consumer Demand for a Unique Cultural Event. Journal of Consumer Research 10(December): 348-361

McFadden D (1974) Conditional Logit Analysis of Qualitative Choice Behaviour. In: Zarembka P (ed) Frontiers in Econometrics, Academic Press, New York

McNair B J, Bennett J and Hensher D A (2011) A comparison of responses to single and repeated discrete choice questions. Resource and Energy Economics (in press)

McNair B J, Bennett J, Hensher D A and Rose J M (2010) Households' willingness to pay for undergrounding electricity networks. Submitted to Energy Policy

Mirrlees J (1971) An exploration in the theory of optimal income taxation. Review of Economic Studies 38: 175-208

NERA Economic Consulting and ACNielsen (2003) Willingness to pay research study. A report for ACTEW Corporation and ActewAGL, September

Plott C R (1996) Rational individual behavior in markets and social choice processes: the discovered preference hypothesis. In: Arrow K, Colombatto E, Perleman M and Schmidt C (ed) Rational foundations of economic behavior, Macmillan, London

Samuelson P A (1954) The pure theory of public expenditure. Review of Economics and Statistics 36: 387-389

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

Scarpa R and Rose J (2008) Design Efficiency for Non-Market Valuation with Choice Modelling: How to Measure it, What to Report and Why. Australian Journal of Agricultural and Resource Economics 52(3): 253-282

Zellman E, Kaye-Blake W and Abell W (2010) Identifying consumer decision-making strategies using alternative methods. Qualitative Market Research 13: 271-286