Strategic response to a sequence of discrete choice questions

Ben McNair[1], Jeff Bennett[2] and David A. Hensher[3]

Paper presented at the 54th Annual Conference of the Australian Agricultural and Resource Economics Society, Adelaide, South Australia, 10-12 February 2010

Abstract According to neoclassical economic theory, the only stated preference elicitation format that can feasibly be employed in field studies to which truthful response can be the dominant strategy for all respondents is a single binary choice between the status quo and one alternative. In studies where the objective is estimation of preferences for multiple attributes of a good, it is preferred (and, in some cases, necessary) based on econometric considerations, to present respondents with a sequence of choice tasks. Economic theory predicts that utility-maximising respondents may find it optimal to misrepresent their preferences in this elicitation format. In this paper, the effect on stated preferences of expanding the number of choice tasks per respondent from one to four is tested using a split sample treatment in an attribute-based survey relating to the undergrounding of overhead electricity and telecommunications wires in the Australian Capital Territory. We find evidence to suggest that presenting multiple choice tasks per respondent decreases estimates of total willingness to pay and that this effect is related to the ordering of cost levels presented over the sequence of choice tasks. Two behavioural explanations can be advanced - a weak cost minimisation strategy, which implies divergence between stated and true preferences, and a 'good deal / bad deal' heuristic, in which stated preferences reflect true preferences that change over the course of the sequence of choice tasks. Preferences stated in the first of a sequence of choice tasks are not significantly different from those stated in the incentive compatible single binary choice task. A key objective of future research will be to establish whether this effect becomes less prevalent as the number of attributes and alternatives per choice task are increased.

Keywords Choice experiments, willingness to pay, incentive compatibility, strategic behaviour, order effects, underground electricity

[1] (presenting author) Crawford School of Economics and Government, The Australian National University, Canberra ACT 0200, Australia. Email address: ben.mcnair@anu.edu.au

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

[3] Institute of Transport and Logistics Studies, Faculty of Economics and Business, The University of Sydney, NSW 2006, Australia.

1. Introduction

Any attempt to achieve the social welfare maximising level of electricity network service quality relies upon estimates of the social valuation of changes in service quality. In the case of public provision, this social valuation forms one half of the classic Samuelson (1956) condition for optimal provision of a public good. In the case of private monopoly provision, this social valuation is used by the regulator to set quality incentive rates, which can result in the optimal social welfare outcome when applied as part of a quality-adjusted price cap with yardstick competition over both cost and quality (which is the culmination of theory developed by Spence (1975), Loeb and Magat (1979), Baron and Myerson (1982) and Shleifer (1985) among others). The literature on optimal private monopoly provision (termed 'the new regulatory economics' by Laffont and Tirole (1993)) is based on the use of mechanism design techniques (Green and Laffont 1979) to analyse regulation as a principal-agent problem where firms have private information about costs. Less attention has been paid in the regulatory economics literature to the fact that consumers have private information about their preferences and to the difficulties introduced by the need to elicit those preferences in order to estimate the social valuation of service quality. Regulators and firms are increasingly employing stated preference surveys to gather detailed information on consumers' preferences with respect to service quality (for example KPMG (2003) and Accent (2008, 2003)). The theory of mechanism design (Mirrlees (1971) and Hurwicz (1972)), which has been used extensively to analyse interactions between firms and regulators, can also be used to analyse whether utility-maximising consumers may find it optimal to misrepresent their preferences in such surveys.

Ideally, regulators would implement survey mechanisms in which truthful response is the dominant strategy for all respondents. That is, truthful response is the utility-maximising response for all respondents regardless of their beliefs about others' responses. Such a survey mechanism is said to be incentive compatible. It has long been recognised that a survey mechanism can be incentive compatible if its elicitation format is in the form of a single binary (SB) choice between the status quo and an alternative (Farquharson 1969).[4] A format comprising repeated binary choices between the status quo and various alternatives can only be incentive compatible where the social choice function is based on a single randomly selected choice task from the sequence (Carson and Groves 2007). This 'random selection' social choice function may be possible in a laboratory environment (Boyle et al. 2004), but in field surveys respondents are unlikely to believe that the agency would discard the majority of the data that they expended resources collecting. Our maintained assumption is that this is the case and therefore the SB elicitation format is the only format that can be incentive compatible in field surveys.

The SB format has successfully been employed where the price of the alternative varies across respondents, but the good is fixed. This is the form of contingent valuation (CV) survey recommended by the NOAA panel in 1993 (Arrow et al. 1993). However, there are a number of difficulties associated with employing the SB format when the good in the alternative varies across respondents. This is the form required to elicit preferences for multiple attributes, which is often the regulator's objective (for instance to estimate preferences for the frequency, duration, advance warning and time of day of electricity supply interruptions). Estimates of willingness to pay from SB data are less statistically significant than those from repeated choice data because of the absence of opportunities for institutional learning (Braga and Starmer 2005) as well as the much lower number of choice observations. Some evidence suggests that it may not be possible to estimate individual-specific taste intensities or the heterogeneity in taste intensities across a population using SB data.[5] For these reasons, formats used to elicit preferences for multiple attributes have tended to involve the presentation of multiple choice tasks per respondent– and in doing so, lose the property of incentive compatibility.

---

[4] A necessary condition is that the agency can credibly claim to be able to force any of the alternatives on any given respondent. However, it is not necessary for the SB choice survey to be binding (Carson et al. 1997) or a full public vote (Green and Laffont 1978).

[5] Estimating random parameters on the single binary choice data is problematic because the models may be unable to disentangle the Gumbel error distribution and the random parameter distributions. Rose et al (2009) found random parameter estimates statistically insignificant where data were a single choice observation per respondent in their study of the impact of the number of choice tasks per respondent.

This paper uses a split sample treatment of elicitation format in a web-based survey relating to the undergrounding of overhead electricity and telecommunications wires in the Australian Capital Territory to assess the effect on stated preferences of presenting multiple choice tasks per respondent. The elicitation formats employed in the survey include a SB choice task and a sequence of four binary choice tasks (RB).[6] The objectives of this paper are to use the data from these two elicitation formats to test:

a) whether stated preferences are affected by presenting four as opposed to one attribute-based choice task per respondent; and,

b) whether stated preferences in the first choice task presented are affected by advance knowledge that four as opposed to one choice tasks will be presented.

Where stated preferences are affected by elicitation format we would also like to identify the response strategies or other effects underlying the difference.

The paper is organised as follows. Section 2 discusses the effects of elicitation format on stated preferences identified in the literature both in theory and empirically. Section 3 sets out the design of the survey mechanism used in this study and the econometric modelling approaches used to analyse the data. The results of the analysis are set out in Section 4 and Section 5 concludes.

2. Elicitation format and stated preferences

There are a number of possible behavioural explanations for differences in preferences stated in a SB choice and a sequence of binary choices. One such explanation is that respondents employ a 'cost minimisation strategy' as predicted by neoclassical economic theory. It has long been recognised that consumers may conceal their true preferences if it enables them to obtain a public good at a lower cost (Samuelson 1954). In choice experiments, this strategy is generally thought to be manifest by the rejection of an

[6] The survey also employed a sequence of four choice tasks containing two alternatives to the status quo (RMN). Data from this format are not analysed in this paper.

alternative that is preferred to the status quo when a similar good was offered at a lower cost in a previous choice task. In doing so, respondents increase the likelihood that their most preferred option across the sequence of choice tasks is implemented. Bateman et al. (2008) differentiate between a 'strong' case, in which respondents always reject a good if it was offered at a lower cost in a previous choice task, and a 'weak' case, in which respondents weigh up this rejection against the perceived risk of the good not being provided at the lower cost. These strategies imply that preferences stated in a sequence of choice tasks may diverge from true underlying preferences.

Another explanation for divergence between preferences stated in SB and sequential choice formats is that respondents discover their preferences as they progress through a sequence of choice tasks (Plott 1996). Bateman et al. (2008) describe a 'good deal / bad deal' heuristic in which respondents revise their preferences on the basis of the cost levels presented as they progress through the sequence of choice tasks. An alternative is more (less) likely to be chosen if its price level is low (high) relative to the levels presented in previous choice tasks. This heuristic could arise where respondents take the average of cost levels presented in the sequence to that point as a signal for the quality of the good. The key difference between this and the cost minimisation strategies is that the cost minimisation strategies assume respondents hold constant, well-formed preferences. The offering of a high-cost alternative would increase the likelihood of acceptance in subsequent choice tasks under the 'good deal / bad deal' heuristic, but not under the strong cost minimisation strategy.[7] According to the 'value learning' and 'good deal / bad deal' heuristic explanations, preferences stated in a sequence of choice tasks do not diverge from true underlying preferences.

A potential consequence of presenting a sequence of choice tasks containing similar goods with large variations in cost levels is that respondents may find some alternatives

[7] In the weak cost minimisation strategy the perceived risk of the good not being provided depends upon the cost levels in the chosen alternatives in the sequence to that point. The offering of a high-cost alternative that is rejected would result in a higher likelihood of acceptance in subsequent choice tasks than the offering of a lower-cost alternative if that alternative is accepted and has the highest cost level of all alternatives chosen in the sequence to that point.

implausible and answer the question as though cost were at a level considered more realistic by the respondent. Bateman et al. (2008) refer to this as a cost averaging strategy. Under such a strategy we would expect alternatives with cost levels from the low (high) end of the range observed by the respondent to be accepted less (more) frequently than in a truthful response. This strategy implies that preferences stated in a sequence of choice tasks diverge from the true underlying preference for the levels actually set out in the choice tasks.

Presenting a sequence of choice tasks also affords respondents with opportunities to learn and become more familiar with the choice task format. Braga and Starmer (2005) refer to this as an institutional learning process in which responses become more accurate as the sequence progresses. This learning process is thought to have the effect of decreasing the variance of the random error component in choice models (or equivalently increasing scale) (Holmes and Boyle 2005). There is also evidence to suggest that respondents may become fatigued and respond less accurately once they proceed beyond a certain point in a survey (Bradley and Daly 1994, Caussade et al. 2005). Fatigue is thought to have the effect of increasing error variance (or equivalently, decreasing scale).

Several studies have tested the effects of expanding the SB format in the fixed good (CV) context. Recognising the large sample sizes required for statistically significant estimation when using the SB choice format, some CV surveys have incorporated a second (or follow-up) question in the elicitation format. A number of studies have found differences in the WTP implied by the first and second questions in this double-bounded CV format (Cameron and Quiggan 1994, Hanemann et al. 1991, McFadden and Leonard 1995). Herriges and Shogren (1996) interpret this difference as starting point bias. Carson et al. (2008) relate the difference to strategic response by showing that responses to the first and second questions are equivalent in the presence of a social choice function in which the outcome of the second question cancels out and replaces the outcome of the first question.

The attribute-based choice experiment format has not been subject to the same degree of testing for strategic response as the CV format.[8] Some studies have compared stated preferences from fixed-good (CV) SB and attribute-based repeated binary formats. For example Cameron et al. (2002) were unable to reject the hypothesis of identical indirect utility-difference functions across these elicitation formats. A number of studies have examined the implications of presenting multiple attribute-based choice tasks per respondent without employing an incentive compatible SB comparator. Carlsson and Martinsson (2006) found no evidence of starting point bias in their split sample treatment of inclusion of a 'good deal' alternative (one with a large improvement in the good at a low cost) in the first choice task of a sequence. Bateman et al. (2008) found evidence of a weak cost minimisation strategy using a split sample treatment of advance knowledge of attribute levels in a sequence of choice tasks. They used the first choice task in a sequence as an incentive compatible comparator where respondents had not been informed that they would be presented with multiple choice tasks. This relies on the assumption that respondents assumed with certainty that the first choice task would be the only choice task presented. If respondents had any uncertainty as to whether this would be the case, then the necessary 'take it or leave it' property of the incentive compatible SB choice is violated.

None of the field studies discussed above make spilt-sample comparisons between preferences stated in a sequence of choice tasks and those stated in an incentive compatible attribute-based SB choice task. We have found only two such studies. The

---

[8] A number of studies have focussed on hypothetical bias by comparing results from hypothetical choice experiments with those from choice experiments with immediate and certain implementation (Alfnes and Steine 2005, Carlsson and Martinsson 2001, Hensher 2009, Lusk and Schroeder 2004). Carson and Groves (2007) distinguish between inconsequential hypothetical surveys (where a respondent believes there is 0 per cent chance of implementation) and consequential hypothetical surveys (where a respondent believes their responses will influence up to some non-zero probability the likelihood of an alternative being implemented by the agency). The same conditions for incentive compatibility apply to the survey mechanism regardless of whether it is a consequential hypothetical survey or a survey with immediate and certain implementation. If the survey is inconsequential, then neoclassical economic theory cannot be used to predict responses. Consistent with this theory, Carson et al. (2006) found a difference between responses to inconsequential hypothetical questions and questions involving 100 per cent probability of actual payment, but, importantly, found equivalence in responses to all questions involving a non-zero (20 per cent, 50 per cent, 80 per cent and 100 per cent) probability of actual payment.

first is a study by Racevskis and Lupi (2008), which found a significant difference between models fitted to data collected from single and repeated binary attribute-based choice tasks. The second is a study by Scheufele and Bennett (2010a), which is being conducted concurrently with this study. This paper adds to this small, but growing, body of research. It builds on previous research by modelling the response strategies employed by respondents when presented with a sequence of choice tasks. Bateman et al. (2008) present models that allow cost sensitivity to change over the course of a sequence of choice tasks by including in their random parameter logit models an interaction between question order and cost. In this paper, we present a more flexible model of the effect of question order on cost sensitivity and develop new models that allow cost sensitivity to change according to the positioning of the cost level relative to levels presented in previous choice tasks. A number of authors, including Bateman et al. (2008) and Carson and Groves (2007), have discussed the potential effects of cost levels presented in previous choice sets, but, to our knowledge, no studies have modelled these effects. We aim to fill this research gap in this paper.

3. Research design and method

The empirical testing was carried out 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.[9] Until around 1990, electricity and telecommunications networks in the ACT were installed as overhead wires supported by poles. Since that time, underground networks have become the accepted service standard in new developments due to a number of advantages over overhead networks. Fires, high winds, ice storms, lightning and other severe weather events can damage overhead networks leading to extended power outages and risks of

[9] Further analysis of these data is to take place, including identification of the appropriate point estimate of mean willingness to pay, before completion of the economic evaluation of an undergrounding program in the ACT.

electrocution by members of the public. The supply reliability of overhead networks is also affected by vegetation coming into contact with power lines. Underground networks lead to more aesthetically pleasing residential areas and allow unobstructed views. Other household benefits include the avoided costs of trimming trees away from power lines and increased flexibility in the use of residential yard space. At the project area level, undergrounding can be analysed as a public good due to the indivisibility of provision. However, we note that the benefits conferred on homeowners in a project area can be reflected in higher property values (McNair 2009). The good therefore has a private element in that it is a property characteristic that can be traded in the property market as part of a bundle of characteristics, albeit subject to high transaction costs. As a consequence, there is a possibility that respondents answered questions not only with their own preferences in mind, but also the preferences of others in the form of a perceived property market value. Almost all participants in pre-testing interviews stated that they did not consider any property value impact when completing the choice tasks. Only 4 per cent of respondents to the main survey answered ‘yes’ to the question, ‘Were any of your choices influenced by what you think other respondents would choose?’ We expect that stated preferences were not greatly influenced by perceived property market value, but this cannot be tested. We therefore make a maintained assumption in the analysis that follows that the preferences underlying responses are not affected by property market considerations. We note that our analysis of the effects of elicitation format on stated preferences does not rely upon this assumption. The assumption relates to the underlying (or initial) demand that is formed before respondents begin completing the choice task(s). The opportunities to cost-minimise or revise this demand over the course of a sequence of choice tasks remain.

The survey employed a hybrid stated preference methodology, combining the attribute-based approach of choice experiments with the project-based dichotomous choice approach of contingent valuation. Three elicitation formats were used in the survey – a single binary choice task (SB), a sequence of four binary choice tasks (RB) and a sequence of four choice tasks containing two alternatives to the status quo (RMN). In each choice task, respondents were presented with a description of their current service and either one or two undergrounding options. Each choice alternative was described in terms of the attributes in Table 1. All of the benefits of undergrounding other than supply reliability benefits are embodied in the *Type* attribute. This includes the amenity and safety benefits that are generally thought to be the major household benefits from undergrounding.

Table 1: Attributes used to describe alternatives in choice tasks

Attribute description in choice tasks
Type of infrastructure (underground or overhead)
Power cuts without warning:
Number of power cuts each 5 years
Average duration of power cuts
Power cuts with 7 days written notice (occurring in normal business hours):
Number of power cuts each 5 years
Average duration of power cuts
Your one-off undergrounding contribution[10]

In the status quo alternative, the 'type of infrastructure' (*Type*) attribute was set at the 'overhead' level and the cost attribute was set at the level $0. Respondents were presented with default supply reliability attribute levels for the status quo and given the opportunity to adjust them to fit with their own experience. The *Type* attribute was set to the 'underground' level for all change (non status quo) alternatives in the design. This ensured that every alternative in the design was meaningful as a SB choice, while allowing the same set of alternatives to be used in all three elicitation formats. The supply reliability levels in the undergrounding alternatives were calculated as a proportion of the status quo level: 'number of power cuts without warning' (0.25, 0.5, 0.75 and 1), 'duration of power cuts without warning' (0.33, 0.66, 1.33, 1.66), 'number of power cuts with notice' (0.2, 0.4, 0.6, 0.8) and 'duration of power cuts with notice' (0.33, 0.66, 1.33, 1.66). Where respondents chose very low status quo levels (1 or less) for the power cut frequency attributes absolute levels were assigned (0, 1 and 2). The cost attribute took 16

[10] Payable either up-front with a 3 per cent discount or in instalments for up to 5 years at a 6.5 per cent per annum interest rate.

levels. Eight were assigned ($1000, $2000, $3000, $4000, $6000, $8000, $12,000 and $16,000) and a further eight were anchored on these levels as part of the experimental design (-$200, -$100, +$100 and +$200).

Eight choice tasks were designed in the RMN elicitation format to maximise the C-efficiency (Scarpa and Rose 2008) of the design using Bayesian priors (derived from pre-testing responses and NERA and ACNielsen (2003)) assuming the default supply reliability levels for the status quo. This approach maximises the statistical significance of the least significant WTP estimate across the attributes of interest (assuming a standard multinomial logit model). The RMN design was used because it was expected that estimates of WTP for supply reliability, which were less statistically significant in the design than the *Type* and cost attributes, would rely heavily on data from that elicitation format. These eight choice sets were blocked into 2 sequences of 4 choice tasks. Each respondent in the RMN sample split received one of these sequences. The RB design was created by splitting each of these sequences into two new sequences giving 4 sequences of 4 choice tasks with only one alternative to the status quo. That is, each of the 16 different (non-status-quo) alternatives present in the RMN design represented a binary choice task in the RB design. Each respondent in the SB sample split received one of these 16 choice tasks. This could be thought of as extreme blocking of the RB design.

The web-based questionnaire was refined based on in-depth interviews with a total of 11 participants. Households were recruited by telephone and screening questions were used to ensure that participating households were owner-occupiers of stand-alone houses serviced by overhead wires.[11] Email invitations were sent to the 2,485 households that agreed to participate. The invitation included some background information on the research and a URL and unique password for accessing the online questionnaire. 1,744 respondents completed the online questionnaire (1,163 in SB and 292 in RB). The questionnaire provided background information on undergrounding before asking respondents to identify the most important benefits and disadvantages of undergrounding

[11] This is the only group that face both the benefits and the costs of both undergrounding and reliability improvements.

to their household. After establishing individual-specific reference levels for power supply reliability attributes, the questionnaire advised respondents of the number of choice tasks that would be presented, the number of alternatives that would be presented in each task and the attributes that would be used to describe each alternative.[12] The questionnaire outlined a suburb-based majority rule social choice function (often referred to as a provision rule or decision rule in the non-market valuation literature) that ensured incentive compatibility in the SB response format. In the RB format the equivalent social choice function was that an undergrounding option would be considered for implementation in a suburb if it was preferred to the status quo by more than 50 per cent of respondents in that suburb. Importantly, the survey program did not allow respondents to navigate back through the sequence of choice tasks. The survey was programmed to cycle through the various blocks and elicitation format sample splits to ensure approximately equal representation across choice observations. The final sections of the questionnaire comprised questions about information processing and the socio-demographic characteristics of the respondent and their household.

Three different modelling approaches are used to analyse the effect of presenting a sequence of binary choice tasks (RB) as opposed to a SB choice. The first is the binary logit model, which estimates the effect of elicitation format on the probability of choosing the undergrounding option in a given choice task. This approach allows examination of bid acceptance curves and total willingness to pay (TWTP) in line with standard analysis of single binary choice data in the literature on CV referenda. The second and third are the multinomial logit (MNL) model and the panel random parameter logit (RPL) model, which are used to estimate marginal willingness to pay (MWTP) for the various attributes of the electricity network service. These models are commonly used to analyse data from surveys in which respondents are presented with a sequence of attribute-based choice tasks (choice experiments or conjoint analysis). The RPL model has been preferred to the MNL model in the recent choice experiment literature due to its generality and its ability to estimate heterogeneity in preferences across the population.

[12] The questionnaire did not include a practice or learning choice task.

However, there is some doubt as to whether a single observation per respondent is sufficient to enable the separate estimation of random parameter and error distributions that characterises the RPL model. We therefore present evidence on TWTP and MWTP from both MNL and RPL models.

All of these models are based on random utility theory, which is built on the assumption that the utility, U, derived by a respondent from an alternative is a function of the attributes of the alternative, the characteristics of the respondent as well as unobserved individual heterogeneity (in the case of the RPL model) and a random element, ε. In any given choice task, respondents choose the alternative that yields the highest utility. The outcome is an index of the observed choice, y. The utility that respondent i derives from alternative j in choice task t is

$$U_{ijt} = \beta_i'\mathbf{x}_{ijt} + \delta_i'\mathbf{z}_{it} + \varepsilon_{ijt}.$$

where $\mathbf{x}_{ijt}$ is a vector of observed variables, $\mathbf{z}_{it}$ is a set of choice invariant characteristics (potentially socio-demographic characteristics of the respondent or characteristics of the choice task such as its position order in any sequence) and β_i and δ_i are vectors of coefficients to be estimated. The models presented in this paper assume ε to be independently and identically distributed (i.i.d.) according to the extreme value type I function. In the case of a binary choice between the status quo and one alternative, the probability that respondent i chooses the alternative (y=1) rather than the status quo (y=0) in choice task t is

$$\Pr(y = 1 \mid x) = \frac{\exp(\beta' x_{it})}{1 + \exp(\beta' x_{it})}$$

where the components of x are defined as the difference in attribute levels between the alternative and the status quo. In the standard multinomial logit (MNL) model, the probability that respondent i chooses alternative j in choice task t is

$$\Pr(j) = \frac{\exp(\beta' x_{ijt})}{\sum_{m=1}^{J} \exp(\beta' x_{imt})}$$

and in the case of the random parameter logit (RPL) model, the probability is

$$\Pr\left(j \mid v_i\right) = \frac{\exp\left(\beta_i' x_{ijt}\right)}{\sum_{m=1}^{J} \exp\left(\beta_i' x_{imt}\right)} \quad \text{where } \beta_i = \beta + \Delta z + \Gamma v_i$$

Γ is a diagonal matrix containing on its diagonal σ_k, the standard deviations of the marginal distributions of β_i, and v_i is distributed with mean 0 and standard deviation 1 according to the distribution specified by the analyst. Correlation between the multiple observations from each respondent (panel data) is accommodated by incorporating in the log-likelihood function the probability of respondent *i*'s observed sequence of choices, which is the product of the probabilities for each choice task in the sequence.

To test whether stated preferences are affected by presenting four as opposed to one attribute-based choice task per respondent, we compare estimates of TWTP from basic binary logit, MNL and (if possible) RPL models estimated on data from the SB format with equivalent models estimated on data from the RB format. We also compare estimates of MWTP from the MNL and RPL models where possible. To identify behavioural explanations for any differences, we examine three additional variable specifications for models on data from the RB format. The first incorporates effects coded variables for the order in which choice tasks were presented as described in Table 2.[13] In the binary logit model, these variables enter the bid acceptance function directly. In the MNL and RPL models, these are interacted with the cost variable. Bateman et al. (2008) used an interaction between cost and log of question order to show that cost sensitivity increases over the course of a sequence of choice tasks. Our specification allows for a non-monotonic relationship between cost sensitivity and question order. We also examine the resulting relationship between question order and WTP, where MWTP for attribute *x* is:

$$MWTP(x) = -\beta_x \,/\, (\beta_{cost} + q1.\beta_{q1*cost} + q2.\beta_{q2*cost} + q3.\beta_{q3*cost})$$

[13] The order in which choice tasks were presented to respondents was cycled so that each choice task was approximately equally represented in each order position over the population.

To test whether advance knowledge that multiple choice tasks will be presented affects stated preferences in the first of a sequence of choice tasks, we compare WTP estimates derived from the basic models on data from the SB format with WTP estimates from the model with question order variables on the RB data, where, consistent with the formula above, MWTP for attribute *x* evaluated at the first question in the sequence is:

$$MWTP(x) = -\beta_x / (\beta_{cost} + \beta_{q1*cost})$$

In the second extended model, we incorporate two variables in addition to the effects coded question order variables. The first is an interaction between cost and a (1,-1) variable indicating whether the cost level is the minimum presented to the respondent in the sequence to that point (*Min*). The second is a similar interaction between cost and a (1,-1) variable indicating whether the cost level is the maximum presented to the respondent in the sequence to that point (*Max*). We use this model to examine whether any relationship between cost sensitivity and question order can be explained by the positioning of the cost level relative to the levels presented to the respondent in previous choice tasks. A positive parameter estimate on the *Min* interaction indicates that cost sensitivity is lower (WTP is higher) when the cost level being presented is the lowest presented in the sequence to that point (with all other variables, including cost, held constant). A negative parameter estimate on the *Max* interaction indicates that cost sensitivity is higher (WTP is lower) when the cost level being presented is the highest presented in the sequence to that point (with all other variables, including cost, held constant). The signs on the *Min* and *Max* interactions implied by the various response strategies discussed above are presented in Table 3.

Table 2: Effects coding of question order variables

Variable	*q1*	*q2*	*q3*
Level in question 1	1	0	0
Level in question 2	0	1	0
Level in question 3	0	0	1
Level in question 4	-1	-1	-1

The third extended model includes effects coded variables accounting for the four possible 'relativities' for the cost level presented in a choice task as described in Table 4.[14] In any given choice task, the cost level presented must be either:

a) both the minimum and the maximum level presented in the sequence to that point (*m11*);

b) the minimum, but not the maximum level presented in the sequence to that point (*m10*);

c) the maximum, but not the minimum level presented in the sequence to that point (*m01*); or

d) neither the minimum nor the maximum level presented in the sequence to that point (*m00*).

Table 3: Parameter signs and relationships implied by response strategies

Response strategy	$\beta_{min*cost}$	$\beta_{max*cost}$	$\beta_{m10*cost}$, $\beta_{m11*cost}$, $\beta_{m00*cost}$, $\beta_{m01*cost}$
Strong cost minimisation	+	= 0	$\beta_{m10*cost}=\beta_{m11*cost}>\beta_{m00*cost}=\beta_{m01*cost}$
Weak cost minimisation	+	? [15]	$\beta_{m10*cost}>\beta_{m00*cost}$ and $\beta_{m11*cost}>\beta_{m01*cost}$
Good deal / bad deal heuristic	+	-	$\beta_{m10*cost}>\beta_{m11*cost}>\beta_{m01*cost}$ and $\beta_{m10*cost}>\beta_{m00*cost}>\beta_{m01*cost}$
Cost averaging	-	+	$\beta_{m10*cost}<\beta_{m11*cost}<\beta_{m01*cost}$ and $\beta_{m10*cost}<\beta_{m00*cost}<\beta_{m01*cost}$
Truthful (with stable preferences)	= 0	= 0	$\beta_{m10*cost}=\beta_{m11*cost}=\beta_{m00*cost}=\beta_{m01*cost}=0$

[14] The order variables are omitted from this model since the *q1* and *m11* variables are essentially identical (since the cost level in first choice task presented is always both the minimum and maximum presented to that point and this is not possible at any other point in the sequence).

[15] It can be shown using simple examples that, for a given cost in the current choice task, acceptance can be more likely or less likely when the cost in the current choice task is the maximum presented in the sequence to that point. The likelihood depends on the level of the maximum cost accepted in the sequence to that point, which depends on the cost levels presented in previous choice tasks as well as the preferences of the respondent. It is difficult to differentiate between the effects of a weak cost minimisation strategy and a 'good deal / bad deal' heuristic in this study as our experimental design does not allow a scope test such as that conducted by Bateman et al. (2008).

The full sample of RB data contained 289, 299, 297 and 271 choice observations for the *m11*, *m10*, *m01* and *m00* cost level relativities, respectively. In the binary logit model, these variables enter the bid acceptance function directly. In the MNL and RPL models, these are interacted with the cost variable. This allows estimation of the relationship between cost sensitivity and the positioning of the cost level relative to the levels presented to the respondent in previous choice tasks. If the parameter estimate for the *m00* interaction is significantly higher than that for the *m01* interaction, this indicates that cost sensitivity is lower (and WTP is higher) when the cost level is within the range of levels presented in previous choice tasks relative to when it is the highest level presented in the sequence to that point (with all other variables, including cost, held constant). This implies, for example, that an alternative with a cost level of $4,000 is more likely to be chosen if previously presented cost levels were $2,000 and $6,000 than if they were $2,000 and $1,000. The relationships between the cost relativity interactions implied by various response strategies are presented in Table 3.

Table 4: Effects coding of cost relativity variables

Variable	*m11*	*m10*	*m01*
Level when cost is both minimum and maximum presented in the sequence to that point	1	0	0
Level when cost is minimum, but not maximum presented in the sequence to that point	0	1	0
Level when cost is maximum, but not minimum presented in the sequence to that point	0	0	1
Level when cost is neither minimum nor maximum presented in the sequence to that point	-1	-1	-1

We also examine the resulting relationship between cost relativity and WTP, where MWTP for attribute x is:

$$MWTP(x) = -\beta_x \,/\, (\beta_{cost} + m11.\beta_{m11*cost} + m10.\beta_{m10*cost} + m01.\beta_{m01*cost})$$

The *m11* cost relativity occurs only in the first question in a sequence and the first question can only take only this cost relativity. Therefore, we can further test whether advance knowledge that multiple choice tasks will be presented affects stated preferences in the first of a sequence of choice tasks by comparing WTP estimates derived from the basic models on data from the SB format with WTP estimates from the model with cost relativity variables on the RB data, where, consistent with the formula above, MWTP for attribute *x* evaluated at the first question in the sequence is:

$$MWTP(x) = -\beta_x / (\beta_{cost} + \beta_{m11*cost})$$

4. Results

The binary logit model results are presented in Table 5. Data are excluded where respondents took less than 5 minutes to complete the SB survey or less than 6 minutes to complete the RB survey. It is expected that these questions were answered without consideration (possibly randomly) solely as a means of qualifying for the prize draw participation incentive. The basic models on the SB and RB formats (Models 1 and 2, respectively) include a constant, the log of the contribution amount, supply reliability attributes and effects coded (-1,1 or -1,0,1) variables for the respondent's sex, age, household income and exposure to media coverage of the issue of undergrounding in Canberra. Other socio-demographic characteristics were tested and omitted from the models after they were found to be statistically insignificant.

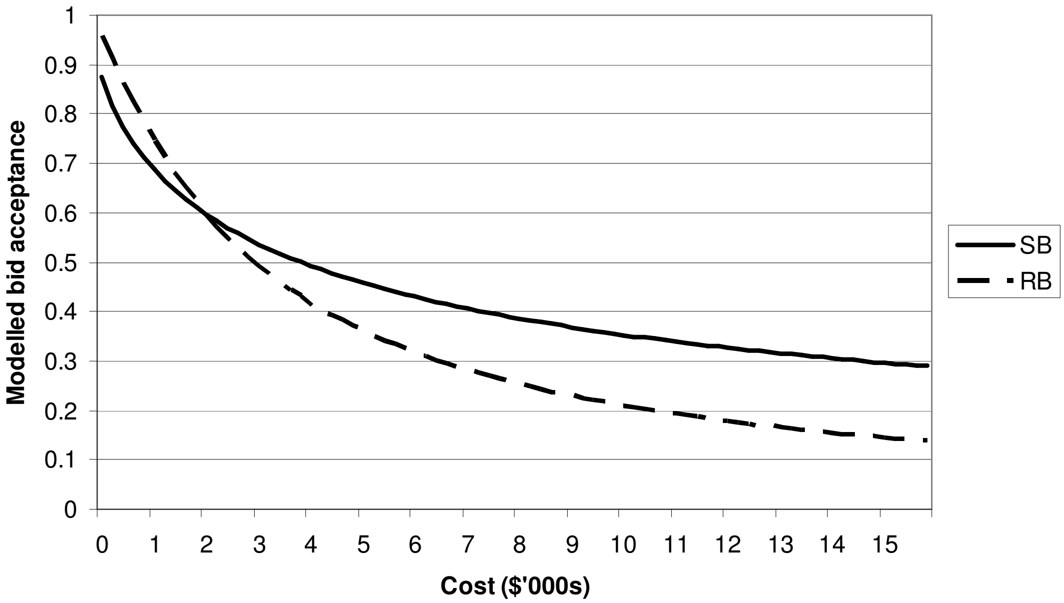
The bid acceptance curves derived from Models 1 and 2 with non-cost variables set at their population means are shown in Figure 1. Bid acceptance is significantly lower in the RB format relative to the incentive compatible SB format for all cost levels except those at the lower end of the range used in the design. This is consistent with predictions based both on the cost minimisation strategy (both the strong and weak cases) and on the 'good deal / bad deal' heuristic. The bid acceptance rate is at least as high in the RB format for choice tasks with cost levels at the lower end of the range used in the design. The effect on the shape of the bid acceptance curve from presenting multiple choice tasks per respondent is the opposite outcome to that predicted by the cost averaging strategy, which was lower bid acceptance at low cost levels and higher bid acceptance at high cost levels relative to the incentive compatible response.

Table 5: Summary of results from binary logit models

Model	Model 1	Model 2	Model 3	Model 4	Model 5
Response format	SB	RB	RB	RB	RB
Parameter estimates:					
Constant	.65063*** (.16489)	.91712*** (.17105)	.92918*** (.17199)	.67703*** (.18144)	.33437 (.21173)
Log of household contribution ($'000s)	-.64969*** (.07768)	-1.1276*** (.08734)	-1.1404*** (.08806)	-.90854*** (.10195)	-.77625*** (.11441)
Change in number of unplanned outages per 5 years	-.04957 (.05533)	-.04692 (.04197)	-.04669 (.04218)	-0.07938* (0.04448)	-.06986 (.04346)
Change in unplanned minutes off supply per 5 years	-.00061 (.00055)	-.000044 (.00046)	.000006 (.00046)	.00014 (.00047)	.00011 (.00047)
Change in planned minutes off supply per 5 years	-0.00034* (0.0002)	-.00062*** (.00020)	-.00064*** (.00020)	-.00052*** (.00019)	-.00053*** (.00019)
Gender (male=1)	.05860 (.06481)	.13980* (.07168)	.14024* (.07190)	.13954* (.07251)	.13319* (.07262)
Age: young (<40=1)	-.12385 (.12163)	-.32247** (.12775)	-.32272** (.12809)	-.32396** (.12852)	-.32022** (.12878)
Age: old (>65=1)	.05994 (.12577)	.32288** (.13863)	.32375** (.13888)	.31080** (.14099)	.30544** (.14109)
Household income: lower (<$52,000 pa =1)	-.18730 (.14051)	-.41362** (.16680)	-.42543** (.16818)	-.39958** (.16883)	-.41485** (.16887)
Household income: upper (>$182,000 pa =1)	.01561 (.10678)	-.00160 (.12240)	.00413 (.12331)	.01748 (.12383)	.02872 (.12407)
Exposure to media (yes=1)	.16884** (.06805)	.11986 (.07352)	.12153 (.07396)	.14361* (.07491)	.15400** (.07514)
Order: question 1 (q1=1)			.29826** (.11844)	.06355 (.19155)	
Order: question 2 (q2=1)			-.12053 (.11917)	-.09576 (.12346)	
Order: question 3 (q3=1)			-0.20429* (0.12023)	-.10884 (.13847)	
Minimum cost in sequence to that point (minimum=1) * log of contribution				.22848*** (.06435)	
Maximum cost in sequence to that point (maximum=1) * log of contribution				-.13564** (.06211)	
Relativity: M11 (minimum and maximum cost in sequence to that point =1)					.35528*** (.11686)
Relativity: M10 (minimum, but not maximum cost in sequence to that point =1)					.49496*** (.14831)
Relativity: M01 (maximum, but not minimum cost in sequence to that point =1)					-.87863*** (.17496)
Model fit:					
Observations	1090	1112	1112	1112	1112
Log-likelihood	-705	-631	-627	-616	-616
Information criterion AIC	1.31	1.15	1.15	1.14	1.13

*, ** and *** indicate statistical significance at the 0.1, 0.05 and 0.01 levels respectively; standard errors are in parentheses.

Figure 1: Bid acceptance curves derived from Models 1 and 2



The descriptive statistics for mean TWTP estimates derived from these models are presented in Table 7.[16] Point estimates for mean TWTP are $6,916 and $5,362 in the SB and RB models, respectively, calculated as the area under the bid acceptance curve truncated at the $16,000 cost level with all non-cost variables set at their population means. The result of a test to establish whether TWTP is significantly lower in the RB format than in the incentive compatible SB format is presented in Table 6. The null hypothesis that mean TWTP in the SB format is less than or equal to that in the RB format is rejected at the .10 level, but not at the .05 level.

Table 6: Tests for differences in TWTP

Null hypothesis (H_0)	$TWTP_{Model\ 1} \leq TWTP_{Model\ 2}$	$TWTP_{Model\ 1} = TWTP_{Model\ 3:q1}$	$TWTP_{Model\ 1} = TWTP_{Model\ 5:m11}$
p-value	0.0993	0.6099	0.8512

[16] Confidence intervals and hypothesis testing were based on a bootstrapping procedure with 1000 random draws of WTP (calculated by drawing from normal distributions for all relevant parameters with moments set at their means and standard errors).

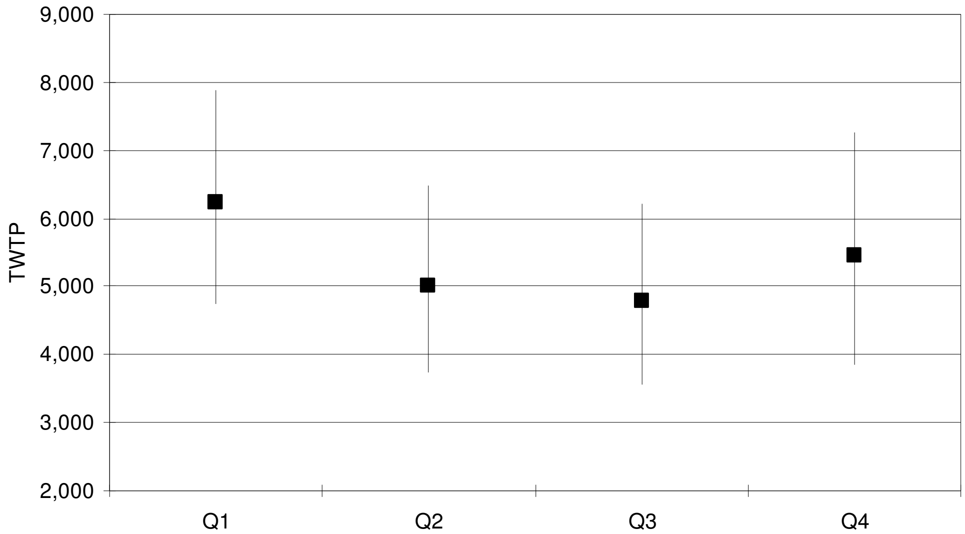
Table 7: Estimates of TWTP derived from binary logit models ($2009)

	Mean	90 per cent CI
Model 1	6,916	5,386 - 8,481
Model 2	5,362	4,120 - 6,752
Model 3: Q1	6,233	4,736 - 7,879
Model 3: Q2	5,013	3,740 - 6,485
Model 3: Q3	4,786	3,551 - 6,207
Model 3: Q4	5,459	3,855 - 7,262
Model 5: M11	6,624	4,667 - 8,708
Model 5: M10	7,104	5,008 - 9,344
Model 5: M01	3,094	1,936 - 4,523
Model 5: M00	5,595	3,546 - 8,033

Model 3 incorporates into the RB model effects coded variables for the order in which choice tasks were presented to the respondent. The parameter estimate for the *q1* variable is positive and significantly higher than the parameter estimates for the *q2* and *q3* variables. The coding of the variables, which is set out in Table 2, means that the implicit parameter estimate for *q4* is the negative of the sum of the parameter estimates for *q1*, *q2*, and *q3*, which is 0.02656. The parameter estimates indicate that bid acceptance is significantly higher in the first question relative to the later questions in the sequence (with all other variables, including cost, held constant). The modelled relationship between question order and TWTP (mean and 90 per cent confidence interval) is presented in Figure 2. The point estimate of TWTP is highest in the first choice task presented. It decreases over the second and third choice tasks before increasing at the fourth and final choice task. This is similar to the finding of Bateman et al. (2008) that cost sensitivity increases over a sequence of choice tasks.

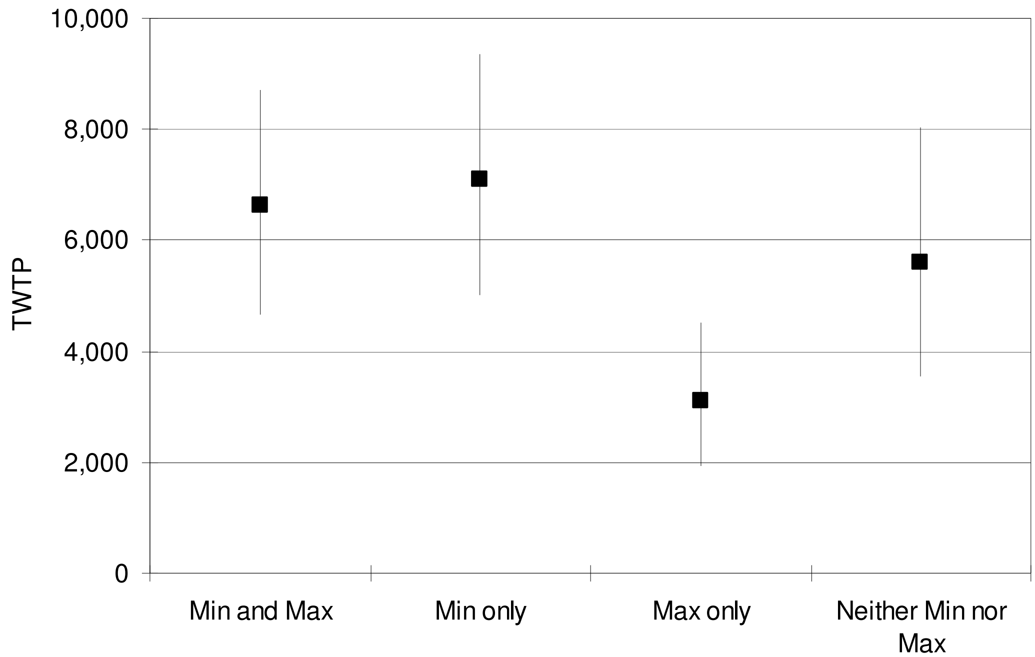
The model predicts mean TWTP in the first choice task at $6,233, which is slightly lower than the point estimate of mean TWTP from the incentive compatible SB format of $6,916. Table 6 presents a test of the statistical significance of this difference. We fail to reject statistical equivalence of these two estimates with a p-value of 0.61. That is, we find no evidence to suggest that advance knowledge that multiple choice tasks would be presented has an effect on stated preferences in the first choice task.

Figure 2: TWTP by question order (estimated in Model 3)



The question order variables are retained in Model 4 and *Min* and *Max* interactions are added. *Min* and *Max* are (1,-1) variables indicating whether the cost level is the minimum and maximum, respectively, presented to the respondent in the sequence to that point. Both of these interactions are statistically significant at the 0.05 level and the AIC value suggests the explanatory power of the model is improved. All of the effects coded variables for question order become insignificant indicating that there is no significant residual order effect once we account for the effect of cost levels presented in earlier choice tasks. This is a key result because it indicates that other possible interpretations for order effects, such as learning and fatigue effects, do not appear to be affecting bid acceptance (and therefore TWTP) in this case, though they may be affecting the error variance (and scale). The positive coefficient on the *Min* interaction indicates that an undergrounding option is less likely to be chosen if an option with a lower cost had been presented earlier in the sequence. The negative coefficient on the *Max* interaction indicates that an undergrounding option is more likely to be chosen if an option with a higher cost had been presented earlier in the sequence. This evidence is consistent with a 'good deal / bad deal' heuristic and possibly a weak cost minimisation strategy. It is clearly contrary to the hypotheses of cost averaging and truthful response with stable preferences.

Figure 3: TWTP by cost relativity (estimated in Model 5)



Model 5 includes effects coded variables for cost relativity as described in Table 4. Cost level relativity was found to have a significant effect on bid acceptance. The relationship between cost level relativity and TWTP (mean and 90 per cent confidence interval) evaluated at population means for all variables is shown in Figure 3. The relativity with the highest point estimate of TWTP is *m10* (minimum, but not maximum). The incentive compatible relativity, *m11*, had the next highest estimate, while *m01* (maximum, but not minimum) had the lowest. This is further evidence to support a 'good deal / bad deal' heuristic hypothesis ($\beta_{m10*cost}>\beta_{m11*cost}>\beta_{m01*cost}$ and $\beta_{m10*cost}>\beta_{m00*cost}>\beta_{m01*cost}$) or a weak cost minimisation strategy hypothesis rather than a strong cost minimisation strategy hypothesis ($\beta_{m10*cost}=\beta_{m11*cost}>\beta_{m00*cost}=\beta_{m01*cost}$), a cost averaging hypothesis ($\beta_{m10*cost}<\beta_{m11*cost}<\beta_{m01*cost}$ and $\beta_{m10*cost}<\beta_{m00*cost}<\beta_{m01*cost}$) or truthful response with stable preferences ($\beta_{m10*cost}=\beta_{m11*cost}=\beta_{m00*cost}=\beta_{m01*cost}=0$). Mean TWTP is $6,624 when evaluated at the incentive compatible relativity, *m11*. This is remarkably similar to the mean TWTP estimate from the incentive compatible SB format of $6,916. Table 6 presents a test of the statistical significance of the difference between these two estimates. We fail to reject statistical equivalence with a p-value of 0.85. That is, we find no evidence that advance knowledge that multiple choice tasks would be presented had an impact on responses to the first choice task.

The MNL model results are presented in Table 9. The basic models on the SB and RB data (Models 6 and 7, respectively) include supply reliability variables and the 'type of infrastructure' and cost variables. All parameter estimates are more statistically significant in the model on the RB data (Model 7) than in the model on the SB data (Model 6) even without accounting for the panel nature of the data from the RB elicitation format. The estimates of TWTP derived from the MNL models are presented in Table 8. Consistent with the binary logit model results, the point estimate of mean TWTP is higher in the SB model than in the RB model. The result of the test of statistical significance of this difference is set out in Table 10. We fail to reject the null hypothesis that TWTP in the SB format is less than or equal to TWTP from the RB format with a p-value of 0.21. This is a higher p-value than that derived from the equivalent test on TWTP estimates from the binary logit models. It appears the difference is mainly due to a larger standard error on the TWTP estimate from the MNL model on the SB data.

Table 8: Estimates of mean TWTP for mean undergrounding scenario (from MNL models) ($2009)

Model	Mean	90 per cent CI
Model 6 (SB)	5,585	2,893 - 8,438
Model 7 (RB)	4,069	2,759 - 5,428
Model 8: Q1	5,864	3,673 - 8,551
Model 8: Q2	3,730	2,499 - 5,028
Model 8: Q3	3,465	2,299 - 4,740
Model 8: Q4	4,030	2,511 - 5,872
Model 10: m11	5,574	2,679 - 9,758
Model 10: m10 [a]	7,948	-25,256 - 47,180
Model 10: m01	3,141	1,676 - 4,765
Model 10: m00	3,260	1,594 - 5,642

[a] Confidence intervals are large because the denominator in WTP calculations (β_{cost} + $\beta_{cost*m10}$) was very close to zero in a number of the random draws. Mean is 10,139 after removing outliers (absolute value > $40,000) from random draws.

Table 9: Summary of results from MNL models

Model	Model 6	Model 7	Model 8	Model 9	Model 10
Response format	SB	RB	RB	RB	RB
Parameter estimates:					
Household contribution ($'000s)	-.10901*** (.01397)	-.19976*** (.01713)	-.21104*** (.01793)	-.16199*** (.02652)	-.16161*** (.02650)
Change in frequency of unplanned outages each 5 years	-.08506 (.05538)	-.08416** (.04284)	-.08801** (.04301)	-.10150** (.04411)	-.09637** (.04366)
Change in unplanned minutes off supply each 5 years	-.00040 (.00054)	.00011 (.00045)	.00013 (.00046)	.00019 (.00047)	.00016 (.00046)
Change in planned minutes off supply each 5 years	-.00022 (.00018)	-.00045** (.00018)	-.00044** (.00018)	-.00044** (.00018)	-.00042** (.00018)
Type of infrastructure (underground=1)	.19822*** (.06331)	.29816*** (.06398)	.31795*** (.06460)	.22916*** (.07521)	.23013*** (.07506)
Interactions with household contribution:					
Order: question 1 (q1=1)			.06251*** (.01611)	-.07499 (.04720)	
Order: question 2 (q2=1)			-.01906 (.01914)	.00069 (.02111)	
Order: question 3 (q3=1)			-0.03636* (0.02011)	.02026 (.02748)	
Minimum cost in sequence to that point (minimum=1)				.09398*** (.02939)	
Maximum cost in sequence to that point (maximum=1)				.01092 (.01690)	
Relativity: M11 (minimum and maximum cost in sequence to that point =1)					.03049 (.02002)
Relativity: M10 (minimum, but not maximum cost in sequence to that point =1)					.09255** (.03870)
Relativity: M01 (maximum, but not minimum cost in sequence to that point =1)					-.05938*** (.02115)
Model fit:					
Observations	1090	1112	1112	1112	1112
Log-likelihood	716	656	648	643	644
Information criterion AIC	1.32	1.19	1.18	1.17	1.17

*, ** and *** indicate statistical significance at the 0.1, 0.05 and 0.01 levels respectively; standard errors are in parentheses.

Table 10: Tests for differences in TWTP (derived from MNL models)

Null hypothesis (H_0)	$TWTP_{Model\ 6} \leq TWTP_{Model\ 7}$	$TWTP_{Model\ 6} = TWTP_{Model\ 8:Q1}$	$TWTP_{Model\ 6} = TWTP_{Model\ 10:m11}$
p-value	0.2122	0.9247	0.9333

The estimates of MWTP derived from the MNL models are presented in Table 11. Point estimates of MWTP are generally lower in the RB model than the SB model. However, the test results presented in Table 12 indicate that the differences are not statistically significant even at the 0.1 level.

Table 11: Estimates of mean MWTP (based on MNL models) ($2009)

Attribute level change	Model	Mean	90 per cent CI
Type of infrastructure (OH to UG)	Model 6 (SB)	3,682	1,681 - 5,929
	Model 7 (RB)	3,000	1,901 - 4,195
Number of unplanned power cuts each 5 years (unit increase)	Model 6 (SB)	-798	-1,711 - 46
	Model 7 (RB)	-426	-793 - -74
Unplanned minutes off supply each 5 years (unit increase)	Model 6 (SB)	-3.77	-12.26 - 4.43
	Model 7 (RB)	0.53	-3.20 - 4.18
Planned minutes off supply each 5 years (unit increase)	Model 6 (SB)	-2.07	-4.91 - 0.64
	Model 7 (RB)	-2.28	-3.87 - -0.80

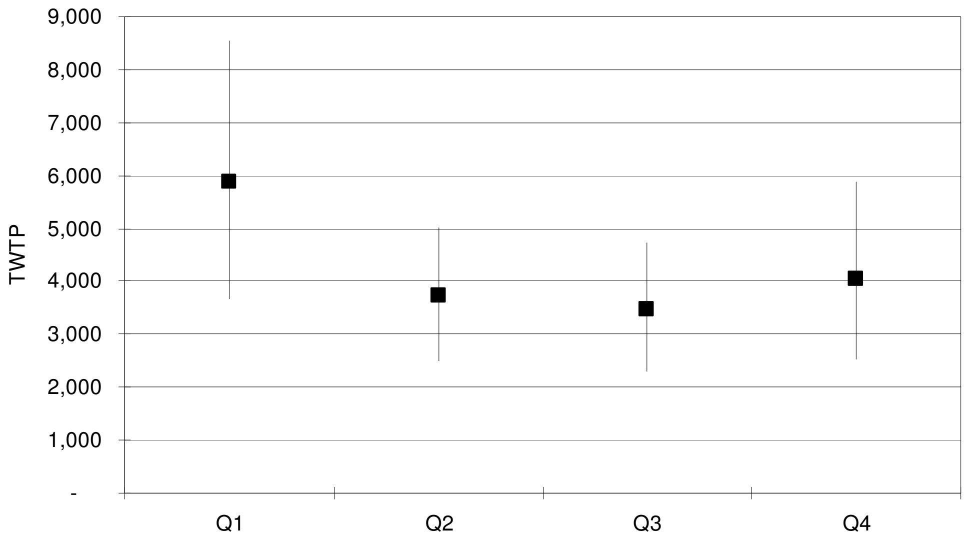
Table 12: Tests for differences in MWTP (derived from MNL models)

Attribute	Type of infrastructure	Number of unplanned power cuts each 5 years	Unplanned minutes off supply each 5 years	Planned minutes off supply each 5 years
Null hypothesis H_0:	$MWTP_{Model\ 6} \leq MWTP_{Model\ 7}$	$-MWTP_{Model\ 6} \leq -MWTP_{Model\ 7}$	$-MWTP_{Model\ 6} \leq -MWTP_{Model\ 7}$	$-MWTP_{Model\ 6} \leq -MWTP_{Model\ 7}$
p-value	0.3370	0.2537	0.2217	0.5344

Model 8 incorporates interactions between the cost variable and the question order variables effects coded as described in Table 2. The positive and significant coefficient on the *q1* interaction indicates that cost sensitivity is relatively low in the first choice task (with all other variables, including cost, held constant). Figure 4 shows that, consistent with the results from the binary logit model, the point estimate of mean TWTP is highest for the first choice task, and decreases over the second and third choice tasks before increasing at the fourth choice task. The estimate of mean TWTP for the first choice task in the RB format is $5,864. Table 10 shows that we fail to reject statistical equivalence between this estimate and the estimate from the incentive compatible SB format of

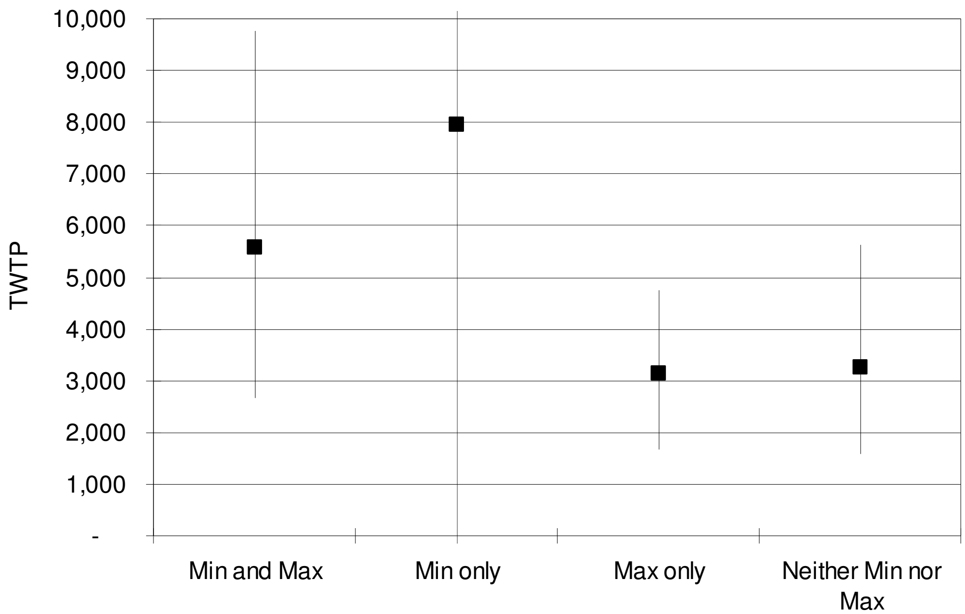
$5,585 with a p-value of 0.92. That is, we find no evidence to suggest that advance knowledge that multiple choice tasks would be presented has an effect on stated preferences in the first choice task.

Figure 4: TWTP by question order (estimated in Model 8)



Consistent with results from the binary logit model (Model 4), the order variables become insignificant once the *Min* and *Max* interactions are incorporated in Model 9, indicating that the majority of the order effect is explained by the effect of cost levels in previously presented choice tasks. Model 10 incorporates interactions between the cost variable and the cost relativity variables effects coded as described in Table 4. Figure 5 shows the means and 90 per cent confidence intervals for TWTP estimated at each of the cost relativity effects. The ranking of mean effects is the same as that derived from the binary logit model (Model 5), which is consistent with a 'good deal / bad deal' heuristic hypothesis ($\beta_{m10*cost}>\beta_{m11*cost}>\beta_{m01*cost}$ and $\beta_{m10*cost}>\beta_{m00*cost}>\beta_{m01*cost}$). However, after accounting for uncertainty around the estimates, the evidence is more supportive of a weak cost minimisation strategy hypothesis ($\beta_{m11*cost}>\beta_{m01*cost}$ and $\beta_{m10*cost}>\beta_{m00*cost}$) or a strong cost minimisation strategy hypothesis ($\beta_{m10*cost}=\beta_{m11*cost}>\beta_{m00*cost}=\beta_{m01*cost}$). The results are contrary to the cost averaging hypothesis ($\beta_{m10*cost}<\beta_{m11*cost}<\beta_{m01*cost}$ and $\beta_{m10*cost}<\beta_{m00*cost}<\beta_{m01*cost}$) and truthful response with stable preferences ($\beta_{m10*cost}=\beta_{m11*cost}=\beta_{m00*cost}=\beta_{m01*cost}=0$). At $5,574, the estimate of mean TWTP evaluated at the incentive compatible relativity (*m11*) is remarkably close to the estimate from the incentive compatible SB format of $5,585. Table 10 shows that we fail to reject statistical equivalence of the two estimates with a p-value of 0.93. Again, we find no evidence to suggest that advance knowledge that multiple choice tasks would be presented has an effect on stated preferences in the first choice task.

Figure 5: TWTP by cost relativity (estimated in Model 10)



Random parameter models accounting for the panel nature of the data from the RB response format (Models 12–15) are presented in Table 14. The estimated effects of question order and cost level relativity are consistent with those from the MNL models. The model results on the SB data (Model 11) do little to dispel the doubts as to whether a single observation per respondent is sufficient to enable the separate estimation of random parameter and error distributions. Very little heterogeneity across respondents is estimated in parameter estimates even for the *Type* parameter. The standard error for mean TWTP in the SB model is more than twice the size of that in the RB model,[17] and, at around $3,000 (after accounting for outliers), the estimate of mean TWTP based on individual-specific conditional distributions for the SB data is somewhat low. Rose et al.

[17] Variance of mean WTP was calculated as $\beta^{-2}[Var(\alpha) - 2\alpha/\beta.Cov(\alpha,\beta) + (\alpha/\beta)^2 Var(\beta)]$ as per Scarpa and Rose (2008).

(2009) had similar problems estimating an RPL model on data with a single choice observation per respondent. Meaningful comparisons between the SB and RB estimates are hampered by these problems. The model results are presented here in order to highlight this key restriction on analysis when using the incentive compatible SB elicitation format.

Estimates of mean TWTP calculated using individual-specific conditional parameter distributions for the mean undergrounding scenario in the survey are set out for both elicitation formats in Table 13. At -$4,782, the mean estimate of TWTP in Model 11 is implausible and closer inspection reveals that it is strongly influenced by an outlier in the mean estimates of conditional WTP. This is a problem that is common when taking the ratio of two unconstrained distributions. One way of obtaining more plausible estimates of mean WTP is to remove or limit outliers at some arbitrary thresholds after estimation.[18] Table 13 presents such estimates based on thresholds set at -$40,000 and $40,000.

Table 13: Estimates of mean TWTP for mean undergrounding scenario (from RPL models)

Model	Model 11 (SB)	Model 12 (RB)
Mean	-$4,782	$8,684
Mean after removing outliers (>$40k or <-$40k)	$2,812	$5,101
Mean after limiting outliers (>$40k or <-$40k)	$3,239	$4,360
Standard error	$917	$369

[18] Modelling solutions to this problem include constraining the distribution on the random cost parameter (Hensher and Greene 2003) or estimating the model in WTP-space rather than utility space (Fiebig et al. 2009, Hensher and Greene 2009, Scarpa et al. 2008, Train and Weeks 2005). Both types of models were not estimable on the SB data. Although they were estimated on the RB data, they are excluded here because no meaningful comparison between SB and RB models is possible.

Table 14: Summary of RPL models

Model	Model 11	Model 12	Model 13	Model 14	Model 15
Response format	SB	RB	RB	RB	RB
Random parameters: means					
Household contribution ($'000s)	-.28094*** (.10286)	-1.48572*** (.27525)	-1.67766*** (.42140)	-1.37517*** (.31462)	-1.42002*** (.32023)
Change in frequency of unplanned outages each 5 years	-.16112 (.09832)	.04288 (.14050)	-.03936 (.20513)	-.03097 (.13026)	-.01380 (.12899)
Change in unplanned minutes off supply each 5 years	-.00166 (.00117)	-.00073 (.00138)	-.00057 (.00196)	-.00064 (.00163)	-.00083 (.00158)
Change in planned minutes off supply each 5 years	-.00048 (.00034)	-.00225** (.00091)	-.00227** (.00105)	-.00201** (.00084)	-.00203** (.00086)
Type of infrastructure (underground=1)	.37109** (.15052)	2.28979*** (.45483)	2.88937*** (.74042)	2.22473*** (.52385)	2.27083*** (.53007)
Random parameters: standard deviations					
Household contribution ($'000s) (triangular)	.82343** (.40680)	2.19786*** (.40319)	2.16636*** (.60178)	2.48494*** (.53236)	2.55728*** (.52548)
Change in frequency of unplanned outages each 5 years (normal)	.16478 (.39574)	.08065 (.20073)	1.25167*** (.39564)	.23629 (.16347)	.23083 (.15508)
Change in unplanned minutes off supply each 5 years (normal)	.00455* (.00269)	.00249 (.00192)	.00763*** (.00290)	.00367 (.00261)	.00371 (.00240)
Change in planned minutes off supply each 5 years (normal)	.00018 (.00054)	.00181 (.00186)	.00244* (.00133)	.00034 (.00095)	.00039 (.00097)
Type of infrastructure (underground=1) (normal)	.24422 (.49869)	2.55384*** (.48749)	3.76972*** (.95573)	2.95092*** (.62700)	2.98149*** (.63927)
Heterogeneity in means					
Household contribution ($'000s):					
Order: question 1 (q1=1)			.27882*** (.09674)	-.26923 (.18912)	
Order: question 2 (q2=1)			-.03957 (.07171)	.01166 (.07176)	
Order: question 3 (q3=1)			-0.16294* (0.08583)	.09254 (.08470)	
Minimum cost in sequence to that point (minimum=1)				.38571*** (.12932)	
Maximum cost in sequence to that point (maximum=1)				.02378 (.07027)	
Relativity: M11 (minimum and maximum cost in sequence to that point =1)					.15152* (.07841)
Relativity: M10 (minimum, but not maximum cost in sequence to that point =1)					.40679** (.16700)
Relativity: M01 (maximum, but not minimum cost in sequence to that point =1)					-.29465*** (.10295)
Model fit:					
Observations	1090	1112	1112	1112	1112
Log-likelihood	709	485	479	470	471
Information criterion AIC	1.32	0.89	0.88	0.87	0.87

*, ** and *** indicate statistical significance at the 0.1, 0.05 and 0.01 levels respectively; standard errors are in parentheses.

Finally, we put the results in context by making some comments on the attribute processing strategies employed by respondents in this study. The models presented herein are based on the assumption that all attributes are strictly processed as full compensatory. However, pre-testing interviews revealed that the supply reliability attributes were often ignored in the SB and RB response formats.[19] Questions about attribute attendance after each choice task revealed that attendance to the supply reliability attributes was generally low in the SB and RB formats, with some respondents effectively treating the choice tasks as dichotomous CV questions by focussing solely on the *Type* and cost attributes.[20] Opportunities for strategic response may have been relatively obvious in this case because the good was viewed as similar from one choice task to another, while cost levels varied significantly.

5. Conclusions

The evidence presented in this paper suggests that estimates of mean TWTP from an elicitation format presenting multiple choice tasks to each respondent may be lower than those from a single choice task format. While there are several possible interpretations for this result, our models show that at least part of this difference is related to the ordering of cost levels over the sequence of choice tasks. In particular, the evidence points to a 'good deal / bad deal' heuristic in which respondents revise their valuation of the good on the basis of cost levels presented over the course of the sequence and/or a 'weak cost minimisation' strategy in which respondents might reject an alternative that is preferred to the status quo if a similar good was offered at a lower cost in a previous choice task.[21] These two behavioural explanations have quite different implications. The latter implies

[19] They were attended to more often in the RMN format as a means of discriminating between the two underground power options presented in each choice task.

[20] Models accounting for reported attribute attendance did not improve the statistical significance of the supply reliability attributes either because there were insufficient observations of attendance to those attributes or because the binary choice design did not induce the trade-offs necessary for the econometric models to isolate the role of these attributes in respondents' choices. These models are not reported here, but are available from the presenting author by request.

[21] Response strategies may differ from one respondent to another, giving a mixed overall result.

divergence between stated and true preferences while the former does not. The design in this study does not allow us to conduct a scope test to differentiate between the two. Regardless, the results are contrary to the standard assumption of truthful response with stable preferences.

We find no evidence to support the hypothesis that stated preferences in the first choice task presented are affected by advance knowledge that four as opposed to one choice tasks will be presented. In fact, there is equivalence in the evidence on TWTP estimates from the first choice task in a sequence and a SB choice task.[22] This goes some way to justifying the use of the first choice task in a sequence as an incentive compatible comparator in studies such as Bateman et al. (2008). We expect that this equivalence would be at least as strong where respondents are not informed about how many choice tasks will be presented. This could be confirmed by further research.

The cost attribute was particularly dominant in respondents' choices in this study due to the experimental design and the attribute processing strategies employed by respondents. Similar goods were offered at quite different cost levels over the course of a sequence of choice tasks, making opportunities for strategic response relatively obvious. We note that Scheufele and Bennett (2010b), in their concurrent and similar study focussing on the case of a pure public good, found similar results to this study using a survey in which the cost attribute was less dominant. The cognitive burden of the choice tasks was also low in their study with binary choices between alternatives described by one cost and two non-cost attributes. A key objective for future research will be to establish whether the response strategies identified in these studies become less prevalent as the cognitive burden of the trade-offs in the choice task (potentially measured by the number of attributes attended to and the number of alternatives per choice task) is increased and opportunities for strategic response become less obvious. While conventional wisdom suggests this would be the case, mechanism design theory suggests that further scope for strategic response is created by presenting two as opposed to one alternative to the status

[22] The web-based questionnaire did not allow respondents to navigate back through the choice tasks. A different result may be derived from a paper-based or other survey in which respondents can review all choice tasks before finalising their responses.

quo in each choice task (Satterthwaite 1975). Further research by these authors will utilise data collected from the RMN elicitation format in the survey described in this paper to examine these issues. It will be important for further research to consider not only the presentation of similar goods at different cost levels, but also the presentation of very different goods at similar cost levels over the course of a sequence.

The evidence from this study supports the notion that the analyst's choice of elicitation format should depend upon the requirements of the study. The SB choice format is best suited to estimation of TWTP for a good with a fixed set of attributes (i.e., contingent valuation). Despite its desirable incentive properties, this format appears to perform poorly when the objective is estimation of MWTP for multiple attributes of a good. Even when the sample size is very large, the single choice task may result in relatively high error variance (due to the lack of institutional learning) and may not provide enough information to allow estimation of heterogeneity in tastes across a population. An elicitation format with multiple alternatives to the status quo in each choice task and multiple choice tasks per respondent may be required to estimate MWTP for multiple attributes of a good. In these cases, analysts should be conscious of the potential for strategic response to a sequence of choice tasks and, if possible, report on any relationship between cost sensitivity and question order or the cost levels presented in previous choice tasks.

Acknowledgements The authors would like to thank Gabriela Scheufele for valuable discussions and exchanges of ideas, David Graham of ActewAGL Distribution for allowing the publication of these results and Andrew Collins and Yang Lan of the Institute of Transport and Logistics Studies for their survey programming work.

References

Accent (2008) Expectations of DNOs & willingness to pay for improvements in service. A report for OFGEM, July

____ (2003) Expectations of electricity DNOs & WTP for improvements in service: stage 1 quantitative research findings. A report for OFGEM, September

Alfnes F and Steine G (2005) None-of-these bias in hypothetical choice experiments. Discussion paper DP-06/05, Department of Economics and Resources Management, Norwegian University of Life Sciences, Aas,

Arrow K, Solow R, Portney P R, Leamer E E, Radner R and Schuman H (1993) Report of the NOAA panel on contingent valuation. Federal Register 58: 4601-4614

Baron D and Myerson R (1982) Regulating a monopolist with unknown costs. Econometrica 50: 911-930

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

Boyle K, Morrison M and Taylor L (2004) Why value estimates generated using choice modelling exceed contingent valuation: further experimental evidence. Paper presented at Australian Agricultural and Resource Economics Society Conference, Melbourne, 11-13 February

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, Poe G L, Ethier R G and Schulze W D (2002) Alternative non-market value-elicitation methods: Are the underlying preferences the same? Journal of Environmental Economics and Management 44: 391-425

Cameron T A and Quiggan 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 (2001) Do hypothetical and actual marginal willingness to pay differ in choice experiments? Journal of Environmental Economics and Management 41: 179-192

____ (2006) 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. Working Papers in Economics no. 191, Department of Economics, School of Business, Economics and Law, Göteborg University

Carson K S, Chilton S M and Hutchinson W G (2008) Decision Rules and Demand Revelation in Decisive Double Referenda. Paper presented at 16th Annual Conference of the European Association of Environmental and Resource Economists, Gothenburg, Sweden, 25-28 June

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. Manuscript,

Carson R T, Groves T and Machina M J (1997) Stated preference questions: context and optimal response. Paper presented at National science foundation preference elicitation symposium, University of California, Berkeley,

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

Farquharson R (1969) Theory of voting. Yale University Press, New Haven

Fiebig D, Keane M, Louviere J J and Wasi N (2009) The generalized multinomial logit: accounting for scale and coefficient heterogeneity. Marketing Science published online before print July 23, DOI:10.1287/mksc.1090.0508

Green J and Laffont J J (1979) Incentives in public decision making. North-Holland, Amsterdam

Green J R and Laffont J J (1978) A sampling approach to the free rider problem. In: Sandmo A (ed) Essays in public economics, Lexington Books, Lexington, MA

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 (2009) Hypothetical bias, choice experiments and willingness to pay. Working paper ITLS-WP-09-01, Institute of Transport and Logistics Studies, The University of Sydney,

Hensher D A and Greene W H (2003) The mixed logit model: The state of practice. Transportation 30(2): 133-176

____ (2009) Valuation of travel time savings in WTP and preference space in the presence of taste and scale heterogeneity. Manuscript, 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

KPMG (2003) Consumer preferences for electricity service standards. A report for Essential Services Commission of South Australia, September

Laffont J J and Tirole J (1993) A theory of incentives in procurement and regulation. Massachusetts Institute of Technology Press, Cambridge, Massachusetts

Loeb M and Magat W (1979) A decentralized method for utility regulation. Journal of Law and Economics 22(2): 399-404

Lusk J and Schroeder T (2004) Are choice experiments incentive compatible? A test with quality differentiated beef steaks. American Journal of Agricultural Economics 86(2): 467-482

McFadden D and Leonard G (1995) Issues in the contingent valuation of environmental goods: Methodologies for data collection and analysis. In: Hausman J A (ed) Contingent Valuation: A Critical Assessment, North-Holland, Amsterdam

McNair B (2009) House prices and underground electricity distribution lines: the case of three selected suburbs in Canberra. Occasional Paper #13, Environmental Management and Development, Crawford School of Economics and Government, The Australian National University

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

NERA 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

Racevskis L and Lupi F (2008) Incentive Compatibility in an Attribute-Based Referendum Model. Paper presented at American Agricultural Economics Association Annual Meeting, Orlando, FL, 27-29 July

Rose J M, Hess S, Bliemer M C J and Daly A (2009) The impact of varying the number of repeated choice observations on the mixed multinomial logit model. Paper presented at European Transport Conference, Leeuwenhorst, The Netherlands, 5-7 October

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

____ (1956) Social indifference curves. Quarterly Journal of Economics 70(1): 1073-1094

Satterthwaite M A (1975) Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory 10(2): 187-217

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

Scarpa R, Thiene M and Train K (2008) Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. American Journal of Agricultural Economics 90(4): 994-1010

Scheufele G and Bennett J (2010a) Effects of alternative elicitation formats in discrete choice experiments. Paper presented at 54th annual Australian Agricultural and Resource Economics Society conference, Adelaide, South Australia, 9-12 February

____ (2010b) Ordering effects and response strategies in discrete choice experiments. Paper presented at World Congress of Environmental and Resource Economists, Montreal, Canada, 28 June - 2 July

Schleifer A (1985) A theory of yardstick competition. Rand Journal of Economics 16: 319-327

Spence M A (1975) Monopoly, quality, and regulation. The Bell Journal of Economics 6(2): 417-429

Train K and Weeks M (2005) Discrete choice models in preference space and willingness-to-pay space. In: Scarpa R and Alberini A (ed) Applications of simulation methods in environmental and resource economics, Springer Publisher, Dordrecht