ISSN 1835-9728

Environmental Economics Research Hub Research Reports

Effects of alternative elicitation formats in discrete choice experiments

Gabriela Scheufele and Jeff Bennett

Research Report No. 52

March 2010

About the authors

Gabriela Scheufele is PhD student in the Crawford School of Economics and Government at the Australian National University.

Jeff Bennett is Professor at the Crawford School of Economics and Government at the Australian National University.

Environmental Economics Research Hub Research Reports are published by The Crawford School of Economics and Government, Australian National University, Canberra 0200 Australia.

These Reports present work in progress being undertaken by project teams within the Environmental Economics Research Hub (EERH). The EERH is funded by the Department of Environment and Water Heritage and the Arts under the Commonwealth Environment Research Facility.

The views and interpretations expressed in these Reports are those of the author(s) and should not be attributed to any organisation associated with the EERH.

Crawford School of Economics and Government
THE AUSTRALIAN NATIONAL UNIVERSITY
http://www.crawford.anu.edu.au

Table of content

ABSTRACT 3
1 INTRODUCTION 4
2 LITERATURE REVIEW 5
3 EMPIRICAL APPLICATION 10
4 RESULTS 17
5 CONCLUSION 25

Abstract

An elicitation format prevalently applied in DCE is to offer each respondent a sequence of choice tasks containing more than two choice options. However, empirical evidence indicates that repeated choice tasks influence choice behavior through institutional learning, fatigue, value learning, and strategic response. The study reported in this paper employs a split sample approach based on field surveys using a single binary elicitation format with a majority vote implementation as the baseline to expand the research on effects of sequential binary DCE formats. We provide evidence for effects caused by institutional learning and either strategic behavior or value learning after respondents answered repeated choice questions. However, we did not find any indications for strategic behavior caused by awareness of having multiple choices. The choice between a sequential and a single elicitation format may thus imply a trade-off between decreased choice accuracy and potentially increased strategic behavior due to an incentive incompatible mechanism. Further research is needed to explore strategic behavior induced by incentive incompatible elicitation formats using alternative approaches that are not compromised by a confounded baseline, that facilitate the differentiation between value learning and strategic behavior, and that allow the use of less restrictive model specifications. Such research should also investigate the effects of varying incentives induced by the order in which choice questions are presented to respondents.

Keywords: discrete choice experiments, split sample approach, elicitation format, incentive compatibility, strategic behavior, learning effects, panel mixed logit models

1 Introduction

Discrete choice experiments (DCE) are being increasingly used to estimate non-market values as inputs in cost-benefit analysis to ensure improved efficiency in resource allocation (Bateman et al. 2006; Bennett and Blamey 2001). DCE involve respondents making trade-offs between attributes that describe non-market goods and services. A variation of attribute levels are bundled in choice options and offered to respondents in choice sets. Choice sets are thus distinguished by differing choice options. The number of choice options and choice sets varies widely across studies. An elicitation format prevalently applied in DCE is to offer each respondent a sequence of choice tasks containing more than two choice options rather than limiting choice to a single binary choice set[1,2]. Increasing the number of choice options and choice questions presented to each respondent is commonly assumed to increase the statistical efficiency of the data for a given number of respondents. However, empirical evidence indicates that repeated choice tasks influence choice behavior through institutional learning, fatigue, value learning, and strategic response. Furthermore, econometric theory suggests that asking respondents a sequence of choice questions introduces correlations of random components across choice tasks. Such correlations have been assumed to affect choice outcomes.

The main objective of the study presented in this paper is to expand the research on effects of alternative DCE formats. We employ a split sample approach based on field surveys using a single binary elicitation format with a majority vote implementation as the baseline. In particular, this paper explores (1) whether a sequential binary elicitation format affects choices, (2) impacts of introduced correlations of error components across choice questions on econometric model results, and (3) whether awareness of having multiple choices influences choice behavior. We hypothesize that the choice between a sequential and a single elicitation format implies a trade-off between decreased choice accuracy and potentially increased strategic behavior due to an incentive incompatible mechanism.

The next section reviews the literature that is concerned with effects associated with alternative choice formats of DCE. This is followed by an overview of the survey logistics, an

[1] A single multiple elicitation format requires respondents to make one choice between more than two choice options presented in one single choice set. A sequential binary elicitation format asks respondents to make repeated trade-offs between two choice options. A sequential multiple elicitation format, finally, offers respondents repeated choices between more than two choice options presented in a sequence of choice sets.
[2] Respondents choose between a zero cost choice option (often the status quo) and one or more choice options with positive cost where the goods and services are assumed to be positively valued. This paper excludes cases where choice options are associated with disutility or where none is the status quo.

explanation of the research design, the formulation of the hypotheses, information about the experimental design, and a discussion of the econometric framework. Results are presented in section four. Finally, in section five, the results are discussed and conclusions drawn.

2 Literature Review

The efficiency of decisions concerning resource allocation depends on individuals truthfully disclosing their privately known preferences. However, revealing true preferences in a DCE might not be an individual's optimal strategy for a given social choice function (see, for example, Mas-Colell et al. 1995). Samuelson (1954) concluded that there exists no mechanism that can guarantee an efficient level of public goods since individuals have a strong incentive to conceal their true preferences. Despite Samuelson's findings, economists have continued to seek incentive compatible demand revealing mechanisms. The analysis of demand revealing mechanisms is the province of mechanism design theory, originally introduced by Hurwicz (1960). Mechanism design theory compares equilibrium outcomes of alternative mechanisms in non-cooperative games of incomplete information with self-interested participants. Hurwicz (1972) defined a mechanism as a communication system used by utility-maximizing participants to reveal private information, such as true or simulated preferences, where the aggregated private information assigns the outcome. The social choice function, called the provision rule in DCE, defines the aggregation process. Accordingly, the provision rule is the link between respondents' choices and the corresponding policy outcome. Whether revealing true preferences is a dominant strategy thus depends on both, the mechanism and the expectations of respondents about the provision rule used to aggregate their choices (Gibbard 1973; Moulin 1994; Satterthwaite 1975). Mazur and Bennett (2010) found that providing respondents with a framing statement for incentive compatibility affects choice behavior in DCE[3]. This evidence suggests that communicating to respondents which provision rule will be used to aggregate choice outcomes is crucial to reduce influences of uncertainty that may confound comparisons between elicitation formats.

A mechanism is defined as incentive-compatible if revealing private information truthfully is a dominant strategy for all participants[4]. The Gibbard-Satterhwaite theorem (Gibbard 1973;

[3] Mazur and Bennett (2010) provided evidence that examined the impact of providing respondents with a framing statement for incentive compatibility in a field survey DCE using a split sample approach. They found that whether the inclusion of a provision rule affects preferences depends on community characteristics.

[4] A widely cited example for an incentive compatible mechanism is a binding referendum between two candidates in an election. Carson and Groves (2007) provided evidence to suggest that replacing the binding character of the referendum by an advisory referendum does not change the incentive compatibility properties of

Satterthwaite 1975) provides a theoretical foundation to analyze the incentive compatibility properties of mechanisms used in stated preference techniques such as DCE. The theorem shows that all non-dictatorial mechanisms other than the single binary choice format are generically incentive incompatible[5,6,7]. A choice format prevalently used in DCE, however, is a sequence of multiple choice options per choice set. One reason behind this is common assumption that sequential multiple choice formats increase the statistical efficiency of the data for a given number of respondents. Such a choice format changes the incentive compatibility properties by firstly asking respondents to choose between more than two options, and secondly by presenting respondents with more than one choice set. Hence, restricting research designs exclusively to the analysis of the latter dimension can reduce influences that may confound effects of repeated choice.

Choice dependencies across respondents are one effect of repeated binary elicitation formats that are based on a plurality vote implementation. The literature on incentive compatibility proposes that respondents who are presented with a repeated binary choice task condition their preferences on expectations about the choices of other survey participants (see, for example, Carson and Groves 2007). Accordingly, the dominant strategy for some respondents is to choose a less preferred option across choice sets if they believe that their most preferred option has no chance of winning[8]. As to our knowledge the effect of such preference conditioning has not been investigated in DCE.

---

the mechanism. Green and Laffont (1978) showed that this also holds for a sample rather than population based referenda. This is important since the majority of choice experiments use statistical samples and, when dealing with public goods, frequently simulates an advisory referendum.

[5] The Gibbard-Satterthwaite theorem also holds for Nash implementations if provision rules are required to be singleton-valued (see Maskin 1977; Muller and Satterthwaite 1985). A non-singleton provision rule may result in potentially incentive compatibility. Many policy decisions that are concerned with the provision of environmental goods and services, however, are confronted with mutually exclusive policy scenarios, that is, the choice of a single scenario is required. Therefore, using a mechanism with a Nash implementation is not a feasible alternative. Carson and Groves (2007) pointed out that in the case of private and quasi-public goods the provision of more than one good may be possible, that is, the provision rule is not singleton-valued. This provides the possibility of an incentive compatible Nash implementation, that is, respondents' incentives to untruthfully reveal their preferences may be reduced.

[6] In laboratory choice experiments, provision rules that are based on a randomly drawn choice question to be binding may introduce incentive compatibility properties in a sequential binary elicitation format, that is, it increases the probability that respondents reveal their true preferences (see, for example, Collins and Vossler 2009). Policy decisions based on random draws, however, raise credibility concerns in the context of public goods valued in field studies (Carson and Groves 2007).

[7] Carson and Groves (2007) suggested that for respondents to disclose private information truthfully, a consequential survey format is required.

[8] This is also true for a single multiple choice format. In that case, a single multiple choice format collapses to a binary choice between the two choice options that the respondent perceives to be other respondents' most preferred choice option if a plurality vote provision rule is applied. However, a single multinomial elicitation format may be potentially incentive compatible if respondents have uniform priors about other respondents' preferred choices (Moulin 1994).

Repeated binary choice formats with a plurality vote implementation additionally imply that respondents may exploit strategic opportunities by including information about previous choice sets and choice decisions (see, for instance, Carson and Groves 2007). As a result, it is optimal for some respondents to choose a less preferred option in one or more binary choice questions. Evidence of such lag effects in sequential binary DCE were presented, for instance, by Holmes and Boyle (2005).

Hence, the literature suggests that both dependencies across respondents and information about previous choice questions may trigger strategic behavior. Their partial effects on strategic response, however, remain unclear.

Bateman et al. (2008) add a further dimension to the discussion about incentive properties of elicitation formats. Their study presented evidence that respondents' awareness of having multiple choices may induce strategic behavior. This could occur through information provided to the respondent before the choice task ('choice set awareness effect')[9] and through a dynamic increase in awareness of strategic opportunities as progress in made through the sequence of choice questions ('ordering effect'). Previous and successive choice sets may contain alternative prices for the same or a similar level of provision of a particular good or, vice versa, the same or similar price for alternative levels of provision of a particular good. This may trigger respondents to either question the credibility of the survey or learn to take advantage of this inconsistent pricing by rejecting a preferred choice option when the same or a similar level of provision was offered in a previous or successive choice question at a lower price. This implies that repeated choice may cause learning about strategic opportunities and how to exploit them. These findings are supported by the concurrently conducted study of McNair et al. (2010) who provide evidence that increasing the number of choice sets per respondent decreases estimates of willingness to pay (WTP), and that this effect may be explained by the ordering of alternative cost levels offered across a sequence of four choice questions.

In comparison to Bateman et al. (2008) who used the first question of a sequential choice task as the incentive compatible baseline to explore sequence effects, Racevskis and Lupi (2008) used a split sample design to explore the effect of asking respondents a single versus a sequence of binary choice questions. Racevskis and Lupi (2008) found a significant difference between fits across the two models based on pooling the data of the two response formats in two different ways: the first model included generic attributes whereas the second

[9] Mail surveys disclose all choice questions before a choice has to be made whereas an internet based survey can be programmed to reveal only one choice question at a time.

model included split sample specific attributes. This study fell short to account for effects of differing sample sizes implied by each choice format and was focused on a comparison of the model fit between the two split samples. It missed the opportunity to explore impacts on further DCE dimensions.

Carson and Groves (2007) discuss an additional dimension of incentive compatibility. They argue that for respondents to disclose private information truthfully, a consequential survey format is required. Consequentiality means that the commodity has to be of relevance to the respondent and respondents have to believe that their choices have an impact on the outcome. In inconsequential surveys, respondents perceive choice options as equally non-beneficial and indistinguishable. Under such circumstances, it remains unknown whether or not respondents reveal their true preferences. Associated drivers postulated to additionally influence choice behavior include the properties of the payment vehicle, plausibility of the choice questions, credibility of the policy scenario, and comprehensibility of the choice task (Carson and Groves 2007). Surveys lacking a payment vehicle that respondents perceive as coercive induce free-riding behavior. Implausible choice task make result in choices that are based on a different choice set than the one presented by the researcher. If respondents are presented with an incredible policy scenario they may be unsure whether presented options will be deliverable. If that is the case, respondents may include their perceived probability of provision into their choice rule. Respondents who misunderstand the choice task may answer the question they think has been asked instead of the one the researcher intended to have answered (Carson and Groves 2007). In an extreme case, respondents may choose without providing any information about their preferences if answering the choice questions lies beyond their capabilities.

Commonly, analysts using DCE assume consequentiality of the survey, plausibility of the choice questions, credibility of the policy scenario, and comprehensibility of the choice task. These assumptions, however, may be violated. Including follow-up questions concerned with exploring these issues is a possible means of investigating to what extent true preferences are disclosed. However, the incentive properties of such follow-up questions are unknown. Hence, the actual opinions of respondents may not be reflected in their answers.

Learning and fatigue are other impacts types of repeated binary choice formats that have been discussed to influence choice behavior. Braga and Starmer (2005) proposed a process where respondents become increasingly familiar with the choice context, the offered good, and the choice task (‘institutional learning’). Typically, ‘institutional learning’ is assumed to affect the accuracy of responses reflected in the scale factor[10] rather than changing preferences. As respondents progress through the choice questions their responses become more accurate (increase in scale factor) until fatigue sets in (decrease in scale factor). In this context, Swait and Adamowicz (2001) discuss ('smaller noise to signal ratio' and 'larger noise to signal ratio', respectively. Plott (1996) proposed that respondents may 'discover' their true underlying preferences through a learning process rather than possessing stable preferences ('value learning'). Such learning processes are expected to change preferences, and thus parameter estimates in DCE. The empirical evidence of Bateman et al. (2008) discussed previously in this paper suggests that the notion of learning additionally includes 'strategic learning', such that respondents become increasingly aware of and learn to exploit strategic opportunities while making progress through the choice task. Such strategic opportunities provide incentives to misstate rather than to disclose truthfully preferences.

This review of the literature suggests that only a few empirical studies have investigated the effects of sequential binary DCE formats and associated strategic behavior. The existing empirical evidence indicates that repeated choice tasks influence choice behavior through a mixture of drivers including institutional learning, fatigue, value learning, and strategic behavior. Differentiating between these drivers challenges the research design and is susceptible for misleading conclusions.

The main objective of this study is to extend the research on this topic by exploring the following research questions:

1. Does a sequential character of a binary elicitation format affect choices?
2. How do introduced correlations of error terms across choice questions impact on econometric model results?
3. Does awareness of having multiple choices influences choice behavior?

In comparison to the research of Bateman et al. (2008) the study reported in this paper employs a split sample approach based on field surveys using a single binary elicitation format with a majority vote implementation as the baseline. We are unaware of any work other than the research of Racevskis and Lupi (2008) and a concurrently conducted research of McNair et al. (2010) that has tested sequence effects focused on the incentive compatibility properties of elicitation formats in DCE using field data and a split sample approach with a single binary elicitation format as a baseline. We expand the approach of Racevskis and Lupi by exploring additional outcome dimensions, trying to separate lag effects from effect

[10] The scale factor is inversely related to the variance of the error distribution (Swait and Louviere 1993).

induced by dependencies across respondents, testing for choice set awareness, and adjusting the number of observations in the choice experiment with a single binary elicitation format to reduce confounding influences. In contrast to the concurrently conducted study of McNair (2010) that is based on a public good with private elements, we investigate incentive compatibility properties of elicitation formats using a pure public good that provides use and non-use values. Finally, in comparison to previous studies, follow-up questions are included to examine the properties of the potentially incentive compatible baseline.

3 Empirical Application

The hypotheses are tested using data from a discrete choice experiment concerned with estimating use and non-use values of a public good, the preservation of a natural area, using Nadgee Nature Reserve as an example. Nadgee Nature Reserve is one of the largest coastal wilderness areas in NSW and covers an area of 17,116 ha. It is pristine and has a high level of landscape diversity. The data set used in this study is derived from a random sample of the population of Sydney drawn from an internet panel[11]. The data were collected using an internet based survey[12].

The survey material was developed using expert opinion and focus groups[13]. A pilot survey was conducted to test the survey material and internet set-up, as well as to obtain parameter priors for the development of the experimental design. The final survey was structured as follows. In the first part, respondents were asked about their socio-demographic characteristics as well as their general experience of visiting protected areas in Australia or worldwide. In the second part respondents were provided with background information including photographs and explanations about the reserve and future management options. The reserve was described in term of the features of Nadgee Nature Reserve, even though it was presented as an *area of land* without revealing its identity. Respondents were told that funds had to be raised to enable the government to purchase the land, and thus conserve the area. A plurality vote was used as provision rule[14]. The third part of the survey asked

[11] Only Australian citizens or permanent residents of Australia 18 years or above qualified.

[12] The overall response rate was 34%: Invited but not participated (55%), participated but below five minutes completion time (2%), participated but dropped out before completion (9%).

[13] Two focus groups are conducted in Canberra. In order to ensure the applicability of the survey material for a sample of the population of Sydney the pilot survey included four follow-up questions at the end of the questionnaire. Respondents were asked if they had any concerns, comments or suggestions with any part of the questionnaire. Obtained information was used to adjust the survey material accordingly.

[14] The management option that receives the greatest support would be implemented and everyone would have to make the payment associated with that management option.’

respondents to make trade-offs between future management options including development and preservation alternatives (see Figure 1).

The management options were described by three attributes with five, four, and two levels, respectively (see Table 1). In order to increase the comprehensibility of the choice task, respondents were presented with an explanation of the outcome of their first choice and given the opportunity to revise it (see Figure 2). This part of the survey was followed by questions designed to check the consequentiality, plausibility, credibility, and understandability of the survey material. The final part of the survey asked additional questions about socio-economic characteristics of the respondents.

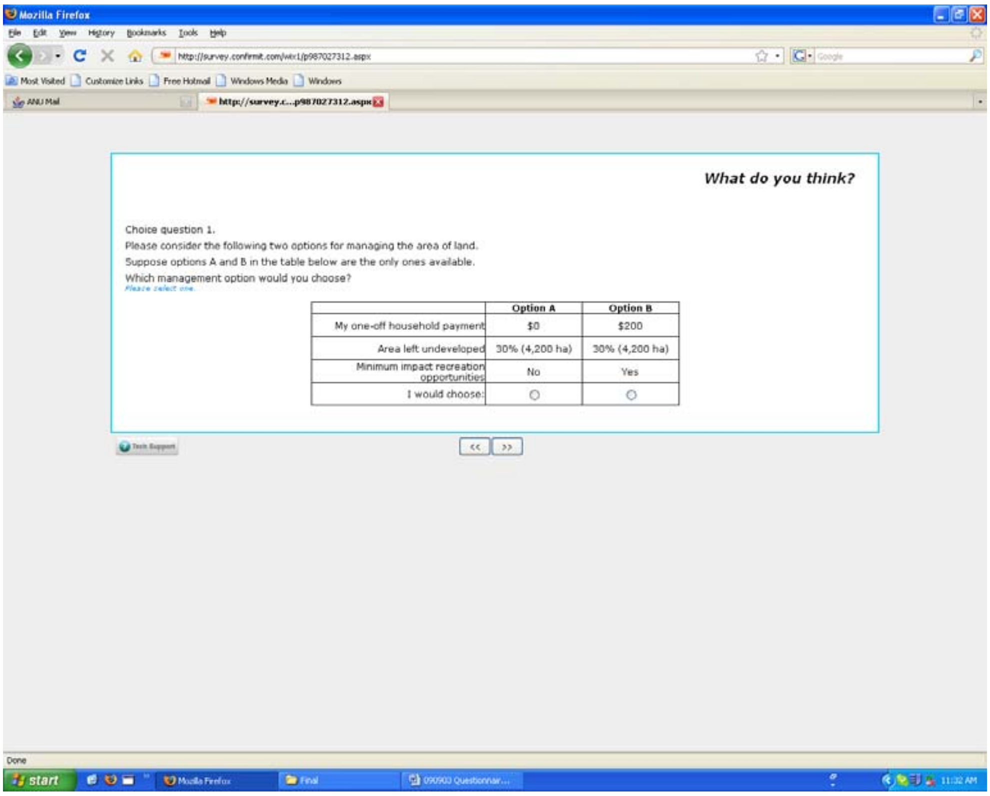

Figure 1: Example of choice set

Table 1: Attributes and attribute levels

Attribute	Attribute level	Coding
Cost	$0 $50 $100 $200 $300	numerical
Area of land	30% (4,200ha) 50% (7,000 ha) 70% (9,800 ha) 100% (14,000 ha)	numerical
Access for minimum impact recreation	yes no	1 -1

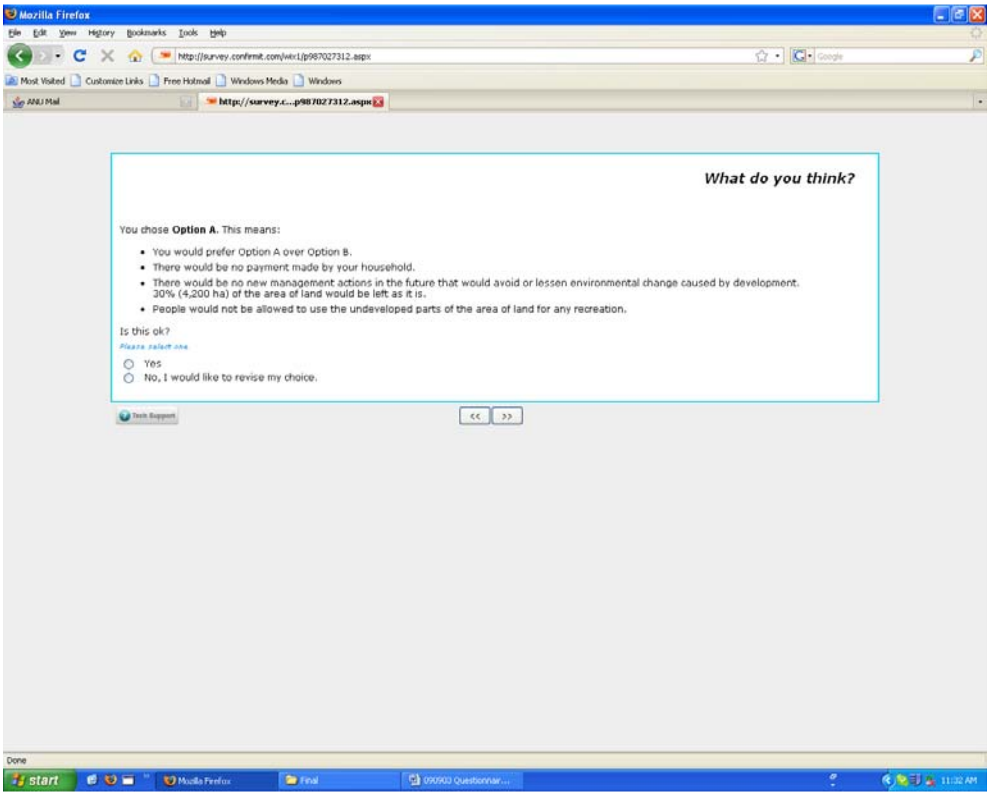

Figure 2: **Example of choice set explanation**

Five split sample treatments were used that differed only in the number of choice sets per respondent, the choice set order, and the wording of some explanations and instructions so necessitated. This study is based on the maintained assumption that the marginal differences in wording of the choice questions do not alter statistically significantly choice incentives across split samples. All split samples were based on the same experimental design with a total of 16 choice sets that contained two choice options each: one invariant zero cost choice option that was available in each choice set and one non-zero cost choice option that varied across choice sets. For the repeated binary choice task split samples (RB1, RB2, RB3, and RB4) the 16 choice sets were divided into four blocks of four choice questions per respondent. For the single binary choice task split sample (SB) each respondent was asked to answer one choice question only. In order to avoid the confounding impacts of having different numbers of observation across split samples, SB was assigned to about four times as many respondents (1444) as each of RB1 (367), RB2 (371), RB3 (369), and RB4 (376)[15]. A sixth split sample (PoolRB) was created by pooling the RB split samples. A seventh split sample treatment (CrossRB1) was obtained by pooling the first choice questions of PoolRB.

[15] Rose et al. (2009) used simulated data to investigate the statistical impact of panel data in discrete choice experiments. They showed that increasing the number of choice observations per respondent, while holding sample size constant results in less biased estimates and larger t-ratios. However, this advantage diminishes with increasing sample size.

The four RB split samples underlying the CrossRB1 were developed to differ systematically in terms of the position choice sets in the sequence. For example, the first (second) choice set in RB1 was the last (first) choice set in RB2, etc. Hence, the presented choice sets are cycled four times such that each choice set is presented in first position approximately the same number of times across the sample. The first choice question of RB1, RB2, RB3 and RB4 were the same as those in SB, with the sole exception being the number of choice tasks presented to each respondent.

The following section specifies the hypotheses and their respective tests to explore the stated research questions.

In order to explore whether a repeated binary elicitation format affects choices we test the three following hypotheses:

H_0^1: Choice outcomes of a single binary choice task (SB) are the same as those of a repeated binary choice task (RB) that contains four choice sets.

To test H_0^1 we compare choice outcomes of RB1, RB2, RB3, and RB4 with those of SB using MNL, MML, panel MML model specifications. We contrast each of the four RB split samples with SB to account for potentially confounding effects of the order in which the choice sets of the RB split samples were presented to the respondents[16]. The prior expectation is a higher acceptance rate of non-zero cost options, increased magnitude of the parameter estimates, a smaller scale factor, and higher WTP in the single as opposed to the repeated choice task.

H_0^2: DCE outcomes of respondents who stated that they answered the questions within a sequence of four binary choice tasks independently from their previous choices are the same as those who stated the opposite.

H_0^3: DCE outcomes of respondents who answered a sequence of four binary choice questions and stated that they considered what other survey participants might choose are the same as those who stated the opposite.

To test the joint hypotheses H_0^2 and H_0^3 two effects coded variables were created: (1) a variable reflecting respondents' subjective view on whether their choices were conditional on expectations about the choices of other survey participants, and (2) a variable indicating

[16] The authors are aware that a completely randomized design may reduce these effects more drastically. However, the comparison of a single binary with the first question of a sequential binary choice format limited the design options.

whether respondents answered the choice questions conditional on previous their previous choices. These variables were interacted with the constant term and included in the MNL and the panel MML model estimation of PoolRB (*oth*con, ind*con*)[17]. We expect the parameter estimates of all interaction variables to be statistically significantly different from zero indicating that whether respondents conditioned their choices on expectations about other respondents' choices/ on previous choices influenced their decision whether to choose a non-zero cost option.

To investigate H_0^2 and H_0^3 further the effects coded variables were interacted with the *cost* attribute and included in the MNL and the panel MML model estimation of PoolRB (*oth*cost, ind*cost*). Prior expectations are statistically significant *oth*cost* and *ind*cost* parameter estimates with a positive sign. A positive sign indicates a higher WTP for respondents who stated that their choices are independent from previous choice sets and expectations about other respondents' choices, respectively, such that

$$wtp = \frac{\beta^k}{-[\beta^m + \beta^{oth*\cos t, ind*\cos t}]}.$$

In order to investigate the impact of correlated error components across choice questions in a DCE on econometric model outcomes we test the following hypothesis:

H_0^4: DCE outcomes estimated using a MML model with and without panel specification are the same.

To examine H_0^4 we compare choice outcomes estimated by MML models with and without panel specification. The prior expectation is an increase in the number of statistically significant attributes and a statistically significantly better model fit in the panel specification. Model fit is evaluated by the Aikaike-Information Criteria (AIC), the Bayesian-Information Criteria (BIC), and by conduction a likelihood-ratio test following the Chi-square distribution (LR).

Finally, to examine whether awareness of having multiple choices influences choice behavior we test the following hypothesis:

[17] The constant term was included in the utility function of the non-zero cost option.

H_0^5: Choice outcomes of a single binary choice task are the same as those of the first choice question of repeated binary choice tasks that contain four choice sets.

H_0^5 is tested by comparing choice outcomes of SB with those of the first choice questions of the PoolRB (CrossRB1) using MNL, MML, panel MML model specifications. A higher acceptance of non-zero cost options, increased magnitude of the parameter estimates, a smaller scale factor, and higher WTP of the single binary as opposed of the first choice question of four DCE using repeated binary choice format are anticipated.

The complete research design is summarized in Table 2.

Table 2: Research design

H_0^1	Comparison: Sequential binary (RB1, RB2, RB3, RB4) - Single binary (SB);
H_0^2	Inclusion of interaction variables *ind*con* and *ind*cost* in econometric models 'I answered the choice questions independently from my previous choices'; effects code: 1 (yes), -1 (no);
H_0^3	Inclusion of interaction variables *oth*con* and *oth*cost* in econometric models 'I did made my choice independent on beliefs about other respondents' choices; effects code: 1 (yes), -1 (no);
H_0^4	Comparison of choice outcomes estimated by MML models with and without panel specification (PoolRB);
H_0^5	Comparison: Single binary (SB) – First questions of PoolRB (CrossRB1)

All choice sets were created using a Bayesian D-efficient design (Bliemer et al. 2008). Bayesian D-efficient designs are statically efficient designs (see, for example, Ferrini and Scarpa 2007; Rose and Bliemer 2008; Rose et al. 2008). Statistically efficient designs aim to maximize the amount of obtained information. A commonly used measure to express the global level of efficiency is the D-error, which minimizes the determinant of variance-covariance matrix. The smaller the D-error, the more statistically efficient is the design. Therefore, a statistically efficient design can be used to increase efficiency while holding the sample size fixed. The Bayesian D-efficient designs (100 Halton draws) used in this study are developed based on the calculation of the Db-error of randomly selected designs (10,000 iterations). Attribute levels are randomly assigned to each attribute in each choice set of the change options while accounting for attribute balance. The base level (zero cost option) is held constant but included in the design process. Priors ware obtained from pilot studies targeting the population of Sydney and Canberra[18]. Following a suggestion of Rose and Bliemer (2005), the rows and columns related to the constant term are excluded from the

[18] The choice sets of the pilot study were created using a Bayesian D-efficient design. Priors were obtained from the focus group choice experiment.

calculation of the Db-error in order to avoid the dominance of the unproportionally large standard errors of the constant. Dominant and redundant choice sets are removed through restrictions and swapping of attribute levels marginally reducing the Db-efficiency (3%). The Bayesian D-efficient designs are developed for multinominal logit (MNL) models without accounting for covariate effects. Estimating different models may alter the design efficiency (Rose and Bliemer 2005). The results that are used to test H_0^4 are based on the assumption that these differences do not statistically significantly influence the comparison of choice outcomes.

There are a range of discrete choice models motivated by random utility theory (McFadden 1974; 1980), which can be used to analyzed discrete choices. In this study, we used multinomial logit (MNL) and a panel mixed multinomial logit (MML) models to analyze the collected data. The MNL model, introduced by McFadden (1974), is restrictive in that is assumes parameter vectors to be fixed across respondents and choice tasks, and the error terms to be independently and identically (IID) extreme value type 1 (EV1) distributed. A behavioral result of the IID assumption is Arrow's (1951) axiom of independence of irrelevant alternatives (IIA). Simonson and Tversky (1992) suggested that, in the context of DCE, the violation of the IIA properties has two dimensions: a lateral dimension that refers to the presence of more than two options within a choice question, and a vertical dimension that refers to the panel character of the elicitation format. This implies that repeated choice tasks introduce correlations of error components across choice sets within respondents, a fact which is frequently ignored in DCE and further explored in this paper.

MML models (see, for example, Brownstone and Train 1999; Greene and Hensher 2006; 2007; Greene 2008; Hensher et al. 2005; Hensher and Greene 2003; Louviere et al. 2000; McFadden and Train 2000) allow for a complete relaxation of these assumptions by disaggregating the error component in a stochastic IID-EV1 error term and error terms that are based on underlying parameter vectors and observed data associated with choice options and respondents.

This relaxation provides the opportunity to model preference heterogeneity associated with preference parameters that are assumed to be distributed continuously over respondents around a fixed or heterogeneous mean, where the assumed distributions, may be specified as heteroscedastic across respondents. In a random parameter specification, preference parameters can be assumed to be random across both respondents and choice tasks (cross-sectional) or across respondents but not choice tasks (panel). Cross sectional data assume a single choice task per respondent whereas panel data assumes repeated choices per respondent. MML models allow accommodating correlated choice tasks within respondents for panel data in two ways. One way is to change the log-likelihood function, presuming that the random effects are the same across choice tasks (Revelt and Train 1998). As such, the log-likelihood function of a cross-sectional specification is replaced by a log-likelihood function that accounts for dependencies across choice options and choice tasks[19].

In the study reported in this paper, we used MNL models to test H_0^1 and H_0^5. Panel MML models with replaced log-likelihood function were employed to test H_0^2, H_0^3, and H_0^4. In all MML models, all choice attributes were defined as random parameters to account for preference heterogeneity. If not stated otherwise, all econometric models were estimated using Nlogit 4.1. Following Greene and Hensher (2006; 2006), a constrained triangular distribution was used for the *cost* parameter to ensure a negative sign. The distributions on the *access* and the *area of land* attributes were not constrained to allow for both positive and negative preferences towards these attributes. A normal distribution was assumed for these attribute parameters. The WTP for all attribute parameters[20] were estimated using a bootstrapping procedure with 1000 draws (Krinsky and Robb 1986).

4 Results

Sample characteristics

A series of chi-square tests were conducted to test for equivalence between the population statistics (ABS 2006) and the split samples. No statistically significant differences at the 5% level with respect to *sex* and *age* were discovered. However, *individual gross income*, *household gross income*, *level of non-school education*, and *highest year of school completed* were statistically significantly different at the 5% level between the population and all split samples. The split samples are therefore not representative of the households of Sydney and care should be taken when interpreting the results on a population level. Additionally, a series of chi-square tests was carried out to test for differences in the socio-demographic characteristics[21] between split samples. No statistically significant differences at the 5% level

[19] A second way to incorporate correlations across choice tasks is to include a first order autoregressive (AR1) error term, assuming that previous choices influence latter choices (see, for example, Greene 2007).
[20] Implicit prices
[21] Sex, age, individual gross income, household gross income, level of non-school education, highest year of school completed.

were found for any of the comparisons. Consequently, it is assumed that there are no varying underlying population structures present that may confound comparisons across split samples.

A range of follow-up questions was included in the questionnaire to check for consequentiality of the survey format, the plausibility of the choice questions, the credibility of the policy scenario, and the comprehensibility of the choice task[22]. The following results are based on adding the percentages of the categories 'strongly agree' and 'agree' chosen by respondents. Sixty seven percent of respondents were interested in the management of the natural area of land. The provided information was understandable for 74%. Seventy seven percent understood the concept of making choices but 16% found making choices confusing. 38% did not believe that recreation – even if it is low impact - would cause only minor environmental changes. The management options made sense for 54%. Thirty nine percent thought their choices would have an impact and 27% believed that the management plan would be implemented. These results indicate that the survey format lacks consequentiality and other associated properties that influence the incentive properties of the surveys. Hence, the theoretically incentive compatible baseline, the single binary elicitation format, may be compromised potentially confounding comparisons across split samples.

Effects of repeated binary choices

Hypothesis 1

In order to test H_0^1 we firstly investigate choice shares of non-zero cost options of SB and RB1, RB2, RB3, and RB4. The percentage of choosing any non-zero cost option is 56% for SB as opposed to 52% for RB1, 46% for RB2, 43% for RB3 and RB4. The difference between SB and the RB split samples is statistically significant at the 5% level (chi-square test) for RB3 (p=0.05) and RB4 (p=0.05) but not for RB1 (p=0.58) and RB2 (p=0.14). Since these four split samples only differ in the order in which the choice set are presented to respondents these results indicate that choice behavior in repeated choice tasks may be affected by choice set ordering[23].

To investigate H_0^1 further, two econometric model specifications were estimated. The results of the MNL model estimation for SB and exemplarily for RB1 are reported in Table 3[24]. The *cost* parameter estimates for all four RB split samples are statistically significantly different

[22] All questions were based on a five point Likert scale: 'strongly disagree', 'disagree', 'neither disagree nor agree', 'agree', 'strongly agree'.
[23] Potential ordering effects induced by this data set were further investigated by Scheufele and Bennett (2010).
[24] For parsimony, the detailed model results of RB2, RB3, and RB4 are not reported in this paper but are available from the authors on request.

from zero and have the expected negative sign indicating that lower cost options are preferred to higher cost options, ceteris paribus. The *area of land* parameter estimates are statistically significantly different from zero and positive as expected suggesting that a larger area of land provides a higher utility than a smaller area, ceteris paribus. The *access* parameter estimates, however, are not statistically significantly different from zero in neither split sample.

The MNL restrictions were relaxed by estimating MML models using Halton draws with 500 replications (Train 2000). Using a MML model specification instead of a MNL model specification did not improve the model fit of SB. The *cost* parameter was the only attribute parameter that was statistically significantly different from zero at the 5% level. Rose et al. (2009), using simulated data, suggested that obtaining only a single choice observation may not allow the discovery of random parameters that are statistically significantly different from zero. A possible explanation is that in the absence of a very large sample it is impossible to disentangle the assumed distribution of random terms associated with preference parameters or alternatives from the assumed EV1 distribution of the remaining random term that is assumed to be IID across alternatives and individuals. This implies that the MML model specification cannot be used to compare SB with RB1, RB2, RB3, and RB4, and CrossRB1. Hence, the analysis of H_0^1 and H_0^5 is restricted to the MNL model specification.

The WTP estimates for each of the four RB split samples and SB are reported in Table 4. A Poe test (Poe et al. 2002; Poe et al. 2005) was conducted to test for equivalence of WTP estimates. We find a statistically significantly higher WTP for SB than for each of the four RB split samples (see Table 4). However, the 95% confidence interval of SB is wider and overlaps partially with the 95% confidence interval of RB1, RB2, RB3, and RB4.

Differences in the attribute and scale factor between SB and each of the RB split samples are explored using the Swait-Louviere test (1993) [25,26]. The results are displayed in Table 5. We find statically significant differences in attribute parameter estimates comparing SB to RB2, RB3, and RB4 with the exception of RB1. Possible explanations for changes in attribute parameters suggested in the literature include value learning and learning to exploit strategic opportunities.

Table 3: MNL model results for RB1, SB and CrossSB1

	Sequential binary	Single binary	First of sequential binary

[25] The relative scale factor was estimated based on a heteroscedastic multinomial logit model in STATA 10 (see, for example, Hensher et al. 1999).
[26] For a detailed discussion about the Swait-Louviere test and the confounding influence of the scale factor in multinominal logit models see Louviere and Eagle (2006).

	(RB1)		(SB)		(CrossRB1)	
Variable	Coefficient	Standard error	Coefficient	Standard error	Coefficient	Standard error
Constant	0.39878*** (0.0039)	0.13803	0.18680 (0.1760)	0.13805	0.00666 (0.9611)	0.13641
Cost	-0.00453*** (0.0000)	0.00057	-0.00218*** (0.0001)	0.00056	-0.00224*** (0.0001)	0.00056
Area of land	0.01342*** (0.0000)	0.00214	0.01273*** (0.0000)	0.00217	0.01291*** (0.0000)	0.00206
Access	0.00059 (0.9914)	0.05430	-0.00537 (0.9211)	0.05426	0.05640 (0.2887)	0.05316
Model statistics						
N (observations)	1468		1445		1483	
LL_{ASC}	-1015.9650		-991.9259		-1279.8540	
LL_{β}	-962.2373		-965.8225		-995.7384	
$\chi^2_{,3}$	107.46 (0.0000)		52.21 (0.0000)		568.23 (0.0000)	
Adjusted ρ^2	0.05		0.02		0.22	
AIC	1.29595		1.34231		1.34826	
BIC	1.33560		1.35692		1.36257	

***=significant at 1% level, **=significant at 5% level, *=significant at 10% level; p-values in parentheses;

Table 4: $WTP_{area\ of\ land}$ estimates for RB1, RB2, RB3, RB4, and SB

	$WTP_{area\ of\ land}$	Standard error	$CI(95\%)_{WTP}$	Poe test RB-SB (p-value)
RB1	$2.98*** (0.0000)	0.61584	$1.88-4.34	0.0566*
RB2	$2.41*** (0.0000)	0.50173	$1.48-$3.46	0.0145**
RB3	$2.51*** (0.0000)	0.52867	$1.62-$3.58	0.01808**
RB4	$2.99*** (0.0000)	0.54493	$2.03-$4.17	0.04636**
SB	$6.44** (0.0203)	2.7705	$3.42-$13.08	-

***=significant at 1% level, **=significant at 5% level, *=significant at 10% level; p-values in parentheses; 95% confidence intervals in parentheses based on the 2.5th and 97.5th percentile of the simulated WTP distribution. In comparison to the delta method, this method does not imply a normal distribution.

A statistically significant difference in an attribute parameter estimate prevents a test for scale factor estimates equality[27]. Hence, solely RB1 was tested in this regard. The hypotheses of equal scales was rejected (p_{scale}=0.0124). The reduced relative scale factor for each of the RB split samples suggests a less accurate choice since the scale factor is inversely related to the variance of the error term. Smaller relative scale factor and larger confidence intervals of SB as opposed to the RB split samples indicate that the difference in WTP is mainly induced by

[27] Parameter vector and scale factor are confounded in MNL models. Hence, having a varying scale factor prevents testing for parameter vector equality.

differences in the variance of the error term; that is, repeated choices increase the choice accuracy.

Overall, this leads to a rejection of H_0^1.

Table 5: Test results for equality between attribute and scale factor of RB1, RB2, RB3, RB4 and SB

LL RB	LL SB	LL Pooled[a]	LR-test[b] (5d.f.)	Reject H_0:β_i=β_j	Scale ratio λ_i/λ_j	LL Pooled[c]	LR-test[d] (1 d.f.)	Reject H_0:λ_i=λ_j
-962.2373 RB1	-965.823	-1931.128	0.2932	no	0.7795	-1934.252	0.0124	yes
-965.627 RB2	-965.823	-1942.752	0.0004	yes	0.6087	-1953.9	NA	NA
-952.739 RB3	-965.823	-1930.921	0.0000	yes	0.4694	-1949.737	NA	NA
-961.4410 RB4	-965.823	-1937.9	0.0007	yes	0.4609	-1957.801	NA	NA

[a] Pooled MNL model allowing varying scale factor;
[b] Log-likelihood ratio test, test statistics $-2(LL_{pool}- (LL_1+LL_2))$ with d.f. k+1, where k is the number of parameters including the constant is asymptotically chi-square distributed;
[c] Pooled MNL model assuming equal scale factor in both split samples;
[d] Log-likelihood ratio test, test statistics $-2(LL_{equalscale}- (LL_{varyingscale}))$ with 1 d.f.; is asymptotically chi-square distributed

Joint Hypotheses 2 & 3

To investigate H_0^2 and H_0^3 we firstly examine the follow-up questions directly. About 23% of the respondents answered that their successive choices were influenced by their previous choices. About 7% did not choose their most preferred management option if they did not think it would be the most popular option. About 93% stated that they always chose their preferred management option, whereof about 59% nevertheless thought about what other people might choose.

The inclusion of the variables *oth*con* and *ind*con* in the model estimation improved the model fit statistically significantly at the 1% level for the MNL model but but not for the panel MNL model[28]. The results are reported in Table 6 and Table 7. The *ind*con* parameter estimate is statistically significantly different from zero at the 1% level and 10% level, respectively. These results suggest that respondents, who did consider previous choices are less likely to choose a non-zero cost option, i.e. behave strategically. The *oth*con* parameter estimate was statistically insignificant in both models providing no indication that respondents' choices were influenced by beliefs about other respondents' choices.

The inclusion of *oth*cost* and *ind*cost* in the MNL and the panel MML model estimation improved the model fit statistically significantly at the 1% level and the 10% level,

[28] Likelihood ratio test (-2*LLr-LLur]).

respectively (LR-test). The results are displayed in Table 6 and Table 7. The parameter estimate for *ind*cost* was statistically significantly different from zero at the 1% level and the 5% level, respectively. The positive signs indicate that respondents who stated they answered each question independently from previous ones have a higher WTP than those who stated the opposite. The parameter estimate for *oth*cost* was not statistically significantly different from zero indicating that respondents WTP was not influenced by beliefs about others respondents' choices. This result is in accordance with the low percentage of respondents who stated that they conditioned their choices on expectations about choice of other survey participants.

Table 6: Model results of a MNL model specification including a variable reflecting subjective views on strategic behavior

	MNL PoolRB *oth*con*, *ind*con*		MNL PoolRB *oth*cost*, *ind*cost*	
Variable	Coefficient	Standard error	Coefficient	Standard error
Parameter				
constant	0.06242 (0.4563)	0.08379	0.16329** (0.0167)	0.06825
cost	-0.00489*** (0.0000)	0.00029	-0.00516*** (0.0000)	0.00040
area of land	0.01318*** (0.0000)	0.00106	0.01321*** (0.0000)	0.00106
access	0.03706 (0.1724)	0.02716	0.03699 (0.1730)	0.02715
oth*con/ oth*cost	0.02264 (0.6794)	0.05477	-0.00015 (0.6047)	0.00030
ind*con/ ind*cost	0.15510*** (0.0000)	0.03242	0.0075*** (0.0000)	0.00018
Model statistics				
N (observations)	5932		5932	
LL_β	-3849.935		-3853.052	
$\chi^2_{,2\ MML}$ Inclusion of oth*cost and ind*cost versus no inclusion in model	23.894*** (0.0000)		17.660*** (0.0001)	
AIC	1.30005		1.30110	
BIC	1.30681		1.30786	

***=significant at 1% level, **=significant at 5% level, *=significant at 10% level; p-values in parentheses;

Table 7: Model results of a panel MML model specification including a variable reflecting subjective views on strategic behavior

	Panel MML PoolRB *oth*con*, *ind*con*		Panel MML PoolRB *oth*cost*, *ind*cost*	
Variable	Coefficient	Standard error	Coefficient	Standard error
Nonrandom parameter				
constant	0.63508***	0.20663	0.77771*	0.13780

	(0.0021)		(0.0000)	
oth*con/ oth*cost	0.04987	0.17542	-0.00144	0.00145
	(0.7762)		(0.3219)	
ind*con/ ind*cost	0.18727*	0.10643	0.00205**	0.00091
	(0.0785)		(0.0234)	
Random parameter				
cost	-0.02035***	0.00160	-0.02033***	0.00205
	(0.0000)		(0.0000)	
area of land	0.05038***	0.00515	0.05090***	0.00519
	(0.0000)		(0.0000)	
access	0.20225***	0.06640	0.20372***	0.06686
	(0.0023)		(0.0023)	
Standard deviation				
cost	0.04886***	0.00407	0.04903***	0.00412
	(0.0000)		(0.0000)	
area of land	0.08804***	0.00785	0.08899***	0.00794
	(0.0000)		(0.0000)	
access	1.23361***	0.17318	1.25044***	0.17310
	(0.0000)		(0.0000)	
Model statistics				
N (observations)	5932		5932	
LL_{β}	-3226.612		-3225.330	
$\chi^2_{,2\ MML}$	3.298		5.862*	
Inclusion of oth*cost and ind*cost versus no inclusion in model	(0.1922)		(0.0533)	
AIC	1.09090		1.09047	
BIC	1.10105		1.10062	

***=significant at 1% level, **=significant at 5% level, *=significant at 10% level; p-values in parentheses;

Hence, these results justify the rejection of H_0^2, whereas H_0^3 cannot be rejected.

Hypothesis 4

To test H_0^4, estimates for PoolRB MML and panel MML models were compared. The results are displayed in Table 8. The model fit improved when using a panel MML instead of a MML model as shown by the differences in the AIC and BIC. All attribute parameter estimates are statistically significantly different from zero at the 1% level and the 5% level, respectively and have the expected sign. The standard deviations of the random parameters are statistically significantly different from zero for (p=0.0000). This suggests considerable unobserved heterogeneity in preferences towards the choice attributes, particularly in case of the *access* parameter, which was statistically insignificant in the MNL model[29]. These results indicate that sequential binary elicitation formats induce correlations of the error components across choice tasks within respondents. Misspecifying the model by ignoring the panel character of

[29] To investigate further the access parameter a LC model was estimated. The preferred model containing three classes was chosen on the basis of the AIC, the BIC and on the significance of class membership probabilities. The model results disclose that about 40% of the respondents value access positively, about 34% prefer to not allow any access opportunities, and for about 27% the access variable is irrelevant. This bipolar distribution thus provides an explanation for the insignificance of the access parameter in the MNL model: positive and negative valued access parameters cancel each other out.

data leads to significantly different results in terms of attribute parameter significance and model fit.

Hence, H_0^4 is rejected.

Effects of awareness of having repeated choice

Hypothesis 5

In order to test H_0^5 we firstly investigate choice shares of non-zero cost options of SB and CrossSB1. The percentage of choosing any non-zero cost option was 56% for SB as opposed to 53% for CrossRB1. This difference of about 5% between the two split samples is not statistically significant at the 5% level (p=0.5917) using a chi-square test. These results do not provide statistically significant evidence of effects introduced by repeated choice tasks.

To further test the effects of choice set awareness, differences in the attribute and scale factors between SB and CrossRB1 were investigated (see Table 9). Comparing the attribute parameter estimates and the scale factors of SB with CrossRB1 yields no statistically significantly difference in either ($p_{attribute}$=0.6413, p_{scale}=1.0000). The WTP estimates for SB and CrossRB1 are reported in Table 3. A Poe test did not reveal statistically significant differences in WTP (p=0.99) between SB ($6.43) and CrossRB1 ($6.20) and the 95% confident interval are similar ($3.42-$13.08; $3.36–$11.56)[30]. These results indicate that the knowledge of the possibility of making multiple choices does not affect choice behavior.

Table 8: Model results for PoolRB using a MML and panel MML specification

	MML PoolRB		Panel MML PoolRB	
Variable	Coefficient	Standard error	Coefficient	Standard error
Nonrandom parameter				
Constant	0.51596** (0.0220)	0.22520	0.78259*** 0.0000)	0.13830
Random parameter				
Cost	-0.01171** (0.0134)	0.00474	-0.02057*** (0.0000)	0.00162
Area of land	0.03002* (0.0178)	0.01267	0.05095*** (0.0000)	0.00521
Access	0.09012 0.1909)	0.06890	0.20451*** (0.0023)	0.06699
Standard deviation				
Cost	0.02365* (0.0589)	0.01252	0.04930*** 0.0000)	0.00410

[30] 95% confidence intervals in parentheses based on the 2.5th and 97.5th percentile of the simulated WTP distribution. In comparison to the delta method, this method does not imply a normal distribution.

Area of land	0.05920** 0.0491)	0.03009	0.08969*** 0.0000)	0.00794
Access	0.59179 0.3176)	0.59217	1.25072*** 0.0000)	0.17506
Model statistics				
N (observations)	5932		5932	
LL_{MNL}	-3861.882		-3861.882	
LL_{β}	-3854.004		-3228.261	
$\chi^2_{,3\ (\text{MNL vs. MML})}$	15.76 (0.0013)			
$\chi^2_{,3\ (\text{MNL vs. panel MML})}$			1276.24 (0.0000)	
AIC	1.30175		1.09078	
BIC	1.30965		1.09867	

***=significant at 1% level, **=significant at 5% level, *=significant at 10% level; p-values in parentheses; 95% confidence intervals in parentheses based on the 2.5^{th} and 97.5^{th} percentile of the simulated WTP distribution. In comparison to the delta method, this method does not imply a normal distribution.

Table 9: Test results for equality for attribute and scale factor for SB and CrossSB1

LL SB	LL CrossRB1	LL Pooled[a]	LR-test[b] (5d.f.)	Reject $H_0:\beta_i=\beta_j$	Scale ratio λ_i/λ_j	LL Pooled[c]	LR-test[d] (1 d.f.)	Reject $H_0:\lambda_i=\lambda_j$
-965.8225	-995.738	-1963.252	0.6413	no	1.0000	-1963.252	1.0000	no

[a] Pooled MNL model allowing varying scale factor;
[b] Log-likelihood ratio test, test statistics $-2(LL_{pool} - (LL_1+LL_2))$ with d.f. k+1, where k is the number of parameters including the constant is asymptotically chi-square distributed;
[c] Pooled MNL model assuming equal scale factors in both split samples;
[d] Log-likelihood ratio test, test statistics $-2(LL_{equalscale} - (LL_{varyingscale}))$ with 1 d.f.; is asymptotically chi-square distributed

Based on the overall results, H_0^5 was not rejected indicating either that awareness of having repeated choices does not induce strategic behavior or that the information given was not sufficient to create respective opportunities.

5 Conclusion

The main objective of this study was to extend the research on effects of alternative elicitation formats in DCE. A split sample approach based on field surveys was conducted using a single binary elicitation format with a majority vote provision rule as the baseline. In particular, this paper explored (1) whether a sequential binary elicitation format affects choices, (2) whether repeated choice tasks per respondent introduce correlated error components across choice questions, and (3) whether awareness of having multiple choices influences choice behavior. The results indicate that repeated choice tasks affect choice. However, the results are ambiguous.

The results of examining choice shares provide no statistically significant evidence of any effects. However, the investigation of the subjective views of the respondents indicates that respondents take previous information and choices into account, and thus may exploit strategic opportunities while progressing through the choice task. However, this study did not find evidence that respondents additionally condition their preferences on the expectations about the choices of other survey participants.

The econometric results obtained by comparing choice experiments based on a single as opposed to a repeated binary format suggest institutional learning rather than strategic behavior. However, the presented econometric results are based on restricted MNL models. Bateman et al. (2008) showed that differences in scale are at least partially a result of preference heterogeneity. That is, a difference in scale induced by preference heterogeneity will vanish if a model specification is used that allows accounting for preference heterogeneity.

Furthermore, the potentially incentive compatible baseline was compromised by reduced consequentiality of the survey format, which may have confounded comparisons across split samples. This provides an alternative explanation for the ambiguous results associated with the question of whether strategic behavior results from repeated choice. However, the incentive properties to answer such follow-up questions truthfully are unknown. That is, the answers may be strategically biased and may thus not reflect the actual opinions of respondents. Further approaches capable of testing these issues are required to answer these questions.

Additionally, this study suggests that ignoring the correlation of error components across choice questions can have a profound impact on model outcomes. Consequently, results that are based on models that do not account for the panel character of the data may be misleading.

This research further implies that awareness of having multiple choices does not affect choice behavior. This result contrasts with findings of Bateman et al. (2008) who found choice set awareness to be significant. The differences may be explained by the different questionnaire designs. Bateman et al. provided respondents with information about all possible attribute levels, whereas in the study presented in this paper only information about the attributes and a note explaining that choice options are based on different attribute levels were given to respondents. One possible explanation is, therefore, that the information provided may not have been detailed enough to trigger measurable strategic behavior. Further testing is needed to explore this issue.

In summary, we provide evidence for effects induced by institutional learning and effects that may be explained by either strategic response or value learning. However, we did not find any indications for strategic behavior caused by choice set awareness. The choice between a repeated and a single elicitation format may thus imply a trade-off between decreased choice accuracy and potentially increased strategic behavior due to an incentive incompatible mechanism. Further research is needed to explore strategic behavior induced by incentive incompatible elicitation formats using alternative approaches that are not compromised by a confounded baseline and that allow the use of less restrictive model specifications. Such research should also investigate the effects of varying incentives induced by the order in which choice questions are presented to respondents.

REFERENCES

ABS, 2006. 'Census data', Australian Bureau of Statistics.

Arrow, K., 1951. *Social Choice and Individual Values*Wiley, New York.

Bateman, I.J., Carson, R.T., Day, B., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P., 2008. *Choice set awareness and ordering effects in discrete choice experiments,* Research Centre for Social and Economic Research on the Global Environment

Bateman, I.J., Day, B., Georgiou, S. and Lake, I., 2006. 'The aggregation of environmental benefit values: welfare measures, distance decay and total WTP', *Ecological Economics,* 60:450-460.

Bennett, J. and Blamey, R., 2001. *The choice modelling approach to environmental valuation*Edward Elgar Publishing Limited, Massachusetts.

Bliemer, M.C.J., Rose, J.M. and Hess, S., 2008. 'Approximation of Bayesian efficiency in experimental choice designs', *Journal of Choice Modelling,* 1(1):98-127.

Braga, J. and Starmer, C., 2005. 'Preference anomalies, preference elicitation and the discovered preference hypothesis', *Environmental and Resource Economics,* 32(1):55-89.

Brownstone, D. and Train, K., 1999. 'Forecasting new product penetration with flexible substitution patterns', *Journal of Econometrics,* 89(1):109-129.

Carson, R.T. and Groves, T., 2007. 'Incentive and informational properties of preference questions', *Environmental and Resource Economics,* 37(1):181-210.

Collins, J.P. and Vossler, C.A., 2009. 'Incentive compatibility tests of choice experiment value elicitation questions', *Journal of Environmental Economics and Management,* 58:226-235.

Ferrini, S. and Scarpa, R., 2007. 'Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study', *Journal of Environmental Economics and Management,* 53:343-362.

Gibbard, A., 1973. 'Manipulation of voting schemes', *Econometrica,* 41:587-601.

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

Greene, W. and Hensher, D., 2006. 'Accounting for heterogeneity in the variance of unobserved effects in mixed logit models.', *Transportation Research Part B: Methodological,* 40:75-92.

Greene, W. and Hensher, D., 2007. 'Heteroscedastic control for random coefficients and error components in mixed logit', *Transportation Research Part E: Logistics and Transportation Review,* 43:610-623.

Greene, W.H., 2007. *Nlogit version 4.0 reference guide*Econometric Software, Inc., Castle Hill.

Greene, W.H., 2008. *Econometric analysis*Pearson/Prentice Hall, Upper Saddle River, N.J.

Greene, W.H., Hensher, D.A. and Rose, J., 2006. 'Accounting for heterogeneity in the variance of unobserved effects in mixed logit models', *Transportation Research Part B: Methodological,* 40(1):75-92.

Hensher, D., Louviere, J. and Swait, J., 1999. ' Combining sources of preference data', *Journal of Econometrics,* 89: 197-221.

Hensher, D., Rose, J.M. and Greene, W.H., 2005. *Applied choice analysis : a primer* Cambridge University Press, Cambridge.

Hensher, D.A. and Greene, W.H., 2003. 'The Mixed Logit model: The state of practice', *Transportation,* 30:133-176.

Holmes, T.P. and Boyle, K.J., 2005. 'Dynamic learning and context-dependence in sequential, attribute-based, stated preference valuation questions', *Land Economics,* 81(1):114-126.

Hurwicz, L., 1960. 'Optimality and informational efficiency in resource allocation processes', in K. Arrow, S. Karlin and P. Suppes (eds), *Mathematical Methods in the Social Sciences*, Stanford University Press

Hurwicz, L., 1972. 'On informationally decentralized systems', in R. Radner and C.B. McGuire (eds), *Decision and Organisation*, Amsterdam

Krinsky, I. and Robb, A.L., 1986. 'On approximating the statistical properties of elasticities', *Review of Economics and Statistics,* 68:715-719.

Louviere, J. and Eagle, T.C., 2006. 'Confound it! That Pesky Little Scale Constant Messes Up Our Convenient Assumptions!', *CenSoC Working Paper No. 06-002*, Centre for the Study of Choice, University of Technology, Sydney.

Louviere, J., Hensher, D. and Swait, J., 2000. *Sated choice methods: analysis and applications*Cambridge University Press, Cambridge, UK ; New York, NY, USA.

Mas-Colell, A., Whinston, M.D. and Green, J.R., 1995. *Microeconomic theory*Oxford University Press, New York Oxford.

Maskin, E., 1977. 'Nash equilibrium and welfare optimality', *Summer workshop of the Econometric Society* Published 1999 in the Review of Economic Studies 66, 23-38, Paris.

Mazur, K. and Bennett, J., 2010. 'Framing for incentive compatibility in choice modelling', *54st Annual Conference of the Australian Agricultural and Resource Economics Society*, Adelaide, SA.

McFadden, D., 1974. 'Conditional logit analysis of qualitative choice behaviour', in P. Zarembka (ed.), *Frontiers in econometrics*, Academic Press, New York

McFadden, D., 1980. 'Econometric models for probabilistic choice among products', *Journal of Business,* 53:13-29.

McFadden, D. and Train, K., 2000. 'Mixed MNL models for discrete response', *Journal of Applied Econometrics,* 15:447-470.

McNair, B., Bennett, J. and Hensher, D., 2010. 'Strategic response to a sequence of discrete choice questions', *54th Annual conferences of the Australian Agricultural and Resource Economics Society*, Adelaide.

Moulin, H., 1994. 'Social choice', in R.J. Aumann and H. S. (eds), *Handbook of game theory with economic applications*, Amsterdam

Muller, E. and Satterthwaite, M.A., 1985. 'Strategy-proofness: the existence of dominant-strategy mechanisms', in L. Hurwicz, D. Schmiedler and H. Sonnenschein (eds), *Social Goals and Social Organization*, Cambridge University Press, Cambridge

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

Poe, G.L., Clark, J., Rondeau, D. and Schulze, W.D., 2002. 'Provision point mechanisms and field validity tests of contingent valuation', *Environmental and Resource Economics,* 23:105-131.

Poe, G.L., Giraud, K.L. and Loomis, J.B., 2005. 'Computational methods for measuring the difference of empirical distributions', *American Journal of Agricultural Economics,* 87(2).

Racevskis, L.A. and Lupi, F., 2008. 'Incentive compatibility in an attribute-based referendum model', *American Agricultural Economics Association Annual Meeting*, Orlando.

Revelt, D. and Train, K., 1998. 'Mixed logit with repeated choices: households' choices of appliance efficiency level', *Review of Economics and Statistics,* 80(4):647-657.

Rose, J.M. and Bliemer, M.C.J., 2005. 'Designing efficient data for stated choice experiments: accounting for socio-dempgraphic and contextual effects in designing stated choice experiments', *11th International Conference on Travel Behaviour Research*, Kyoto, Japan.

Rose, J.M. and Bliemer, M.C.J., 2008. 'Stated preference experimental design strategies', in D.A. Hensher and K.J. Button (eds), *Handbook of transport modelling*, Pergamon

Rose, J.M., Bliemer, M.C.J., Hensher, D.A. and A., C., 2008. 'Designing efficient stated choice experiments in the present of reference alternatives', *Transportation Research Part B,* 42 395–406.

Rose, J.M., Hess, S., Bliemer, M.C.J. and Daly, A., 2009. 'The impact of varying the number of repeated choice observations on mixed multinomial logit models'.

Samuelson, P., 1954. 'The pure theory of public expenditure', *Review of Economics and Statistics,* 36:387-389.

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,* 10:187-217.

Scheufele, G. and Bennett, J., 2010. 'Ordering effects and response strategies in discrete choice experiments', *54th Annual Conference of the Australian Agricultural and Resource Economics Society*, Adelaide, SA.

Simonson, I. and Tversky, A., 1992. 'Choice in context: trade-off context and extremeness aversion', *Journal of Marketing Research,* 29:281-295.

Swait, J. and Adamowicz, W., 2001. 'Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice', *Organizational Behavior and Human Decision Processes,* 86(2):141–167.

Swait, J. and Louviere, J., 1993. 'The role of the scale parameter in the estimation and comparison of multinominal logit models', *Journal of Marketing Research,* 30(3):305-314.

Train, K., E, 2000. 'Halton Sequences for Mixed Logit', *Paper E00-278*, Institute of Business and Economics, University of California, Berkeley.