The Australian Journal of Agricultural and Resource Economics

Journal of the Australian Agricultural and Resource Economics Society

The Australian Journal of Agricultural and Resource Economics, 57, pp. 214–233

Effects of alternative elicitation formats in discrete choice experiments*

Gabriela Scheufele and Jeff Bennett†

An elicitation format prevalently applied in discrete choice experiments (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 outcomes through order effects. The study reported in this article employs a split sample approach based on field surveys to expand the research on effects of repeated-binary DCE elicitation formats. A single-binary elicitation format is used as the baseline. Our results indicate that choice outcomes may vary across the single-binary and repeated-binary elicitation formats. The choice between the two elicitation formats may imply a trade-off between decreased choice precision in the single-binary and order effects that may be explained by strategic misrepresentation of preferences, cost uncertainty effects and reference effects in the repeated-binary elicitation format.

Key words: discrete choice experiments, elicitation format, incentive compatibility, order effects.

1. Introduction

Discrete choice experiments (DCE) are increasingly being used to estimate nonmarket values as inputs in cost–benefit analysis to ensure improved efficiency in resource allocation (Bennett and Blamey 2001; Bateman *et al.* 2006). DCE involve respondents making trade-offs between attributes that describe nonmarket goods and services. A variety 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 (e.g. Hensher and Rose 2007; Horne 2008; Kontoleon and Yabe 2008). 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

* Funding for the research presented in this article was provided by the Australian Department of Sustainability, Environment, Water, Population and Communities under the Commonwealth Environment Research Facility.

† Gabriela Scheufele, Jeff Bennett (email: gabriela.scheufele@anu.edu.au) are at Crawford School of Public Policy, The Australian National University, Canberra, ACT, Australia.

doi: 10.1111/j.1467-8489.2012.00612.x

single-binary choice set.[1] However, empirical evidence suggests that repeated choice tasks may influence choice outcomes through order effects (e.g. Morrison 2000; Carlsson and Martinsson 2001; Swait and Adamowicz 2001; McNair *et al.* 2011; Day *et al.* 2012). The main objective of the study presented in this article is to expand the research on order effects through an examination of effects of alternative DCE elicitation formats focusing on effects of repeated-binary formats. We employ a split sample approach based on field surveys to compare the effects of using repeated-binary with single-binary elicitation formats. Specifically, the following research questions are explored:

1. Does the elicitation format (repeated-binary versus single-binary) affect choice outcomes?
2. Does in advance awareness of having repeated choices influence choice outcomes?

We hypothesise that choice outcomes differ across the two alternative elicitation formats and that these differences may be caused by order effects.

The choice between the two alternative elicitation formats may thus imply a trade-off between decreased choice precision in the single-binary and increased *strategic misrepresentation of preferences*, *cost uncertainty effects* and *reference effects* in the repeated-binary elicitation format.

The next section reviews the literature on effects associated with alternative choice formats of DCE and derives the hypotheses. This is followed by an outline of the survey logistics, empirical application, research design, experimental design, econometric model and hypotheses testing. The subsequent section is used to present the results. In the last section, the results are discussed and conclusions drawn.

2. Literature review

The repeated choice format is the commonly used style of elicitation used in DCE. Under this format, respondents face a sequence of choice tasks. Repeated choice formats increase the statistical efficiency of the data for a given number of respondents. However, an increasing number of empirical studies suggest that repeated choice formats may induce order effects (e.g. Morrison 2000; Carlsson and Martinsson 2001; Swait and Adamowicz 2001; Holmes and Boyle 2005; McNair *et al.* 2011; Day *et al.* 2012). Day *et al.* (2012) provide a comprehensive discussion of order effects. They define order effects as '[ ]; a term that embraces a variety of phenomena in which system-

[1] A single-binary elicitation format asks respondents to choose between two choice options once. A single–multiple elicitation format requires respondents to make one choice between more than two choice options presented in one single choice set. A repeated-binary elicitation format asks respondents to make repeated trade-offs between two choice options. A repeated-multiple elicitation format offers respondents repeated choices between more than two choice options presented in a sequence of choice sets.

atic changes in expressed preferences are observed along a sequence of valuation tasks' (Day *et al*. 2012, p. 73). They differentiate between position-dependent order effects (effects relating to the position in the sequence of choice tasks) and precedent-dependent order effects (effects relating to the nature of options in previous choice tasks). Moreover, they discuss a range of theories that may explain precedent-dependent order effects (*strategic misrepresentation of preferences*, *anchoring effects* and *reference effects*) and position-dependent order effects (*institutional learning*, *fatigue effects*, *failing credibility*, *cost uncertainty effects* and *preference learning*) in repeated choice formats.

The efficiency of policy decisions concerning public good resource allocation depends on information based 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 (Mas-Colell *et al*. 1995) and may result in *strategic misrepresentation of preferences*. The analysis of demand revealing mechanisms is the province of mechanism design theory (Hurwicz 1960). A mechanism is defined as incentive compatible if revealing private information truthfully is the dominant strategy for all participants.[2] The theorems of Gibbard (1973) and Satterthwaite (1975) provide a theoretical foundation to analyse the incentive-compatibility properties of mechanisms used in DCE. Both theorems show that all nondictatorial mechanisms other than the single-binary choice format are generically incentive incompatible.[3,4,5] That is, choice formats that ask respondents to choose between more than two options per choice task and/or to respond to a sequence of choice tasks are not incentive compatible.

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 'multiple-binary' elicitation format, that is, it increases the probability that respondents reveal their true preferences (see, e.g. Collins and Vossler 2009). Using random draws as a base for policy decisions

[2] A binding referendum between two contestants in an election is an example for an incentive-compatible mechanism. Carson *et al*. (1997) suggested that using an advisory referendum instead of a binding referendum maintains the incentive-compatibility properties of the mechanism. Green and Laffont (1978) showed that this is true not only for population-based, but also for sample-based referenda. These two findings are critical for choice experiments. Choice experiments are mostly based on sample data and, in cases eliciting preferences for public goods, often simulate an advisory referendum.

[3] 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 article excludes cases where choice options are associated with disutility or where none is the status quo.

[4] A DCE with a binary elicitation format differs from a contingent valuation (CV) study in that in the former all attribute levels vary across choice sets whereas in the latter variation is limited to the cost variable.

[5] If the provision rule does not require only one option to be selected and a multiple option provision rule can be applied, potential incentive compatibility can be achieved (see Maskin 1977; Muller and Satterthwaite 1985). However, policy decisions about the provision of public goods often require the choice of a single option from a number of mutually exclusive alternatives.

concerned with public goods, however, may reduce their credibility (Carson and Groves 2007).

Choice dependency across respondents is one effect of repeated-binary elicitation formats. The literature on incentive compatibility proposes that respondents who are presented with repeated-binary choice tasks condition their preferences on expectations about the choices of other survey participants (see, e.g. Carson and Groves 2007). Accordingly, some respondents may choose a less-preferred option across choice sets if they believe that their most preferred option has little chance of being selected for policy implementation.[6] Repeated-binary elicitation formats with a plurality vote implementation additionally imply that respondents may exploit strategic opportunities by including information about previous choice sets and choice decisions into subsequent choices (see, for instance, Carson and Groves 2007). As a result, some respondents may choose a less-preferred option in one or more binary choice questions. Evidence of such an effect in repeated-binary DCE is presented, for instance, by Holmes and Boyle (2005). Carson and Groves (2007) suggest that previous and successive choice sets may contain alternative prices for the same or similar level of provision of a particular good or, vice versa, the same or similar cost for alternative levels of provision of a particular good. As argued by Carson and Groves (2007), this may trigger respondents to 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 cost. Hence, repeated choice may cause respondents to become aware of and use strategic opportunities to misrepresent their preferences. Alternatively, as further suggested by Carson and Groves (2007), respondents may react to 'inconsistent' pricing by questioning either survey credibility or cost certainty.[7] As argued by Day *et al*. (2012), *failing credibility* may result in respondents increasingly favouring the status-quo choice option and/or increasing randomness of choice, whereas *cost uncertainty effects* may lead to a decreasing marginal willingness to pay (WTP) along the sequence of choice tasks.

Carson and Groves (2007) add to the discussion about incentive-compatible elicitation formats by arguing 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 respon-

[6] This also holds for a single–multiple elicitation format where respondents are offered one single choice set containing three or more choice options. If a plurality vote is used, such an elicitation format becomes a single-binary elicitation format with the two remaining choice options being those that the respondent perceives to be the options most preferred by other respondents. However, as argued by Moulin (1994), if these perceptions about other respondents' preferences are the same across all respondents, a single–multiple elicitation format may be potentially incentive compatible.

[7] As pointed out by an anonymous referee, the effects of non-credible designs and inconsistent pricing can be addressed by design constraints, though at an efficiency cost.

dent, and respondents have to believe that their choices have an impact on the outcome. If this is not the case, respondents may regard choice options as equally nonbeneficial and may or may not disclose their true preferences. Carson and Groves (2007) additionally suggest associated factors that may affect respondents' choices, including the properties of the payment vehicle, plausibility of the choice questions, credibility of the policy scenario and comprehensibility of the choice tasks. Surveys lacking a payment vehicle that respondents perceive as coercive induce free-riding behaviour. Implausible choice tasks may result in the respondent inferring a different choice set to that which is presented in the questionnaire. If respondents are presented with a noncredible 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. Once again, respondents may answer a question they think has been asked instead of the one the researcher intended to have answered (Carson and Groves 2007).

Braga and Starmer (2005) proposed a process whereby respondents become more and more acquainted with the choice context, the offered good and the choice task as they progress through the sequence of choice questions (*institutional learning*). Typically, *institutional learning* is assumed to affect the precision of responses in representing preferences as reflected by the scale parameter[8] rather than changing preferences. As respondents progress through the choice questions, their responses are hypothesised to become more precise in expressing their preferences (increase in the scale parameter) until fatigue sets in and precision declines (decrease in the scale parameter). Swait and Adamowicz (2001) discuss this as 'smaller noise to signal ratio' and 'larger noise to signal ratio', respectively. That is, *institutional learning* and *fatigue effects* may be indicated by decreasing/increasing randomness of choice, respectively (Day *et al*. 2012).

Plott (1996) proposed that respondents may 'discover' their true underlying preferences through a learning process rather than possessing stable preferences (*preference learning*). Such learning processes are expected to change preferences and thus parameter estimates in DCE. As suggested by Day *et al*. (2012), *preference learning* may manifest itself in changes in preferences for noncost attributes and decreasing randomness in choice. A range of studies has explored learning effects in stated preference elicitation (e.g. Morrison 2000; Carlsson and Martinsson 2001).

Anchoring effects and *reference effects* are alternative theories that may explain order effects in repeated choice formats (Day *et al*. 2012). If respondents have ill-formed and malleable preferences (Ready *et al*. 1995; Kooten *et al*. 2001), they may think that the first choice task provides cues on which they may anchor their preferences (Ariely *et al*. 2003, 2006). Effects associated with such choice behaviour are *anchoring effects*. The theory behind

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

reference effects is based on Wicksteed (1910) who suggested that consumer choice is influenced by consumers' judgments of what are 'good deals' and 'bad deals'. Respondents in a DCE may form preferences based on previous choice tasks. That is, respondents' choices in subsequent choice tasks may be influenced by references to previous choice tasks.

The evidence for order effects induced by repeated choice formats has been growing. Day *et al*. (2012) investigate position-dependent and precedent-dependent order effects of repeated-binary choice formats. They find robust evidence for both types of ordering effects. Moreover, they analyse whether advanced disclosure (in contrast to stepwise disclosure) of choice tasks influences the probability of order effects. They find precedent-dependent order effects in both the advanced and the stepwise disclosure treatments, but position-dependent order effects are observed primarily in the latter.

The research of McNair *et al*. (2011) suggests that repeated-binary choice decreases estimates of expected WTP compared with a single-binary choice format. Furthermore, they find that even though preferences stated in the first choice task of a repeated-binary choice format are not significantly different from a single-binary choice format, subsequent choices may be influenced by cost levels observed in previous choice tasks. This effect may be explained by the ordering of alternative cost levels offered across a sequence of four choice questions. They further discuss three possible behavioural explanations for these effects: strategic misrepresentation of preferences, reference point revision and cost-driven value learning.

Racevskis and Lupi (2008) used a split-sample design to explore the effect of a single-binary versus a repeated-binary elicitation format. They pooled the data from the two elicitation formats in two different ways and found a significant difference between the two models. The first model included generic attributes (a restricted model assuming that the attribute parameter estimates are equal across the two sample splits), whereas the second model included sample split–specific attributes (an unrestricted model assuming that the attribute parameter estimates are not equal across the two sample splits). Their study was focused on a comparison of the model fit between the two sample splits and did not explore differences in attribute or scale parameter estimates.

The study presented in this article complements the existing research of order effects induced by repeated-binary DCE formats.

The fist research question 'Does the elicitation format (repeated-binary versus single-binary) affect choice outcomes?' is tested using the following hypothesis:

H_0^1: Choice outcomes obtained by a single-binary elicitation format are the same as those obtained by a repeated-binary elicitation format that contains four choice sets.

The second research question 'Does in advance awareness of having repeated choices influence choice outcomes?' is tested using the following hypothesis:

H_0^2: Choice outcomes obtained by a single-binary elicitation format are the same as those obtained by the first choice question of repeated-binary elicitation formats that contain four choice sets.

In contrast to Day *et al*. (2012) who investigate order effects using repeated-binary elicitation formats only, we employ a split-sample approach comparing a single-binary with repeated-binary DCE.

Furthermore, we are unaware of any work other than the research of Racevskis and Lupi (2008) and McNair *et al*. (2011) that has tested order 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 (2008) by exploring additional outcome dimensions, testing for choice set awareness and adjusting the number of observations in the choice experiment with a single-binary elicitation format to reduce potentially confounding influences of varying sample sizes. While McNair *et al*. (2011) focuses on a public good with private elements, we investigate a pure public good context. As the incentive properties of elicitation formats related to public and private goods are not the same,[9] we contribute to the literature by analysing the effects of alternative elicitation formats used in DCE that are focused on pure public goods. Finally, this study uniquely includes follow-up questions to examine whether the potentially incentive-compatible baseline, the single-binary elicitation format, has been affected by limited consequentiality of the survey format, the plausibility of the choice questions, the credibility of the policy scenario and the comprehensibility of the choice task.

3. Empirical application

The hypotheses are tested using data from a DCE concerned with estimating the preservation values of a natural area, using Nadgee Nature Reserve as the context. The reserve was described to respondents as a natural area of land using the features of Nadgee Nature Reserve, even though its identity was not revealed.[10] The identity was not revealed because the scenario depicted in the questionnaire was incompatible with the actual setting. Respondents were told that the area of land is a pristine wilderness located on the South Coast of New South Wales, which covers an area of 17,116 ha. A random sample was drawn from an internet panel of the Sydney population.[11] The overall response rate was 34 per cent. Nonresponse was defined as 'invited but not participated' (55 per cent), 'participated but below five minutes completion

[9] For a discussion see, for example, Carson and Groves (2007).

[10] This may have increased the hypothetical nature of the questionnaire and thus may have affected the results.

[11] Quotas with respect to age and gender based on ABS statistics were used to select the samples. Difficulties in achieving the quota levels in the samples of low and high age groups resulted in some statistically significant differences across all sample splits in some sociodemographic variables. This may have influenced the results.

Table 1 Attributes and attribute levels

Attribute	Attribute level	Coding
Cost	\$0	Numerical
	\$50	
	\$100	
	\$200	
	\$300	
Area of land	30% (4,200 ha)	Numerical
Described as the percentage (hectares) of area of land that remains in its current undeveloped condition.	50% (7,000 ha)	
	70% (9,800 ha)	
	100% (14,000 ha)	
Access for minimum impact recreation[a]	Yes	1
Described as whether or not people are allowed to use the undeveloped parts of the area of land for minimum impact recreation opportunities.	No	−1

Notes: [a]The 'access' variable has been effects-coded to avoid confounding the base level of the attribute with the constant term.

time' (2 per cent) and 'participated but dropped out before completion' (9 per cent). The survey material was developed using expert opinion, focus groups and a pilot survey (sample size: 150 respondents). In the survey, respondents were told that funds had to be raised through additional government taxes and charges to enable the government to purchase the land and thus conserve the area; otherwise, development would be allowed. A plurality vote was used as the provision rule: '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 were not allowed to go back to a previous choice question. The programming of the online survey prevented respondents from revising their choices. The management options offered in the DCE were described by three attributes with five, four and two levels, respectively (see Table 1). The survey included questions designed to check the consequentiality, plausibility, credibility and understandability of the survey material (Table 2).[12]

Split-sample treatments were used. The splits only differed in the number of choice sets per respondent, the choice set order and the wording of some explanations and instructions so necessitated. It is assumed that this change of wording of the choice questions across the sample splits[13] did not cause any statistically significant difference in the choice outcomes. The results are conditional on this maintained assumption.

All sample splits used the same experimental design with a total of 16 choice sets each containing two choice options: One invariant zero cost choice

[12] Details are provided in the results section.

[13] 'In the next question we want you to make a choice [ ].'/'In the next four questions we want you to make some choices [ ].'; 'When making your choice [ ].'/'When making your choices [ ].'

Table 2 Consequentiality, plausibility, credibility and comprehensibility of DCE survey[a]

	SD (%)	D (%)	N (%)	A (%)	SA (%)		SD (%)	D (%)	N (%)	A (%)	SA (%)
	'I am interested in the future management of the area of land'						'I understood all the information'				
RB	1	5	27	53	14	RB	1	5	21	58	15
SB	1	4	26	55	14	SB	1	5	20	59	16
	'I understood the concept of making choices between alternative management options'						'I found making choices between alternative management options confusing'				
RB	1	4	19	60	17	RB	11	41	32	16	1
SB	1	2	18	63	16	SB	11	41	33	14	1
	'I do not believe that recreation – even if it is low impact – would cause only minor environmental changes'						'I found that the available management options made sense'				
RB	5	27	31	30	7	RB	2	11	37	46	5
SB	5	25	30	33	7	SB	2	13	36	44	5
	'I believe that my choices will have an impact on how the area of land will be managed in the future'						'I believe that the chosen plan for management will be implemented'				
RB	4	18	40	33	5	RB	2	14	57	24	3
SB	4	19	37	34	6	SB	3	17	55	23	3

Notes: [a]All questions were based on a five-point Likert scale: 'strongly disagree' (SD), 'disagree' (D), 'neither disagree nor agree' (N), 'agree' (A), 'strongly agree' (SA).

option that was available in each choice set (do-nothing option) and one nonzero cost choice option that varied across choice sets (change option). Figure 1 provides a choice set example.[14]

For the single-binary elicitation format sample split (SB), respondents were randomly assigned to one of the 16 choice sets by dividing the experimental design into 16 blocks each containing one choice set. That is, each respondent was asked to answer one choice question only. For the repeated-binary elicitation format sample split (RB), respondents were randomly assigned to four of the same 16 choice sets by dividing the experimental design into four blocks of four choice questions. That is, each respondent received four choice questions. The order of the choice sets in each block was rotated four times to account for the potentially confounding influences of differences in choice set order when comparing SB and RB sample splits.[15] This rotation resulted in four RBx sample splits each containing four blocks (16 choice sets) but in a different order (RB1, RB2, RB3, RB4). Respondents in both the SB and the RBx sample splits were told in advance how many choice questions they would receive and received information on the attribute types. Information on attribute levels was not provided in advance.

A sixth sample split (FirstRB) was created by pooling the first choice question of RB1, RB2, RB3 and RB4. The choice set rotation described

[14] The questionnaire is available upon request.

[15] As pointed out by an anonymous reviewer, the initial choice made in the RB sample split may have affected the comparisons with the SB results through this difference in choice set order.

	Option A	Option B
My one-off household payment	$ 0	$ 200
Area left undeveloped	30% (4,200 ha)	30 % (4,200 ha)
Recreation opportunities	No	Yes
I would choose:	◯	◯

Figure 1 Choice set example.

previously guaranteed that each choice set was presented in the first position approximately the same number of times across RB1, RB2, RB3 and RB4. Consequently, the choice questions in FirstRB were the same as those in SB. The complete research design is illustrated in Table 3.

To avoid the confounding effects of having different numbers of observations across the SB, RBx and FirstRB sample splits, the number of respondents surveyed for the SB sample split (1444) was about four times that for sample splits RB1 (367), RB2 (371), RB3 (369) and RB4 (376).[16]

A Bayesian efficient experimental design (Sándor and Wedel 2001; Ferrini and Scarpa 2007) was used to generate the choice sets (D-efficiency). The data were analysed using conditional logit (CL) models (McFadden 1974).[17,18] A bootstrapping procedure applying 1000 draws (Krinsky and Robb 1986) was used to estimate marginal WTP (implicit prices) for the attribute parameters.

4. Hypothesis testing

To test H_0^1, we compare choice outcomes of a single-binary with a repeated-binary choice format using conditional logit (CL) models (McFadden 1974). Choice outcomes are defined as the attribute parameter estimates, scale parameter estimates and welfare measures (marginal WTP estimates).

H_0^2 is tested by comparing single-binary choice outcomes with the choice outcomes derived from the first choice questions of a repeated-binary elicitation format using CL models. Choice outcomes are defined as the attribute parameter estimates, scale parameter estimates and welfare measures (marginal WTP estimates).

[16] Rose *et al.* (2009) explored the statistical impact of panel data in discrete choice experiments by means of simulated data. They showed that an increase in the number of choice observations per respondent leads to less-biased estimates and larger *t*-ratios while holding sample size constant. An increase in sample size, though, decreases this advantage.

[17] Econometric models were estimated using NLOGIT4.1 (Greene 2007) and STATA10 (StataCorp 2007).

[18] Adding a quadratic term into the utility function did not improve the model fit.

Table 3 Research design – choice sets

Block (SB)	SB	Block (RBs)	RB1	RB2	RB3	RB4	First RB			
1	1	1	1	2	3	4	1	2	3	4
2	2		2	3	4	1				
3	3		3	4	1	2				
4	4		4	1	2	3				
5	5	2	5	6	7	8	5	6	7	8
6	6		6	7	8	5				
7	7		7	8	5	6				
8	8		8	5	6	7				
9	9	3	9	10	11	12	9	10	11	12
10	10		10	11	12	9				
11	11		11	12	9	10				
12	12		12	9	10	11				
13	13	4	13	14	15	16	13	14	15	16
14	14		14	15	16	13				
15	15		15	16	13	14				
16	16		16	13	14	15				

The analysis of both hypotheses is restricted to an estimation of a CL model specification. The estimation of a random parameter logit model for data obtained in the single-binary elicitation format is not feasible.[19]

5. Results

A series of chi-square tests were conducted to check for equivalence between each sample split and the population statistics using the 2006 census data (Australian Bureau of Statistics 2009). The sample splits are not representative of the households of Sydney, and care should be taken when interpreting the results on a population level.

Follow-up questions were 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. The results are displayed in Table 2. These results indicate that the incentive compatibility of the SB elicitation format may be compromised. That is, observed differences in choice outcomes between the SB and RBx sample splits may be influenced by limited consequentiality of the survey format, plausibility of the choice questions, credibility of the policy scenario and comprehensibility of the choice task.[20]

[19] A single choice observation per respondent obtained in the single-binary elicitation format may not allow the discovery of random parameters that are statistically significantly different from zero (see, e.g. Rose *et al*. 2009). 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.

[20] All respondents have been kept in the samples because defining the extent of inconsequentiality that would warrant exclusion is regarded as being too subjective.

5.1. Effects of repeated-binary choice – Hypothesis 1

To investigate H_0^1, CL models were estimated. The results for SB and RB1 (representing the RBx sample splits) are reported in Table 4. For all sample splits, the *cost* parameter estimates are statistically significantly different from zero (1 per cent significance level) and have the expected negative signs 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 (1 per cent significance level) and positive as expected, suggesting that a larger area of land provides higher utility than a smaller area, *ceteris paribus*. The *access* parameter estimate, however, is not statistically significantly different from zero in any but the RB2 sample split (5 per cent significance level).[21] The constant term is statistically significantly different from zero (1 per cent significance level) only in RB1 and has a positive sign. As the constant term was included into the do-nothing option, we find no evidence that respondents favour the do-nothing option more in the RBx than in the SB sample splits.

The marginal $WTP_{\text{area of land}}$ estimates[22] for SB and all RBx sample splits are reported in Table 5. Equivalence of marginal $WTP_{\text{area of land}}$ estimates was tested using a procedure suggested by Poe *et al*. (2002, 2005). We find a statistically significant higher marginal $WTP_{\text{area of land}}$ for SB than for RB1, RB2, RB3 and RB4.

Differences in the attributes and scale parameter estimates between SB and the RBx sample splits were explored using a test procedure suggested by Swait and Louviere (1993).[23] For a detailed discussion about the Swait–Louviere test and the confounding influence of the scale parameter in conditional logit models, see Louviere and Eagle (2006). The results are displayed in Table 6. We find statistically significant differences in attribute parameter estimates comparing SB to RB2, RB3 and RB4. The parameter vector and the scale parameter are confounded in CL models. A statistically significant difference in an attribute parameter estimate prevents a test for scale parameter estimate equality. Hence, only RB1 was tested in this regard. The hypothesis of equal scales was rejected ($p_{\text{scale}} = 0.0124$). The reduced relative scale parameter for RB1 suggests a less-precise choice as the scale parameter is inversely related to the variance of the error term. Whether this is also the case for the RB2, RB3

[21] As pointed out by an anonymous reviewer, one possible reason for the access parameter estimates not being statistically significantly different from zero may be not having revealed the name of the area of land.

[22] The marginal WTP estimate for the access parameter in RB2 is not statistically significantly different from zero.

[23] The relative scale parameter was estimated based on a heteroscedastic conditional logit model in STATA10 (see, e.g. Hensher *et al*. 1999).

Table 4 CL model results for RB1, SB and FirstRB

Variable	Repeated-binary (RB1)		Single-binary (SB)		First of repeated-binary (First RB)	
	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*** (< 0.0001)	0.00057	−0.00218*** (0.0001)	0.00056	−0.00224*** (0.0001)	0.00056
Area of land	0.01342*** (< 0.0001)	0.00214	0.01273*** (< 0.0001)	0.00217	0.01291*** (< 0.0001)	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 (< 0.0001)		52.21 (< 0.0001)		568.23 (< 0.0001)	
Adjusted ρ^2	0.05		0.02		0.22	
AIC	1.29595		1.34231		1.34826	
BIC	1.33560		1.35692		1.36257	

Notes: Variables significant at *10%, **5% and ***1%. p-values are displayed in parentheses. AIC, Akaike Information Criteria; BIC, Bayesian Information Criteria.

Table 5 Marginal WTP estimates for RB1, RB2, RB3, RB4, SB and First RB

	$WTP_{area\ of\ land}$	Standard error	CI $(95\%)_{WTP}$	*P*-value RB-SB First RB-SB (Test procedure by Poe)
RB1	\$2.98***	0.61584	\$1.88–\$4.34	0.0566*
RB2	\$2.41***	0.50173	\$1.48–\$3.46	0.0145**
RB3	\$2.51***	0.52867	\$1.62–\$3.58	0.01808**
RB4	\$2.99***	0.54493	\$2.03–\$4.17	0.04636**
SB	\$6.44**	2.7705	\$3.42–\$13.08	–
First RB	\$6.20***	2.3187	\$3.36–\$11.56	0.9905

Notes: Variables significant at *10%, **5% and ***1%. 95% confidence intervals in parentheses are based on the 2.5th and 97.5th percentile of the simulated marginal WTP distribution. In comparison with the delta method, this method does not imply a normal distribution.

and RB4 sample splits remains unclear. Hence, SB and RB2, RB3 and RB4 may differ in parameter estimates, scale parameters or both.

Larger confidence intervals for SB as opposed to RB1, RB2, RB3 and RB4 indicate that the difference in marginal $WTP_{area\ of\ land}$ may be induced by differences in the variance of the error term. Put simply, repeated choices may increase choice precision through decreased choice randomness. However, the difference in the scale parameters may be a consequence of a difference in the magnitude of the cost parameter estimate. That is, a change in the scale parameter does not necessarily have to be responsible for the differences in $WTP_{area\ of\ land}$ across SB and RBx.[24]

Overall, these findings support the rejection of H_0^1.

5.2. Effects of in advance awareness of having repeated choice - Hypothesis 2

To test H_0^2, differences in the attribute and scale parameter estimates between SB and FirstRB were investigated (see Table 6). No statistically significantly difference was found in either ($p_{attribute} = 0.5785$, $p_{scale} = 0.5179$). The marginal $WTP_{area\ of\ land}$ estimates for SB and FirstRB are reported in Table 5. The test procedure suggested by Poe *et al.* (2005) did not reveal statistically significant differences in marginal $WTP_{area\ of\ land}$ ($p = 0.9905$) between SB (\$6.44) and First RB (\$6.20), and the 95 per cent confident intervals are similar (SB:\$3.42–\$13.08; First RB:\$3.36–\$11.56). These results indicate that awareness of having the possibility of making repeated choices does not affect choice outcomes.

These findings do not justify a rejection of H_0^2. We find no evidence that awareness of having repeated choices without specific information about the attribute levels may influence choice outcomes.

[24] We are grateful to an anonymous reviewer for pointing this out.

Table 6 Test results for attribute and scale factor equality

RB	LL RB	LL SB	LL Pooled[a]	LR-test[b] (5 d.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:$\lambda_i = \lambda_j$
RB1	−962.237	−965.823	−1931.128	0.2932	No	0.7795	−1934.252	0.0124	Yes
RB2	−965.627	−965.823	−1942.752	0.0004	Yes	0.6087	−1953.900	NA	NA
RB3	−952.739	−965.823	−1930.921	< 0.0001	Yes	0.4694	−1949.737	NA	NA
RB4	−961.441	−965.823	−1937.900	0.0007	Yes	0.4609	−1957.801	NA	NA
First RB	−995.738	−965.823	−1963.461	0.5785	No	1.1062	−1963.252	0.5179	No

Notes: [a]Pooled CL model allowing varying scale parameter. [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 CL model assuming equal scale parameter in both sample splits. [d]Log-likelihood ratio test, test statistics $-2(LL_{equalscale} - (LL_{varyingscale}))$ with 1 d.f. is asymptotically chi-square distributed.

6. 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, we explored (i) whether a repeated-binary elicitation format affects choice outcomes and (ii) whether in advance awareness of having repeated choices influences choice outcomes.

With respect to the first research question, we find statistically significant differences in choice outcomes comparing choice data obtained by single-binary and repeated-binary elicitation formats.

The literature suggests these differences may be caused by position- and precedent-dependent order effects present in the RBx sample splits. Possible explanations for these order effects include *strategic misrepresentation of preferences*, *anchoring effects*, *reference effects*, *institutional learning*, *fatigue effects*, *failing credibility*, *cost uncertainty effects* and *preference learning*.

Our results suggest that marginal $WTP_{\text{area of land}}$ estimates obtained using a single-binary choice format are statistically higher compared with those estimated using a repeated-binary elicitation format. Yet, the larger confidence intervals derived from single-binary choice data suggest that the differences in marginal $WTP_{\text{area of land}}$ may be induced by differences in the variance of the error term. That is, repeated choices may increase choice precision, indicating *institutional learning* in the RBx sample splits. However, the difference in the scale parameters may be a consequence of the difference in magnitude of the cost parameter estimate.

An alternative explanation of the difference in marginal $WTP_{\text{area of land}}$ estimates is *strategic misrepresentation of preferences*. Respondents may become aware of and exploit strategic opportunities, trying to get the most desired option at the lowest cost. As most of the changes are centred on the cost parameter, *strategic misrepresentation of preferences* seems to be a more likely explanation than *preference learning* because there is no real difference in the other attribute parameter estimates (see footnote 24). However, the difference in marginal $WTP_{\text{area of land}}$ estimates may also be explained by *reference effects*. Respondents may form preferences about what they perceive constitutes a good deal or a bad deal based on deals offered in previous choice tasks.

Effects of *failing credibility* and *fatigue effects* in the RBx sample splits are unlikely to be an explanation for the observed differences in choice outcomes across the two elicitation formats as we find no evidence that respondents favour the do-nothing option more in the RBx than in the SB sample split. The lack of evidence for *fatigue effects* is not surprising as respondents were only presented with four choice tasks each.

However, *cost uncertainty effects* may be an explanation for the observed lower marginal marginal $WTP_{\text{area of land}}$ estimate in the RBx compared with the SB sample split. A decreasing marginal $WTP_{\text{area of land}}$ estimate along the

sequence of choice tasks implies an overall lower marginal $WTP_{area\ of\ land}$ estimate as no differences in the marginal $WTP_{area\ of\ land}$ estimate across SB and FirstRB sample split were observed.

Finally, *anchoring effects* cannot be excluded as a possible explanation of the differences in choice outcomes across the two elicitation formats. That is, respondents may have anchored their preferences on the first choice question in the RBx sample splits. Unfortunately, our experimental design did not allow to test whether *anchoring effects* may have influenced the comparison across the two elicitation formats.

The results presented here are based on restricted CL models due to the econometric limitations outlined earlier. Consequently, the panel character of the repeated-binary elicitation format and possible preference heterogeneity could not be taken into account. Hence, possible model misspecifications may have influenced the results.

Furthermore, the limited consequentiality of the survey format, plausibility of the choice questions, credibility of the policy scenario and comprehensibility of the choice task may have confounded the comparison between single-binary and repeated-binary sample splits. On the other hand, the incentive properties to answer follow-up questions truthfully are unknown. That is, the answers to the follow-up questions may themselves be strategically biased and may thus not reflect the actual opinions of respondents. Further testing of these issues is required.

With respect to the second research question, we found no evidence that awareness of having multiple choices without specific information about the attribute levels provided in advance affects choice outcomes in a repeated-binary DCE. That is, respondents may need specific information about the choice tasks (in particular the attribute levels) they face to be able to exploit that information strategically (*strategic misrepresentation*) or to express *failing credibility* or *cost uncertainty* in their choices. This result complements the findings of Day *et al.* (2012) who find differences in the probability of position-dependent order effects related to advanced compared to stepwise disclosure of choice tasks. The differences across the two studies may be explained by differences in the questionnaire design. Day *et al.* (2012) provided respondents in the advance disclosure treatment with information about the number of choices, the attributes and all possible attribute levels, whereas in the study presented in this article only information about the number of choices and the attribute types (including a note explaining that choice options are based on different attribute levels) was given to respondents. Hence, the results set out here, in combination with the results of Day *et al.* (2012), may indicate that providing substantive information may influence responses to the first (and possibly subsequent choice questions), whereas only providing information on the number of choices and the attribute types is likely to not have such an effect.

In summary, we provide evidence for differences in choice outcomes when comparing single-binary with repeated-binary choice tasks, which may be

explained by *institutional learning*, *strategic misrepresentation of preferences*, *cost uncertainty effects* and *reference effects*. The choice between the two elicitation formats may imply a trade-off between decreased choice precision in the single-binary and order effects that may be explained by *strategic misrepresentation of preferences*, *cost uncertainty effects* and *reference effects* in the repeated-binary elicitation format.

However, our results also suggest that the first question in a repeated-binary choice task yields the same choice outcome as a single-binary choice task if no specific information on attribute levels is provided in advance. That is, we did not find any indications for *strategic misrepresentation of preferences*, *cost uncertainty effects* and *reference effects* caused by awareness of having repeated choices.

References

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

Ariely, D., Loewenstein, G. and Prelec, D. (2006). Tom Sawer and the construction of value, *Journal of Economic Behaviour and Organization* 60, 1–10.

Australian Bureau of Statistics (2009). 2006 Census data. Available from URL: http://www.abs.gov.au [accessed 21 September 2012].

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 Ltd., Massachusetts.

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

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.

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

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.

Day, B., Bateman, I., Carson, R., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R. and Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies, *Journal of Environmental Economics and Management* 63, 73–91.

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 Sandmo, A. (ed.), *Essays in Public Economics*. Lexington Books, Lexington, pp. 95–114.

Greene, W.H. (2007). *Nlogit Version 4.0 Reference Guide*. Econometric Software, Inc., Castle Hill.

Hensher, D. and Rose, J.M. (2007). Development of commuter and non-commuter mode choice models for the assessment of new public transport infrastructure project: a case study, *Transportation Research Part A* 41, 428–443.

Hensher, D., Louviere, J. and Swait, J. (1999). Combining sources of preference data, *Journal of Econometrics* 89, 197–221.

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

Horne, P.. (2008). Use of choice experiments in assessing the role of policy instruments in social acceptability of forest biodiversity conservation in Southern Finland, in Birol, E. and Koundouri, P. (eds), *Choice Experiments Informing Environmental Policy*. Edward Elgar Publishing Ltd., Cheltenham UK, Northampton, USA, pp. 178–197.

Hurwicz, L.. (1960). Optimality and informational efficiency in resource allocation processes, in Arrow, K., Karlin, S. and Suppes, P. (eds), *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford, pp. 27–46.

Kontoleon, A. and Yabe, M. (2008). Latent market segmentation analysis of choice experiment data and implications for the EU's GM labelling policy, in Birol, E. and Koundouri, P. (eds), *Choice Experiments Informing Environmental Policy*. Edward Elgar Publishing Ltd., Cheltenham UK, Northampton, USA, pp. 146–177.

Kooten, G.C.v., Krcmar, E. and Bulte, E.H. (2001). Preference uncertainty in non-market valuation: a fuzzy approach, *American Journal of Agricultural Economics* 83, 487–500.

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.

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 Paris, Published 1999 in the Review of Economic Studies* 66, 23–38.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour, in Zarembka, P. (ed.), *Frontiers in Econometrics*. Academic Press, New York, pp. 105–142.

McNair, B., Bennett, J. and Hensher, D. (2011). A comparison of responses to single and repeated discrete choice questions, *Resource and Energy Economics* 33, 554–571.

Morrison, G.C. (2000). WTP and WTA in repeated trial experiments: learning or leading?, *Journal of Economic Psychology* 21, 57–72.

Moulin, H. (1994). Social choice, in Aumann, R.J. and Hart, S. (eds), *Handbook of Game Theory with Economic Applications*, North-Holland, Amsterdam, pp. 1091–1123.

Muller, E. and Satterthwaite, M.A. (1985). Strategy-proofness: the existence of dominant-strategy mechanisms, in Hurwicz, L., Schmiedler, D. and Sonnenschein, H. (eds), *Social Goals and Social Organization*. Cambridge University Press, Cambridge, pp. 131–171.

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. (eds), *Rational Foundations of Economic Behavior*. Macmillan, London, pp. 225–250.

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, 353–365.

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

Ready, R.C., Whitehead, J.C. and Blomquist, G.C. (1995). Contingent valuation when respondents are ambivalent, *Journal of Environmental Economics and Management* 29, 181–196.

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, *European Transport Conference*, Leiden, The Netherlands.

Sándor, Z. and Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs, *Journal of Marketing Research* 38, 430–444.

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.

StataCorp. (2007). *Stata Release 10*. Stata Press, Texas.

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, 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, 305–314.

Wicksteed, P.H. (1910). *The Common Sense of Political Economy*. Routledge, London.