Choice Certainty and Consistency in Repeated Choice Experiments

Roy Brouwer · Thijs Dekker · John Rolfe · Jill Windle

Abstract The main objective of this study is to examine how repeated choice affects preference learning in stated preference experiments. We test different hypotheses related to preference learning by analyzing response patterns and asking respondents in a choice experiment to report their experienced certainty when going through the choice tasks. In a split-sample test, we show that follow-up choice certainty questions are procedural invariant. The self-reported certainty results indicate that learning occurs, but econometric testing procedures do not identify any significant impact of learning effects on parameter estimates or variance across choice tasks. Additional tests of choice consistency suggest that preferences in the choice experiment are stable and coherent.

1 Introduction

One of the main assumptions underlying micro-economic theory and stated preference (SP) methods is that respondents know their preferences and that these preferences are stable and coherent (Rabin 1998; Brown et al. 2008). This implies that individuals consistently know their preference ordering for a set of goods or services and the rate at which they are willing to trade off good characteristics, such as price and quality. Hence, from a set of alternatives

R. Brouwer (✉) · T. Dekker
Department of Environmental Economics, Institute for Environmental Studies (IVM), VU University,
De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands
e-mail: roy.brouwer@ivm.vu.nl

J. Rolfe · J. Windle
Centre for Environmental Management, CQ University, Bruce Highway, Rockhampton,
QLD 4702, Australia

the individual is assumed to be capable of selecting the most preferred one based on its characteristics.

The described accuracy of this selection process may well apply to everyday shopping decisions in a supermarket. However, public environmental goods presented in SP surveys are often abstract and involve hypothetical changes in their provision level, including quality. Based on pre-existing preferences, respondents are assumed to be able to compare, rank and trade-off the benefits of different levels of increases in environmental quality against the costs involved in terms of personal money income they are willing to give up. Lack of familiarity and experience with changes in environmental quality and its valuation may undermine these a priori assumptions (e.g. Shaik et al. 2007).

Hoeffler and Ariely (1999) show that preference stability is positively correlated with choice experience (single versus repeated choice) and choice effort (higher versus lower effort). In the latter case, a higher level of effort leads to more stable preferences. On the other hand higher effort also reduces preference strength, meaning that respondents facing a hard choice are less certain of their choice than respondents facing an easy choice. Through repetition respondents are expected to be capable of making more precise and consistent decisions, because they learn about the survey format, the associated (hypothetical) market environment and their own preferences (List 2003).

To illustrate the process described above, we model preferences in terms of McFadden's (1974) random utility model (RUM), allowing for a separation of utility (U_{ij}^c) into a deterministic part (V_{ij}^c) and a stochastic part (ε_{ij}^c). Choice experiments fall in the class of attribute-based methods in which the deterministic part of utility for individual i for good j in choice situation c is described in (1) as a function of its attributes X_{ij}^c (Train 2003).

$$U_{ij}^c = V_{ij}^c + \varepsilon_{ij}^c = \beta^c X_{ij}^c + \varepsilon_{ij}^c \tag{1}$$

The superscript c indicates that the vector of preference parameters β, which determine the rate of trade-off between environmental quality characteristics, may vary over the choice sequence. Changes in the parameters of the deterministic part of utility reflect preference learning effects, since respondents alter their trade-off rates between the good characteristics based on past choices. Preference learning can also be measured through the stochastic term as making more precise choices reduces the variance of the stochastic term. Highly uncertain respondents are expected to make random choices, reflected by widely distributed utility functions (Holmes and Boyle 2005).

The main objective of this study is to examine how repeated choice affects preference learning. More specifically, we test the following hypotheses based on existing theories of preference learning and available empirical evidence:

> H_0^1: *preference precision increases with experience (preference refinement)*
> Preference precision, measured through utility variance, is expected to increase (variance decreases) with repeated choice. To this end, respondents are asked in an experiment to state their preferences for 3 multi-attribute alternatives in 5 repeated choice tasks.
> H_0^2: *choices are consistent (stable)*
> Choice behavior at the start of the experiment is expected to be consistent with choice behavior at the end of the experiment. To this end, respondents are presented with the same choice task at the beginning and the end of the choice experiment.
> H_0^3: *self-reported choice certainty increases with experience*
> Consistent with hypothesis 1, self-reported choice certainty is expected to increase with repeated choice due to learning and experience. To this end, respondents are

asked after each choice task how certain they are about their choice on a 0-10 rating scale.

H_0^4: *self-reported choice certainty is procedurally invariant*

Asking respondents to consider and report how certain they are about their choices is not expected to affect choice behavior. To test for potential procedural variance bias, a split sample was used, one with and one without the follow-up certainty questions.

H_0^5: *self-reported choice certainty decreases as the utility differential between alternatives decreases*

Respondents are expected to become less certain if the utility differential between available alternatives decreases. To this end, variation in self-reported certainty is regressed on the estimated RUM functions for the choice alternatives.

More details of the experimental design, the hypothesis testing procedure, and underlying econometric model are provided in the next sections. Section 2 includes a further overview of the relevant literature in this field on which we base our hypotheses, while Sect. 3 involves a description of the econometric model and test procedure underlying the choice experiment. The practical experimental design is presented in Sect. 4, followed by the presentation of the results in Sect. 5. Finally, conclusions are drawn in Sect. 6.

2 Theoretical Background

Repeated choice has been applied in several SP studies, mainly contingent valuation (CV) and contingent ranking (CR), to test preference stability and consistency. The 'discovered preference hypothesis' (DPH) postulates that preferences may be ill-defined when respondents come to a hypothetical market, but that a set of stable intrinsic values already exists within each individual (Braga and Starmer 2005). Through repetition, individuals will gradually discover their exact preferences and adjust their behavior accordingly. In the case of 'coherent arbitrariness' (CA) (Ariely et al. 2003), preferences are also assumed to converge to a stable level allowing the respondent to make coherent and consistent choices, but here the convergence level is dependent on contextual cues and other arbitrary framing effects arising in the survey. Bateman et al. (2008) find support for the DPH and show that learning effects through repeated choices may overturn the wide-spread value anchoring biases found in double bounded CV, hence minimizing the influence of arbitrary effects in SP valuation research.

In this study, we look into choice consistency and preference learning in choice experiments, that is, convergence of model and scale parameters over a series of similar choice tasks. We test for choice consistency by comparing the identical first and last choice task. Furthermore, we measure the development of choice model variance and compare this with self-reported certainty over the choice sequence. No conditions were in place in the experiment to test for the influence of value cues or framing effects, i.e. contrasting DPH and CA.

Choice experiments (CE) are ideal to test preference stability, learning effects and choice consistency. In CE, respondents are asked a sequence of preference questions, varying from 2 to sometimes up to 100 choice tasks, related to different alternatives that are made up of several characteristics or attributes. Typically, practical applications are limited to 4–8 cards (Bennett and Blamey 2001). Brazell and Louviere (1998) identified that significant learning effects can occur compared to the earlier choice tasks in repeated choice sets. In order

to reduce the cognitive burden of repeated choice experiments, avoid fatigue and improve response rates, Louviere et al. (2000) suggest to limit the choice task to 4 preference questions. The number of alternatives, the number of attributes and the correlational structure of information in a choice set are also important determinants of choice task complexity and have been shown to affect choice consistency. Consistency decreases as choice complexity increases (e.g. DeShazo and Fermo 2002; Rose et al. 2009). However, the use of CE to test preference learning, preference stability and choice consistency is limited (e.g. Brazell and Louviere 1998; Dellaert et al. 1999; DeShazo and Fermo 2002; Holmes and Boyle 2005; Ladenburg and Olsen 2008; Savage and Waldman 2008; Rose et al. 2009).

Contrary to the attention it has received in the CV literature (e.g. Li and Mattsson 2005; Alberini et al. 2003; Chang et al. 2007), self-reported choice uncertainty is underinvestigated in repeated choice experiments (Lundhede et al. 2008). Using a 5-level Likert scale (very certain, certain, uncertain, very uncertain, don't know), Lundhede et al. (2008) use self-reported preference uncertainty to calibrate (recode) choices similar as in the CV literature. Of particular interest is their test of the 'Wang-hypothesis' (Wang 1997) that choice uncertainty is highest when the utility difference between alternatives is small. This hypothesis is a transformation from the CV literature to choice experiments. In the original hypothesis it is argued that respondents become more uncertain if the bid level gets closer to the respondent's true willingness to pay (WTP) (Loomis and Ekstrand 1998). In one of the very few empirical models trying to explain self-reported preference uncertainty, Loomis and Ekstrand (1998) show that respondent familiarity with the environmental good in question is the only significant explanatory factor together with the bid level. A significant quadratic effect was found for the bid level, implying that self-reported certainty is highest around the lowest and highest bids, and lowest for intermediate bid levels closest to a respondent's maximum WTP. In a CE, Dellaert et al. (1999) examined the effect of variations in attribute level differences on consumer choice consistency, and found that choice consistency decreases as bid level differences increase and absolute bid levels increase.

The focus of this paper is to examine preference refinement through learning effects in a repeated CE, helping to explain issues of choice uncertainty in the process. To this end, we interpret preference uncertainty as an integral part of the stochastic term in the utility function.[1] Increasing preference uncertainty consequentially increases choice uncertainty. In addition, choice certainty is measured in this choice experiment on a more informative scale than the one used in Lundhede et al. (2008), allowing us to extend the analysis of the self-reported certainty answers. There is a possibility that the presence of an option to express response uncertainty allows for strategic behavior and scaling down of stated WTP (Samnaliev et al. 2006). Therefore, a split sample approach was used where a control group was given the same choice tasks without the certainty questions in order to be able to test the influence of these questions on the choice model parameters.

3 Econometric Modeling of Preference Refinement in Choice Experiments

As described in the introduction, preferences for each presented alternative in the CE are modeled in accordance with McFadden's (1974) random utility model. Indirect utility can be expressed as a function of the vector of attributes of the proposed policy X_{ij}^c:

---

[1] Note that our interpretation of preference learning differs from the decay of behavioral anomalies as discussed under the DPH (Braga and Starmer 2005). Here preference uncertainty is treated as a source of random behavior instead of systematic violations of model predictions.

$$U_{ij}^c = V\left(X_{ij}^c, \beta^c\right) + \varepsilon_{ij}^c \quad \forall j \in D^c, c = 1, \ldots, T \tag{2}$$

where U_{ij}^c is the indirect utility for individual i for the proposed policy j ($j = 1, 2, \ldots, J$) in choice number c ($c = 1, 2, \ldots, T$), β^c and ε_{ij}^c are both choice number specific and represent a vector of preference parameters and a stochastic error term. In each choice number c the respondent is presented with a limited set of policy proposals D^c. If utility is linear with respect to the policy attributes and includes a set of alternative specific constants, as denoted in (3) where α_j^c represents the alternative specific constant (ASC), then the probability that individual i selects alternative k over the other alternative policies can be described by (4).

$$U_{ij}^c = \alpha_j^c + \beta^c X_{ij}^c + \varepsilon_{ij}^c \quad \forall j \in D^c, c = 1, \ldots, T \tag{3}$$

Assuming that the stochastic terms are IID extreme value type 1 distributed within each choice and that they are associated with scale parameters $\lambda^1, \ldots, \lambda^T$, this gives the following conditional logit probability of selecting alternative k over the alternative policy options:

$$P_{ik}^c = \frac{\exp\left[\lambda^c\left(\alpha_k^c + \beta^c X_{ik}^c\right)\right]}{\sum_{j \in D^c} \exp\left[\lambda^c\left(\alpha_j^c + \beta^c X_{ij}^c\right)\right]}, \quad \forall k \in D^c, c = 1, \ldots, T \tag{4}$$

To correct for preference heterogeneity across respondents we allow the vector of preference parameters to vary over individuals by specifying a mixing distribution for β_i^c. More specifically we apply constrained triangular distributions, because they can be bounded such that the random parameter has the same sign across respondents, and excessively large coefficients as occurring under normal and lognormal distributions are avoided (Train 2003). $\Delta(\beta|b)$ in (5) represents the constrained triangular density for β, with mean b and restricted to the domain [0,2b].

$$P_{ik}^c = \int \left(\frac{\exp\left[\lambda^c\left(\alpha_k^c + \beta_i^c X_{ik}^c\right)\right]}{\sum_{j \in D^c} \exp\left[\lambda^c\left(\alpha_j^c + \beta_i^c X_{ij}^c\right)\right]}\right) \Delta(\beta|b)\, d\beta, \quad \forall k \in D^c, c = 1, \ldots, T \tag{5}$$

Louviere et al. (2000) explain that the scale parameter is inversely related to the variance of the error term.[2] If the scale increases, variance decreases: that is, people are making a more accurate choice between the presented alternatives. Obtaining estimates for the scale parameter therefore provides more insight into preference refinement during a choice sequence (Holmes and Boyle 2005). Our first hypothesis of preference refinement predicts an increase in the scale parameter over the choice sequence, while our second hypothesis concerns the stability of the preference parameters. However, it is not possible to compare directly both sets of parameters between choice numbers, because the scale parameter and preference parameters are jointly estimated and therefore confounded (Louviere et al. 2000).

To identify the potential learning effects of interest, a sequential testing procedure is applied in line with Swait and Louviere (1993). First, we test for changes in the preference parameters β^{cn} by allowing for varying scale parameters λ^{cn} between choice tasks. Second, a test for scale parameter equality is performed. The latter test can only be performed if the preference parameters are equal between the analyzed choice sets, because the confoundedness prevents the attribution of observed differences to both differences in scale and preference or only to preference parameter inequality and equal scales (Swait and Louviere 1993). We illustrate these tests by comparing two choice tasks at a time, but the analysis can be easily generalized to a larger set of choice tasks.

---

[2] $(\sigma^c)^2 = \frac{\pi^2}{6(\lambda^c)^2}$.

In the first step, we start by estimating a separate mixed logit model for two choice tasks,[3] which provides us with efficient estimates for $\lambda^1 \beta_i^1$ and $\lambda^2 \beta_i^2$ and a likelihood for both choices. For identification purposes, the scale parameter of the first choice task is normalized to $\lambda^1 = 1$. This normalization implies that estimates of scale should be interpreted as relative scale parameters to choice task 1 (i.e. $\frac{\lambda^2}{\lambda^1}$). A pooled model is then estimated across the two choice tasks, which has the effect of imposing preference parameter equality $\left(\beta_i^1 = \beta_i^2 \text{ and } \alpha_j^1 = \alpha_j^2 \right)$. A search procedure over a range of relative scale parameters is applied to estimate the combination of scale and (pooled) preference parameters providing the best model fit (Swait and Louviere 1993; Louviere et al. 2000). At each possible relative scale parameter the data for the 2nd choice task are rescaled such that a mixed logit model can be estimated to obtain an estimate for β_i and an associated log likelihood value. After the best fit model has been identified, a chi-square test using the log likelihood of each estimated model can be used to test the difference in the preference parameters for the policy attributes under the null hypothesis of equality between the two choice tasks. The standard chi-square distributed Likelihood Ratio (LR) test is as follows:

$$-2 \left(LL^{\text{pooled}} - \left(LL^{c1} + LL^{c2} \right) \right) \text{ with d.f. } |\beta| - 1 \qquad (6)$$

where $|\beta|$ is the number of imposed parameter restrictions.

If the LR test concludes that preference parameters are equal across two choice tasks, it is then possible to test for differences in scale parameters. This second step of the procedure requires the estimation of a mixed logit model for the same pooled model as in step 1, but with equality imposed on both preference and scale parameters this time $\left(\beta_i^1 = \beta_i^2, \alpha_j^1 = \alpha_j^2 \text{ and } \lambda^1 = \lambda^2 \right)$. A LR test can be applied again to compare the log likelihood of the estimated model to the log likelihood of the pooled model with varying scale parameters:

$$-2 \left(LL^{\text{equalscale}} - \left(LL^{\text{pooled}} \right) \right) \text{ with d.f. } 1 \qquad (7)$$

Rejection of equality of scale parameters allows for the identification of learning effects during the choice sequence, particularly if choice variance decreases over the sequence. The equation below shows that if the relative scale parameter for choice task 2 increases, the relative variance of choice task 2 compared to choice task 1 decreases.

$$\left(\frac{\lambda^2}{\lambda^1} \right)^2 = \frac{\left(1/\lambda^1 \right)^2}{\left(1/\lambda^2 \right)^2} = \frac{\pi^2/6 \left(\lambda^1 \right)^2}{\pi^2/6 \left(\lambda^2 \right)^2} = \frac{\sigma_1^2}{\sigma_2^2} \qquad (8)$$

An increase in the scale parameter would indicate that choice behavior becomes less random due to reductions in preference uncertainty. We now turn to the choice experiment.

4 The Experiment

The CE focuses on water scarcity in Rockhampton, a regional city with a population of approximately 60,000, in central Queensland, Australia, where due to climate change water scarcity problems are expected to intensify in the future due to more extreme weather conditions (droughts) and an overall decrease in annual rainfall (Garnaut 2008). Survey respondents were selected at random following a cluster sampling approach. Different sections of

---

[3] Choice task refers to a choice card here. Each choice card represents a choice task where the respondent is asked to indicate his or her most preferred alternative.

the city were identified and streets selected at random within each section. Interviewers were instructed to approach every third house in the street.[4] If agreeable, face-to-face interviews were conducted at the residence with any adult who had knowledge about the household budget and payment of bills. The CE is based on a common design developed in a cross-country study in Southern-Europe aiming to test the international transferability of choice models for similar climate change and water-related problems. In the experiment, respondents were asked to choose between two possible policy alternatives in which water supply in the next 10 years is secured for different uses, including the environment, compared to a baseline situation where water scarcity problems remain the same or aggravate.

The choice experiment design consists of three attributes: restrictions on domestic outdoor water use, environmental quality of waterways, and a cost price (increase in household water bill). Respondents were asked to indicate their preferences for different trade-offs between these three attributes. The base level for outdoor water use was a frequency of restrictions being imposed during the dry season in 4 of the next 10 years, with potential improvements reducing this to 1, 2 or 3 years as a result of water saving measures. The other principal attribute, the environmental condition of waterways, had poor condition as its base level, corresponding to the situation in many parts of the Fitzroy river basin, i.e. the area where the survey was conducted (Rolfe and Windle 2005). Potential improvement levels were moderate, good and very good conditions. The consequences of the improvement levels for both these two attributes were described and visualized during the face-to-face interviews with the help of show cards to help respondents assimilate the information. Six price levels were used, varying from AU$35 to $210.[5]

Respondents from the city of Rockhampton were presented with five choice cards, randomly assigned from six different choice sets, with three alternatives per card (including a constant baseline as described above), described in terms of the three attributes. Each choice set consists of four different cards. The combination of attributes and alternatives in the choice cards across the six choice sets were based on a fractional factorial design. Both the choice set and the first card in the choice set were each time randomly selected and shown to a respondent in order to avoid procedural variance. Rotating the order in which the choice tasks appeared in the questionnaire meant that a full experimental design was associated with each choice task. The first card was shown again at the end of the experiment without telling the respondent. The survey furthermore consisted of two versions: one version with and one without the choice certainty questions after each choice task. An example choice card is presented in Fig. 1. The base level option described the situation that currently existed with zero additional cost. The two alternatives in each card showed possible improvements at an associated cost within a 10 year time frame.

The common CE format was slightly modified to better fit local circumstances in the study area and pretested in 60 face-to-face interviews, resulting in additional minor changes in wording. The main survey was conducted in September 2008. Only residents who had lived in Rockhampton for more than 10 years were asked to participate in the survey as some questions requested information about respondents' experience with water restriction over the past 10 years. Three hundred completed surveys were collected. One hundred and ten residents refused to participate and 28 were ineligible because they lived less than 10 years in the area. The response rate hence was 73%. The 300 surveys were equally distributed over the 2 versions and 6 choice sets (i.e. 250 choice observations per choice set).

---

[4] The pretest and main survey were carried out by hired professional interviewers from the Population Research Laboratory, Centre for Social Science Research, Central Queensland University.

[5] In September 2008, one AU$ equaled approximately US$0.82.

Fig. 1 Example choice card

5 Results

5.1 Sample Characteristics

The socio-demographic characteristics of the survey sample closely match those of the population from which they were drawn in most categories (Table 1). There were statistically fewer single person dwellings included in the survey sample compared with the population, which means that the average household size is slightly higher in the sample. There were also more women and more respondents with higher education levels in the sample compared with the population as a whole.

Respondents were asked a number of general questions to gauge their attitudes to water scarcity and the environmental condition of the waterways in their area as well as their experience with water use restrictions. The results indicate that 11% of the respondents considered water shortage the most important problem in the area. Although the majority of respondents (59%) said that they did not experience water restrictions before, most people expected water use restrictions in the future. In reality, there have been water restrictions in Rockhampton, but these restrictions were not enforced, which is why many respondents may have been unaware that they existed. In the interviews, it was emphasized that the likelihood of water restrictions would increase in the next 10 years, even if people never experienced restrictions so far. Sixty-four percent of the sample believed that they would face restrictions in the future with an average expected occurrence of one in every 4.4 years. Only 12% of the respondents indicated that they thought the choice alternatives were not credible.

Finally, there exists a significant positive correlation between recreational water use and respondent belief that water availability affects the environmental condition of waterways (half of the respondents use freshwater areas for recreational use). In times of water shortages,

Table 1 Socio-demographic characteristics for the sample and population

	Sample statistics	Population statistics[a]
Average age (range 18–91 years) (years)	48	46
% Female	59	50
Average household size	2.9	2.5
Households with children under 18	40%	–
Education		
% With non-school qualification	47	39
% With tertiary education	24	19
Employment		
% Employed	65	61
% Unemployed	3	3
% Not in workforce	32	35
Average annual household income (gross)	$59,556	$56,836

[a] Rockhampton. Source: Australian Bureau of Statistics 2006 Census

respondents thought agriculture should be given higher priority for limited water than the environment. However, the majority of respondents nevertheless believe that the environment also has a right to be protected.

5.2 Preference Refinement and Choice Consistency

To test for preference learning and refinement across the choice tasks, the tests outlined in Sect. 3 were applied to pairs of choice tasks. There was potential for changes in preferences to be gradual, so that significant differences only emerged after several successive tasks. To circumvent this problem, each choice has been contrasted to the other choices on an individual one-to-one basis. Hence, based on the available five choice tasks, 10 sets of comparisons were made of preference and scale parameters.

Table 2 presents the results of the basic pooled mixed logit model over the five choice cards, in which we controlled for the panel structure of the data by allowing the random parameters to vary over individuals, but not over the choice sequence. Heterogeneity was furthermore captured by interacting demographic and other respondent characteristics with the ASC. Interaction effects between individual respondent characteristics and the CE attributes were not statistically significant. Note that the ASC is not included on its own as it was not statistically significant at the 10% level. The outcome of the LR-test and the adjusted R-square imply that the model fits the data well. As expected, utility for the policy alternatives increases if the frequency of water use restrictions decreases and environmental conditions of the waterways (water status) improve, while price has a significant negative effect on choice probability. Random taste variation is detected for both the outdoor use restriction and water status attribute. The presented lower and upper bound values are based on the triangular mixing distribution where the mean is bound to [0, 2b]. In addition to the presented mixed logit specification we also control for preference heterogeneity through the inclusion of individual respondent characteristics. Only those demographic and other characteristics are presented that have a statistically significant impact on choice behavior at the 10% level, such as respondent experience with water use restrictions, perception of the credibility of the presented policy alternatives and attitude towards environmental conservation in general.

Table 2 Mixed logit pooled model parameter results

Variable	Value range		Coefficient	SE	p-Value	WTP[a]
Attributes						
Use restriction	1–4	Mean	−1.365	0.648	0.035	−62
		Lower bound	−2.730	1.296	0.035	(−108 to −16)
Water status	1–4	Mean	1.913	0.181	0.001	87
		Upper bound	3.825	0.362	0.001	(73 to 101)
Cost price	$35–210	Mean fixed	−0.022	0.002	0.001	
Respondent characteristics						
Gender	1 = female	Mean fixed	1.529	0.380	0.001	
Age	18–91	Mean fixed	−0.017	0.01	0.098	
Income	$10–100 × 10^3	Mean fixed	0.120	0.057	0.036	
Experience	1 = yes	Mean fixed	−1.310	0.383	0.001	
Credibility	0–4	Mean fixed	0.551	0.152	0.001	
Importance	0–4	Mean fixed	0.705	0.198	0.001	
LL(restricted)			−1, 538.057			
LL(unrestricted)			−886.463			
LR-test (χ^2)			1,303.189		0.001	
Adjusted R^2			0.422			
Observations			1,475			

[a] AU$ per household per year, 95% confidence interval between brackets

Table 3 presents the results from the Swait and Louviere (1993) procedure for testing the hypotheses regarding the equality of preference and scale parameters. For all choice task comparisons, the results show that the preference parameters are stable throughout the sequence (column 5 in Table 3). Thus, respondents do not seem to change their rates of trade-off between the attributes making up the policy alternatives as they become more familiar with the policy scenarios and more experienced in trading-off the attributes. Our second hypothesis that preferences are the same and hence stable over all five choice tasks is not rejected at the 10% significance level.[6]

Examining the changes in preference uncertainty as measured by the relative scale parameter (column 6), the results suggest reductions in uncertainty over the choice sequence. However, comparing the log-likelihood values of the pooled models with and without constraints on the equality of the scale parameter does not identify a significant difference between the pooled models at the 10% significance level (column 10). This is despite a reduction in the relative variance (column 7) by 52% when comparing the change in relative scale parameter based on cards 1 & 2 and cards 1 & 5. Hence, an increase in preference precision during the choice sequence is suggested, but the differences in relative scale parameters cannot be proven to be statistically significant. The same pattern is observed when comparing the change in variance for cards 2 & 3 and 2 & 5 (reduction of variance by 22%) and 3 & 4 and 3 & 5 (a further reduction of variance by 5%). So, although our analysis provides indications of

---

[6] Note that in estimating the pooled model for two choice cards we allowed the mean of the random parameters to vary over individuals and the choice tasks. The increase in efficiency gained by varying the mean of the random parameters only over the individual would have prevented the identification of the stability of preference parameters.

Table 3 Test results for equality of preference and scale parameters between choice tasks

Choice task	1	2	3	4	5	6	7	8	9	10
	LL 1st card	LL 2nd card	LL joint [a]	LR-test (8 df) [b]	Reject H_0: $\beta_i = \beta_j$	Rel. scale (μ_j/μ_i)	Rel. var. (σ_j^2/σ_i^2)	LL joint [c]	LR-test (1 df) [d]	Reject H_0: $\mu_i = \mu_j$
1&2	−212.822	−230.381	−444.143	1.879	No	0.70	2.041	−445.24	2.193	No
1&3	−212.822	−215.107	−429.599	3.339	No	0.90	1.235	−429.689	0.180	No
1&4	−212.822	−211.569	−425.006	1.230	No	0.96	1.085	−425.018	0.023	No
1&5	−212.822	−213.125	−426.521	1.147	No	1.01	0.980	−426.523	0.003	No
2&3	−230.381	−215.107	−446.501	2.025	No	1.27	0.620	−447.113	1.224	No
2&4	−230.381	−211.569	−442.151	0.402	No	1.37	0.533	−443.147	1.991	No
2&5	−230.381	−213.125	−444.622	2.231	No	1.44	0.482	−445.881	2.518	No
3&4	−215.107	−211.569	−427.504	1.655	No	1.09	0.842	−427.58	0.151	No
3&5	−215.107	−213.125	−428.844	8.300	No	1.12	0.797	−428.984	0.279	No
4&5	−211.569	−213.125	−425.619	1.850	No	1.07	0.873	−425.66	0.082	No

[a] Pooled mixed logit model allowing scale parameters to vary

[b] Critical value for chi-square test with 8 degrees of freedom and 90% confidence is 13.362

[c] Pooled mixed logit model keeping scale parameters constant

[d] Critical value for chi-square test with 1 degree of freedom and 90% confidence is 2.706

preference refinement during the choice sequence such that more respondents seem to make accurate decisions, the outcome of the econometric tests rejects the statistical significance of the observed pattern, and thereby rejects our first hypothesis.

The additional test of choice consistency compares responses between cards 1 and 5 (the latter being identical to the first). Seventy-three percent of the respondents chose the same policy alternative on both cards, and no significant difference in choice variance could be detected (row 4, column 10 in Table 3). Hence, in line with the finding that preference parameters remain stable, our second hypothesis of choice consistency cannot be rejected.

5.3 Self-reported Choice Certainty and Procedural Invariance

In this section, we turn to the self-reported choice certainty results. The approach presented in the previous section allowed us to investigate two potential preference refinement effects. We also compared this to respondent feedback, with an additional question added after each choice task to identify how certain respondents felt about their choices. The responses were identified on a semi-itemized 0–10 rating scale, where 0 means completely uncertain and 10 completely certain. Of interest here is to see whether respondents felt they became more confident and hence certain as a result of experience and learning as they went through the choice sequence as opposed to the findings presented in the previous section. The results for self-reported choice certainty are summarized in Fig. 2.

Significant differences between stated choice certainty are found across the choice tasks based on the non-parametric Kruskal–Wallis test (χ^2 (4df) = 11.709; $p = 0.020$). Examining the differences between individual choice tasks, self-reported choice certainty increases significantly when going from the first to the second and from the fourth to the fifth choice task (based on the non-parametric Mann–Whitney test[7]). However, no significant differences can be detected (at the 10% level) between the second and third and the third and fourth choice tasks. Most importantly perhaps is the finding that choice certainty is significantly higher for the fifth and final choice task compared to the first (Mann–Whitney (MW) $Z = -3.353$; $p = 0.001$). This suggests learning occurs as a respondent goes from the first to the fifth choice task and confirms the third hypothesis. An interesting related finding is that respondents who changed their choice during the final choice task compared to their first
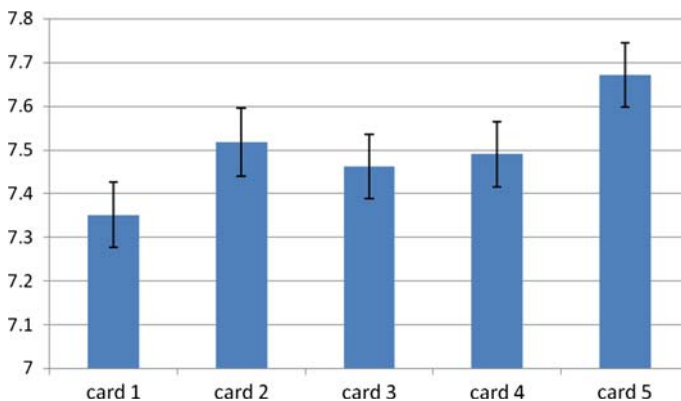

Fig. 2 Average self-reported choice certainty in the repeated choice experiment

---

[7] Test results are available from the authors.

Table 4 Procedural invariance test results

Test	Result
LL—without preference certainty question	−331.654
LL—with preference certainty question	−327.199
LL joint [a]	−659.536
LR-test (8 df)[b]	1.367
p-Value	0.995
Relative scale (μ_j/μ_i)	0.820
Relative variance (σ_j^2/σ_i^2)	1.487
Reject H$_0$: $\beta_i = \beta_j$?	No
LL joint [c]	−660.509
LR-test (1 df)[d]	1.945
p-Value	0. 163
Reject H$_0$: $\mu_i = \mu_j$?	No

[a] Pooled mixed logit model allowing scale parameters to vary
[b] Critical value for chi-square test with 8 degrees of freedom and 90% confidence is 13.362
[c] Pooled mixed logit model keeping scale parameters constant
[d] Critical value for chi-square test with 1 degree of freedom and 90% confidence is 2.706

choice (27% of the sample) reported a significantly higher choice certainty than respondents who displayed consistent choice behavior (MW $Z = -3.186$; $p = 0.001$).

Even though the choice certainty question was asked as a follow-up to the actual choice, there is a possibility that it may have affected individual choice behavior, that is, choices made by respondents may be dependent on the survey procedure. Using a version with and without the choice certainty follow-up questions in the CE, we tested if the estimated utility functions are procedural invariant, i.e. independent of these additional questions. The Swait and Louviere (1993) procedure outlined in Sect. 3 was again used to test if the observations from both versions can be combined (Table 4).

The outcome of the first LR-test in Table 4 indicates that the preference parameters are equivalent in both sub-samples, with the scale parameter ratio estimated at 0.820. This provides a first indication that procedural invariance is not violated by asking a part of the sample the choice certainty follow-up question. The second LR-test of the pooled models with and without constraints on the equality of the scale parameter also fails to reject the null-hypothesis of equal scale parameters at the 10% level. Hence, there is no indication of procedural variance. Respondents do not seem to answer the choice question differently if they are presented with a choice certainty follow-up question, confirming the fourth hypothesis.

5.4 Modelling Determinants of Respondent Certainty

Another issue addressed in the analysis is what drives consistency and changes in choice certainty. To explore this, self-reported choice certainty from the survey was regressed on a number of possible explanatory factors in an ordered probit model. This allows assessment of whether the observed increase in average self-reported certainty during the choice sequence (Fig. 2) remains statistically significant whilst controlling for possible influencing factors, such as experimental design factors and individual respondent characteristics.

Like Lundhede et al. (2008), we were also interested to provide further empirical evidence of Wang's (1997) hypothesis that choice uncertainty increases as alternatives become less distinguishable from each other in terms of the utility they generate. For each respondent, the absolute difference between the utility derived from the alternative chosen and the second best alternative was therefore calculated and included as a regressor in the analysis. Given

Table 5 Self-reported preference certainty determinants (ordered probit regression results)

Explanatory factors	Value range	Coefficient estimate	SE	*p*-Value
Respondent characteristics				
Gender	1 = female	−0.261	0.071	0.001
Age	18–91	−0.010	0.002	0.001
Household income	$10–100 $\times 10^3$	0.224×10^{-5}	0.116×10^{-5}	0.053
Use restriction experience	1 = yes	0.261	0.072	0.001
Familiarity with information	0–4	0.097	0.034	0.004
Credibility policy alternatives	0–4	0.156	0.037	0.001
Utility difference	0.73–9.45	0.051	0.022	0.018
Design characteristics				
Use restriction (attribute)	1–4	−0.524	0.427	0.220
Water status (attribute)	1–4	0.318	0.044	0.001
Cost price (attribute)	$35–210	−0.014	0.003	0.001
Cost price^2 (squared)		0.496×10^{-4}	0.109×10^{-4}	0.001
Choice task number	1–5	0.024	0.024	0.314
LL(restricted)		−1, 872.890		
LL(unrestricted)		−1, 790.558		
LR-test (χ^2)		164.66		0.001
Observations		968		

the fact that no structural change in preferences could be detected during the choice sequence (see Sect. 5.2), this procedure is based on the estimated RUM presented in Table 2 excluding the respondent characteristics as these are accounted for separately in the ordered probit model. The results of the ordered probit regression are presented in Table 5.

Self-reported choice certainty is significantly influenced by respondent gender (women tend to be less certain than men), age (older respondents are less certain), and disposable household income (respondents earning more are more certain). Education level also has a significant positive impact on choice certainty (higher educated respondents are more certain), but correlates with household income and was therefore excluded from the analysis to avoid multicolinearity. Whether or not respondents ever experienced water use restrictions, respondent familiarity with the information supplied during the survey and the credibility of the presented policy alternatives all have a significant positive impact on choice certainty. Experience increases, as expected, self-reported certainty, and also the more familiar someone is with the provided information and the more credible the policy alternatives are, the higher respondent certainty. The estimated utility difference between the chosen and second-best alternative is also included, and significantly influences self-reported certainty in a positive way, supporting the results reported in Lundhede et al. (2008). The larger the utility difference, the more certain respondents are about their choice.

New in our analysis and of particular interest here are the impacts of the design characteristics on self-reported choice certainty. Two out of the three attributes of the chosen alternative significantly influence self-reported certainty. Higher environmental quality conditions result ceteris paribus in more confident choices, while a higher price results in less certainty. Also in this study a small, but statistically significant positive quadratic price effect is found, implying a U-shaped curve when plotting the predicted certainty against the price

levels. That is, certainty is highest around the lowest and highest price levels and lowest for intermediate levels where maximum WTP is expected to be. Together with the positive effect for the utility difference, this confirms the fifth and last hypothesis.[8]

Finally, choice task sequence has no significant impact on preference certainty, which is in line with the results presented in Sect. 5.2 based on the estimated RUM. We included this variable also as a dummy variable (four dummies for each choice task with the first choice task as the baseline level), but also in that case the variables remain statistically insignificant. Choices further down the choice sequence do not appear to result in a significant improvement of choice certainty when we control for experimental design and respondent characteristics.

6 Conclusions

The main objective of this study was to examine how repeated choice affects preference learning. More specifically, we tested different hypotheses related to preference refinement and asked respondents at the same time to also report their own experienced certainty when going through the choice tasks. This comparison between RUM preference structure dynamics and self-reported choice certainty has to our knowledge not taken place in the CE literature before. We show that asking respondents how certain they are about their choice does not affect choice behavior, taking away concerns expressed in the literature about the impact of such follow-up questions on procedural variance, and providing legitimacy to our comparison.

The results of self-reported choice certainty suggested that a learning process takes place during the course of the choice experiment, with self-reported certainty increasing across repeated choice tasks. This confirms a priori expectations that learning effects do occur. Respondents felt significantly more confident and certain about their choice at the end of the choice experiment than they were at the beginning. However, whilst controlling for both experimental design and respondent characteristics, we have to reject the hypothesis of preference refinement based on the self-reported choice certainty results. Moreover, econometric testing procedures did not identify any significant impact of learning effects on parameter estimates or variance across the choice tasks. Hence, the lack of econometric evidence of significant reductions in utility variance during the choice sequence is confirmed by respondent stated choice certainty.

Based on the outcomes of choice consistency tests, we conclude that choices and underlying preferences are stable and coherent in this specific water conservation case study. This is exactly the opposite of what was found in previous studies referred to in Sect. 2, which all provide empirical evidence of preference instability and choice inconsistency. It is hard if not impossible to generalize our results to other decision-making contexts or environmental domains. Rose et al. (2009) show that choice complexity is context dependent and varies across countries and cultures. However, an important explanation for the results found in this study may have been that the choice task in the experiment presented here was not too complicated (Hoeffler and Ariely 1999). The number of alternatives and attributes was limited to three, and respondents were only submitted to five repeated choice tasks. In previous studies either the number of attributes or the number of choice tasks are considerably higher. In terms of choice tasks, our study is closest to the one conducted by Holmes and Boyle (2005). They use eight attributes of which seven are statistically significant at the 10% level in the

---

[8] No indication was found that the variables were correlated. The highest correlation level is found between the estimated utility difference and cost price ($r = 0.31$). The correlation between the utility difference and the attribute water use restriction is -0.22 and between the utility difference and water status -0.13.

context of forest management, which may have been an important driving force behind the preference structure dynamics found in their study. The number of attributes used in other repeated choice studies with more than five choice tasks is quite high (5–9) even though most applications concern private good consumer decisions related to, for example, internet access, commuting or apartment time-sharing in a Mexican holiday resort. Given our simple choice design, we tend to agree with DeShazo and Fermo (2002), who argue that the negative impacts of choice complexity on choice consistency predicted by behavioral economics can be minimized in the design phase of SP studies through pre-testing to asses 'optimal' alternative and attribute numbers. Given the variety (heterogeneity) of respondents usually found in social surveys, we are not convinced that there exists an 'optimal' level of complexity in CE, nor do we believe that we managed to find this 'optimal' level in our specific case study. However we are confident that the CE in our case study was well understood and meaningful to respondents as a result of thorough pre-test procedures. More research is needed in both private and public decision-making contexts, varying the degree of choice complexity as per DeShazo and Fermo (2002), to test for preference stability and choice consistency in order to be able to justify equal treatment of repeated choices in pooled choice models.

References

Alberini A, Boyle K, Welsh M (2003) Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. J Environ Econ Manag 45:40–62

Ariely D, Loewenstein G, Prelec D (2003) Coherent arbitrariness: stable demand curves without stable preferences. Q J Econ 118(1):73–105

Bateman IJ, Burgess D, Hutchinson WG, Matthews DI (2008) Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. J Environ Econ Manag 55:127–141

Bennett JW, Blamey RK (2001) The choice modelling approach to environmental valuation. Edward Elgar, Cheltenham

Braga J, Starmer C (2005) Preference anomalies, preference elicitation and the discovered preference hypothesis. Envir Res Econ 32:55–89

Brazell JD, Louviere JJ (1998) Length effects in conjoint choice experiments and surveys: an explanation based on cumulative cognitive burden. Department of Marketing, The University of Sydney

Brown TC, Kingsley D, Peterson GL, Flores NE, Clarke A, Birjulin A (2008) Reliability of individual valuations of public and private goods: choice consistency, response time, and preference refinement. J Public Econ 92:1595–1606

Chang J-I, Yoo S-H, Kwak S-J (2007) An investigation of preference uncertainty in the contingent valuation study. Appl Econ Lett 14:691–695

Dellaert BG, Brazell JD, Louviere JJ (1999) The effect of attribute variation on consumer choice consistency. Mark Lett 10(2):139–147

DeShazo JR, Fermo G (2002) Designing choice sets for stated preference methods: the effects of complexity on choice consistency. J Environ Econ Manag 44:123–143

Garnaut R (2008) The garnaut climate change review. Cambridge University Press, Cambridge

Hoeffler S, Ariely D (1999) Constructing stable preferences: a look into dimensions of experience and their impact on preference stability. J Consum Psychol 8(2):113–139

Holmes TP, Boyle KJ (2005) Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Econ 81:114–126

Ladenburg J, Olsen SB (2008) Gender-specific starting point bias in choice experiments: evidence from an empirical study. J Environ Econ Manag 56(3):275–285

Li C-Z, Mattsson L (1995) Discrete choice under preference uncertainty: an improved structural model for contingent valuation. J Environ Econ Manag 28:256–269

List JA (2003) Does market experience eliminate market anomalies. Q J Econ 118:41–72

Loomis J, Ekstrand E (1998) Alternative approaches for incorporating respondent uncertainty when estimating willingness to pay: the case of the Mexican spotted owl. Ecol Econ 27:29–41

Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and application. Cambridge University Press, Cambridge

Lundhede TH, Olsen SB, Jacobsen JB, Thorsen J (2008) Choice experiments and certainty in choice: a test of the influence of utility difference on self-reported certainty levels and evaluations of three recoding approaches to handle uncertrain responses. Paper presented at the 16th annual conference of the european association of environmental and resource economists (EAERE), June 25–28, 2008, Gothenburg

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

Rabin M (1998) Psychology and economics. J Econ Lit 36:11–46

Rolfe J, Windle J (2005) Option values for reserve water in the fitzroy basin. Aust J Agric Res Econ 49(1):91–114

Rose JM, Hensher DA, Caussade S, Ortuzar J, Jou R-C (2009) Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. J Transp Geogr 17(1):21–29

Samnaliev M, Stevens TH, More T (2006) A comparison of alternative certainty calibration techniques in contingent valuation. Ecol Econ 57:507–519

Savage SJ, Waldman DM (2008) Learning and fatigue during choice experiments: a comparison of online and mail survey modes. J Appl Econ 23(3):351–371

Shaik SL, Sun L, van Kooten GC (2007) Treating respondent uncertainty in contingent valuation: a comparison of empirical treatments. Ecol Econ 62:115–125

Swait J, Louviere J (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. J Mark Res 30(3):305–314

Train K (2003) Discrete choice methods with simulation. Cambridge University Press, Cambridge

Wang H (1997) Treatment of don't know responses in contingent valuation surveys: a random valuation model. J Environ Econ Manag 32(2):219–232