University of St. Andrews

Discussion papers in Environmental Economics

http://www.st-andrews.ac.uk/gsd/research/envecon/eediscus/

Paper 2017-03

The relationship between perceived difficulty and randomness in discrete choice experiments: Investigating reasons for and consequences of difficulty

T. Börger, O. Frör and S. Weiß

Keywords: Discrete choice experiment, scale heterogeneity, perceived difficulty, river management

JEL codes: H4, Q25, Q51

The relationship between perceived difficulty and randomness in discrete choice experiments: Investigating reasons for and consequences of difficulty

Tobias Börger[a], Oliver Frör[b], Sören Weiß[b]

[a] School of Geography and Sustainable Development, University of St Andrews, UK
[b] Institute of Environmental Sciences, University of Koblenz-Landau, Germany

Abstract: Discrete choice experiments to value environmental goods and services constitute a complex and demanding task for survey respondents. This study looks at the effect of perceived difficulty with the choice tasks on choice consistency and preferences. The choice data come from two parallel surveys valuing river management outcomes in Germany. Results show that perceived difficulty decreases response scale, an indicator of the relative weight of the explained over the random component of indirect utility of a choice alternative. The reasons for this effect have more to do with the design of the actual task in the choice experiment than with the content and topic of the valuation exercise. Results also show only a very limited effect on preferences and willingness to pay for aspects of river management. The proposed econometric strategy manages to effectively separate the effect of difficulty on inter-individual differences of preference and scale. Based on these results, we recommend (i) to rigorously test the attribute design to allow only meaningful trade-offs as perceived by respondents and (ii) to put greater emphasis on the explanation of the choice tasks.

Keywords: Discrete choice experiment, scale heterogeneity, perceived difficulty, river management
JEL codes: H4, Q25, Q51

1. Introduction

As a stated preference method, discrete choice experiments (DCE) have become one of the main techniques to value environmental non-market goods since their first applications in this field in the 1990s (Adamowicz et al. 1994, Boxall et al. 1996, Adamowicz et al. 1998, Hanley et al. 1998). Yet, while in comparison to the more traditional contingent valuation method (Carson and Hanemann 2005) DCE produce more detailed value information on the specific characteristics of an environmental good, the difficulty and complexity of the task survey respondents are given is far higher (DeShazo and Fermo 2002, Boxall et al. 2009). Respondents typically have to complete a whole series of choice tasks which often differ only slightly and they are presented with and have to trade off a set of choice attributes which are varied across choice alternatives and occasions.

There are a number of aspects that contribute to the potential perceived difficulty of responding to DCE as perceived by the respondent. Firstly, and in the field of environmental economics in particular (as opposed to transport), respondents are required to read through or listen to as well as understand often complex and new information regarding the good to be valued, its attributes, their levels and the mechanisms and measures how the proposed change in these attribute levels can be achieved (Mathews et al. 2006). They are then required to apply this newly acquired knowledge regarding the attributes and the choice process in the subsequent choice tasks. Secondly, respondents are required to understand 'the rules of the game', i.e. what exactly they are presented with and how to read this information; that they have to identify and indicate their most preferred alternative in each and every choice; that there are repeated choice tasks, which nevertheless are to be evaluated and completed independently; that attributes are to be considered to vary independently from each other; that the selection of an alternative is associated with a potential real cost to the respondent in the future. Lastly, the nature of environmental goods to be valued has in the recent years developed from fairly well-known and familiar goods, such as landscape characteristics (Adamowicz et al. 1998, Campbell 2007) and recreational sites (Scarpa and Thiene 2005), over the preservation of animal and plant species (Jacobsen et al. 2008, Yao et al. 2014) to remote and complex marine ecosystem services (McVittie and Moran 2010, Börger et al. 2014, Jobstvogt et al. 2014, Norton and Hynes 2014, Aanesen et al. 2015). The more remote and less well known the good is to the respondent, the more difficult is the choice task expected to be. This begs the question whether the specifics of the choice task (the 'rules of the game') or the nature of the good to be valued cause the most difficulty for respondents. This study attempts to address this question by comparing the effect of respondent-perceived difficulty on choice behaviour in two related surveys with goods of different complexity. It is well established that there exist limitations in information processing when people make choices (Heiner 1983), which suggests a general gap between the cognitive abilities of a respondent in DCE and the requirements set by the choice tasks. Early empirical evidence shows more likely use of decision heuristics as the complexity of the choice situation increases (Mazzotta and Opaluch 1995, Swait and Adamowicz 2001). In that sense, increased choice complexity can be expected to lead to choice behaviour which is less consistent with rational utility maximisation in the random utility model (RUM). Consequently, a number of studies have investigated the influence of the design elements and the complexity of the choice task on information processing and choice behaviour, preferences and willingness to pay (WTP) estimates. The considered aspects of complexity are usually the number of choice situations presented to one respondent, the number of alternatives in each choice task, the number of attributes within each alternative, the number of levels of any one attribute and the range of levels that any one attribute can take (e.g. DeShazo and Fermo 2002, Caussade et al. 2005, Louviere et al. 2008). Rather than a constructed measure of choice complexity, this study uses difficulty of the choice task as perceived by the respondent as its main focus. We believe this approach is an important addition to the literature because it (a) takes into account the respondent's perspective on the choice task and (b) interprets difficulty more broadly than just relating to the above-mentioned characteristics of the choice tasks encompassing also the whole process consisting of introduction of valuation scenario and attributes, task instructions, exemplary choice card and actual choice tasks.

The study uses a heteroskedastic, i.e. scaled, random parameters logit (RPL) model to test the effect of perceived difficulty on the scale parameter. The latter is inversely related to the variance of the error term in the indirect utility function in the RUM with higher scale indicating a higher weight of the explained component of indirect utility compared to the error term. Recent studies have identified a number of determinants of scale in stated choice surveys, such as prior knowledge about (LaRiviere et al. 2014) or experience with the good to be valued (Czajkowski et al. 2015), response time (Börger 2016) or level of education (Czajkowski et al. 2014). Building on this literature, we will assess the effect of perceived difficulty of the choice task on preferences and scale.

The study uses two related DCE surveys on river restoration measures in Germany to investigate the role of perceived difficulty of choice tasks and the underlying reasons. The surveys were planned and administered in a coordinated manner to comparable samples living in the same area around the river Queich in the German state of Rhineland-Palatinate. The overall aim of the surveys is to answer the question on which level respondents can express their preferences for aspects of river restoration. Are they able to consider detailed, e.g. ecology-related aspects, or can they only make trade-offs on a more aggregate level e.g. comparing ecological, recreational and flood protection characteristics of rivers. One important component to answer this question is to investigate how difficult the tasks in the respective surveys are perceived to be and what effect perceived difficulty has on preferences and choice consistency.

Results show a significantly and robustly negative effect of perceived difficulty on the scale parameter, indicating that respondents with more difficulties make more random choices. Simultaneous effects on preferences can be found in the survey with a more general view on river management, yet these are limited. This is further reflected in the effects on estimates of marginal WTP, which only differ slightly across respondents perceiving the choice tasks to be easy or difficult. Yet, perceived difficulty has a major detrimental effect on the precision of these welfare estimates. The remainder of the paper is structured as follows. Subsequent Section 2 provides an overview of the literature. Section 3 introduces the methodology and data. Section 4 presents the results, and discussion and conclusions are provided in Section 5.

2. The effect of choice complexity on stated choices

The effects of the complexity of the choice task on choice responses have attracted a fair amount of interest from choice modelling practitioners in recent years. Yet evidence regarding the effect of complexity on preferences and WTP is mixed. In the context of consumption choices between types of fruit juice, Swait and Adamowicz (2001) investigate the relationship between complexity of the choice task and preferences and find a strong effect of the former. They also present evidence for respondents using simplifying choice heuristics (i.e. focussing less on choice attributes but more on alternative [brand] names) as complexity cumulates over the course of multiple choice tasks, which hints at decreases in choice consistency as complexity increases. Hensher (2004) finds that more complex DCE designs in terms of number of choice sets, alternatives within each set, attributes within each alternative and number and levels for each attribute generally elicit higher WTP for travel time savings. In contradiction to that, Boxall et al. (2009) show that the number of differing attributes across choice alternatives and the overall number of choices to complete exacerbates status quo bias, consequently reducing WTP. Supporting these mixed findings, Rose et al. (2009) fail to find universal effects of choice design complexity on valuations of travel time savings in different countries. These differences lead the authors to conclude that such effects are quite localised and not transferrable across countries.

When it comes to choice consistency (or inversely related error variance), four potential reasons for a reduction in choice consistency due to choice complexity have been proposed (DeShazo and Fermo 2002): Given the existence of a high level of complexity, the preference ordering of the respondent might not be complete with respect to the range of choice alternatives displayed, i.e. the preference relationship with one or more attribute levels might not be defined; (2) the preference ordering might not be transitive and therefore appear less consistent in the random utility model; (3) while being complete and transitive, there is a chance that the preference ordering is sub-optimal in the sense that the respondent lacks information to perfectly and correctly rank all alternatives; and (4) a respondent might be indifferent between any two choice alternatives and hence prefer one randomly.

While some studies detect a negative effect of the number of choice sets on scale (Caussade et al. 2005, Bech et al. 2011), Louviere et al. (2013) find a slightly positive relationship between number of choice sets and scale. The latter conclude that adding more choice sets does not impair data quality. This supports earlier findings both in transport (Arentze et al. 2003) and energy contexts (Carlsson and Martinsson 2008). The relationship between number of choice alternatives and scale has been found to be inversely U-shaped, i.e. increasing scale at comparably few alternatives but decreasing it when the number of alternatives exceeds a certain threshold (Caussade et al. 2005, DeShazo and Fermo 2002). Caussade et al. (2005) further find mixed effects of differences of attribute levels between alternatives with comparably narrow differences reducing and wider differences increasing error variance. This confirms earlier findings that larger price differences between alternatives reduce scale (Dellaert et al. 1999). In a large design of designs framework, Louviere et al. (2008) find that more complex experimental designs, i.e. having a higher number of attributes and/or higher number of levels leads to less consistent responses as indicated by lower scale. This means that despite the higher statistical efficiency of more complex designs, such designs reduce "cognitive efficiency" by inducing respondents to make choices more at random. These results confirm earlier suggestions of a reduced ability to choose when faced with higher choice complexity (de Palma et al. 1994).

While many of the above studies have used the heteroskedastic multinomial logit (MNL) model, the choice modelling literature has recently seen a further rise in the application of scaled random parameter models to identify determinants of the scale factor. Most of these applications use variants of the generalised multinomial logit (GMNL) model (Fiebig et al. 2010), which encompasses the heteroskedastic MNL. It has been found that the scale factor is positively affected by the familiarity with the good to be valued (LaRiviere et al. 2014, Czajkowski et al. 2015, 2016), level of education of the respondent (Czajkowski et al. 2014) and response time (Börger 2016). In addition to that, stated choice certainty has been found to positively affect scale (Bech et al. 2011, Beck et al. 2013, Börger 2016).

This work on systematic differences in choice scale has thus far been linked to the choice complexity as a constructed measure of objective design characteristics. The number of studies looking at the relationship between difficulty perceived by the respondent and choice consistency is still comparably small. To the best of our knowledge only Bech et al. (2011) and Ruokamo et al. (2016) test this link empirically and consistently find a detrimental effect of perceived difficulty on scale. Our study builds on these findings and extends this literature in several respects. In the empirical application we (i) compare two parallel surveys with differing content; (i) investigate the underlying causes of perceived difficulty and their effect on choice consistency; and (iii) assess the potential impact of difficulty on welfare measures. We expect any effect on scale to be behavioural rather than topic-specific, and therefore potential scale effects to be consistent across surveys with differing topical focus. This follows Louviere et al. (2008) who find differences in scale (and its determinants) only between designs of different levels of complexity but not between surveys with differing topics.

3. Data and methods

3.1. Data

The data come from two parallel surveys on aspects of restoration of the river Queich, a tributary to the Rhine in Germany. Both surveys deal with the valuation of aspects following a subsidy scheme for river restoration projects in the federal state of Rhineland-Palatinate, the so-called "Aktion Blau Plus". Under this scheme local projects of river restoration are subsidised with a maximum of 90% of the project costs depending on the specific measures and benefits to be achieved. In the valuation survey respondents were presented either one of two potential new river restoration projects along the Queich River. The project "River Management" contains as attributes a number of rather aggregate elements like flood protection, ecological status, access to the river and development of towns and villages along the river (see Table 1 for a description of the attributes). In contrast, the project "River Ecology" zooms in on only ecological aspects of the restoration like water quality, river structure and species protection (see Table 2). The idea of using these two projects and survey versions is to compare the way respondents make choices depending on the respective level of aggregation of attributes. Consequently, one half of the survey sample is asked to trade off the more general attributes of river management whereas the other half needs to focus in more detail on the specific ecological aspects only. The choice alternatives also contain a cost attribute.

Table 1: Choice attributes of survey "River Management"

Attribute	Description	Levels
Flood protection	The attribute levels represent the level of flood protection whereby "every 25 years" means that there is protection against floods that occur statistically once in any 25 year period (low level). Practical implementation: Natural measures like flooded meadows and polder areas, technical flood protection measures in towns.	*Every 25 years (status quo)*; every 150 years (FLOOD150); every 200 years (FLOOD200)
Ecological status	The attribute levels represent the ecological status of the river in close analogy to the EU Water Framework Directive. Practical implementation: (1) an improvement of water quality by reducing agricultural fertilisers, pesticides and waste water; and (2) changing the physical structure of the banks and shape of the river	*Unsatisfactory to poor (status quo)*; moderate (ECO_MODERATE); good (ECO_GOOD); very good (ECO_VERYGOOD)
Out-of-town river access	The attribute levels represent the accessibility to the river and opportunities to stay, outside of towns. Practical implementation: Access points should be created preferably at flat river shores to allow a direct river experience with additional benches, information boards and playgrounds.	*Few access points (status quo)*; regular access points (ACCESS_REG); regular access points with benches, information boards and playgrounds (ACCESS_PLUS)
Development of towns and villages	The attribute levels represent the attractiveness of rivers in towns. Practical implementation: The creation of promenades, parks or playgrounds along the river.	*No development (status quo)*; development supported (TOWN_DEV)
Cost	The attribute levels display the different costs for the scenarios, whereby the costs are related to the waste water fee and "water cent" which have to be paid by every household living in the watershed area per year.	*€0 (status quo)*; €10; €20; €40; €80

In the choice task respondents were presented a sequence of 9 choice sets, each containing the status quo as well as two different alternatives. Respondents were asked to choose from each subsequent choice set their preferred alternative or the status quo. For the construction of the choice sets of each survey version the software Ngene (ChoiceMetrics 2012) was used to generate a D-efficient design (Scarpa and Rose 2008) based on zero priors.

After completion of the choice task perceived difficulty of the task was assessed in a follow-up question asking respondents "Was it rather difficult or easy for you to decide between the status quo and any of the change alternatives?". Responses on the scale "Very difficult", "Rather difficult", "Rather easy", "Very easy" were coded 4-3-2-1 into the variable DIFFICULTY. Alternatively, the binary variable DIFF was created indicating very or rather difficult responses (DIFF=1) as opposed to very or rather easy decisions (DIFF=0). Respondents who perceived the choice task to be very or rather difficult were subsequently asked "What exactly was the problem?". Responses to this open-ended question were sorted into thematic categories, and a set of problem dummies were created. These are used to get further insights into the exact nature of the issues around the choice task and their respective effect on scale. The remainder of the survey elicited a number of attitude and socio-demographic variables to be used in the regression analyses.

Table 2: Choice attributes of survey "River Ecology"

Attribute	Description	Levels
Water quality	The attribute levels represent the change in water quality in close analogy to the EU Water Framework Directive. Practical implementation: Change in water quality by reducing immissions of agricultural fertilisers and pesticides as well as chemical waste water	*Moderate (status quo)*; good (WAT_GOOD); very good (WAT_VERYGOOD)
Structure	The attribute levels represent the shape of the water body. Practical implementation: Measures to improve the structure of the water body by creating buffers and meanders of different extent.	*Narrow buffers & straight path (status quo)*; narrow buffers & slightly meandering path (STRUC_1); narrow buffers & moderately meandering (STRUC_2); wide buffers & strongly meandering (STRUC_3)
Species	The attribute levels represent the abundance of rare and endangered species in and around the water body. Practical implementation: Measures to improve the living conditions of endangered species.	*Low abundance (status quo)*; high abundance (SPEC_HI)
Cost	The attribute levels display the different costs for the scenarios. Whereby the costs are related to the waste water fee and "water cent" which have to be paid by every household living in the watershed area per year.	*€0 (status quo)*; €10; €20; €40; €80

3.2. Modelling approach

The analysis makes use of the GMNL modelling framework (Fiebig et al. 2010), which is generally based on the RUM (McFadden 1974). In any choice situation t the utility that respondent n derives from alternative i is

$$U_{nit} = \beta_n' x_{nit} + \varepsilon_{nit} \quad , \tag{1}$$

where x_{nit} is a vector containing variables specific to the alternative (and potentially the respondent), and β_n is a vector of parameters to be estimated. The elements of β_n, referred to as utility weights, are respondent specific and hence indexed with n. Utility of any one choice alternative for respondent n consists of the representative component $\beta_n{}' x_{nit}$, which can be observed by the analyst, and the idiosyncratic error term ε_{nit}. Under the assumption that respondents choose the alternative that maximises their utility and further that ε_{nit} is independently and identically distributed according to the Type I Extreme Value distribution, the probability that respondent n chooses alternative i in choice situation t is

$$P_{nit}|\beta_n = \frac{exp(\beta_n{}' x_{nit})}{\sum_{j=1}^{J} exp\left(\beta_n{}' x_{njt}\right)} \quad \forall\, j = 1, \dots, i, \dots, J\,;\, t = 1, \dots, T\,;\, n = 1, \dots, N. \tag{2}$$

Building on the RUM, heterogeneity of preferences across respondents is modelled by specifying β_n, the vector of the individual coefficients in the sample as

$$\beta_n = \sigma_n(\beta + \eta_n)\,, \tag{3}$$

where σ_n is the respondent-specific scale factor which shifts the coefficient vector β. η_n is a random variable with mean zero and estimable standard deviation and captures unobserved preference heterogeneity.

This general GMNL framework allows for several different specifications of β_n which express different assumptions regarding the structure of respondents' preference patterns. The well-known multinomial logit (MNL) model results from assuming constant scale, i.e. $\sigma_n = 1$, and no random preferences heterogeneity, i.e. $Var(\eta_n) = 0$. Assuming constant scale but allowing for random preference parameters, i.e. estimating $Var(\eta_n)$, leads to $\beta_n = \beta + \eta_n$, which is the random parameters logit (RPL) model (Revelt and Train 1998).[1] While a number of distributional forms of η_n are possible the analysis below uses the normal distribution, i.e. $\eta_n \sim N(0, \sigma)$, for all choice attributes, as is common in the literature.

Finally, allowing for heterogeneity in scale and preferences defines the generalized multinomial logit (GMNL) model. In this model, σ_n is estimated on a respondent-specific level. This scale factor shifts the whole vector of a respondent's preference weights up or down in magnitude compared to the unexplained component of the utility function (1). Consequently, it is inversely correlated to the variance of the error term ε_{nit} and provides a relative measure of the extent to which choice probability is explained by the (observable) attributes rather than being random. A higher individual scale factor indicates that a respondent's choices are determined to a larger extent by the attributes he sees on the choice cards, i.e. the choices are more consistent within the assumptions of the RUM. Analogously, a low scale factor is a signal for choices that were made mainly at random, without the influence of the depicted attribute levels.

To identify which respondent-specific characteristics drive the scale factor σ_n can be parameterised as

$$\sigma_n = exp(\bar{\sigma} + \theta' z_n + \tau\varepsilon_{0n}). \tag{4}$$

$\bar{\sigma}$ is a normalising constant; $\theta' z_n$ constitutes the systematic component of scale variation consisting of a vector of respondent-specific variables z_n and an associated coefficient vector θ'. In this general specification, $\tau\varepsilon_{0n}$, with $\varepsilon_{0n} \sim N(0,1)$, is the random component of scale. However, it has been criticised that despite its claims, the GMNL cannot separate random preference heterogeneity (captured in the unexplained variation of preference weights, $Var(\eta_n)$) and random scale heterogeneity (as expressed in $\tau\varepsilon_{0n}$) because of the multiplication of σ_n and β_n (Hess and Rose 2012). We therefore constrain the coefficient of the random component of scale, τ, to zero. Consequently, the model effectively becomes a heteroskedastic RPL, essentially an RPL model with respondent-specific scale determined by the elements of z_n. In our study perceived difficulty of the choice task is used as covariate of scale to explain more or less random choice behaviour across respondents. Model runs are performed in Stata 14 (Hole 2007, Gu et al. 2013) using 1,000 Halton draws to simulate the likelihood.

To investigate the effect of perceived difficulty of stating choices, we employ the following strategy for analysis. In a first step, the variable measuring perceived difficulty on a 4-point scale is included as covariate of scale in the way explained above. Subsequently, responses to the question "What exactly was the problem?", which was put to those who perceived the choice tasks as rather/very difficult, are sorted into a number of topical categories to construct dummy variables. The dummy variables indicating different sets of problems respondents perceived in the choice tasks are included in a second set of models. The base category in those models will be perceiving the choices

[1] Note that both in the MNL and RPL there is no need to define γ as it drops from equation (3).

as rather/very easy (i.e. not having been asked the follow-up question), so the dummy coding will allow to identify whether any one problem category has a particularly strong effect on scale. In a third step, the potential impact of perceived difficulty on preferences for river restoration programmes will be studied by interacting attribute levels with the difficulty dummy in non-scaled and heteroskedastic RPL models. This step is performed to effectively separate the effect of difficulty on preference and scale. In a last step, we simulate marginal WTP for the choice attributes separately for respondents who found the choice tasks easy and difficult.

4. Results

4.1. Sample characteristics and perceived difficulty

The surveys were implemented in the summer of 2015, and interviews were conducted in person by a group of trained enumerators in respondents' homes. Both surveys used the same sampling frame based on the same underlying study population. Table 3 reports sample characteristics for both data sets. Comparing means and standard deviations across samples, these turn out to be highly comparable.

Table 3: Characteristics of survey samples "River Management" and "River Ecology"

	River Management			River Ecology		
Variable (unit)	**N**	**Mean**	**Std. dev.**	**N**	**Mean**	**Std. dev.**
Age (years)	315	49.30	17.27	324	51.68	16.59
Male (share)	314	0.45		323	0.48	
Degree (share)	315	0.31		313	0.28	
Time in the region (years)	311	27.03	22.65	324	29.09	22.32
Rent[a] (years)	314	0.32	0.47	323	0.28	0.45
Household size (persons)	309	2.53	1.20	319	2.60	1.28
Income (refused) (share)	315	0.22		323	0.23	
Income (mean) (EUR)	247	2,876.52	1,647.84	248	3,059.48	1,609.29

Notes: [a] Respondent lives in rented accommodation;

4.2. Perceived difficulty and response scale

With respect to the perceived difficulty question there are no obvious differences in the response pattern between samples (Table 4). Only a small minority of 4 % perceived the tasks to be very difficult in both surveys. Approximately half of respondents and another quarter found the choice tasks rather easy or very easy.

Table 4: Responses to the perceived difficulty question in both surveys

	River Management		River Ecology	
	N	**Share**	**N**	**Share**
Very difficult	12	4 %	14	4 %
Rather difficult	71	23 %	79	25 %
Rather easy	150	48 %	153	48 %
Very easy	77	25 %	75	23 %
TOTAL	310	100 %	321	100 %

Notes…

Categorising open-ended responses to the subsequent question "What exactly was the problem?" provides some insights as to the reasons for subjective difficulty. Comparing the specifically mentioned problems between the two surveys in Table 5 reveals some systematic differences regarding the types of difficulties respondents had when performing the choice tasks. In the

Management sample making trade-offs, either among the project attributes themselves or between the project attributes and the cost attribute, constitute the main types of problems. The other difficulty categories were less but equally mentioned (8.4 % - 10.8 %), whereas the difficulty of having too little time to complete the choice tasks was only mentioned by 6 % of respondents with stated difficulties. This distribution is different in the Ecology sample: almost one quarter of respondents with difficulties stated that they had problems in making trade-offs between the project attributes followed by equal percentages (16 % – 17 %) for categories (2), (3) and (4). It is also striking that respondents felt the lack of knowledge more severe in the Ecology sample and also found the changes of conditions in the scenarios and the attribute level combinations more unrealistic. As in the Management sample ca. 10 % of respondents with difficulties stated that the choice task was unclear. While the percentage differences are not large, this comparison shows the tendency that in the Ecology sample respondents with difficulties had these predominantly in the process of understanding and comparing the different project related attributes and levels so that making the explicit trade-off between the project attributes and choosing on each choice set the preferred alternative was not as straightforward as it should be. The trade-off against costs, however, was mentioned slightly less often than in the Management sample.

Table 5: Categorised responses to the question "What exactly was the problem?"

	River Management		River Ecology	
Specific problem (Category)	**N**	**Share**	**N**	**Share**
(1) Making trade-offs among project attributes	14	16.9 %	22	23.7 %
(2) Unrealistic changes / combinations	8	9.6 %	16	17.2 %
(3) Lack of knowledge / information	9	10.8 %	15	16.1 %
(4) Trading off benefits against costs	18	21.7 %	16	17.2 %
(5) Too little time to complete the choice tasks[a]	5	6.0 %	2	2.2 %
(6) Optimal combination of attributes missing[a]	7	8.4 %	2	2.2 %
(7) Unclear (choice) task	8	9.6 %	10	10.8 %
(8) Other / No response	14	16.9 %	10	10.8 %
TOTAL	83	100.0 %	93	100.0 %

[a] Due to very few responses, these two categories were combined for the regression models of the Ecology sample.

We now turn to the regression analysis of choice responses and focus specifically on the effects of the difficulty variables on the scale parameter as a measure of more systematic vs. more random choice behaviour. The model on the left-hand side in Table 6 is a heteroskedastic RPL model for the Management sample. In terms of preference weights, results show that respondents have significant preferences for an improvement of the ecological status of the Queich (ECO_MODERATE, ECO_GOOD, ECO_VERYGOOD) and measures of town and village development (TOWN_DEV), but are indifferent towards out-of-town river access (ACCESS_REG, ACCESS_REP). As to the ecological status, it is remarkable that respondents value a good status more strongly than a very good status. Respondents do not value measures against floods occurring every 150 years (FLOOD150) and have negative preferences with respect to measures against floods occurring once in 200 years (FLOOD200). This result is counter-intuitive at first glance, but may be a consequence of the already rather high flood protection level along the Queich and the associated aversion of spending public money without foreseeable benefits for the population there.

Most importantly in this model, the coefficient of DIFFICULT, as covariate of scale, is significantly negative. Hence, perceived difficulty explains lower scale. Using the binary difficult dummy DIFF as covariate of scale also yields a significantly negative coefficient (not shown in table 6).

Table 6: Heteroskedastic RPL models of the River Management sample

	het. RPL (1)			het. RPL (2)		
	Coef.		**Std. Err.**	**Coef.**		**Std. Err.**
Mean of coefficients						
ASC_CHANGE	5.498	***	(1.344)	4.539	***	(0.914)
FLOOD150	-0.086		(0.290)	-0.054		(0.219)
FLOOD200	-1.104	***	(0.402)	-0.820	***	(0.288)
ECO_MODERATE	1.668	***	(0.430)	1.181	***	(0.270)
ECO_GOOD	4.282	***	(0.708)	3.247	***	(0.282)
ECO_VERYGOOD	3.447	***	(0.762)	2.525	***	(0.423)
ACCESS_REG	-0.352		(0.309)	-0.317		(0.222)
ACCESS_PLUS	0.094		(0.304)	0.019		(0.227)
TOWN_DEV	1.322	***	(0.354)	1.031	***	(0.226)
COST	-0.099	***	(0.017)	-0.074	***	(0.008)
Standard deviation of coefficients						
ASC_CHANGE	-8.652	***	(1.527)	6.523	***	(0.652)
FLOOD150	0.719		(0.538)	-0.260		(0.603)
FLOOD200	-0.322		(0.464)	-0.485		(0.304)
ECO_MODERATE	1.589	***	(0.460)	1.373	***	(0.302)
ECO_GOOD	0.476		(0.787)	-0.336		(0.557)
ECO_VERYGOOD	-2.283	***	(0.444)	1.603	***	(0.253)
ACCESS_REG	1.325	***	(0.506)	-0.256		(0.420)
ACCESS_PLUS	-1.741	***	(0.383)	1.231	***	(0.213)
TOWN_DEV	1.379	***	(0.326)	-0.963	***	(0.222)
COST	-0.070	***	(0.012)	-0.053	***	(0.006)
Covariates of scale						
DIFFICULT	-0.172	**	(0.068)			
prob_e1				-0.339		(0.229)
prob_e2				-0.373		(0.335)
prob_e3				-0.297		(0.271)
prob_e4				0.088		(0.239)
prob_e5				-0.802	**	(0.326)
prob_e6				-0.664	**	(0.301)
prob_e7				-0.634	**	(0.308)
prob_e8				-0.288		(0.248)
/tau	0		(constr.)	0		(constr.)
LL_0	-2,993			-2,993		
LL_m	-1,663			-1,657		
Observations	8,370			8,370		
Respondents	310			310		
Halton draws	1,000			1,000		
Parameters	21			28		
Adjusted R^2	0.437			0.437		
BIC	3,454			3,486		

*** indicate 1%-level of confidence. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model and the intercept-only model, respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC is preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

The left-hand column in Table 7 reports the same model for the Ecology sample. Here we find all preference coefficients to be significant. Respondents have preferences for improvements in water quality (WAT_GOOD, WAT_VERY), the different proposed changes to the structure of the course of the river and its banks (STRUC_1, STRUC_2, STRUC_3) as well as high abundance of rare and endangered species (SPEC_HI). While the different proposed river structures are valued with approximately the same intensity, respondents value very good water quality more strongly than merely good quality. As in the corresponding model in the Management sample, perceived difficulty (DIFFICULT) negatively affects scale. Here too, the effect of the binary difficult variable DIFF is significant and negative, as well.[2]

Table 7: Heteroskedastic RPL models of the River Ecology sample

	het. RPL (1)			het. RPL (2)		
	Coef.		**Std. Err.**	**Coef.**		**Std. Err.**
Mean of coefficients						
ASC_CHANGE	5.752	***	(1.197)	4.618	***	(0.698)
WAT_GOOD	1.119	***	(0.275)	0.849	***	(0.159)
WAT_VERY	1.343	***	(0.270)	1.010	***	(0.137)
STRUC_1	1.780	***	(0.369)	1.322	***	(0.186)
STRUC_2	1.974	***	(0.381)	1.497	***	(0.185)
STRUC_3	1.981	***	(0.419)	1.489	***	(0.220)
SPEC_HI	1.745	***	(0.325)	1.311	***	(0.144)
COST	-0.077	***	(0.015)	-0.058	***	(0.007)
Standard deviation of coefficient						
ASC_CHANGE	-9.155	***	(1.902)	-6.574	***	(0.731)
WAT_GOOD	-0.013		(0.405)	0.055		(0.297)
WAT_VERY	-1.093	***	(0.271)	-0.887	***	(0.169)
STRUC_1	-1.790	***	(0.402)	-1.321	***	(0.206)
STRUC_2	1.095	***	(0.417)	-0.804	***	(0.281)
STRUC_3	-2.318	***	(0.481)	-1.705	***	(0.253)
SPEC_HI	1.936	***	(0.375)	1.436	***	(0.167)
COST	-0.088	***	(0.016)	0.066	***	(0.007)
Covariates of scale						
DIFFICULT	-0.165	**	(0.074)			
prob_e1				-0.575	***	(0.214)
prob_e2				0.494		(0.402)
prob_e3				-0.194		(0.274)
Prob_e4				-0.107		(0.279)
prob_e56				-0.593		(0.592)
prob_e7				-0.269		(0.320)
prob_e89				-0.186		(0.323)
/tau	0		(constr.)	0		(constr.)
LL_0	-2,957			-2,957		
LL_m	-1,904			-1,903		
Observations	8,667			8,667		
Respondents	321			321		
Halton draws	1,000			1,000		
Parameters	17			23		
Adjusted R^2	0.350			0.349		
BIC	3,914			3,948		

*** indicate 1%-level of confidence. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model and the intercept-only model, respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC is preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

The right-hand columns in Tables 6 and 7 display the heteroskedastic RPL using dummies denoting the specific reasons for the difficulty as covariates of scale. For these dummy variables, the base category is perceiving the choice task as rather/very easy since only those respondents who found it

[2] Results of these alternative models are available from the authors on request.

rather/very difficult have answered the follow-up question. Therefore, the inclusion of these dummy variables allows to study which problems affect scale most compared to those respondents who found the tasks easy.[3] In the Management sample problems 5 "Too little time to complete the choice tasks", 6 "Optimal combination of attributes missing" and 7 "Unclear (choice) task" have significantly negative effects on scale. It is striking that all of these problems refer to the choice task rather than issues related to the topic of the survey or the description of the attributes. In the Ecology sample, however, we find that only problem 1 "Making trade-offs between all attributes" has a significant effect on scale. This problem was also the one most often mentioned by the respondents reporting perceived difficulty (Table 5).

While the above results strongly suggest that perceived difficulty decreases the scale parameter and thus leads to more random choice responses from a random utility perspective, it is still theoretically possible that the scale effect is in fact caused by a simultaneous and coincidental shift of all preference coefficients when moving from easy to difficult. Therefore, in the following the effect of perceived difficulty on preferences and scale is assessed simultaneously.

4.3. Separating systematic scale and preference heterogeneity

To test any potential differences in preferences between respondents who perceive the choice tasks as easy and difficult, we first run unscaled RPL interacting all attributes with DIFF, the binary version of the difficulty variable. They are displayed in two columns on the left-hand side of Tables 8 and 9. In the Management sample, the preference pattern is virtually the same as in the non-interacted models in Table 6 for the 'easy' group (i.e. interactions with EASY), but differs to some extent for the 'difficult' group (i.e. interacted with DIFF). In general, all but the cost coefficient are smaller for respondents stating that the choices were difficult. The differences for FLOOD150, FLOOD200, ECO_MODERATE, ECO_GOOD, ECO_VERYGOOD are statistically significant using a Wald-test of equality of coefficients. Notably however, the cost coefficient does not differ between these two groups (Wald-tests for COST: $p = 0.861$). In the Ecology sample, none of the differences between the two groups of respondents are significant. We can therefore conclude that difficulty (as measured by DIFF) does not affect preferences in this sample. In this sample as well, the cost coefficient is virtually the same across the two groups (Wald-test for COST: $p = 0.620$). When moving on to a heteroskedastic RPL in which DIFF is included as covariate of scale, we therefore exclude COST from interaction with DIFF.

In these models, reported in the right-hand columns in Tables 8 and 9, DIFF is included as covariate of scale and all but the cost attribute are interacted with DIFF. The most important finding in these models is that the significantly negative effect of difficulty on scale persists even when systematic differences in preferences are controlled for via the interactions with DIFF. Based on the finding that the COST coefficients do not differ between the easy and difficult groups, we can be sure that this is a genuine scale effect rather than a disguised shift in preferences. Consequently, comparing respondents that find the choice tasks easy or difficult, the differences in scale come in addition to any potential differences in preferences caused by this distinction. The differences in preferences between the two groups found in the Management sample disappear in the heteroskedastic model. Using Wald-tests, none of the differences between easy and different turns out to be significant. In the River Ecology sample, the lack of an effect of the easy-difficult distinction on preferences found in the unscaled RPL model is corroborated in the heteroskedastic model.

[3] Due to the very low numbers of cases, categories 5 "Too little time to complete the choice tasks" and 6 "Optimal combination of attributes missing" are combined in the Ecology sample. Consequently only one dummy variable is used to indicate both categories.

Table 8: Unscaled and heteroskedastic RPL models including interactions with the difficulty dummy in the River Management sample

	RPL		het. RPL	
	Coef.	**Std. Err.**	**Coef.**	**Std. Err.**
Mean of coefficients				
EASY*ASC_CHANGE	3.689 ***	(1.039)	4.125 ***	(0.990)
EASY*FLOOD150	0.301	(0.268)	0.125	(0.258)
EASY*FLOOD200	-0.353	(0.344)	-0.635 *	(0.325)
EASY*ECO_MODERATE	1.583 ***	(0.326)	1.451 ***	(0.314)
EASY*ECO_GOOD	3.511 ***	(0.335)	3.607 ***	(0.343)
EASY*ECO_VERYGOOD	3.371 ***	(0.546)	3.029 ***	(0.503)
EASY*ACCESS_REG	-0.187	(0.270)	-0.379	(0.259)
EASY*ACCESS_PLUS	0.243	(0.278)	0.020	(0.263)
EASY*TOWN_DEV	1.308 ***	(0.277)	1.214 ***	(0.275)
EASY*COST	-0.069 ***	(0.009)		
DIFF*ASC_CHANGE	4.993 ***	(1.319)	5.817 ***	(1.555)
DIFF*FLOOD150	-0.571 *	(0.326)	-0.429	(0.462)
DIFF*FLOOD200	-1.477 ***	(0.439)	-1.408 ***	(0.507)
DIFF*ECO_MODERATE	0.595 *	(0.354)	1.060 *	(0.577)
DIFF*ECO_GOOD	1.928 ***	(0.383)	2.993 ***	(0.681)
DIFF*ECO_VERYGOOD	0.923	(0.569)	2.201 **	(0.935)
DIFF*ACCESS_REG	-0.244	(0.334)	0.096	(0.540)
DIFF*ACCESS_PLUS	-0.170	(0.358)	0.122	(0.492)
DIFF*TOWN_DEV	0.427	(0.332)	0.922 *	(0.545)
DIFF*COST	-0.067 ***	(0.010)		
COST			-0.077 ***	(0.008)
Standard deviation of coefficients				
EASY*ASC_CHANGE	6.358 ***	(0.679)	6.607 ***	(0.743)
EASY*FLOOD150	0.353	(0.519)	0.268	(0.562)
EASY*FLOOD200	-0.595 **	(0.273)	0.429	(0.280)
EASY*ECO_MODERATE	1.224 ***	(0.334)	1.239 ***	(0.331)
EASY*ECO_GOOD	0.591	(0.588)	-0.809 **	(0.371)
EASY*ECO_VERYGOOD	1.891 ***	(0.293)	1.962 ***	(0.305)
EASY*ACCESS_REG	-0.624	(0.400)	0.065	(0.687)
EASY*ACCESS_PLUS	1.251 ***	(0.242)	1.253 ***	(0.234)
EASY*TOWN_DEV	0.987 ***	(0.247)	-1.132 ***	(0.237)
EASY*COST	0.063 ***	(0.008)		
DIFF*ASC_CHANGE	-5.026 ***	(0.922)	-7.972 ***	(1.984)
DIFF*FLOOD150	0.567	(0.442)	-1.112	(0.746)
DIFF*FLOOD200	0.254	(0.401)	-0.257	(0.735)
DIFF*ECO_MODERATE	-0.612	(0.718)	1.463 **	(0.606)
DIFF*ECO_GOOD	-1.163 ***	(0.423)	1.097	(0.772)
DIFF*ECO_VERYGOOD	1.186 ***	(0.314)	1.492 ***	(0.519)
DIFF*ACCESS_REG	0.552	(0.791)	1.148	(0.973)
DIFF*ACCESS_PLUS	-1.230 ***	(0.339)	1.675 ***	(0.479)
DIFF*TOWN_DEV	-0.697 *	(0.382)	1.521 ***	(0.496)
DIFF*COST	-0.033 ***	(0.007)		
COST			-0.059 ***	(0.007)
Covariates of scale				
DIFF			-0.393 **	(0.181)
/tau			0	(constr.)
LL_m	-1,652		-1,650	
Observations	8,370		8,370	
Halton draws	1,000		1,000	
Parameters	40		39	
Adjusted R^2	0.435		0.436	
BIC	3,550		3,540	

*** indicate 1%-level of confidence. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model and the intercept-only model, respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC is preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

Table 9: Unscaled and heteroskedastic RPL model including interactions with the difficulty dummy in the Ecology sample

	RPL			het. RPL		
	Coef.		**Std. Err.**	**Coef.**		**Std. Err.**
Mean of coefficients						
EASY*ASC_CHANGE	4.808	***	(0.816)	5.091	***	(0.883)
EASY*WAT_GOOD	0.860	***	(0.185)	0.915	***	(0.186)
EASY*WAT_VERY	0.935	***	(0.157)	0.976	***	(0.154)
EASY*STRUC_1	1.131	***	(0.214)	1.136	***	(0.216)
EASY*STRUC_2	1.372	***	(0.205)	1.442	***	(0.210)
EASY*STRUC_3	1.335	***	(0.261)	1.491	***	(0.265)
EASY*SPEC_HI	1.199	***	(0.161)	1.332	***	(0.171)
EASY*COST	-0.052	***	(0.008)			
DIFF*ASC_CHANGE	3.135	***	(0.839)	4.601	***	(1.335)
DIFF*WAT_GOOD	0.655	*	(0.241)	0.777	**	(0.320)
DIFF*WAT_VERY	1.064	***	(0.209)	1.294	***	(0.294)
DIFF*STRUC_1	1.421	***	(0.265)	1.935	***	(0.464)
DIFF*STRUC_2	1.455	***	(0.276)	1.819	***	(0.442)
DIFF*STRUC_3	1.557	***	(0.326)	1.858	***	(0.421)
DIFF*SPEC_HI	1.296	***	(0.211)	1.671	***	(0.333)
DIFF*COST	-0.058	***	(0.009)			
COST				-0.060	***	(0.008)
Standard deviation of coefficients						
EASY*ASC_CHANGE	-6.764	***	(0.809)	6.736	***	(0.895)
EASY*WAT_GOOD	0.148		(0.289)	-0.158		(0.291)
EASY*WAT_VERY	-0.966	***	(0.202)	0.917	***	(0.189)
EASY*STRUC_1	-1.418	***	(0.248)	1.506	***	(0.253)
EASY*STRUC_2	-0.830	*	(0.319)	-0.938	***	(0.324)
EASY*STRUC_3	-2.097	***	(0.317)	-1.980	***	(0.293)
EASY*SPEC_HI	1.455	***	(0.188)	-1.432	***	(0.177)
EASY*COST	0.068	***	(0.008)			
DIFF*ASC_CHANGE	5.796	***	(1.093)	7.577	***	(1.974)
DIFF*WAT_GOOD	0.213		(0.902)	0.268		(0.780)
DIFF*WAT_VERY	-0.539	*	(0.280)	0.592		(0.370)
DIFF*STRUC_1	0.615	**	(0.309)	-0.907	**	(0.423)
DIFF*STRUC_2	-0.319		(0.524)	-0.049		(0.981)
DIFF*STRUC_3	0.752	**	(0.331)	0.736		(0.670)
DIFF*SPEC_HI	-1.110	***	(0.215)	1.484	***	(0.371)
DIFF*COST	0.044	***	(0.008)			
COST				-0.068	***	(0.008)
Covariates of scale						
DIFF				-0.324	*	(0.188)
/tau				0		(constr.)
LL_0	-2,957			-2,957		
LL_m	-1,891			-1,898		
Observations	8,667			8,667		
Respondents	321			321		
Halton draws	1,000			1,000		
Parameters	32			31		
Adjusted R^2	0.350			0.348		
BIC	3,981			3,987		

*** indicate 1%-level of confidence. Adjusted R^2 is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where LL_m and LL_0 are the log-likelihoods of the full model and the intercept-only model, respectively, and k the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with N denoting the number of respondents. The use of BIC is preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model.

4.4. Welfare estimation

In the last step we analyse the consequences of the effects of the difficulty variables as described in the regression models in Tables 8 and 9 on the estimates of marginal willingness to pay (WTP) for the project attributes in the two surveys. Since in all models the cost coefficients are normally distributed, estimates of marginal WTP for attribute k cannot be computed simply as $mWTP_k = -\beta_k\beta_{cost}^{-1}$. This results from the fact that moments of the resulting distribution of estimates would not be defined (Meijer and Rouwendal 2006, Carson and Czajkowski 2013). Instead, a bootstrapping approach is used to simulate marginal WTP for different choice attributes. This approach is a modified version of the WTP simulation in Czajkowski et al. (2016). As a first step, 10^4 draws are taken from the multivariate normal distribution described by the mean of the coefficients and the asymptotic variance-covariance matrix of the maximum likelihood model. For each of the 10^4 draws, the individual-specific coefficients, conditional on individual stated choices, are calculated (Revelt and Train 2000, Campbell 2007). These individual coefficients per draw in Step 1 ($n = 310$ in the Management sample and $n = 321$ in the Ecology sample) are used in Step 2 to compute marginal WTP for attribute k as $mWTP_k = -\beta_k\beta_{cost}^{-1}$, and the medians are stored. Finally, the means and 95%-confidence intervals of the 10^4 simulated medians of marginal WTP are computed and reported in Tables 10 and 11.

Table 10: Simulated marginal WTP of choice attributes in the River Management sample

	Based on coefficients from unscaled RPL (Table 8)					
	easy			difficult		
Variable	**Mean**	**CI (95%)**	**Size**	**Mean**	**CI (95%)**	**Size**
FLOOD150	4.18	[-1.73 - 9.94]	11.68	-8.40	[-16.20 - -0.59]	15.61
FLOOD200	-5.16	[-12.88 - 2.49]	15.38	-21.92	[-30.30 - -13.08]*	17.22
ECO_MODERATE	21.30	[13.27 - 28.35]	15.08	9.23	[0.38 - 19.15]	18.77
ECO_GOOD	49.28	[40.05 - 58.59]	18.54	29.25	[19.62 - 40.37]	20.76
ECO_VERYGOOD	42.89	[30.46 - 54.44]	23.99	14.98	[-0.09 - 32.65]	32.74
ACCESS_REG	-2.32	[-8.18 - 3.41]	11.59	-3.28	[-11.09 - 4.84]	15.93
ACCESS_PLUS	2.35	[-3.65 - 8.33]	11.98	-2.13	[-10.30 - 7.04]	17.34
TOWN_DEV	17.83	[11.68 - 24.34]	12.67	6.62	[-1.63 - 16.13]	17.76

	Based on coefficients from heteroskedastic RPL (Table 8)					
	easy			difficult		
Variable	**Mean**	**CI (95%)**	**Size**	**Mean**	**CI (95%)**	**Size**
FLOOD150	1.60	[-3.61 - 6.89]	10.49	-5.29	[-15.09 - 4.41]	19.49
FLOOD200	-7.89	[-14.17 - -1.43]	12.74	-17.76	[-27.60 - -7.60]	20.00
ECO_MODERATE	18.27	[10.73 - 25.73]	15.00	13.25	[1.40 - 25.13]	23.73
ECO_GOOD	46.39	[39.33 - 54.04]	14.71	37.89	[24.85 - 50.92]	26.07
ECO_VERYGOOD	34.95	[24.80 - 45.93]	21.13	27.65	[8.60 - 46.44]	37.84
ACCESS_REG	-4.57	[-9.63 - 0.55]	10.18	1.75	[-9.81 - 13.45]	23.26
ACCESS_PLUS	-0.25	[-4.94 - 5.01]	9.95	1.59	[-8.36 - 11.54]	19.90
TOWN_DEV	15.03	[9.23 - 21.19]	11.96	11.43	[0.68 - 22.72]	22.04

All figures in EUR. Simulations based on models in Table 8. CI - confidence interval. * Confidence intervals of 'easy' and 'difficult' do not overlap.

WTP estimates in the Management sample differ to some extent between the easy/difficult groups with respondents who find the choice tasks difficult generally having a lower WTP for the respective attributes. However, only the difference for FLOOD200 is significant, i.e. the 95% confidence intervals of the bootstrapped WTP estimates do not overlap. In the Ecology sample, WTP estimates in the easy group are slightly smaller, yet none of these differences is significant. In both samples, the confidence intervals are substantially bigger for the 'difficulty' group. This confirms the existence of a scale effect of difficulty as found in Tables 6 and 7. Since the weight of the systematic component of the utility function relative to the error component is larger for the 'easy' group, bootstrapped WTP estimates have got a higher level of precision for the 'easy' group as compared to the 'difficult' group.

Table 11: Simulated marginal WTP of choice attributes in the Ecology sample

	Based on coefficients from unscaled RPL (Table 9)					
		easy			difficult	
Variable	**Mean**	**CI (95%)**	**Size**	**Mean**	**CI (95%)**	**Size**
WAT_GOOD	12.56	[8.48 - 16.84]	8.35	10.84	[4.54 - 17.42]	12.88
WAT_VERY	14.04	[10.33 - 17.87]	7.54	18.17	[12.96 - 23.88]	10.92
STRUC_1	18.09	[12.95 - 23.60]	10.65	23.81	[15.91 - 32.94]	17.03
STRUC_2	19.18	[14.58 - 24.19]	9.61	24.76	[17.00 - 33.71]	16.71
STRUC_3	15.35	[10.14 - 20.57]	10.43	25.93	[18.62 - 33.80]	15.19
SPEC_HI	17.49	[14.19 - 21.05]	6.86	22.08	[16.70 - 28.17]	11.47

	Based on coefficients from heteroskedastic RPL (Table 9)					
		easy			difficult	
Variable	**Mean**	**CI (95%)**	**Size**	**Mean**	**CI (95%)**	**Size**
WAT_GOOD	10.42	[7.13 - 13.83]	6.69	8.57	[2.78 - 14.42]	11.64
WAT_VERY	10.89	[8.09 - 13.79]	5.70	14.37	[9.61 - 19.09]	9.48
STRUC_1	13.55	[9.92 - 17.63]	7.71	21.86	[14.19 - 29.57]	15.38
STRUC_2	15.57	[11.65 - 19.83]	8.19	20.23	[12.96 - 27.42]	14.46
STRUC_3	13.65	[9.58 - 18.01]	8.43	20.48	[13.50 - 27.40]	13.91
SPEC_HI	14.44	[11.56 - 17.62]	6.06	18.33	[13.13 - 23.54]	10.41

All figures in EUR. Simulations based on models in Table 9. CI - confidence interval

A very similar picture arises when WTP is bootstrapped based on the heteroskedastic RPL models in Tables 8 and 9. In the Management sample perceived difficulty leads to slightly lower WTP, albeit not significant, and in the Ecology sample none of the WTP differences is significant, either. In both samples, the confidence intervals of the WTP estimates are again substantially larger for the respondents who find the choice tasks difficult. These persisting size differences corroborate the finding of a scale effect of the easy-difference distinction even when the effect of the latter on preferences is controlled for. Note that the difference in the size of the confidence intervals somewhat increases when the scale effect of DIFF is controlled for in the Management sample (Table 10). While the confidence intervals for the 'easy' group slightly decrease in size, the intervals for the 'difficult' group are somewhat bigger in this model. In the Ecology sample, the sizes of the confidence intervals for both group decrease marginally when moving from the unscaled to the heteroskedastic RPL.

5. Discussion and conclusions

The present study investigates the role of perceived difficulty of the choice task for choice consistency and preference heterogeneity. We used a subjective measure of difficulty as opposed to objective, constructed measures as done in earlier studies in the literature (e.g. Caussade et al. 2005, Louviere et al. 2008). As the data show, less than 30% of respondents perceived the choice task as difficult, which demonstrates that there exists considerable heterogeneity in difficulty perception. Also, the reasons for the task being perceived as difficult are rather heterogeneous. Consequently, we see considerable merit in using a perceived instead of or in addition to a constructed difficulty measure to gain further insight into the potential improvement of the DCE approach.

Our main result is that difficulty as perceived by survey respondents in the form of a self-report measure has a negative effect on the scale parameter in a scaled RPL model indicating less consistent and more random choices by respondents when the task is perceived as difficult. This result is in agreement with earlier findings on the negative effect of perceived difficulty on choice consistency (Bech et al. 2011, Ruokamo et al. 2016). Going beyond those studies, our results corroborate their finding that self-reported difficulty reduces choice consistency regardless of the survey topic. This finding provides support for the hypothesis that the scale effect of difficulty is generally a behavioural rather than a topic-specific phenomenon (Louviere et al. 2008).

However, when comparing the more specific effects of difficulty between our two samples Management and Ecology we find that the nature of attribute trade-offs seems to matter. While in the Management sample some respondents had difficulties due to making trade-offs, unrealistic choices and missing information, the difficulty aspects that significantly influenced the scale parameter were not these but the procedural aspects of the choice tasks like the time constraint, the missing optimal combination or unclear choice task. This was found to be different in the Ecology sample where only the aspect of making trade-offs between the project attributes (not money!) significantly influenced scale, the procedural aspects played no role. This points to severe problems with the design of the chosen attributes in the Ecology sample. From the results we deduct that a considerable fraction of respondents (here almost one quarter) did not perceive the attributes as suitable for trade-offs but, as we suppose, considered ecology as a whole within which trade-offs do not make sense. Forcing such respondents into making trade-offs in such a situation leads to significantly less consistent and, therefore, more noisy choice responses as demonstrated by the influence on the scale factor.

Consequently, overall we seem to find that the topical context in which trade-offs have to be made may be decisive for perceived difficulty and, thus, for response quality. The attributes of the Management sample seemed sufficiently closed in the sense that they stand for separate aspects of rivers that, in principle, could be valued independently from each other like ecology vs. flood protection vs. river access vs. village development. Zooming into one specific attribute, here ecology, however, seemed to have created problems in the sense that the single ecology attributes are perceived to belong to one single sphere (environment) within which respondents find it more difficult or even improper to make trade-offs since all of the corresponding attributes are considered valuable jointly. From this it follows that while we do not find evidence that a specific valuation topic may create perceived difficulty or not, the level of aggregation of attributes appears to be important in DCE survey design. Attributes must be sufficiently closed in their definition and independently valuable as perceived by respondents to allow them to make meaningful trade-offs and, thus, choices of alternatives.

The analysis also finds some limited evidence for differences in preferences between respondents who perceive the choice tasks as easy or difficult, yet these differences are confined to the Management sample. While preferences for some attributes differ significantly between the 'easy' and 'difficult' in the unscaled model, these differences are insignificant when the scaled effect of difficulty is controlled for in the heteroskedastic model. This effect if puzzling and future research into the interaction between scale and preference heterogeneity could look into this aspect further.

Differences between the 'easy' and 'difficult' groups in terms of marginal WTP for choice attributes are not significant (with the exception of FLOOD200 in the unscaled RPL model). This result is not surprising for two reasons. Firstly, the scale factor cancels out when marginal WTP is computed as the ratio of attribute and cost coefficient. Secondly, despite the significantly different coefficients of some attributes in the unscaled RPL model in the Management sample, the cost coefficients are not significantly different between 'easy' and 'difficult'. The difference in precision of the welfare estimates as evidenced by the substantially smaller confidence intervals for the 'easy' group, however, provides further support to the finding of a robustly significant negative scale effect of perceived difficulty.

In terms of methodology, a potential weakness of the analysis lies in the categorisation of the reasons for difficulty (Table 5). While most open-ended responses could be clearly sorted into one of the categories in a straightforward way, some statements were ambiguous. We dealt with this potential ambiguity by conducting sensitivity analyses using different variants of classification, but we found no significant differences. Future studies might use closed-ended questions to further investigate the sources of perceived difficulty.

On the positive side, this study demonstrates a research plan that manages to disentangle the influence of a respondent-specific variable (perceived difficulty) on interpersonal differences in scale and preferences. It therefore builds on and further develops the approach in Czajkowski et al. (2016) to investigate systematic heterogeneity in preferences and scale. Unlike previous studies, however, our analysis tests for the systematic difference of marginal utility of the price attribute before excluding it from interaction with the difficulty dummy in the heteroskedastic RPL models. Taking this approach we can avoid a situation in which a significant scale effect of the dummy variable is potentially a disguised preference effect of the non-interacted attribute. The presented findings are therefore more strongly suggesting the existence of a scale effect compared to basing the selection of the attribute not to be interacted merely on theory. We recommend future studies on scale and preference effects of respondent-specific or treatment variables to take a similar econometric strategy.

The above results provide the basis for further research questions. (i) An obvious one is the relationship between constructed measures of task complexity and perceived difficulty and how this interplay affects choice behaviour. The relationship between increased complexity, in terms of e.g. number of attributes or choice alternatives, and perceived difficulty is not necessarily linear. That means that it is possible to determine an optimal mix of complexity characteristics which does not exceed a specific threshold in perceived difficulty (Bech et al. 2011). (ii) Another question is the mechanism that makes choices which are perceived to be more difficult more random. Following the latent class approach in (Swait and Adamowicz 2001), recent research into choice heuristics (e.g. Sandorf et al. in press) could be used to explore whether respondents perceiving difficulties are more likely to ignore particular attributes or choice alternatives in an attempt to simplify the choice task. We leave these issues for future research.

Acknowledgements

The work was supported through the project „Possibilities and constraints of assessing the social value of river rehabilitation measures in monetary units - the example of the Queich and Simmerbach" funded by the Ministry of Environment, Agriculture, Food, Viticulture and Forestry Rhineland-Palatinate, Germany (Ministerium für Umwelt, Landwirtschaft, Ernährung, Weinbau und Forsten (MULEWF) Rheinland-Pfalz); project number 103-92 151/2014-1#10.

References

Aanesen, M., C. Armstrong, M. Czaijkowski, J. Falk-Petersen, N. Hanley and S. Navrud (2015) Willingness to pay for unfamiliar public goods: Preserving cold-water coral in Norway. Ecological Economics 112, 53-67.

Adamowicz, W., J. Louviere and M. Williams (1994) Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. Journal of Environmental Economics and Management 26, 271-292.

Adamowicz, W., P. Boxall, M. Williams and J. Louviere (1998) Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation. American Journal of Agricultural Economics 80, 64-75.

Arentze, T., A. Borgers, H. Timmermans and R. DelMistro (2003) Transport stated choice responses: effects of task complexity, presentation format and literacy. Transportation Research Part E: Logistics and Transportation Review 39, 229-244.

Bech, M., T. Kjaer and J. Lauridsen (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. Health Economics 20, 273-286.

Beck, M.J., J.M. Rose and D.A. Hensher (2013) Consistently inconsistent: The role of certainty, acceptability and scale in choice. Transportation Research Part E: Logistics and Transportation Review 56, 81-93.

Börger, T. (2016) Are fast responses more random? Testing the effect of response time on scale in an online choice experiment. Environmental and Resource Economics 65, 389-413.

Börger, T., C. Hattam, D. Burdon, J.P. Atkins and M.C. Austen (2014) Valuing conservation benefits of an offshore marine protected area. Ecological Economics 108, 229-241.

Boxall, P. C., W.L. Adamowicz, J. Swait, M. Williams and J. Louviere (1996) A comparison of stated preference methods for environmental valuation. Ecological Economics 18, 243-253.

Boxall, P., W.L. Adamowicz and A. Moon (2009) Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. Australian Journal of Agricultural and Resource Economics 53, 503-519.

Campbell, D. (2007) Willingness to Pay for Rural Landscape Improvements: Combining Mixed Logit and Random-Effects Models. Journal of Agricultural Economics 58, 467-483.

Carlsson, F. and P. Martinsson (2008) How Much is Too Much? Environmental and Resource Economics 40, 165-176.

Carson, R.T. and M. Czajkowski (2013) A new baseline model for estimating willingness to pay from discrete choice models. International Choice Modelling Conference 2013, Sydney.

Carson, R.T. and W.M. Hanemann (2005) Contingent valuation, in: K.-G. Mäler and J.R. Vincent (eds.) Handbook of Environmental Economics. Elsevier, 821-936

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

ChoiceMetrics (2012) Ngene 1.1.1 User Manual and Reference Guide.

Czajkowski, M., N. Hanley and J. LaRiviere (2015) The Effects of Experience on Preferences: Theory and Empirics for Environmental Public Goods. American Journal of Agricultural Economics 97, 333-351.

Czajkowski, M., N. Hanley and J. LaRiviere (2016) Controlling for the Effects of Information in a Public Goods Discrete Choice Model. Environmental and Resource Economics 63, 523-544.

Czajkowski, M., T. Kądziela and N. Hanley (2014) We want to sort! Assessing households' preferences for sorting waste. Resource and Energy Economics 36, 290-306.

de Palma, A., G.M. Myers and Y.Y. Papageorgiou (1994) Rational Choice Under an Imperfect Ability To Choose. The American Economic Review 84, 419-440.

Dellaert, B.G.C., J.D. Brazell and J.J. Louviere (1999) The Effect of Attribute Variation on Consumer Choice Consistency. Marketing Letters 10, 139-147.

DeShazo, J.R. and G. Fermo (2002) Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency. Journal of Environmental Economics and Management 44, 123-143.

Fiebig, D.G., M.P. Keane, J. Louviere and N. Wasi (2010) The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Marketing Science 29, 393-421.

Gu, Y., A.R. Hole and S. Knox (2013) Estimating the generalized multinomial logit model in Stata. The Stata Journal 13, 382-397.

Hanley, N., R.E. Wright and V. Adamowicz (1998) Using choice experiments to value the environment. Environmental and Resource Economics 11, 413-428.

Heiner, R.A. (1983) The origin of predictable behavior. American Economic Review 73, 560-595.

Hensher, D.A. (2004) Identifying the Influence of Stated Choice Design Dimensionality on Willingness to Pay for Travel Time Savings. Journal of Transport Economics and Policy 38, 425-446.

Hess, S. and J.M. Rose (2012) Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation 39, 1225-1239.

Hole, A.R. (2007) Fitting mixed logit models by using maximum simulated likelihood. The Stata Journal 7, 388-401.

Jacobsen, J.B., J.H. Boiesen, B.J. Thorsen and N. Strange (2008) What's in a name? The use of quantitative measures versus 'Iconised' species when valuing biodiversity. Environmental and Resource Economics 39, 247-263.

Jobstvogt, N., N. Hanley, S. Hynes, J. Kenter and U. Witte (2014) Twenty thousand sterling under the sea: Estimating the value of protecting deep-sea biodiversity. Ecological Economics 97, 10-19.

LaRiviere, J., M. Czajkowski, N. Hanley, M. Aanesen, J. Falk-Petersen and D. Tinch (2014) The value of familiarity: Effects of knowledge and objective signals on willingness to pay for a public good. Journal of Environmental Economics and Management 68, 376-389.

Louviere, J. J., R.T. Carson, L. Burgess, D. Street and A.A.J. Marley (2013) Sequential preference questions factors influencing completion rates and response times using an online panel. Journal of Choice Modelling 8, 19-31.

Louviere, J J., T. Islam, N. Wasi, D. Street and L. Burgess (2008) Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? Journal of Consumer Research 35, 360-375.

Mathews, K.E., M. L. Freeman and W.H. Desvousges (2006) How and how much? Pages 111-133 *in* B. J. Kanninen, editor. Valuing environmental amenities using stated choice studies. Springer, Dordrecht.

Mazzotta, M.J. and J.J. Opaluch (1995) Decision Making When Choices Are Complex: A Test of Heiner's Hypothesis. Land Economics 71, 500-515.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. Pages 105-142 in P. Zarembka (ed.) Frontiers in econometrics, economic theory and mathematical economics. Academic Press, New York.

McVittie, A. and D. Moran (2010) Valuing the non-use benefits of marine conservation zones: An application to the UK Marine Bill. Ecological Economics 70, 413-424.

Meijer, E. and J. Rouwendal (2006) Measuring welfare effects in models with random coefficients. Journal of Applied Econometrics 21, 227-244.

Norton, D. and S. Hynes (2014) Valuing the non-market benefits arising from the implementation of the EU Marine Strategy Framework Directive. Ecosystem Services 10, 84-96.

Revelt, D. and K. Train. 2000. Customer-specific taste parameters and mixed logit: Households' choice of electricity supplier. University of California, Berkeley.

Rose, J.M., D.A. Hensher, S. Caussade, J.d.D. Ortúzar and R.-C. Jou (2009) Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. Journal of Transport Geography 17, 21-29.

Ruokamo, E., M. Czajkowski, N. Hanley, A. Juutinen and R. Svento (2016) Linking perceived choice complexity with scale heterogeneity in discrete choice experiments: Home heating in Finland. Discussion Papers in Environmental Economics. University of St Andrews, St Andrews.

Sandorf, E.D., D. Campbell and N. Hanley (in press) Disentangling the influence of knowledge on attribute non-attendance. Journal of Choice Modelling. Doi: 10.1016/j.jocm.2016.09.003.

Scarpa, R. and J. M. Rose (2008) Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. Australian Journal of Agricultural and Resource Economics 52, 253-282.

Scarpa, R. and M. Thiene (2005) Destination Choice Models for Rock Climbing in the Northeastern Alps: A Latent-Class Approach Based on Intensity of Preferences. Land Economics 81, 426-444.

Swait, J. and W. Adamowicz (2001) The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. Journal of Consumer Research 28, 135-148.

Yao, R.T., R. Scarpa, J.A. Turner, T.D. Barnard, J.M. Rose, J.H.N. Palma and D.R. Harrison (2014) Valuing biodiversity enhancement in New Zealand's planted forests: Socioeconomic and spatial determinants of willingness-to-pay. Ecological Economics 98, 90-101.