Logar I, Brouwer R & Campbell D (2020) Does attribute order influence attribute-information processing in discrete choice experiments? *Resource and Energy Economics*, 60, Art. No.: 101164.

DOI:https://doi.org/10.1016/j.reseneeco.2020.101164

Does Attribute Order Influence Attribute-Information Processing in Discrete Choice Experiments?

Abstract

The existing empirical evidence shows that both contingent valuation and discrete choice experiment (DCE) methods are susceptible to various ordering effects. However, very few studies have analysed attribute-ordering effects in DCEs, and no study has investigated their potential influence on information-processing strategies, such as attribute non-attendance (ANA). This paper tests for attribute-ordering effects and examines whether the order of attributes describing the alternatives affects respondents' propensity to attend to or ignore an attribute. A split-sample approach is used, where one sample received a DCE version in which the positions of the first and last non-monetary attributes are switched across the sequence of choice tasks compared with the other sample. The results show that attribute order does not affect welfare estimates in a significant way under the standard assumption of full attribute attendance, thus rejecting the notion of procedural bias. However, the welfare estimates for the attributes whose order was reversed and the share of respondents who ignored them differ significantly between the two attribute-ordering treatments once ANA behaviour is accounted for in the estimated choice models. These results highlight the important role of information-processing strategies in the design and evaluation of DCEs.

Keywords: Ordering effects; Information processing; Attribute non-attendance; Discrete choice experiment; Stated preferences; Convergent validity

JEL codes: Q25, Q51, C25, D12

1 Introduction

There is an increasing interest in the discrete choice experiment (DCE) literature in trying to identify the behavioural rules that respondents adopt when processing the information provided in DCEs which ultimately affect their choices. A standard assumption in the neoclassical theoretical framework underlying DCEs is that individuals, as rational economic agents, consistently choose alternatives that maximise the utility they derive from goods with different characteristics (Rabin, 1998; McFadden, 2001). This implies that survey respondents are able to process all the information provided in a rational manner, i.e. they make trade-offs between each and every attribute associated with each alternative and choose their most preferred alternative in a choice set. It has, however, been demonstrated that, when making a choice, individuals often use a number of simplifying decision strategies or choice heuristics in processing the information contained in the attributes which describe alternatives. Examples include attribute non-attendance (ANA), elimination by aspects, attribute aggregation, and parameter transfer between common-metric attributes (Hensher and Greene, 2010; Erdem et al., 2014). Of these, ANA has received particular attention in the DCE literature (Hensher and Rose, 2009; Alemu et al., 2013; Johnston et al., 2017). It refers to a situation in which one or more attributes, and their associated levels, are ignored by a respondent when evaluating alternatives in a choice set (Hensher et al., 2005). In this study, we test to what extent the order in which attributes are presented in a choice task influences attribute attendance. The link between attribute ordering and ANA has, as far as we know, not yet been investigated in the DCE literature.

There is a substantial amount of evidence that stated preferences are susceptible to various ordering effects (e.g. Carson and Mitchell, 1995; Herriges and Shogren, 1996;

Halvorsen, 1996; Holmes and Boyle, 2005; Day et al., 2010; 2012; Carlsson et al., 2012). A number of visual ANA studies suggest that the order in which attributes are presented in a choice task in a DCE may affect both respondents' choices and their level of attendance to an attribute (Balcombe et al., 2015; Spinks and Mortimer, 2016; Selivanova and Krabbe, 2018). However, formal tests of attribute-ordering effects are scarce. Although especially the monetary attribute is of interest for the economic valuation purposes in DCE studies, Kjær et al. (2006), Boyle and Özdemir (2009) and to some extent Glenk (2007) and Krucien et al. (2017) have already investigated attribute-ordering effects due to the positioning of the price attribute, albeit without controlling for ANA. For this reason, our study focuses on the positioning of the non-monetary attributes.

This study contributes to the DCE literature in environmental economics in two significant ways. First, it tests for attribute-ordering effects; and, secondly, it examines whether the attribute order within alternatives results in distinct attribute (non)-attendance patterns. To this end, we use a split-sample approach. There were two versions of the questionnaire, each with an identical DCE design. However, in one of the versions, the positions of the first and last non-monetary attributes was reversed across the entire sequence of choice tasks. This allowed us to examine whether the order of the attributes in the choice set affects stated choices and the estimated willingness-to-pay (WTP) values, thus providing a type of convergent validity test of the DCE method. The main novelty of this paper lies in testing whether the order in which the attributes are presented within a choice task leads to any systematic differences in observed ANA behaviour. An additional contribution to the literature consists of estimating and comparing the results of the combined latent class discrete mixtures model and the combined latent class mixed logit model, which allows testing the robustness of the results. Although these advanced models are considered state-of-the art for analysing inferred ANA, their application in the existing literature is limited, possibly because they are rather complex and computationally challenging. Nevertheless, the ability to more fully (though not completely) separate preference heterogeneity and processing heterogeneity means that these models have the potential to yield much more informative results.

The remainder of this article is organized as follows. Section 2 reviews the existing literature on attribute-ordering effects and ANA. Section 3 describes the study design, the hypotheses to be tested, and the econometric approach. Section 4 presents the results, and Section 5 concludes.

2 Previous research

2.1 Attribute-ordering effects

Ordering effects are not a new phenomenon in the stated preference literature. Several CV studies report significant anchoring and sequencing effects (e.g. Herriges and Shogren, 1996; Halvorsen, 1996). The introduction of the DCE method to the stated preference literature has added new dimensions of ordering effects that have come increasingly under scrutiny. Ordering effects in DCEs describe a diverse array of possible phenomena (Day et al., 2012) and may be observed as a consequence of an order in which a choice task, alternative, or attribute appears in a questionnaire. The most commonly studied ordering effects are those related to the position of the choice tasks in a choice sequence (e.g. Holmes and Boyle, 2005; Day et al., 2010; 2012; Carlsson et al., 2012). They show that preferences are affected by the choice task order and that the features of preceding choice tasks influence individuals' choices in subsequent tasks. In addition, significant ordering effects have been reported when posing an open-ended WTP question before or after a DCE (Metcalfe et al., 2012; Brouwer et al., 2017).

Although the empirical evidence suggests that stated preference methods are prone to various ordering effects, insufficient attention has been paid to examining whether the sequence of information presented to respondents within a choice task, such as the order of attributes, influences their stated preferences and welfare estimates. To the best of our knowledge, Glenk (2007) and Boyle and Özdemir (2009) are the only authors who have analysed attribute-ordering effects in the field of environmental economics. Glenk (2007) presented the list of attributes in reverse order to half of the respondents, with the cost attribute appearing either at the top or at the bottom of the choice task. His findings suggest the presence of recency effects, implying that respondents assign a relatively greater weight to the attribute placed at the bottom position. He assumes that the extent of recency effects depends on the relative importance respondents ascribe to the attributes, but did not test this formally. Boyle and Özdemir (2009) found that placing the monetary attribute first instead of last in the list of attributes does not affect preference parameters or welfare estimates in a significant way.

Insights from DCEs applied in the health economics literature are mixed. Using alternatives described by a list of five attributes, Farrar and Ryan (1999) swapped the top-two attributes with the bottom-two in half of their questionnaires. They found no evidence of an attribute-ordering effect. Scott and Vick (1999) show that positioning the most important attribute as the first or as the last one causes an ordering effect and significantly influences preferences for this particular attribute. The respondents in their study expressed stronger preferences for the most important attribute if it was presented last rather than first in the attribute sequence. Kjær et al. (2006) found that placing the price attribute either as the first or the last significantly influences respondents' relative weighting of the price compared with other attributes. Their results indicate that respondents exhibit higher price sensitivity when the price attribute is placed at the end of the policy description.

Analogous to the ordering effect in DCEs is the position effect in best-worst scaling surveys, where respondents are asked to choose the best and the worst item from a list of items. In a study that focused on the consumers' trust in agents concerning information about nanotechnology and its use in food production, Campbell and Erdem (2015) found that the choices made by approximately half of the sample were subject to a position effect, which was more prominent among male respondents. They also showed that the institution positioned at the top of the choice task stands a significantly higher chance of being identified as the most trustworthy.

2.2 Attribute non-attendance

Ignoring attributes in the choice task violates the continuity axiom in the multi-attribute consumer theory underlying DCEs. Without a trade-off between each pair of attributes, no matter how much the level of an ignored attribute is improved, the improvement will fail to compensate for worsening in the level of the other attributes (Spash, 2000; Sælensminde, 2002; Rekola, 2003; Campbell et al., 2008; Scarpa et al., 2009). In such cases estimating the marginal rates of substitution between attributes and WTP values is problematic (Lancsar and Louviere, 2006). Ignoring the fact that some respondents base their choice only on a subset of attributes, and treating them in the same way as respondents who consider all attributes, may lead to erroneous and biased estimates (Scarpa et al., 2009), and, consequently, to misleading policy recommendations.

Two main approaches, based on stated and inferred data, have been developed to identify and model ANA behaviour. Self-reported ANA can be elicited either at the choice task level (choice-task specific stated ANA) or after the entire choice task sequence (serial stated ANA). The reliability of the stated ANA approach has been questioned as respondents' statements are often inconsistent with the results from statistical models (e.g. Carlsson et al., 2010; Hess and Hensher, 2010; Scarpa et al., 2012; Kragt, 2013; Bello and Abdulai, 2016; Chalak et al., 2016; Tarfasa et al., 2017). Choice-task stated ANA seems to be more congruent with inferred ANA than serial stated ANA (Caputo et al., 2018). Another problem with using self-reported ANA is that it can potentially introduce endogeneity bias into the model (Hess and Hensher, 2013). Therefore, most researchers focus on inferring ANA behaviour from suitable analytical models and advancing the ability of these models to describe such behaviour (Campbell et al., 2008; 2011; Scarpa et al., 2009; 2010; Hensher and Rose, 2009; Cameron and DeShazo, 2010; Balcombe et al., 2011; Hole, 2011; Hensher et al., 2012; Hess et al., 2013; Glenk et al., 2015; Thiene et al., 2015). Heterogeneity is usually captured by allowing the coefficients to vary *between* different ANA classes. Some studies suggest that ANA might be confounded with regular taste heterogeneity, where low attribute importance may imply that some respondents do not ignore the attribute, but simply put less weight on it (Carlsson et al., 2010; Hess et al., 2013). Therefore, the models that do not allow the parameters to vary across respondents *within* a class may incorrectly assign low attribute importance to ANA and hence overestimate the ANA shares. To overcome this problem, Hess et al. (2013) propose to use choice models that are able to capture ANA and taste heterogeneity simultaneously by assuming continuous distributions of random parameters within a class. The main limitation of these models is that they are computationally demanding and identification problems can occur. The choice of distribution is, moreover, expected to affect the weight of ANA classes (Hess et al., 2013). In general, the choice of the inferred ANA model and the model specification can influence the results.

The third and most recent approach is visual ANA, relying mainly on eye-tracking technology. Visual ANA is usually measured in terms of the duration and number of eye fixations. An exception is Mattmann et al. (2017), who applied mouse-tracking to measure visual ANA. The advantage of visual ANA is that it allows ANA behaviour to be directly observed. The studies on visual ANA conclude that low visual attention does not necessarily imply that respondents ignore an attribute or attach a low importance to it, and *vice versa* (Balcombe et al., 2015; Lewis et al., 2016; Mattmann et al., 2017; Van Loo et al., 2018). Visual ANA is often found to be inconsistent with stated and inferred ANA measures (Balcombe et al., 2015; Van Loo et al., 2018). Grebitus et al. (2015), Spinks and Mortimer (2016), Grebitus and Roosen (2018), and Selivanova and Krabbe (2018) show that higher choice task complexity increases visual ANA. The findings of Spinks and Mortimer (2016) suggest that ANA differs across alternatives, indicating possible left-right ordering effects. Selivanova and Krabbe (2018) confirm this, as the respondents in their study paid more visual attention to, i.e. fixated their eyes longer and more often on, alternatives presented on the left side. Interestingly, several studies on visual ANA randomised the choice attributes to avoid potential ordering effects (e.g. Mattmann et al., 2017; Van Loo et al., 2018), or varied the number of attributes and hence their relative position in the choice task (Spinks and Mortimer, 2016), which is why they are less informative about the link between visual ANA and attribute order. Most other studies (e.g. Balcombe et al., 2015; Yegoryan et al., 2018) suggest that visual attention based on eye tracking is not distributed equally across attributes. Balcombe et al. (2015) showed that next to the last (monetary) attribute, the first and the last non-monetary attributes have the highest visual ANA. Chavez et al. (2018) also demonstrated that the last attribute (price) is the most attended to. Krucien et al. (2017) found that cost was the most (least) visually attended attribute when it was placed at the top (bottom) of the choice tasks. On the contrary, price is the least attended attribute in Grebitus et al. (2015), who hypothesise that respondents spend a shorter time processing familiar attributes such as price compared with unfamiliar attributes, for which no information is yet stored in their memory and new associations need to be created. The main limitations of visual ANA studies based on eye-tracking are the rather small sample sizes and the often non-representative samples of respondents (e.g. the student population).

A majority of studies investigating ANA suggest that incorporating ANA behaviour into empirical models improves model fit and significantly affects marginal willingness-to-pay (MWTP) estimates. The results concerning the direction of the impact on WTP values are, however, mixed. While most of the studies report significantly lower welfare estimates when ANA is considered in the choice models (e.g. Hensher et al., 2005; Scarpa et al., 2009; Campbell et al., 2011), Hensher and Greene (2010) find the opposite and Carlsson et al. (2010) detect no significant differences. The literature has identified several factors that seem to influence the degree of ANA, including hypothetical bias mitigation strategies like textual information or honesty stimuli (Bello and Abdulai, 2016), respondents' knowledge about or familiarity with the good being valued (Sandorf et al., 2017; Heidenreich et al., 2018), and the level of choice task complexity (Jonker et al., 2018). Jonker et al. (2018) argue that reducing the choice task complexity, e.g. by using the same levels for several attributes across different alternatives and highlighting differences between alternatives with colours, also reduces ANA. The underlying reasons for ANA have not yet been fully understood, although Alemu et al. (2013) made a first step in this direction.

2.3 Attribute attendance and ordering effects

A systematic review of 28 different ANA studies published between 2005 and 2019, conducted in support of the study presented here (see Appendix A1), reveals that in around one-third of these existing studies (9 studies) the non-monetary attribute that is placed last (or second-last, before price) in the list of attributes is consistently less attended to than the non-monetary attribute which appears first. In 6 studies, the reverse result is found (i.e. the last non-monetary attribute received more attention than the first one), while two studies find equal attendance to the first and last positioned non-monetary attributes. Eleven other studies find mixed results depending on, for example, inferred or stated ANA. Although one-third of the reviewed ANA studies here suggest that the last non-monetary attribute typically receives consistently less attention in DCEs irrespective of the modelling approach that was used, the available empirical evidence over the past 1-2 decades seems inconclusive. The review we conducted is furthermore not without difficulty due to limited information provided in the studies about ANA shares for all attributes, the exact order of the attributes and whether or not the order was fixed. Consequently, the results described here have to be interpreted with the necessary care.

3 Study design, hypothesis testing, and econometric modelling approach

3.1 Study and experimental design

The Swiss government recently decided to restore 4,000 km of rivers in the country over the next 80 years (FOEN, 2012). This implies that 50 km of rivers would need to be restored each year. As a result, there is an increasing number of river restoration projects in the country. We focus on two of them here: the restoration of the rivers Thur and Töss, which are both tributaries of the river Rhine, located in the north-eastern part of Switzerland at approximately 15 km distance from each other. The two rivers provide the same ecosystem services. Certain sections of these rivers have already been restored during the last few decades. These restoration measures have increased species richness at both river sites (Paillex et al., 2017). Another positive effect is an increase in recreational opportunities and the attraction of greater numbers of visitors to the restored sites (Woolsey et al., 2007). This study elicits the preferences and the WTP of the local population for further restoration measures at the degraded river sections using a DCE containing the effects of restoration that are expected to be the same as for the already restored river stretches. The data collected in this study informed the cost-benefit analysis of river restoration projects in Switzerland (reference omitted for the purpose of a blind review).

The DCE included three different labelling treatments. Preferences for the restoration of the rivers Thur and Töss were elicited independently using two identical unlabelled DCEs. Each unlabelled DCE thereby focuses on one river only and does not refer to the existence of the other river. In the third labelled DCE, the respondents had to choose directly between the restoration of either the river Thur or the river Töss, where the first alternative was always labelled as the restoration of the Thur and the second alternative as the restoration of the Töss. The respondents were randomly assigned to either one of the two unlabelled DCEs or the labelled DCE version.

The DCE was part of a survey, which collected information on respondents' river use, awareness, knowledge and perception of river restoration projects, and their socioeconomic characteristics. The questionnaire was thoroughly pretested in two rounds of in-person interviews carried out by professional interviewers familiar with the study area and hired from a marketing agency specialised in public surveys. The interviewers were thoroughly instructed by the first two authors. Debriefing sessions, in which the interviewers provided feedback from the field, took place after each pretest round. The pretest results led to adjustments in the levels of two attributes (biodiversity and price). The survey was administered in person in March 2015 using a spatially stratified sampling approach, which targeted randomly selected households living within a 35 km radius of the river sites that would be restored in the future. The survey included a map of the study area, showing the locations of the restored river sections and the river sections that may be restored in the future. The map also helped the respondents to determine how far they live from the two river sites. The DCE was designed in collaboration with natural scientists who evaluated the ecological effects of previous restoration projects. They helped to select the choice attributes and define attribute levels that describe the expected changes of further river restoration measures. The resulting DCE design is presented in Table 1 in (reference omitted for the purpose of a blind review).

The choice attribute ‘length of the river section that would be restored’ reflects the extent of potential future river restoration projects, measured in kilometres. Three attributes capture the effect of river restoration on recreational opportunities: walking along the river, swimming in the river, and barbecuing on the river bank. Their levels are binary, implying that an option to undertake the activity when visiting the river either exists or does not exist. The ‘biodiversity’ attribute measures species richness, which is expected to increase if further restoration measures would take place. Its levels are defined in terms of the current number of plant and animal species found in and around the river compared with their maximum potential number. Low, medium and high biodiversity levels correspond, respectively, to 60, 75 and 100% of the potential number of species. The biodiversity level in the status quo situation is low for both rivers. The price attribute is defined as an increase in the annual cantonal taxes per person. This was considered the most credible payment vehicle because most taxes in Switzerland are paid once per year and river restoration projects fall under the jurisdiction of the cantonal authorities.

To test for the effect of attribute ordering, two versions of the labelled and unlabelled DCEs were created and randomly allocated to respondents in two equally-sized samples. The first sample received the DCE version with the attribute order as shown in Figure 1A, while the second sample received the version where the order of the 'length' and the 'biodiversity' attributes was reversed, as shown in Figure 1B. Hence, in the latter sample, biodiversity appeared as the first attribute at the top of the choice alternatives and the length of the river section to be restored at the bottom of the alternatives, just above the cantonal tax. Note that we used exactly the same pictogram for these two choice attributes to avoid any potential variation in attribute attendance that might occur due to their visual representation. The order of the other attributes was exactly the same in both versions.

[INSERT FIGURES 1A AND 1B HERE]

A D-efficient experimental design was generated in the software Ngene (ChoiceMetrics, 2014), using prior estimates of the coefficient values derived from the survey's pretest. This design minimizes the D-error, ensuring more reliable parameter estimates given the number of choice observations (e.g. Rose et al., 2008). The resulting DCE design consisted of 36 different choice tasks. They were blocked into six choice sets comprising six choice tasks each, which were randomly distributed to the respondents. Each respondent hence faced six choice tasks. The choice tasks comprised three alternatives: two hypothetical alternatives describing the improvements which would result from the implementation of further river restoration measures; and the opt-out alternative representing the status quo situation. The respondents had to choose their most preferred alternative in each choice task. An example of a choice task is presented in Figures 1A and 1B. After the DCE, respondents were asked which attribute was most important for guiding their choices. Protest responses were identified on the basis of a follow-up question, in which those respondents who chose the opt-out alternative in all six choice tasks were asked for their underlying reasons. Around 4% of all the choices across the two samples were classified as protest responses and were excluded from the choice data analysis, which is common practice in the stated preference literature (Brouwer and Martin-Ortega, 2012).

3.2 Hypotheses testing

The main objectives of this study give rise to two hypotheses. The first hypothesis tests whether an attribute-ordering effect has occurred, i.e. if the placement of the two non-monetary attributes (length and biodiversity) as the first or second-last in the list of attributes describing the alternatives affects the respondents' choices, and hence MWTP estimates. To test this hypothesis, we examine the equality of marginal utilities associated with each choice attribute between the two samples, ~~who~~ which received different attribute ordering treatments *g*:

$$H_0^1:\ E\left(MWTP_{g=1}^{\text{Model 1}}\right) = E\left(MWTP_{g=2}^{\text{Model 1}}\right), \qquad (1)$$

where treatment $g=1$ denotes the sample of respondents who answered the questionnaire version in which the length attribute appeared first, and treatment $g=2$ the sample where biodiversity was positioned as the first attribute. We use the welfare estimates derived from the choice models that assume full attribute attendance and apply the Poe et al. (2005) test procedure, using the gizmo library in R (Sandorf, 2019). A failure to reject this hypothesis means that there is no attribute-ordering effect and hence no procedural bias.

The second hypothesis tests whether the order in which the choice attributes appear in an alternative treatment results in distinct patterns of ANA behaviour. This second hypothesis is tested based on the outcomes of the choice models that account for ANA behaviour. This is done in two ways. First, we verify the assumption that the probability of ignoring each choice attribute, ignoring all the attributes, and considering all of them is equal among the two samples ~~who~~ which were exposed to different attribute-ordering treatments:

$$H_0^{2a}:\ \pi_{\mathrm{g}=1}=\pi_{\mathrm{g}=2}. \tag{2}$$

Secondly, we test whether the equality of MWTP estimates holds when using choice models that take ANA into account:

$$H_0^{2b}:\ E\left(MWTP_{g=1}^{\text{Model 2,3}}\right)=E\left(MWTP_{g=2}^{\text{Model 2,3}}\right). \tag{3}$$

A failure to reject these hypotheses implies that positioning of the non-monetary attribute at the top or the bottom in a choice task does not result in distinct ANA behaviour. The hypotheses tests in the case of latent class models are based on the expected values of MWTP (i.e., $E(MWTP)$), which involve weighting the class-specific MWTP with the (unconditional) class probabilities. We admit that this negates the fact that we have identified heterogeneity in MWTP across the sample, but it does enable a more straightforward comparison and testing of hypotheses. MWTP estimates for the same latent class are also compared between the ordering treatments in the Appendix. We also point out that MWTP estimates obtained from ANA models apply only to the subset of respondents who actually considered the price attribute and the relevant non-monetary attribute, since only in those cases there is substitutability between the attributes and, therefore, a computable marginal rate of substitution.

3.3 Econometric modelling approach

Ignoring the fact that some respondents base their choice only on a subset of attributes and treating them in the same way as respondents who consider all attributes will lead to erroneous and biased estimates (Scarpa et al., 2009). In this paper, we are interested in relaxing the assumption that all respondents consider all attributes, and in identifying the share of respondents who ignore attribute(s). Such (unobserved) attribute processing heterogeneity can be accommodated by applying the combined latent class discrete mixtures model with finite (discrete) distributions or the combined latent class mixed logit model proposed in Hess et al. (2013). We estimate both models, which allows us to compare the results and test their robustness.

We acknowledge the similarity between the discrete mixtures model and the latent class logit model, which also assumes finite representations of heterogeneity. In fact, both models are formally equivalent, the main difference being that in discrete mixtures models the focus is usually on segmenting on a per parameter basis and not on the basis of the full set of parameters, which is typically the case in latent class models. Indeed, both specifications can be estimated using a number of equality constraints. We favour the behavioural appeal of retrieving probabilistic estimates for each parameter directly, afforded by the discrete mixtures approach, and the fact that estimates of ANA can be retrieved using fewer parameters.

The number of possible ANA classes with *K* attributes is $Q = \prod_{k=1}^{K} M_k$, where, in this case, *M* = 2 to allow $\beta_k^1 \neq 0$ and $\beta_k^0 = 0$ to recognise ANA. Each ANA class, $q = \{1, 2, 3, \ldots, Q\}$, implies a different combination of attribute marginal utilities for each of the *K* attributes. The ANA classes represent probabilities of membership in 2^K different classes describing all possible combinations of ANA behaviour. Our experimental design with six attributes (see next section) leads to 64 different ANA classes. Those respondents whose choice strategies match that of the specified pattern of ANA have a higher predicted (unconditional) probability of belonging to that class. The unconditional probability that an attribute has been ignored is calculated as the sum of unconditional probabilities of membership across various classes that describe ANA to that attribute. The unconditional probability of observing combination *q*, denoted using ϕ_q, is the product of the probabilities of observing the respective processing rules for each attribute, π_k^m, that describe combination *q*. For example, the probability of observing β_1^1, β_2^0 and β_3^0 is given by $\phi_q = \pi_1^1 \times \pi_2^0 \times \pi_3^0$. In this paper, we refine this specification. In particular, we derive the unconditional probability associated with respondents: (i) considering all attributes; (ii) ignoring all attributes; and, (iii) ignoring a subset of attributes (but not all), where ϕ_q for $q = \{2,3,...,Q-1\}$ are normalised to assure that all probabilities sum to one.

The probability of a sequence of choices $y_n = \left[i_{n1}, i_{n2}, ..., i_{nT_n} \right]$ made by respondent *n* over the T_n choice occasions can then be written as:

$$\Pr(y_n | \beta_g, C, \lambda, \phi, Q, x_n) = \sum_{q=1}^{Q} \phi_{q_g} \prod_{t=1}^{T_n} \left(\frac{\exp\left[\lambda_h \left(\beta_g^q x_{nit} + C_{i_g} \right) \right]}{\sum_{j=1}^{J} \exp\left[\lambda_h \left(\beta_g^q x_{njt} + C_{j_g} \right) \right]} \right), \tag{4}$$

where β_g is an estimated parameter of marginal utility for attribute *x*; *C* is an alternative specific constant (ASC); *g* is the attribute-ordering treatment; *h* is the labelling treatment (see Section 3.1); and λ are the scale factors, defined relative to the baseline treatment *h,* where both rivers Thur and Töss are included in the choice set. The scale parameters control for any potential differences in error variance that may exist between the three labelling treatments (two unlabelled and a labelled DCE) and are estimated relative to the baseline treatment, i.e. the labelled DCE. Hess and Train (2017) argue that scale parameters may capture not only scale heterogeneity, but can be confounded with other differences in the data, including preferences. To avoid perfect multicollinearity, we arbitrarily set one of the ASCs to zero. It is also important to consider preference heterogeneity due to potential confounding. In a latent class framework, this can be accomplished by further segmenting on the basis of preferences:

$$\Pr(y_n|\beta_g, C, \lambda, \tau, S, \phi, Q, x_n) = \sum_{s=1}^{S} \tau_{s_g} \sum_{q=1}^{Q} \phi_{q_g} \prod_{t=1}^{T_n} \left(\frac{\exp\left[\lambda_h\left(\beta_g^{qs} x_{nit} + C_{i_g}\right)\right]}{\sum_{j=1}^{J} \exp\left[\lambda_h\left(\beta_g^{qs} x_{njt} + C_{j_g}\right)\right]} \right), \quad (5)$$

where τ denotes the unconditional probabilities associated with latent classes S, which capture preference heterogeneity.

In the combined latent class mixed logit model, each element in β follows a distribution θ, which allows additional sources of unobserved heterogeneity to be captured, most notably unobserved preference heterogeneity:

$$\Pr(y_n|\beta_g, \theta, C, \lambda, \phi, Q, x_n) = \sum_{q=1}^{Q} \phi_{q_g} \int \prod_{t=1}^{T_n} \left(\frac{\exp\left[\lambda_h\left(\beta_g^{qs} x_{nit} + C_{i_g}\right)\right]}{\sum_{j=1}^{J} \exp\left[\lambda_h\left(\beta_g^{qs} x_{njt} + C_{j_g}\right)\right]} \right) f\left(\beta|\theta\right) \mathrm{d}\beta \, . \quad (6)$$

We estimate three types of choice models. Model 1 represents the ‘standard’ latent class (LC-FAA) model, which accommodates only preference heterogeneity (i.e., it relies on the standard assumption of full attribute attendance (FAA) and, thus, ignores the existence of ANA). The model has two latent classes, which capture taste heterogeneity among individual respondents. Model 2 (LC-ANA) is the combined latent class discrete mixtures model described in Eq. 5 that accounts for both taste heterogeneity and ANA behaviour. Here too, the two latent classes capture regular taste heterogeneity. Since the respondents in each latent class are assumed to display different patterns of ANA behaviour, we furthermore distinguish between 64 possible ANA classes within each latent class, i.e. 128 ANA classes in total. Since allowing for two preference classes only is a limiting and potentially unrealistic assumption, in Model 3 we use continuous distributions to better accommodate the heterogeneity in preferences and processing strategies and reduce confounding concerns. Model 3 (LC-MXL) is the combined latent class mixed logit model, defined in Hess et al. (2013), which accommodates both taste heterogeneity and ANA behaviour in a more flexible way than Model 2. While we recognise that our models do not entirely overcome the confounding between preferences and processing strategies, accommodating for both simultaneously in this manner improves our understanding of their separate influence on choices.

To assess the impact of taking ANA into account in the empirical analysis, we compare Models 1 with Models 2 and 3 in terms of model performance and MWTP estimates. To analyse the potential effect of attribute ordering on the ANA behaviour, we furthermore use Models 2 and 3 to compare the shares of respondents who ignore the attributes and the MWTP estimates between the two attribute-ordering treatment groups.

All models are coded and estimated using the maxlik library in R (see Henningsen and Toomet (2011) and R Core Team (2014) for further details) using maximum likelihood estimation. We are mindful of their vulnerability to local maxima. To reduce the possibility of reaching a local rather than a global maximum, we started the estimation iterations from a variety of random starting points.

4 Results

4.1 Standard choice models that assume full attribute attendance

Table 1 reports estimation results obtained from the standard LC models that assume full attribute attendance. The first model in Table 1 (Model 1a) relates to the treatment where the *length* and *biodiversity* attributes are listed as, respectively, the first and last non-monetary attributes in the choice set (treatment $g = 1$), whereas in the second model (Model 1b) the placement of these two attributes is reversed (treatment $g = 2$). The order in which the variables are presented in Table 1 corresponds to the order in which they were presented to respondents in treatment $g = 1$ and is kept the same for both ordering treatments, for ease of comparison.

[INSERT TABLE 1 HERE]

The main findings derived from the standard LC models are similar for both ordering treatments. The most notable difference in preferences between the two latent classes is that the respondents in class 1 prefer river restoration over the status quo, as indicated by the significant positive ASCs, while the respondents in class 2 prefer the status-quo option, as indicated by the negative ASCs that are significant in three out of four cases. Differences in the magnitudes of the estimated parameters for the price attribute between the two latent classes indicate that respondents in the second class are more price-sensitive and hence willing to pay substantially less for river restoration than respondents belonging to the first class. The option to walk along the river is the most highly valued feature of river restoration among respondents in both latent classes. This finding is supported by the stated attribute importance, where respondents reported that walking was the most important attribute for their choices (see Appendix A2).

An interesting outcome is that the estimated coefficient for the length attribute is insignificant in class 1 when that attribute is placed at the top of the choice task ($g = 1$), but not when it is placed at the bottom ($g=2$). The same pattern is observed in class 1 for improving biodiversity to a high level, where the coefficient associated with this attribute level is insignificant in the treatment where the biodiversity attribute appears first ($g=2$), and significant when it appears at the bottom ($g=1$). The estimated coefficients for the medium biodiversity level turn out to be insignificant in the two latent classes in both ordering treatments. The probability of class membership is approximately 50% in both treatment groups. Finally, we do not find any evidence to suggest that the scale parameters are significantly different across the three labelling treatments for $g=1$. For $g=2$ a significant difference is detected between the unlabelled DCE for the river Thur and the labelled DCE, but only at the 10% level. This means that either differences in variance or among the utility coefficients exist between these two labelling treatments.

4.2 Testing attribute-ordering effects

The comparison of individual parameter estimates between the various models is not straightforward, since they can be subject to different scaling of the parameter estimates. What is potentially of greater interest to policy analysts are the MWTP estimates, since the scale effect is neutralised when dividing the marginal utilities by the marginal price coefficient. Of particular relevance in this paper is the difference between the MWTP estimates for the length and biodiversity attributes, given that the order of these two attributes was switched. In Table 2 we report the MWTP estimates derived from Models 1 and 2 for both ordering treatments. They represent the weighted MWTP estimates between two latent classes, where class membership probabilities serve as weights. The last two columns in Table 2 present the results of the Poe et al. (2005) test. The individual MWTP estimates for each latent class and ordering treatment are presented in Appendix A3. The Poe et al. (2005) test results based on these estimates are shown in Appendix A4. The confidence intervals around the MWTP estimates are estimated using the Krinsky and Robb (1986) procedure based on 10,000 replications.

[INSERT TABLE 2 HERE]

In order to verify our first hypothesis concerning the presence of an attribute-ordering effect, we compare the MWTP estimates of the two attribute-ordering treatments based on Models 1a and 1b that assume full attribute attendance, and apply the Poe et al. (2005) test procedure. On average, those respondents who received the DCE version in which the length attribute appeared first attach *ceteris paribus* a higher value to all choice attributes. The only exception is a negligibly lower MWTP estimate for the length attribute. The Poe et al. (2005) test results show, however, that the differences in the MWTP estimates between the two treatment groups obtained from the standard LC model are not statistically significant. Therefore, our first hypothesis cannot be rejected. These findings suggest that MWTP values are not sensitive to the positioning of the non-monetary attributes in the choice task when full attribute attendance is assumed. These results corroborate previous findings by Boyle and Özdemir (2009) and Farrar and Ryan (1999) about the absence of attribute-ordering effects.

4.3 Choice models that account for attribute non-attendance

While the estimates retrieved under the standard LC models give important insight into the sample's preferences for different river restoration options, they are based on the premise that respondents considered all the attributes in their decision making. In order to test the hypothesis of the presence of attribute-processing strategies, we turn our attention to the choice models that allow for ANA behaviour. The estimated results of the combined latent class discrete mixtures models are presented in Table 3 and those of the combined latent class mixed logit models in Table 4. Despite the fact that estimating additional support points and the probabilities associated with these support points comes at a high parametric cost, Models 2 and 3 lead to a much better model fit than Model 1, as evidenced by the AIC, BIC and R-squared, confirming previous findings in the literature. Moreover, Models 3a and 3b outperform Models 2a and 2b in terms of BIC and R-squared, i.e. the measures of fit that take into consideration their varying number of parameters.

[INSERT TABLES 3 AND 4 HERE]

Overall, Models 2 and 3 display similar results for the two attribute-ordering treatments. More significantly than in the choice models that assume full attribute attendance, we find that, as expected, the respondents prefer policy outcomes that restore longer river stretches, provide opportunities for walking, swimming and barbecuing, and increase the biodiversity in and around the river. All else being constant, respondents prefer cheaper (relative to more expensive) policy options, which is also an expected finding. Models 2 and 3 show some inconsistencies in the results concerning the scale parameters. In Model 2 significant difference in scale parameters is detected at the 5% significance level under the ordering treatment $g=1$ between the unlabelled DCE for the river Töss and the labelled DCE, indicating that there is a significant scale or preference heterogeneity between the two treatments. In Model 3 the scale parameter is only weakly significant (at the 10% level) under the ordering treatment $g=2$ for the river Thur. Interestingly, the standard deviations of the random parameters in Model 3 indicate that preferences for the price attribute and the non-monetary attribute that appeared at the bottom position (i.e. biodiversity in $g=1$ and length in $g=2$) are heterogeneous and differ significantly across individual respondents. For the remaining attributes no taste heterogeneity has been detected, except for the walking attribute in the ordering treatment $g=2$.

However, when interpreting the parameters estimated in Models 2 and 3, it is important to recognise the unconditional probabilities of ANA. Both models suggest that over 70% of the respondents ignored the swimming and barbecuing attributes, more than 50% did not consider the biodiversity attribute, and over 40% ignored the walking attribute. For the remaining attributes, Model 3 indicates lower ANA shares than Model 2. Model 3 suggests that 26% to 34% respondents ignored the price attribute and that 0% ignored all attributes. The corresponding shares are somewhat higher in Model 2. Model 3 also indicates a lower share of respondents who ignored the length attribute (0%) and a higher share of those who considered all attributes (23%) than Model 2, albeit only for the ordering treatment $g=2$. The results of Model 3 for the ordering treatment $g=1$ are similar to the results of Model 2, which show that at least 48% of respondents ignored the length attribute and that 0.7% to 6.3% of them considered all attributes. Possible explanations for the large overall shares of respondents who ignored the attributes are discussed later on.

It is worth mentioning that the inferred ANA shares match the respondents' stated responses about attribute importance rather well. The walking and price attributes have the lowest inferred ANA shares, implying that these are the least ignored attributes in the choice process. Correspondingly, the largest portion of respondents in both samples indicated that walking and price were the most important attributes for their choices (30% and 19%, respectively). Therefore, the choice attributes that are considered important by respondents are also ignored to a lesser extent, and *vice versa*. As this finding might indicate confoundedness between ANA and regular taste heterogeneity, in particular low attribute importance (Carlsson et al., 2010; Hess et al., 2013), it justifies the use of choice models that consider regular taste heterogeneity in addition to ANA. The most substantial deviation between the inferred ANA shares and the stated attribute importance is found for the biodiversity attribute. Although on average 18% of the respondents selected biodiversity as the most important attribute, the corresponding ANA shares are relatively high. This suggests that the respondents either overstated its importance or the models over-estimated the inferred ANA shares. It is also conceivable that other simplifying choice strategies prevailed when processing information related to this attribute, such as eliminating all the alternatives with low and medium biodiversity attribute levels. The stated attribute importance reveals a few differences between the two attribute-ordering treatments. A higher share of respondents who received the DCE version in which length was positioned as the first attribute stated that walking and biodiversity were the most important attributes in making their choices compared to the other treatment. In the treatment where the biodiversity attribute was placed first, more respondents stated that length and price attributes were the most important ones. Therefore, positioning the non-monetary attribute at the top of the choice tasks is associated with a lower stated attribute importance.

Among those respondents who are predicted to have considered the non-monetary attributes, we find some differences in the implicit ranks compared with models that do not consider ANA. In particular, Model 2 (the second latent class in both ordering treatments) and Model 3a (ordering treatment $g=1$) indicate that improving biodiversity in and around the river to a high level provides the highest marginal utility to respondents. According to Model 3b (ordering treatment $g=2$), swimming has the highest marginal utility. The significant negative coefficient estimates for the ASCs (in Model 2 for both rivers in the first latent class and in Model 3 for the river Thur) suggest that, without consideration of the choice attributes, the respondents prefer the status quo to the policy options that imply river restoration. Respondents seem to be indifferent between river restoration and the status quo in the second class in Model 2 and for the river Töss in Model 3 in both treatment groups.

The major difference between the two attribute-ordering treatments based on Model 2 is observed in the preferences of respondents who belong to the two different latent classes. Specifically, respondents who belong to the second class in the treatments $g=1$ and $g=2$ are willing to pay significantly more for all choice attributes than respondents in class 1 in the ordering treatment where the length attribute appears first ($g=1$) (see columns 4 and 6 in Appendix A4). Respondents in class 2 in the treatment $g=1$ are also willing to pay significantly more for all choice attributes except swimming and barbecuing than respondents in class 1 in the treatment $g=2$. However, in the ordering treatment where the biodiversity attribute appears first, respondents in the second class in Model 2b do not show sensitivity to the price attribute and, as a result, their MWTP values are statistically indistinguishable from zero (see Table 3 and Appendix A3). Despite the fact that a considerable share also ignored price in the second class in Model 2a, the price coefficient in this model is significantly negative. Moreover, there are differences in the probabilities of membership in the two latent classes between the two treatment groups. The probability of membership in class 1 is below 50% in the ordering treatment where the length attribute is placed first, and over 60% where biodiversity comes first. A significant difference in scale parameters is furthermore detected at the 5% significance level under the ordering treatment $g=1$ between the unlabelled DCE for the river Töss and the labelled DCE, indicating that there is a significant scale or preference heterogeneity between the two treatments.

The results of Model 3 on preference parameters for the two treatment groups are similar, except that the highest utility is not provided by the same attribute. The most notable differences between the two attribute-ordering treatments in Model 3 are observed in preference heterogeneity and inferred ANA shares. Preference heterogeneity differs significantly across respondents in both treatment groups for the monetary attribute and the non-monetary attribute that appeared last. Therefore, it is possible that the attribute order is affecting preference heterogeneity, but one cannot formally test this. Differences in the shares of respondents who ignored the attributes between the two ordering treatments are discussed in the next section.

4.4 Testing the effect of attribute order on ANA behaviour

4.4.1 The effect of attribute order on the probability to ignore an attribute

Inspection of the marginal utility parameters π retrieved under the models that take ANA into account shows that a sizeable proportion of the respondents in both samples ignored the attributes. Table 5 reports the Poe et al. (2005) test results for the equality of probabilities of ignoring each attribute, ignoring all the attributes, and considering all the attributes (i.e. the aggregated class membership probabilities) between the two ordering treatments, as well as between the two latent classes within the same ordering treatment for Model 2.

In Model 2 differences in ANA shares are more prominent between the first and the second latent class that capture taste heterogeneity than between the same latent classes across the two ordering treatments. The results indicate that the share of respondents who did not consider one or several choice attributes is generally higher in the second latent class, independently of the attribute-ordering treatment. Only non-attendance to the swimming attribute in $g = 2$ and to all attributes in both treatment groups is higher among respondents in the first latent class. Based on Models 2a and 2b, the second hypothesis specified in Eq. 3 is tested for four possible combinations of between-class comparisons for each choice attribute (columns 4 to 7 in Table 5). The hypothesis of equal ANA shares between the two treatment groups is rejected once for the length attribute, twice for the swimming attribute, three times for price and biodiversity, and twice for the class describing non-attendance to all choice attributes. More specifically, the outcome of the Poe et al. (2005) test shows that a significantly higher share of respondents ignored the length (biodiversity) attribute in the second latent class in the treatment where this attribute was placed at the top of the choice tasks relative to the first (both) latent class(es) in the other ordering treatment. Although these findings suggest that an attribute might receive less attention if it is placed first in the sequence of attributes, we also find evidence pointing in the opposite direction. The share of respondents who did not consider the biodiversity attribute is significantly higher in the second latent class in the treatment group where that attribute appears at the bottom compared with the first latent class in the treatment group where that attribute is placed at the top. The equality in the ANA shares is also rejected in approximately half of the cases between the two latent classes within the same ordering treatment (columns 2 and 3 in Table 5). Therefore, the observed differences in ANA behaviour in Model 2 between the two attribute-ordering treatments seem to be driven more by taste heterogeneity among respondents in the two latent classes than by the position of the non-monetary attribute in the choice task.

[INSERT TABLE 5 HERE]

According to Model 3, the share of respondents who ignored the attributes is in general lower in the ordering treatment where the biodiversity attribute is placed at the top position, except for the attributes that capture recreational activities. Differences in ANA shares between the two treatment groups are, however, significant only for the non-monetary attributes whose order was reversed. Therefore, based on Model 3, the second hypothesis on the equality of ANA shares is rejected for the length and biodiversity attributes (column 8 in Table 5). Although this is an interesting finding, we cannot conclude whether it is driven mainly by the attribute order or about the direction of its effect because the share of respondents who ignored both attributes is significantly lower in the same attribute-ordering treatment, where the biodiversity attribute is placed at the top of the choice tasks.

4.4.2 The effect of attribute order on MWTP estimates based on the ANA models

Turning to the MWTP estimates derived from Models 2 and 3 that take both ANA and taste heterogeneity into account, it has to be noted that these estimates apply only to the subset of respondents who actually considered the cost attribute and the relevant environmental attribute level. This is because for respondents who ignored the cost attribute it is not possible to derive the MWTP values. The welfare estimates obtained from the models corrected for ANA are sensitive to the assumptions made about those respondents who did not make full trade-offs between the cost and the non-monetary attributes. We adopt the most common approach in the ANA literature and assume that such respondents have a zero WTP.

The MWTP estimates derived from Model 3 are presented in Table 6. Since the welfare estimates in this model have an underlying distributions, apart from the mean MWTP we also report the MWTP estimates for the 25th, 50th and 75th percentile. The results of the Poe et al. (2005) test on the equality of the MWTP estimates between the two treatment groups based on Model 3 are shown in the last four columns in Table 6. The MWTP estimates and the Poe et al. (2005) test results for Models 1 and 2 are reported in Table 2. These represent weighted MWTP values between the two latent classes, where the class membership probabilities are used as weights. The unweighted MWTP estimates per latent class are presented in Appendix A3 and the corresponding test results in Appendix A4. Overall, the median MWTP values derived from Model 3 fall within the range of the

MWTP estimates obtained from Models 1 and 2 and seem to be of a more reasonable magnitude than the mean MWTP estimates.

[INSERT TABLE 6 HERE]

The Poe et al. (2005) test results for Model 2 show that only the MWTP estimates for the walking attribute and for the medium biodiversity level are significantly different between the two ordering treatments at the 5 and 10 per cent level, respectively. This means that, based on Model 2, the second hypothesis specified in Eq. 4 is rejected in these two cases. The MWTP value for the walking attribute is significantly higher and for the medium biodiversity level significantly lower in the treatment where the length attribute appeared at the top position. The differences for the high biodiversity level and river length to be restored are only significant at the 11 and 14 per cent level, respectively. The high biodiversity level is also valued higher when positioned at the top instead of at the bottom. This finding suggests that the positioning of a non-monetary attribute at the top of the choice task results in higher MWTP values than placing it at the bottom. This applies to medium and high biodiversity levels and to the length attribute, although the differences between the latter two are not significant at the 10 per cent level. An important disclaimer is, however, that the MWTP values derived from Model 2b have relatively large confidence intervals and are hence not very accurate, which means that caution is required in the interpretation of the results.

Based on Model 3 the differences in mean MWTP estimates between the two ordering treatments are insignificant, implying that the second hypothesis on the equality of the mean MWTP estimates cannot be rejected for any of the choice attributes. However, significant differences in the MWTP estimates for the length and biodiversity attributes, including both the medium and high biodiversity levels, are found between the two attribute-ordering treatments for the 25th and 50th percentiles. For the 75th percentile only the difference in the MWTP estimates for the high biodiversity level is significant at the 10% significance level. In these cases, the second hypothesis on the equality of MWTP estimates is rejected. The MWTP estimates are significantly lower in the ordering treatment where the biodiversity attribute appears at the top of the choice tasks. This is inconsistent with the outcome of Model 2, but consistent with the findings of Model 3 on ANA shares, which are also significantly lower in this treatment for the same attributes. Therefore, in Model 3 lower ANA shares seem to be associated with lower MWTP values. The lower MWTP estimates might result from a slightly higher share of opt-out choices in the treatment $g=2$ (38%) compared to the treatment $g=1$ (35%). However, this does not explain the lower ANA shares. Apart from the attribute order, a possible explanation for the lower ANA shares in the treatment where the biodiversity attribute appears first is a higher price sensitivity of respondents in this treatment group, which is also supported by the stated attribute importance. To conclude, although the mean MWTP estimates for all choice attributes are identical between the two ordering treatments, the MWTP distributions for the two non-monetary attributes whose position was switched are not symmetrical and display significant differences between the two subsamples. The results of Model 3 hence imply that positioning of the non-monetary attribute in a choice task affects the distribution of welfare estimates over the sample in a significant way.

The findings of Models 2 and 3 suggest that when ANA behaviour is considered, welfare estimates become sensitive to the positioning of the non-monetary attributes in the choice task. Therefore, while the order in which the choice attributes are presented to respondents does not seem to be of concern under the standard assumption of full attribute attendance, it might become an issue once ANA behaviour is taken into account.

5 Discussion and Conclusions

Despite the rich literature on ordering effects in stated preference research, very few studies have so far investigated attribute-ordering effects in DCEs, and there is only one in the field of environmental economics. This paper contributes to this limited strand of literature in two distinct ways. First of all, we test whether the positioning of a non-monetary choice attribute as the first or the penultimate one in the list of attributes presented to respondents affects welfare estimates. Secondly, we examine whether the attribute order influences respondents' propensity to attend to or ignore an attribute. The link between attribute order and ANA behaviour has not been explored before and hence constitutes the principal novelty of this paper. Moreover, we estimate the two most sophisticated choice models for analysing ANA, compare the results and test their robustness. The results of this study are in line with the existing empirical evidence, which demonstrates that choice models which account for ANA display considerable improvements in model fit compared with the standard models which assume full attribute attendance. They also confirm previous findings, which show that a considerable proportion of respondents ignores one or several choice attributes.

We do not find evidence, however, that an attribute-ordering effect exists, neither in the marginal utilities associated with the choice attributes whose position was switched in the DCE nor in the marginal utilities of the other choice attributes. Our results therefore reject the notion of procedural bias and show convergent validity of stated preferences derived from the DCE. This conforms to the findings reported in Boyle and Özdemir (2009) and Farrar and Ryan (1999).

The results of this study show that the order in which the attributes are presented to respondents in the choice set can affect attribute (non-)attendance. First, we detect significant differences in the shares of respondents who neglected the non-monetary choice attributes whose order was reversed between the two ordering treatment groups. These differences are found in both ANA models, but are more pronounced in the combined latent class mixed logit model (Model 3). However, we cannot conclude whether placing the non-monetary attribute first in the list of attributes decreases or increases ANA, because the results point in different directions.

Secondly, significant differences between the two attribute-ordering treatments are found in the MWTP estimates for two choice attributes that were presented to the respondents in a reversed order when using the choice models that account for ANA. Here too, the differences are more prominent in Model 3 and the results concerning the direction in which the attribute order influences the MWTP estimates are ambiguous. Therefore, while the attribute order does not seem to affect the welfare estimates derived under the assumption of full attribute attendance, significant differences emerge once ANA behaviour is acknowledged in the choice modelling framework. This suggests that attribute order can impact ANA behaviour and is thus a relevant issue to consider when analysing strategies that respondents use when processing attribute information in DCEs, such as ANA, and other attribute information-processing strategies.

Comparing the outcomes of the two choice models that take ANA into account lead to the~~a~~ conclusion that the results are not very robust. The combined latent class discrete mixtures model (Model 2) tends to overestimate the inferred ANA probabilities compared to Model 3. Model 2 suggests that placing the non-monetary attribute first instead of last in most cases inflates both the shares of respondents who ignored that attribute and the MWTP estimates. On the other hand, Model 3 shows that ANA shares and the MWTP estimates for the length and biodiversity attributes are significantly higher in the ordering treatment where the biodiversity attribute is placed at the top position. Moreover, Model 2 indicates further significant differences between the two treatment groups for other attributes, which are not detected in Model 3. In fact, significant differences in Model 2 are more common between the latent classes than between the ordering treatments, suggesting that regular taste heterogeneity between respondents might be driving differences in ANA behaviour more so than the order of attributes. The outcomes of Model 3 are more straightforward because they clearly indicate that the only differences are those between the non-monetary attributes that were presented in a reversed order. However, the direction in which the attribute order drives ANA shares and MWTP estimates is less clear. Both models imply that lower ANA shares lead to lower welfare estimates. This is, obviously, context specific and may be an artefact of the relative difference in magnitudes between the respective marginal utilities depending on how preference and attribute processing heterogeneity were, and were not, accommodated. Specifically, it could be a signal that respondents who are predicted to have traded-off price against the non-monetary attributes are relatively price sensitive.

It should be noted that our study relies on the inferred ANA models, which ~~can~~ ~~be~~are based on some strong assumptions. As discussed above, the results may be sensitive to the type of inferred ANA model applied, and to the model specification (e.g. the number of classes included). Moreover, as the number of attributes grows, these models will often have problems with inferior local optima and classes collapsing to the same value (Hess et al., 2013). Constraining the number of latent classes to two in Model 2 represents an important limitation of our study as it reduces the flexibility of the model to capture preference heterogeneity and attribute processing strategies. This could be the reason why this model seems to overestimate the inferred ANA shares. The findings of this model need to be taken with caution since increasing the number of latent classes might affect the results. Adding an additional layer of taste heterogeneity through random parameters in Model 3 ensures greater flexibility of the model, leads to a model improvement and results in lower inferred ANA shares, in particular for the treatment in which the biodiversity attribute is placed at the top of the choice tasks. For these reasons, we are also more confident about its findings.

This does not mean, however, that ANA shares inferred from Model 3 represent the ‘real’ non-attendance behaviour. Rather, they describe the real ANA behaviour slightly better than Model 2. The share of respondents who ignored certain attributes (e.g. swimming and barbecuing) remains very high in Model 3. We believe that this can be partly explained by respondents’ true preferences, i.e. that they indeed attach a relatively low importance to these attributes, which is confirmed by their stated attribute importance. Another possible reason for the relatively high ANA shares found in this study could be the relatively high overall proportion of opt-out choices (37%), which do not require making trade-offs between each pair of attributes. The high shares of inferred ANA and opt-out choices can have several reasons. First, they might be related to the fact that nearby sections of both rivers have already been restored and are considered substitutes for our study sites, in particular for recreational activities. This could diminish the importance of the walking, swimming, barbecuing, but also the biodiversity (e.g. bird watching) attributes in our study. Beyond local substitute sites, Switzerland is generally a country rich in freshwater resources like rivers and lakes, which means that the overall number of potential substitutes is high. Secondly, these are local restoration projects and the sites possibly do not possess sufficiently unique features, such as the presence and conservation of charismatic or endangered species, to draw a lot of attention and importance. This may result in respondents not perceiving their restoration as essential. Furthermore, it has to be acknowledged that the combined latent class mixed logit models reduce, but do not entirely eliminate the confoundedness between preference heterogeneity and attribute processing strategies. As a result, Model 3 might still overestimate the shares of respondents who ignored the attributes. Finally, the models applied in this study do not consider other simplifying choice strategies, which respondents might have used when processing attribute information and which may drive the inferred ANA shares upwards.

There are several implications of our results for DCE design and evaluation. First of all, analysts should be aware of the potential ordering effects in DCEs introduced by the order in which attributes, alternatives or choice tasks are presented in a survey. To avoid potential attribute-ordering effects and their impact on attribute (non)-attendance and welfare estimates, we recommend randomising the attribute order across respondents and choice tasks whenever possible. The second-best solution is ~~R~~randomization across respondents only (i.e. without randomization across choice tasks), which can already ~~in case of a large enough sample~~ ensure sufficient variation to average out any ordering effect at the sample level. If randomisation of choice attributes across respondents is not possible (e.g. due to practical challenges when conducting in-person interviews instead of a web-based survey), another option would be to consider the use of split-samples focusing on two or more different attribute orders and test for ordering effects *ex-post* as is the case in this study, and, in case they turn out to be significant, control for them in the data analysis.

Based on existing experimental research on ordering effects in DCEs, randomising the positions of choice alternatives is worth considering too. Randomising choice tasks is something we always recommend to avoid potential path-dependency issues. Another key aspect to take into account when designing DCEs ~~is~~ are the visual properties of the attributes, such as their saliency, which may influence respondents’ propensity to attend to or ignore an attribute. To minimize any potential bias, researchers may consider presenting the different attributes in as similar a format as possible. Given that the existing ANA literature, including our study, indicates that a substantial share of respondents ignores one or several choice attributes, estimating standard choice models that assume full attribute attendance could generate biased estimates. Therefore, using choice models that account for ANA behaviour might have to become the norm.

This study calls for more research on ordering effects and information processing strategies in DCEs and on testing whether and how they are related. In particular, the order in which choice attributes and choice alternatives are presented in the choice tasks merits more attention. This includes testing whether presenting the attributes/alternatives horizontally or vertically has any potential impact on the results. The seminal work of Sandorf et al. (2018) shows that it does, which urges a search for further empirical evidence. The valuation literature could benefit from additional studies on ordering effects related to the monetary attribute. An important topic for future research would be to explore the link between attribute (non-)attendance and the visual properties the attributes, including, for example, the presence or absence of pictures or pictograms, their size, numerical or textual information, font size, and colour. Research from other fields provides evidence that the visual properties of products, in particular saliency, do affect consumers’ visual attention and consequently also their choices (van der Lans et al., 2008;

Milosavljevic et al., 2012; Towal et al., 2013; Jonker et al., 2018). There is still a lot to learn about the reasons why respondents ignore ~~the~~ attributes. Finally, the ~~topic~~ research questions~~of~~ from our study could be ~~repeated~~ answered while randomising the order of the attributes, as this would provide a more comprehensive insight into the link between attribute order and ANA. In general, this may be most feasible with rather simple DCE designs with relatively few attributes, since each additional attribute exponentially increases the number of classes in the ANA models, which is likely to increase the incidence of identification problems, especially if applying a combined latent class mixed logit model.

References

Alemu, M.H., Mørkbak, M.R., Olsen, S.B., Jensen, C.L., 2013. Attending to the reasons for attribute non-attendance in choice experiments. Environmental and Resource Economics 54(3):33-359.

Balcombe, K., Burton, M., Rigby, D., 2011. Skew and attribute non-attendance within a Bayesian mixed logit model. Journal of Environmental Economics and Management 62(3): 446-461.

Balcombe, K., Fraser, I., McSorley, E., 2015. Visual attention and attribute attendance in multi-attribute choice experiments. Journal of Applied Econometrics 30 (3): 447-467.

Bello, M., Abdulai, A., 2016. Impact of ex-ante hypothetical bias mitigation methods on attribute non-attendance in choice experiments. American Journal of Agricultural Economics 98(5): 1486-1506.

Boyle, K.J., Özdemir, S., 2009. Convergent validity of attribute-based, choice questions in stated-preference studies. Environmental and Resource Economics 42(2): 247-264.

Brouwer, R., Logar, I., Sheremet, O., 2017. Choice consistency and preference stability in test-retests of discrete choice experiment and open-ended willingness to pay elicitation formats. Environmental and Resource Economics 68(3): 729-751.

Brouwer, R., Martín-Ortega, J., 2012. Modeling self-censoring of polluter pays protest votes in stated preference research to support resource damage estimations in environmental liability. Resource and Energy Economics 34(1): 151-166.

Cameron, T.A., DeShazo, J.R., 2010. Differential attention to attributes in utility-theoretic choice models. Journal of Choice Modelling 3(3): 73-115.

Campbell, D., Erdem, S., 2015. Position bias in best-worst scaling surveys: A case study on trust in institutions. American Journal of Agricultural Economics 97(2): 526-545.

Campbell, D., Hutchinson, W.G., Scarpa, R., 2008. Incorporating discontinuous preferences into the analysis of discrete choice experiments. Environmental and Resource Economics 41(3): 401-417.

Campbell, D., Hensher, D.A., Scarpa, R., 2011. Non-attendance to attributes in environmental choice analysis: A latent class specification. Journal of Environmental Planning and Management 54(8): 1061-1076.

Caputo, V., Van Loo, E.J., Scarpa, R., Nayga Jr., R.M., Verbeke, W., 2018. Comparing serial, and choice task stated and inferred attribute non-attendance methods in food choice experiments. Journal of Agricultural Economics 69(1): 35-57.

Carlsson, F., Kataria, M., Lampi, E., 2010. Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. Environmental and Resource Economics 47(1): 65-89.

Carlsson, F., Mørbak, M.R., Olsen, S.B., 2012. The first time is the hardest: A test of ordering effects in choice experiments. Journal of Choice Modelling 5(2): 19-37.

Carson, R.T., Mitchell, R.C., 1995. Sequencing and nesting in contingent valuation surveys. Journal of Environmental Economics and Management 28(2): 155-173.

Chalak, A., Abiad, M., Balcombe, K., 2016. Joint use of attribute importance rankings and non-attendance data in choice experiments. European Review of Agricultural Economics 43(5): 737-760.

Chavez, D., Palma, M., Collart, A., 2018. Using eye-tracking to model attribute non-attendance in choice experiments. Applied Economics Letters 25(19): 1355-1359.

ChoiceMetrics, 2014. Ngene 1.1.2 User manual & reference guide. Choice Metrics Pty Ltd.

Day, B., Pinto Prades, J.-L., 2010. Ordering anomalies in choice experiments. Journal of Environmental Economics and Management 59(3): 271-285.

Day, B., Bateman, I., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P., 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. Journal of Environmental Economics and Management 63(1): 73-91.

Erdem, S., Campbell, D., Thompson, C., 2014. Elimination and selection by aspects in health choice experiments: Prioritizing health service innovations. Journal of Health Economics 38: 10-22.

Farrar, S., Ryan, M., 1999. Response-ordering effects: A methodological issue in conjoint analysis. Health Economics Letters 8(1): 75-79.

FOEN, Federal Office for the Environment, 2012. Revitalization of running waters – Strategic Planning (in German). Available online at: https://www.bafu.admin.ch/dam/bafu/de/dokumente/wasser/uv-umwelt-vollzug/revitalisierung_fliessgewaesserstrategischeplanung.pdf. Accessed 20 February 2017.

Glenk, K., 2007. A split sample experiment to test for effects of attribute order in choice experiments, in: Meyerhoff, J., Lienhoop, N., Elsasser, P. (Eds.), Stated Preference Methods for Environmental Valuation: Applications from Austria and Germany. Metropolis, Marburg, pp. 81-104.

Glenk, K., Martin-Ortega, J., Pulido-Velazquez, M., Potts, J., 2015. Inferring attribute non-attendance from discrete choice experiments: Implications for benefit transfer. Environmental and Resource Economics 60(4): 497-520.

Grebitus, C., Roosen, J., Seitz, C.C., 2015. Visual attention and choice: A behavioral economics perspective on food decisions. Journal of Agricultural & Food Industrial Organization 13(1): 73-81.

Grebitus, C., Roosen, J., 2018. Influence of non-attendance on choices with varying complexity. European Journal of Marketing 52(9/10): 2151-2172.

Heidenreich, S., Watson, V., Ryan, M., Phimister, E., 2018. Decision heuristic or preference? Attribute non-attendance in discrete choice problems. Health Economics 27(1): 157-171.

Halvorsen, B., 1996. Ordering effects in contingent valuation surveys. Environmental and Resource Economics 8(4): 485-499.

Henningsen, A., Toomet, O., 2011. Maxlik: A package for maximum likelihood estimation in R. Computational Statistics. 26(3): 443–458.

Hensher, D.A., Rose, J., Greene, W.H., 2005. Implications on willingness to pay of respondents ignoring specific attributes. Transportation 32(3): 203-222.

Hensher, D.A., Rose, J., 2009. Simplifying choice through attribute preservation or non-attendance: Implications for willingness to pay. Transportation Research Part E 45(4): 583-590.

Hensher, D.A., Greene, W.H., 2010. Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. Empirical Economics 39(2): 413-426.

Hensher, D.A., Rose, J., Greene, W.H., 2012. Inferring attribute non-attendance from stated choice data: Implications for willingness to pay estimates and a warning for stated choice experiment design. Transportation 39(2): 235-245.

Herriges, J.A., Shogren, J.F., 1996. Starting-point bias in dichotomous choice valuation with follow-up questions. Journal of Environmental Economics and Management 30(1): 112-131.

Hess, S., Hensher, D., 2010. Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. Transportation Research Part B 44(6):781-790.

Hess, S., Hensher, D., 2013. Making use of respondent reported processing information to understand attribute importance: A latent variable scaling approach. Transportation 40(2):397-412.

Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., Caussade, S., 2013. It's not that I don't care, I just don't care very much: Confounding between attribute non-attendance and taste heterogeneity. Transportation 40(3): 583-607.

Hess, S., Train, K., 2017. Correlation and scale in mixed logit models. Journal of Choice Modelling 23:1-8.

Hole, A.R., 2011. A discrete choice model with endogenous attribute attendance. Economic Letters 110(3): 203-205.

Holmes, T., Boyle, K., 2005. Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. Land Economics 81(1): 114-126.

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C.A., 2017. Contemporary guidance for stated preference studies. Journal of the Association of Environmental and Resource Economists 4(2): 319–405.

Jonker, M.F., Donkers, B., de Bekker-Grob, E.W., Stolk, E.A., 2018. Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. Value in Health 21(7): 767-771.

Kjær, T., Bech, M., Gyrd-Hansen, D., Hart-Hansen, K., 2006. Ordering effect and price sensitivity in discrete choice experiments: Need we worry? Health Economics 15(11):1217-1228.

Kragt, M.E., 2013. Stated and inferred attribute attendance models: A comparison with environmental choice experiments. Journal of Agricultural Economics 64(3): 719-736.

Krinsky, I., Robb, A.L., 1986. On approximating the statistical properties of elasticities. The Review of Economics and Statistics 68(4): 715-719.

Krucien, N., Ryan, M., Hermens, F., 2017. Visual attention in multi-attribute choices: What can eye-tracking tell us? Journal of Economic Behavior & Organization 135: 251-267.

Lancsar, E., Louviere, J., 2006. Deleting ‘irrational’ responses from discrete choice experiments: A case of investigating or imposing preferences? Health Economics 15(8): 797-811.

Lewis, K.E., Grebitus, C., Nayga Jr, R.M., 2016. The impact of brand and attention on consumers’ willingness to pay: Evidence from an eye tracking experiment. Canadian Journal of Agricultural Economics 64(4): 753-777.

Mattmann M., Logar, I., Brouwer, R., 2017. A comparison of attribute non-attendance in discrete choice experiments based on stated, inferred, and mouse-tracking data. Paper presented at the 23rd Annual Conference of the European Association of

Environmental and Resource Economists. Available online at: http://fleximeets.com/eaere23/getpaper.php?fid=1795. Accessed 20 March 2019.

McFadden, D., 2001. Economic choices. The American Economic Review 91(3): 351-378.

Metcalfe, P.J., Baker, W., Andrews, K., Atkinson, G., Bateman, I.J., Butler, S., Carson, R.T., East, J., Gueron, Y., Sheldon, R., Train, K., 2012. An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales. Water Resources Research 48(3): W03526, doi:10.1029/2010WR009592.

Milosavljevic, M., Navalpakkam, V., Koch, C., Rangel, A., 2012. Relative visual saliency differences induce sizeable bias in consumer choice. Journal of Consumer Psychology 22(1): 67-74.

Paillex, A., Schuwirth, N., Lorenz, A.W., Januschke, K., Peter, A., Reichert, P., 2017. Integrating and extending ecological river assessment: Concept and test with two restoration projects. Ecological Indicators 72: 131-141.

Poe, G.L., Giraud, K.L., Loomis, J.B., 2005. Computational methods for measuring the difference of empirical distributions. American Journal of Agricultural Economics 87(2): 353-365.

R Core Team, 2014. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Rabin, M., 1998. Psychology and economics. Journal of Economic Literature 36(1): 11–46.

Rekola, M., 2003. Lexicographic preferences in contingent valuation: A theoretical framework with illustrations. Land Economics 79(2): 277-291.

Rose, J.M., Bliemer, M.C.J., Hensher, D.A., Collins, A.T., 2008. Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research Part B 42(4): 395-406.

Sandorf, E.D., Campbell, C., Hanley, N., 2017. Disentangling the influence of knowledge on attribute non-attendance. Journal of Choice Modelling 24: 36-50.

Sandorf, E.D., dit Sourd, R.C., Mahieu, P.-A., 2018. The effect of attribute-alternative matrix displays on preferences and processing strategies. Journal of Choice Modelling 29: 113-132.

Sandorf, E.D., 2019. gizmo: A Collection of Utility Functions for Choice Modeling and Statistical Analysis. Available online at: https://rdrr.io/github/edsandorf/gizmo/. Accessed 16 January 2020.

Sælensminde, K., 2002. The impact of choice inconsistencies in stated choice studies. Environmental and Resource Economics 23(4): 403-420.

Scarpa, R., Gilbride, T.J., Campbell, D., Hensher, D.A., 2009. Modelling attribute non-attendance in choice experiments for rural landscape valuation. European Review of Agricultural Economics 36(2): 151-174.

Scarpa, R., Thiene, M., Hensher, D.A., 2010. Monitoring choice task attribute attendance in nonmarket valuation of multiple park management services: Does it matter? Land Economics 86(4): 817-839.

Scarpa, R., Zanoli, R., Bruschi, V., Naspetti, S., 2012. Inferred and stated attribute non-attendance in food choice experiments. American Journal of Agricultural Economics 95(1): 165-180.

Scott, A., Vick, S., 1999. Patients, doctors and contracts: An application of principal-agent theory to the doctor-patient relationship. Scottish Journal of Political Economy 46(2): 111-134.

Selivanova, A., Krabbe, P.F.M., 2018. Eye tracking to explore attendance in health-state descriptions. PLoS One 13(1): e0190111.

Spash, C.L., 2000. Ecosystems, contingent valuation and ethics: The case of wetland recreation. Ecological Economics 34(2):195-215.

Spinks, J., Mortimer, D., 2016. Lost in the crowd? Using eye-tracking to investigate the effect of complexity on attribute non-attendance in discrete choice experiments. BMC Medical Informatics and Decision Making 16(1): 14.

Tarfasa, S., Brouwer, R., Sheremet, O., Bouma, J., 2017. Informing water harvesting technology contract design using choice experiments. Water Resources Research 53(10): 8211–8225.

Thiene, M., Scarpa, R., Louviere, J.J., 2015. Addressing preference heterogeneity, multiple scales and attribute attendance with a correlated finite mixing model of tap water choice. Environmental and Resource Economics 62: 637-656.

Towal, R.B., Mormann, M., Koch, C., 2013. Silmultaneous modeling of visual saliency and value computation improves predictions of economic choice. Proceedings of the National Academy of Sciences of the United States of America 110(40): E3858-E3867.

van der Lans, R., Pieters, R., Wedel, M., 2008. Competitive brand salience. Marketing Science 27(5): 922-931.

Van Loo, E.J., Nayga, R.M., Campbell, D., Seao, H.-S., Verbeke, W., 2018. Using eye tracking to account for attribute non-attendance in choice experiments. European Review of Agricultural Economics 45(3): 333-365.

Woosley, S., Capelli, F., Gonser, T., Hoehn, E., Hostmann, M., Junker, B., Paetzold, A., Roulier, C. Schweizer, S., Tiegs, S.D., Tockner, K., Weber, C., Peter, A., 2007. A strategy to assess river restoration success. Freshwater Biology 52 (4): 752-769.

Yegoryan, N., Guhl, D., Klapper, D., 2018. Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. Journal of Business Research (In Press), doi.org/10.1016/j.jbusres.2019.01.061.

Figure 1A. Choice task example for the sample receiving ‘length’ as the first attribute in an unlabelled DCE

Figure 1B. Choice task example for the sample receiving ‘biodiversity’ as the first attribute in an unlabelled DCE

Table 1. Estimation results for the latent class choice models that assume full attribute attendance (LC-FAA)

	Order g = 1 (Length top)		**Order g = 2 (Biodiversity top)**	
	Model 1a		**Model 1b**	
Variable	**Class 1**	**Class 2**	**Class 1**	**Class 2**
β_{Length}	0.146	0.407**	0.205**	0.457***
	(0.102)	(0.176)	(0.105)	(0.163)
$\beta_{Walking}$	1.198***	1.085***	1.013***	0.720***
	(0.144)	(0.230)	(0.154)	(0.189)
$\beta_{Swimming}$	0.557***	0.454**	0.347**	0.336
	(0.128)	(0.196)	(0.143)	(0.227)
$\beta_{Barbecuing}$	0.431***	0.531***	0.320**	0.398*
	(0.123)	(0.189)	(0.133)	(0.226)
$\beta_{Mid-biodiversity}$	0.031	0.179	0.026	0.228
	(0.195)	(0.322)	(0.227)	(0.366)
$\beta_{High-biodiversity}$	0.523***	0.190	0.248	0.190
	(0.198)	(0.305)	(0.213)	(0.359)
β_{Price}	–0.002***	–0.017***	–0.003***	–0.016***
	(0.000)	(0.003)	(0.001)	(0.003)
ASC_{Thur}	1.350***	–1.531*	1.299**	–1.502*
	(0.491)	(0.811)	(0.544)	(0.880)
$ASC_{Töss}$	1.547***	–1.648**	1.671***	–1.349
	(0.493)	(0.829)	(0.548)	(0.878)
$\lambda_h = Thur$	1.179		1.266*	
	(0.148)		(0.140)	
$\lambda_h = Töss$	0.889		0.994	
	(0.091)		(0.111)	
Pr($class$)	0.500***	0.500***	0.490***	0.510***
	(0.030)	(0.030)	(0.036)	(0.036)
N	2106		2106	
Log likelihood	–1,574.98		–1,570.63	
AIC	3191.97		3183.26	
BIC	3307.51		3298.80	
Adjusted R^2	0.310		0.312	

Notes: Standard errors in parenthesis. All estimated standard errors are robust and clustered at the individual level. *, ** and *** indicate statistical significance at, respectively, the 10%, 5% and 1% level. The reported scale factor estimates are relative to the baseline labelling treatment h where both rivers Thur and Töss are included in the choice set (the statistical significance asterisks are with respect to one).

Table 2. Marginal willingness to pay estimates (in Swiss Francs per household per year) for the two ordering treatments, derived from the standard model assuming full attribute attendance (LC-FAA) and the combined latent class discrete mixtures model that accounts for ANA (LC-ANA)[a]

	Order $g = 1$ (Length top)		**Order $g = 2$ (Biodiversity top)**		**Significance (p-values) of Poe tests of MWTP equality between order treatments**	
Attribute	**Model 1a (LC-FAA)**	**Model 2a (LC-ANA)**	**Model 1b (LC-FAA)**	**Model 2b (LC-ANA)**	**Model 1 (LC-FAA)**	**Model 2 (LC-ANA)**
Length	43.31 (2.16–151.62)	70.49 (34.93–152.65)	45.24 (12.62–123.88)	53.81 (–32.75–170.40)	0.279	0.133
Walking	288.12 (179.12–625.00)	142.21 (83.37–283.83)	174.22 (102.55–375.08)	135.86 (–148.22–468.36)	0.359	0.041
Swimming	132.44 (61.03–328.87)	154.45 (104.40–281.98)	62.49 (16.43–176.37)	146.68 (–154.63–477.47)	0.417	0.209
Barbecuing	107.80 (47.41–266.43)	164.45 (114.10–298.32)	60.37 (17.80–164.89)	138.32 (–66.01–387.66)	0.484	0.418
Medium biodiversity	11.99 (–50.49–185.61)	125.03 (62.17–271.88)	11.14 (–40.71–143.89)	198.67 (–910.38–1,494.05)	0.485	0.076
High biodiversity	117.34 (25.28–387.24)	210.15 (111.37–430.05)	43.04 (–15.89–189.47)	297.74 (–1,404.50–2,380.18)	0.351	0.105

Notes: 95 per cent confidence intervals in parentheses.

[a]The reported MWTP estimates are weighted MWTP values between two latent classes, where class membership probabilities serve as weights.

1156**Table 3**. Estimation results for the combined latent class discrete mixtures models (LC-ANA)

Variable	Order $g = 1$ (Length top) Model 2a		Order $g = 2$ (Biodiversity top) Model 2b	
	Class 1	**Class 2**	**Class 1**	**Class 2**
β_{Length}	3.029*** (0.740)	3.176*** (0.891)	2.563*** (0.751)	1.376** (0.621)
$\beta_{Walking}$	5.759*** (1.288)	6.500*** (1.069)	3.639*** (0.928)	4.388*** (1.657)
$\beta_{Swimming}$	6.281*** (0.748)	7.053*** (0.766)	6.892*** (1.735)	3.781** (1.540)
$\beta_{Barbecuing}$	9.670*** (1.142)	6.736*** (0.910)	5.231*** (1.579)	3.975** (1.828)
$\beta_{Mid-biodiversity}$	4.594*** (0.918)	5.836*** (1.402)	4.797*** (1.042)	6.586* (3.555)
$\beta_{High-biodiversity}$	5.668*** (1.173)	10.342*** (1.754)	5.669*** (0.919)	10.361** (4.612)
β_{Price}	–0.099*** (0.020)	–0.030*** (0.006)	–0.078*** (0.016)	–0.016 (0.010)
ASC_{Thur}	–3.117*** (0.620)	–1.272 (2.244)	–2.216*** (0.717)	–0.224 (0.998)
$ASC_{Töss}$	–4.564*** (0.831)	–0.029 (2.387)	–1.624*** (0.480)	–0.283 (0.845)
$\lambda_h = Thur$	1.094 (0.332)		1.499 (0.537)	
$\lambda_h = Töss$	0.662** (0.145)		1.315 (0.251)	
Pr($class$)	0.457*** (0.083)	0.543*** (0.083)	0.614*** (0.051)	0.386*** (0.051)
π^0_{Length}	0.479*** (0.127)	0.843*** (0.077)	0.597*** (0.174)	0.629*** (0.144)
$\pi^0_{Walking}$	0.229 (0.183)	0.423*** (0.066)	0.454*** (0.106)	0.433*** (0.112)
$\pi^0_{Swimming}$	0.759*** (0.113)	0.775*** (0.051)	0.993*** (0.079)	0.660*** (0.118)
$\pi^0_{Barbecuing}$	0.846*** (0.066)	0.757*** (0.054)	0.841*** (0.079)	0.715*** (0.145)
$\pi^0_{Biodiversity}$	0.660*** (0.112)	0.853*** (0.038)	0.593*** (0.146)	0.993*** (0.077)
π^0_{Price}	0.162 (0.111)	0.662*** (0.059)	0.337*** (0.074)	0.500*** (0.181)
$\pi_{Ignored\ all}$	0.167*** (0.062)	0.015 (0.023)	0.130** (0.055)	0.007 (0.080)
$\pi_{Considered\ all}$	0.056 (0.035)	0.036* (0.020)	0.063** (0.026)	0.007 (0.083)

N	2106	2106
Log likelihood	–1,321.19	–1,337.44
AIC	2716.38	2750.89
BIC	2922.38	2956.89
Adjusted R^2	0.413	0.406

Notes: Standard errors in parenthesis. All estimated standard errors are robust and clustered at the individual level. *, ** and *** indicate statistical significance at, respectively, the 10%, 5% and 1% level. The reported scale factor estimates are relative to the baseline labelling treatment *h* where both rivers Thur and Töss are included in the choice set (the statistical significance asterisks are with respect to one).

Table 4. Estimation results for the combined latent class mixed logit models (LC-MXL)

	Order g = 1 (Length top)	**Order g = 2 (Biodiversity top)**
Variable	**Model 3a**	**Model 3b**
β_{Length}	2.334***	0.944***
	(0.490)	(0.281)
$\beta_{Walking}$	5.313***	4.310***
	(0.721)	(0.832)
$\beta_{Swimming}$	5.546***	5.710***
	(0.792)	(0.996)
$\beta_{Barbecuing}$	5.419***	4.038***
	(0.743)	(0.732)
$\beta_{Mid-biodiversity}$	5.372***	2.387***
	(1.165)	(0.472)
$\beta_{High-biodiversity}$	9.304***	2.945***
	(2.318)	(0.513)
β_{Price}	–3.216***	–3.569***
	(0.189)	(0.217)
ASC_{Thur}	–1.198**	–0.813**
	(0.482)	(0.386)
$ASC_{Töss}$	–0.698	–0.491
	(0.434)	(0.367)
$\lambda_h = Thur$	0.107	0.730*
	(0.201)	(0.392)
$\lambda_h = Töss$	–0.181	0.488
	(0.136)	(0.311)
σ_{Length}	0.063	1.057***
	(0.815)	(0.316)
$\sigma_{Walking}$	0.005	2.697***
	(0.808)	(0.621)
$\sigma_{Swimming}$	0.011	0.733
	(0.641)	(1.213)
$\sigma_{Barbecuing}$	0.002	0.669
	(0.444)	(0.904)
$\sigma_{Mid-biodiversity}$	0.003	0.143
	(1.159)	(0.727)
$\sigma_{High-biodiversity}$	3.645**	0.050
	(1.443)	(0.579)
σ_{Price}	1.191***	1.439***
	(0.165)	(0.127)
π^0_{Length}	0.779***	0.363
	(0.061)	(0.228)
$\pi^0_{Walking}$	0.440***	0.527***
	(0.048)	(0.108)

$\pi^0_{Swimming}$	0.776***	1.000***
	(0.041)	(0.000)
$\pi^0_{Barbecuing}$	0.743***	0.844***
	(0.041)	(0.049)
$\pi^0_{Biodiversity}$	0.881***	0.666***
	(0.034)	(0.066)
π^0_{Price}	0.338***	0.261***
	(0.057)	(0.061)
$\pi_{Ignored\ all}$	0.000	0.000
	(0.000)	(0.000)
$\pi_{Considered\ all}$	0.034	0.233***
	(0.023)	(0.044)
N	2142	2106
Log likelihood	–1,340.32	–1,343.22
AIC	2732.63	2738.45
BIC	2880.04	2885.41
Adjusted R^2	0.419	0.408

Notes: Standard errors in parenthesis. All estimated standard errors are robust and clustered at the individual level. *, ** and *** indicate statistical significance at, respectively, the 10%, 5% and 1% level. The reported scale factor estimates are relative to the baseline labelling treatment *h* where both rivers Thur and Töss are included in the choice set (the statistical significance asterisks are with respect to one).

Table 5. Significance (p-values) of the Poe test results of equality of membership probabilities across ANA classes within and between ordering treatments

	Order g=1	Order g=2	Order g=1 vs. g=2				
	Model 2		Model 2				Model 3
Attribute ignored	Class 1 vs. class 2	Class 1 vs. class 2	Class 1 (g=1) vs. class 1 (g=2)	Class 2 (g=1) vs. class 2 (g=2)	Class 1 (g=1) vs. class 2 (g=2)	Class 2 (g=1) vs. class 1 (g=2)	
Length	0.009	0.423	0.321	0.111	0.225	0.093	0.039
Walking	0.146	0.434	0.134	0.480	0.166	0.401	0.774
Swimming	0.458	0.010	0.048	0.178	0.263	0.011	0.503
Barbecuing	0.152	0.226	0.487	0.388	0.212	0.192	0.928
Biodiversity	0.053	0.009	0.364	0.060	0.009	0.044	0.001
Price	0.000	0.213	0.090	0.194	0.055	0.000	0.178
Ignored all	0.010	0.009	0.345	0.442	0.054	0.030	0.353
Considered all	0.303	0.261	0.441	0.365	0.291	0.202	0.491

Table 6. Marginal willingness to pay estimates (in Swiss Francs per household per year) for the two ordering treatments, derived from the combined latent class mixed logit model that accounts for ANA (LC-MXL)

Attribute	**Order *g* = 1 (Length top)**				**Order *g* = 2 (Biodiversity top)**				**Significance (*p*-values) of Poe tests of MWTP equality between order treatments**			
	Model 3a (LC-MXL)				**Model 3b (LC-MXL)**				**Model 3 (LC-MXL)**			
	25th percentile	**50th percentile**	**75th percentile**	**Mean**	**25th percentile**	**50th percentile**	**75th percentile**	**Mean**	**25th percentile**	**50th percentile**	**75th percentile**	**Mean**
Length	23.41 (0.000)	55.45 (0.000)	130.21 (0.000)	124.53 (0.000)	3.15 (0.244)	23.74 (0.001)	86.99 (0.000)	99.10 (0.003)	0.013	0.034	0.166	0.323
Walking	58.74 (0.000)	132.65 (0.000)	302.95 (0.000)	286.44 (0.000)	37.85 (0.001)	127.15 (0.000)	387.24 (0.000)	452.98 (0.000)	0.126	0.427	0.725	0.813
Swimming	61.82 (0.000)	139.36 (0.000)	316.92 (0.000)	299.54 (0.000)	72.84 (0.000)	197.64 (0.000)	532.73 (0.000)	589.51 (0.000)	0.730	0.900	0.936	0.937
Barbecuing	60.65 (0.000)	136.42 (0.000)	309.24 (0.000)	291.87 (0.000)	50.48 (0.000)	138.24 (0.000)	375.45 (0.000)	415.90 (0.000)	0.260	0.521	0.723	0.793
Medium biodiversity	58.86 (0.000)	134.95 (0.000)	311.58 (0.000)	296.83 (0.000)	29.93 (0.000)	82.30 (0.000)	225.59 (0.000)	251.70 (0.000)	0.034	0.094	0.239	0.393
High biodiversity	88.98 (0.000)	219.42 (0.000)	524.74 (0.000)	510.52 (0.000)	38.63 (0.000)	103.78 (0.000)	279.67 (0.000)	308.87 (0.000)	0.021	0.029	0.086	0.192

Notes: Significance (*p*-values) of MWTP estimates based on Poe test results in parentheses.

Appendix A1. Overview of ANA studies and reported ANA shares for the first and last non-monetary attributes

Study	Reference	Study characteristics	ANA shares First non-monetary attribute	 Last non-monetary attribute	Higher ANA
1	Hensher et al. (2005)	Stated ANA, 6 attributes	0.08	0.37	Last
1	Hensher et al. (2005)	Stated ANA, 5 attributes	0.05	0.32	Last
1	Hensher et al. (2005)	Stated ANA, 4 attributes	0.16	0.28	Last
2	Campbell et al. (2008)	Stated ANA	0.23	0.26	Last
3	Puckett and Hensher (2008)	Stated ANA - transporter	0.10	0.11	Last
3	Puckett and Hensher (2008)	Stated ANA - shipper	0.27	0.07	First
4	Scarpa et al. (2009)	Inferred ANA – model 2	0.07	0.23	Last
4	Scarpa et al. (2009)	Inferred ANA – model 3	0.06	0.20	Last
4	Scarpa et al. (2009)	Inferred ANA – model 4	0.07	0.20	Last
5	Carlsson et al. (2010)	Stated ANA	0.11-0.13	0.11	Equal
6	Hess and Hensher (2010)	Stated ANA	0.13	0.30	Last
6	Hess and Hensher (2010)	Inferred ANA	0.16	0.29	Last
7	Scarpa et al. (2010)	Inferred ANA; serial	Higher	Lower	First
7	Scarpa et al. (2010)	Inferred ANA; choice-task specific	Lower	Higher	Last
8	Balcombe et al. (2011)	Stated ANA	0.15	0.10	First
9	Campbell et al. (2011)	Inferred ANA; LC model no scale	0.04	0.08	Last
9	Campbell et al. (2011)	Inferred ANA; LC model scale-adjusted	0.09	0.05	First
10	Hole (2011)	Inferred ANA	0.40	0.40	Equal
11	Hensher et al. (2012)	Inferred ANA	0.45	0.37	First
12	Scarpa et al. (2012)	Stated ANA - beef	0.32	0.37	Last
12	Scarpa et al. (2012)	Stated ANA – chicken	0.15	0.22	Last
12	Scarpa et al. (2012)	Inferred ANA; beef; CLC model	0.06	0.04	First
12	Scarpa et al. (2012)	Inferred ANA; beef; HH model	0.26	0.00	First
12	Scarpa et al. (2012)	Inferred ANA; chicken; CLC model	0.51	0.06	First
12	Scarpa et al. (2012)	Inferred ANA; chicken; HH model	0.50	0.64	Last
13	Hess et al. (2013)	Inferred ANA; LC confirmatory model, 1st study	0.53	0.73	Last
13	Hess et al. (2013)	Inferred ANA; LC-MMNL, 1st study	0	0.60	Last
13	Hess et al. (2013)	Inferred ANA; LC confirmatory model, 2nd study	0.42	0.91	Last

13	Hess et al. (2013)	Inferred ANA; LC-MMNL, 2nd study	0.11	0.82	Last
13	Hess et al. (2013)	Inferred ANA; LC confirmatory model, 3rd study	0.64	0.81	Last
13	Hess et al. (2013)	Inferred ANA; LC-MMNL, 3rd study	0.47	0.72	Last
14	Alemu et al. (2013)	Stated ANA	0.17	0.15	First
15	Kragt (2013)	Stated ANA	0.33	0.12	First
15	Kragt (2013)	Inferred ANA	0.64	0.26	First
16	Glenk et al. (2015)	Inferred ANA; Guadalquivir River Basin	0.63	0.66	Last
16	Glenk et al. (2015)	Inferred ANA; Serpis River Basin	0.55	0.45	First
17	Balcombe et al. (2015)	Visual ANA	0.025	0.125	Last
18	Bello and Abdulai (2016)	Stated ANA; baseline treatment	0.20	0.17	First
18	Bello and Abdulai (2016)	Stated ANA; honesty priming treatment	0.03	0.03	Equal
18	Bello and Abdulai (2016)	Stated ANA; cheap talk treatment	0.09	0.11	Last
18	Bello and Abdulai (2016)	Inferred ANA; baseline treatment, EAA model	0.43	0.54	Last
18	Bello and Abdulai (2016)	Inferred ANA; baseline treatment, MEAA model	0.55	0.41	First
18	Bello and Abdulai (2016)	Inferred ANA; honesty priming treatment, EAA model	0.18	0.28	Last
18	Bello and Abdulai (2016)	Inferred ANA; honesty priming treatment, MEAA model	0.11	0.08	First
18	Bello and Abdulai (2016)	Inferred ANA; cheap talk treatment, EAA model	0.37	0.45	Last
18	Bello and Abdulai (2016)	Inferred ANA; cheap talk treatment, MEAA model	0.09	0.34	Last
19	Spinks and Mortimer (2016)	Visual ANA	Lower	Higher	Last
20	Chalak et al. (2016)	Stated ANA; with risk attribute	0.16	0.03	First
20	Chalak et al. (2016)	Stated ANA; without risk attribute	0.07	0.35	Last
21	Sandorf et al. (2017)	Inferred ANA	0.45-0.57	0.00-0.31	First
22	Tarfasa et al. (2017)	Stated ANA; without visualization	0.85	0.95	Last
22	Tarfasa et al. (2017)	Stated ANA; with visualization	0.75	0.99	Last
22	Tarfasa et al. (2017)	Inferred ANA; without	0.000	0.000	Equal

		visualization			
22	Tarfasa et al. (2017)	Inferred ANA; with visualization	0.000	0.000	Equal
23	Caputo et al. (2018)	Stated ANA, serial	0.50	0.71	Last
23	Caputo et al. (2018)	Stated ANA, choice-task specific	0.42	0.44	Last
23	Caputo et al. (2018)	Inferred ANA	0.45	0.43	First
24	Grebitus and Roosen (2018)	Visual ANA; 3-attribute design	0.29	0.17	First
24	Grebitus and Roosen (2018)	Visual ANA; 5-attribute design	0.31	0.29	First
25	Heidenreich et al. (2018)	Inferred ANA; familiar respondents	0.37	0.34	First
25	Heidenreich et al. (2018)	Inferred ANA; unfamiliar respondents	0.12	0.35	Last
26	Selivanova and Krabbe (2018)	Visual ANA	Lower	Higher	Last
27	Chavez et al. (2018)	Visual ANA	0.34	0.46	Last
28	Yegoryan et al. (2019)	Visual ANA; coffee makers	Higher	Lower	First
28	Yegoryan et al. (2019)	Inferred ANA; coffee makers	Lower	Higher	Last
28	Yegoryan et al. (2019)	Visual ANA; laptops	Lower	Higher	Last
28	Yegoryan et al. (2019)	Inferred ANA; laptops	Lower	Higher	Last

Appendix A2. Stated attribute importance (the share of respondents who selected the choice attribute as the most important one for making choices in the DCE)

Attribute	Order $g = 1$ (Length top)	Order $g = 2$ (Biodiversity top)
	%	%
Length	8.33	13.11
Walking	32.44	27.74
Swimming	9.23	10.06
Barbecuing	11.61	12.50
Biodiversity	20.24	16.46
Price	18.15	20.12

Appendix A3. Marginal willingness-to-pay estimates per latent class derived from Models 1 and 2

	Model 1 (LC-FAA)			
	Order 1 (g = 1)		Order 2 (g = 2)	
Attribute	**Class 1**	**Class 2**	**Class 1**	**Class 2**
Length	62.34 (–17.81–277.72)	24.27 (2.83–58.23)	62.68 (–0.10–223.20)	28.52 (7.48–59.33)
Walking	511.37 (297.75–1,197.05)	64.69 (33.98–117.93)	308.96 (165.24–738.58)	44.96 (19.97–84.01)
Swimming	237.75 (100.51–631.82)	27.05 (3.14–65.00)	105.79 (15.43–348.17)	20.95 (–6.98–60.38)
Barbecuing	183.87 (66.82–507.62)	31.66 (8.60–68.81)	97.41 (14.50–320.34)	24.83 (–2.66–65.43)
Medium biodiversity	13.32 (–105.74–355.63)	10.66 (–23.79–64.72)	7.93 (–87.00–291.46)	14.21 (–30.26–70.36)
High biodiversity	223.27 (44.05–771.53)	11.33 (–21.54–62.40)	75.52 (–38.20–386.94)	11.89 (–31.41–66.29)
	Model 2 (LC-ANA)			
	Order 1 (g = 1)		**Order 2 (g = 2)**	
Attribute	**Class 1**	**Class 2**	**Class 1**	**Class 2**
Length	30.54 (21.75–39.10)	104.17 (38.74–219.45)	32.92 (19.86–42.90)	87.00 (–149.65–392.85)
Walking	58.06 (42.79–74.94)	213.16 (113.82–407.81)	46.75 (32.05–60.93)	277.48 (–474.33–1,128.22)
Swimming	63.33 (42.29–105.11)	231.29 (144.90–405.98)	88.53 (42.61–165.24)	239.11 (–480.98–1,171.69)
Barbecuing	97.49 (78.19–134.53)	220.90 (125.71–412.53)	67.20 (28.18–121.80)	251.34 (–252.23–892.23)
Medium biodiversity	46.32 (29.05–73.01)	191.40 (84.99–386.37)	61.62 (47.57–78.14)	416.48 (–2,629.61–3,961.19)
High biodiversity	57.15 (39.32–81.08)	339.17 (190.70–618.34)	72.82 (60.26–94.24)	655.18 (–3,795.47–6,170.79)

Notes: 95 per cent confidence intervals in parentheses.

	Model 1 (LC-FAA)					
	Order 1 vs. 2		**Order 1 (class 1) vs. order 2 (class 2)**	**Order 1 (class 2) vs. order 2 (class 1)**	**Class 1 vs. 2**	
Attribute	**Class 1**	**Class 2**			**Order 1**	**Order 2**
Length	0.499	0.426	0.073	0.024	0.225	0.191
Walking	0.144	0.287	0.000	0.000	0.000	0.000
Swimming	0.129	0.383	0.000	0.008	0.000	0.048
Barbecuing	0.192	0.381	0.001	0.008	0.006	0.064
Medium biodiversity	0.479	0.458	0.421	0.459	0.480	0.471
High biodiversity	0.163	0.499	0.004	0.124	0.010	0.182

	Model 2 (LC-ANA)					
	Order 1 vs. 2		**Order 1 (class 1)**	**Order 1 (class 2)**	**Class 1 vs. 2**	
Attribute	**Class 1**	**Class 2**	**vs. order 2 (class 2)**	**vs. order 2 (class 1)**	**Order 1**	**Order 2**
Length	0.460	0.258	0.003	0.051	0.000	0.115
Walking	0.160	0.475	0.004	0.002	0.000	0.055
Swimming	0.329	0.327	0.002	0.139	0.000	0.117
Barbecuing	0.409	0.360	0.003	0.263	0.001	0.225
Medium biodiversity	0.341	0.453	0.036	0.030	0.002	0.313
High biodiversity	0.336	0.434	0.028	0.047	0.000	0.185

Appendix A4. Significance (*p*-values) of the Poe test results of equality of MWTP estimates between latent classes in Models 1 and 2

1210