Environ Resource Econ
DOI 10.1007/s10640-015-9905-1

# Are Fast Responses More Random? Testing the Effect of Response Time on Scale in an Online Choice Experiment

**Tobias Börger**[1]

Accepted: 8 March 2015


**Abstract** Scepticism over stated preference surveys conducted online revolves around the concerns over "professional respondents" who might rush through the questionnaire without sufficiently considering the information provided. To gain insight on the validity of this phenomenon and test the effect of response time on choice randomness, this study makes use of a recently conducted choice experiment survey on ecological and amenity effects of an offshore windfarm in the UK. The positive relationship between self-rated and inferred attribute attendance and response time is taken as evidence for a link between response time and cognitive effort. Subsequently, the generalised multinomial logit model is employed to test the effect of response time on scale, which indicates the weight of the deterministic relative to the error component in the random utility model. Results show that longer response time increases scale, i.e. decreases choice randomness. This positive scale effect of response time is further found to be non-linear and wear off at some point beyond which extreme response time decreases scale. While response time does not systematically affect welfare estimates, higher response time increases the precision of such estimates. These effects persist when self-reported choice certainty is controlled for. Implications of the results for online stated preference surveys and further research are discussed.

**Keywords** Attribute non-attendance · Choice experiment · Generalised multinomial logit · Offshore windfarm · Online survey · Response time · Scale heterogeneity

**JEL Classification** C81 · Q51 · H43

✉ Tobias Börger
tobo@pml.ac.uk

1 Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK

Published online: 18 March 2015

## 1 Introduction

Internet surveys have become increasingly popular as a mode for environmental valuation studies. An increasing number of stated preference surveys (Olsen 2009; Lindhjem and Navrud 2011; Windle and Rolfe 2011) are now conducted online (Lindhjem and Navrud 2011). This survey mode offers the potential to reach a wide range of respondents at relatively low cost and data can be collected and managed quite rapidly (Malhotra 2008). Online questionnaires can be tailored to adjust the question set based on prior responses and involve sophisticated graphical and multimedia material. Yet, despite these advantages critics of this survey mode state that biases might result from the use of pre-recruited panels of respondents who complete surveys on a regular basis because these respondents are more likely to be motivated by the incentive provided upon completion of the survey than by truthfully reporting their attitudes and preferences regarding the particular survey topic. Following on from this, it is often suggested that respondents might rush through the questionnaire without sufficiently considering all information provided, a phenomenon which has been described as "quick-and-dirty" responses or numbers (Schwappach and Strasmann 2006; Olsen 2009). Such response behaviour might result in survey data of reduced quality (Malhotra 2008; Campbell et al. 2013) and thus reduce the overall validity of the survey. This study aims to provide evidence in this respect and investigates whether fast answers to a discrete choice experiment are more random than choice responses over which more time is taken.

One convenient feature of internet surveys is that response time can be accurately measured. Online surveys offer the possibility to record the time to complete certain questionnaire sections down to the second, which would be much more burdensome in face-to-face or self-administered surveys. The present study makes use of such response time data and tests the effects of the time respondents need to complete the choice tasks on the randomness of their choices. If—as is often the case in environmental valuation surveys—the survey topic is rather complex and new to the majority of respondents, it is safe to assume that respondents need some time to familiarise themselves with the good to be valued and the procedure of the choice task. Similarly, it can be expected that respondents need at least a minimum amount of time to complete the actual choice experiment. Consequently, it has been conjectured, and also partially tested empirically, that very quick respondents are not able to consider all information provided, which results in their responses being more random than those of respondents who take sufficient time to complete the task at hand (e.g. Haaijer et al. 2000; Rose and Black 2006; Campbell et al. 2013). In such studies, response time is usually interpreted as a proxy for cognitive effort of the respondent. Yet, this relationship is not that straightforward because a long response time in online surveys might have other reasons than simply a high level of concentration on the part of the respondent. It is possible that a respondent not only completes the survey but simultaneously engages in other activities on the computer or in her immediate environment. Therefore, the first objective of the present analysis is the exploration of the link between response time and self-reported and inferred attribute attendance. If longer response time increases attribute attendance it can be interpreted as a proxy for cognitive effort in the choice experiment.

Following from that, the second objective of this paper is an investigation of the impact of response time on choices and particularly on scale, which indicates the weight of the deterministic relative to the error component in the random utility model. The central hypothesis is that respondents who spend more time completing the choice tasks give less random responses, i.e. responses which are determined more strongly by the attributes displayed on the choice cards. Technically this means that with longer response time the deterministic part of utility is expected to increase compared to the error component. To test this hypothesis the generalised multinomial logit (GMNL) model (Fiebig et al. 2010; Gu et al. 2013) is employed on a dataset of an online choice experiment survey valuing different impacts of the installation of an offshore windfarm in the UK. The GMNL model has recently been used in similar choice experiment settings to investigate the effect of familiarity with the good to be valued (Czajkowski et al. 2015) or level of education (Czajkowski et al. 2014b) on scale. Response time is measured and modelled as a determinant of systematic inter-respondent differences in scale. These scale differences are taken as evidence of different respondents making more, or less, random choices. In addition to that, the simultaneous effect of choice certainty on scale is tested, as an example of another variable evaluating the choice task from the perspective of the respondent.

The remainder of the paper is structured as follows: Sect. 2 reviews the literature on online stated preference surveys and response time models and develops some research questions; Sect. 3 describes the survey, the measurement of response time and the econometric approach; Sect. 4 presents the results and Sect. 5 provides the discussion and conclusions.

## 2 Choice Experiments, Mode Effects and the Influence of Response Time

### 2.1 Survey Mode Effects and Online Surveys

An overview of mode effects of internet surveys can be found in Lindhjem and Navrud (2011). A major concern of studies investigating these effects is that welfare estimates might differ systematically across survey modes. Olsen (2009) compares choice experiments conducted online and by mail looking at six different criteria. Among those he finds higher response rates and fewer protest respondents in the online survey, whereas willingness to pay (WTP) estimates do not differ significantly across modes. Windle and Rolfe (2011) compare a complex choice experiment as paper-based and online surveys. These authors do not find significantly different welfare estimates for an improvement of the condition of the Great Barrier Reef. They do, however, detect different sets of demographic variables that influence stated choices. The applicability of internet surveys for other environmental valuation methods, such as contingent valuation (Lindhjem and Navrud 2011) and the travel cost method (Fleming and Bowden 2009) has been tested as well. In a contingent valuation study Taylor et al. (2009) cannot detect significant WTP differences between panel-based online surveys and mail and phone surveys. Similarly, Nielsen (2011) does not find different WTP for a reduction of air pollution in online and face-to-face interviews.

In online surveys, respondents are often recruited through existing panels held by market research companies, which brings about a series of sampling problems. Members of such panels usually actively opt in to join a panel after being contacted or attracted by an advertisement. The effects of this potential self-selection of panel members on survey responses are still subject to ongoing research (Baker et al. 2010), but probability-based sampling increasingly becomes available. Independent of the recruitment procedure it is possible that panel members complete surveys with different topics regularly and are remunerated for their effort. This procedure has been criticised to produce "professional respondents" (Dennis 2001; Hillygus et al. 2014) whose attitudes might be changed through their experience of being a regular survey taker. This experience may make them unrepresentative of the basic population. Such respondents are also likely to rush through the questionnaire and not to give questions an appropriate amount of consideration (Malhotra 2008). Yet, Hillygus et al.

(2014) find that frequent survey takers took more time to complete an online questionnaire. Independent of the experience in taking surveys the question arises whether respondents who spend little time on the questionnaire or some of its core parts give lower quality responses than respondents who take sufficient time. Studies investigating the effect of response time on responses in stated preference surveys are reviewed in the following section.

## 2.2 Response Time Models

Before the recent increase in the use of online surveys, a number of studies investigated the effect of externally provided time-to-think on responses in stated preference surveys (e.g. Whittington et al. 1992; Cook et al. 2007, 2012; Svedsäter 2007). In such surveys, respondents are interviewed face-to-face in their homes, given a certain time to think about the valuation scenario and their responses (typically one night) and contacted again on the second day to state their WTP or discrete choices. Results frequently show that time to think decreases WTP for the good as well as self-reported uncertainty about the responses given. Rather than merely forcing response time to be higher by this approach, Cook et al. (2012) also assess how long a respondent has reflected upon the valuation task during the additional time to think. Their results show that self-reported reflection time does in some occasions explain lower demand for the good to be valued.

Since online and computer-assisted stated preference surveys have become more common, studies investigating the role of objectively measured response time on stated choices have also become more numerous. One of the first of such studies is Holmes et al. (1998) who find that response time systematically affects the preference structure for rainforest protection in an adaptive conjoint analysis. Their results show that response time increases preference intensity, i.e. how strongly respondents prefer one good over another. Further they find that respondents who took little time to answer the conjoint questions did not respond in ways that conform to underlying economic theory. This study tries to explain variance of the error term by response time, but the results are inconclusive. Explaining the latter is the objective of a study by Haaijer et al. (2000) which shows a better fit to the data of a multinomial probit (MNP) model that accounts for response time compared to a model that does not. This study also detects that longer response time leads to lower variance of the error term and thus more systematic choices. Rose and Black (2006) extend this approach and find evidence for the impact of response time on both coefficient means and variances in a mixed logit model. They conclude that response time is not only influencing preference structure but also the variance of randomly distributed taste parameters.

In a laboratory comparison exercise by Brown et al. (2008) involving private and public goods and amounts of money similar to the above studies, response time was found to decrease over repeated comparison tasks. Responses also became more consistent over time. The authors suspect that respondents enter valuation tasks of public goods with relatively undefined preferences and that preference statements become more precise with increasing experience with the task. Vista et al. (2009) record response time in an online choice experiment survey and regress it on respondents' demographic variables. They find response time to be unrelated to any demographic or attitudinal variable except age. Further they run separate conditional logit models grouped by different response times and find evidence of systematically different preference parameters. Contrary to that, Bonsall and Lythgoe (2009) record response time data in a choice experiment survey and investigate its determinants. Their results show that, apart from being affected by demographic variables such as age and level of education, response time increases with perceived scenario complexity, and decreases with choice order. In a study looking at the effect of the number of choice sets Bech et al. (2011)

apply the heteroskedastic conditional logit (HCL) model to test for systematic differences in error variances in an online health related choice experiment survey. They find that among other variables, response time increases error variance, i.e. decreases scale.

A latent class model which allows for different scale across classes is used by Campbell et al. (2013). In their approach response time is modelled as a determinant of class membership, and different classes are allowed to have different levels of scale. Results show that on average error variance (which is negatively proportional to scale) decreases with response time, i.e. choices become more deterministic when respondents take more time to complete them. However, this model limits the number of classes to two, thus only allowing scale to be high in one class and comparatively lower in the other. It is not clear from these results if there might be a continuous effect of response time on scale, i.e. reduced choice randomness.

The problem of assessing the impact on scale heterogeneity of respondent engagement in the survey task is tackled by Hess and Stathopoulos (2013). In their hybrid model, response time is a function of engagement in the survey task which is a latent variable. Survey engagement is further linked to several socio-demographic variables. Results confirm the relationship between latent engagement and both positive statements about survey engagement and longer response time.

### 2.3 Research Questions

Rather than modelling response time as determined by a latent variable that indicates respondent engagement and also affects choices, as in Hess and Stathopoulos (2013) this study follows Campbell et al. (2013) and models it as explanatory variable of choices. However, unlike the latter approach, continuous differences in scale are studied rather than just a two-class latent class model. As suggested by some of the studies reviewed in Sect. 2.2 above and by Haaijer et al. (2000) in particular, it can be expected that choice randomness decreases with increased response time, if the necessary time to consider the choice options and state the choice is an indicator of the diligence or cognitive effort with which a respondent engages in that task. As the link between response time and cognitive effort is not entirely straightforward (Campbell et al. 2013), the first step of the analysis investigates the relationship between attribute attendance and response time. Upon completion of the choice tasks respondents are asked whether they considered each choice attribute (e.g. Alemu et al. 2013; Hole et al. 2013). It is hypothesised that respondents who state that they have considered several or all attributes when stating their choices have taken more time for this task. Further, an equality-constrained latent class model is used to infer attribute attendance from choice data and test the influence of response time on class membership (Scarpa et al. 2009, 2012). These analyses are intended to shed greater light on the question of whether response time is an appropriate proxy for cognitive effort in stated preference surveys.

It is further hypothesised that the less time respondents take to complete the choice tasks the higher the choice randomness as indicated by the inverse of the scale parameter. This hypothesis follows from earlier work on this topic discussed above and is the central focus of this study. If response time is indeed an indicator of cognitive effort it can be expected that more engaged respondents take longer to complete the choice tasks and thus state choices which are influenced more strongly by displayed attribute levels.

Besides response time, there might be other factors which affect choice randomness. It is standard practice in stated preference surveys to ask respondents how certain they are of their choices (or responses in general) (e.g. Li and Mattsson 1995; Brouwer et al. 2010). It is straightforward to expect that self-rated choice certainty should have a positive effect on scale. Choices of respondents who are more certain that they would actually make the choice as stated in the questionnaire can be expected to have a larger deterministic than random component of utility (Lundhede et al. 2009; Beck et al. 2013). The present study therefore investigates whether accounting for this potential source of scale heterogeneity moderates the hypothesised effect of response time on scale.

## 3 Methods

### 3.1 The Survey

This study uses data from a stated preference survey that investigates welfare impacts of the installation of an offshore windfarm in UK waters of the Irish Sea. A sample of residents in Northwest England and Wales were interviewed in an online survey regarding the ecological and amenity impacts of the Rhiannon windfarm to be built between the Welsh island of Anglesey and the Isle of Man.[1] While the decision to build the windfarm has already been made, the valuation scenarios involve different construction designs which would lead to different ecological and amenity impacts. The installation of wind turbines might have positive effects on local biodiversity through so-called “artificial reef” effects, whereby species such as brown crabs, mussels and seaweed colonise the turbine foundations, contributing to the attraction of further mobile species such as fish. Therefore the first choice attribute is an increase in the number of species expected to live in the confines of the windfarm. The second attribute relates the different levels of turbine height and their visibility from different sections of the Irish Sea coast. While the size and future electricity output of the windfarm has already been fixed, it can either consist of a greater number of comparably small turbines (180 m high) or a smaller number of higher turbines (240 or 300 m high) (Celtic Array Limited 2012). The third attribute concerns the potential impacts of electromagnetic fields from subsea cables and their mitigation. Depending on the depth at which these cables are buried, an impact on marine life is either still possible or entirely prevented. Unlike previous studies that focused exclusively on the visual impact (e.g. Ladenburg and Dubgaard 2007; Krueger et al. 2011), this survey includes both amenity and ecological impacts of a windfarm, thus also assessing the relative importance of these issues from a welfare perspective. The fourth attribute is the cost component specified as additional tax to be paid annually by every household in the region. Choice attributes and their respective levels are displayed in Table 1.

The questionnaire was developed based on insights from 60 semi-structured interviews with members of the general public in the region and four focus group meetings. The experimental design was developed based on results of a pilot survey conducted online ($n = 90$). Coefficients indicating the influence of the attributes on choices were calculated by applying a mixed logit model (cf. Sect. 3.3) and used as priors when developing the choice experiment design. A Bayesian D-efficient design (Scarpa and Rose 2008) in the software package Ngene (ChoiceMetrics 2012) was used. The design was restricted to exclude policy options which yield the status quo level for each non-monetary attribute at positive cost, since this option would be dominated by the status quo at zero cost. The design included 24 choice tasks separated into four blocks. Respondents were randomly allocated to one block and just completed the six choice tasks in that respective block. Each choice task contained two

[1] The respondent panel underlying the sampling process had not been constructed in a probability-based way, but panellists were recruited through online advertisements. As discussed in Sect. 2.1, this increases the likelihood of self-selection and resulting problems around “professional respondents”.

**Table 1** Choice attributes of the Irish Sea survey (status quo in Italics)

| Attribute | Description | Levels |
|---|---|---|
| Enhanced biodiversity (SPEC) | Number of additional species to settle in and around the new wind farm | *No change*, 10 more species, 30 more species |
| Turbine height and visibility (HEI) | Height of the wind turbines in metres and coastal strips from which they would be visible | *180 m—visible from Anglesey and the Isle of Man,* 240 m—visible from Anglesey, the Isle of man and Cumbria, 300 m—visible from Anglesey, the Isle of man, Cumbria and Liverpool |
| Electromagnetic fields from cabling (EMI) | Depending on how deep the cabling is buried under the seafloor there is or is not any impact on marine mammals | *Impact—cabling buried at 1 m*, no impact—cabling buried at 2 m |
| Additional tax (COST) | Additional tax to be paid annually by every household to fund alternative wind farm design | *£0*, £5, £10, £25, £50 |

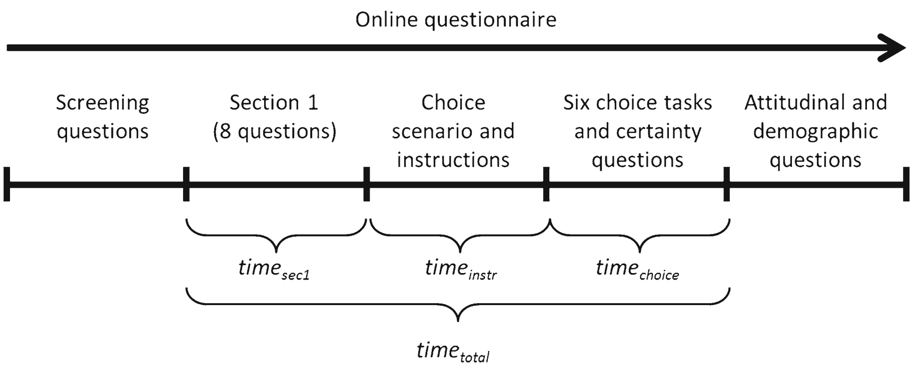


**Fig. 1** Stylised structure of the questionnaire and measurement of relevant time variables

alternative windfarm options at different cost levels (Options A and B, subsequently referred to as change options) and a ‘no change’ option at zero cost (Option C).

## 3.2 Measurement of Response Time

Online surveys allow for exact measurements of the time respondents need to read web-pages and complete response tasks. Different time measures were taken for this study. Time stamps indicating the elapsed time in seconds up to that point in the questionnaire were recorded after each question and each instruction page so that the following time variables could be computed (Fig. 1). The time respondents need to complete general questions about the topic in Sect. 1 is recorded as $time_{sec1}$. $time_{instr}$ denotes how long it takes respondents to read the scenario description and the instructions for the choice tasks. $time_{choice}$ is the time to complete the six choice tasks and the choice certainty questions following each stated choice. Since the order of the choice cards was randomised across respondents it was not possible to record the time needed to complete each single choice task. The total time respondents needed for these three sections is reported as $time_{total}$. Upon completion of the choice tasks the questionnaire continues with sets of attitudinal and demographic questions, however, no response time was recorded for this final section. As additional variables the shares of the time respondents devote to specific sections relative to the total time is computed as e.g. $ts_{choice} = time_{choice}/time_{total}$.

This study measures total response time for all choices made by one respondent rather than assessing it for each choice task separately. It has been shown that response time decreases significantly with the number of choices as respondents learn about the task and their preferences (Haaijer et al. 2000; Rose and Black 2006). It has also been suggested that response time to a single choice task is further influenced by task complexity, which is likely to differ between tasks presented to a single respondent. These effects are expected to cancel out when a series of six choice tasks is analysed in aggregate. Consequently, measuring response time for the whole set of choices allows the researcher to circumvent these potentially biasing effects and might thus assess cognitive effort of the respondent more accurately (Malhotra 2008).

It should further be noted that this study does not control for potential differences in loading time of webpages. As Campbell et al. (2013) remark there might be differences in how long it takes for a new page of the questionnaire to appear on a respondent's screen. While this is certainly an area for further research, it is believed that differences in loading time are unlikely to be very large and thus cannot be a biasing factor in a country with very high internet penetration and transmission speed such as the UK.

Beyond the choice tasks the questionnaire contains additional questions which are used to assess choice behaviour in a wider sense. After each choice respondents are asked to state how certain they are of their choice on a 5-point scale from "Not certain at all" to "100 % certain". In the choice models the sum of all certainty statements over the six choice tasks, ranging from 6 to 30, is used as an additional explanatory variable (CERTAINTY).

### 3.3 Equality-Constrained Latent Class (ECLC) and Generalised Multinomial Logit (GMNL) Models

The analysis of the choice data is based on the random utility model (McFadden 1974). The two econometric models employed in this study are the equality-constrained latent class (ECLC) (Scarpa et al. 2009, 2012) and the generalised multinomial logit (GMNL) models (Fiebig et al. 2010; Gu et al. 2013). In all models, utility of respondent $n$ from choosing option $i$ in choice occasion $t$ is assumed to consist of representative utility explained by the vector $x_{nit}$ consisting of alternative- and respondent-specific attributes and its coefficient vector $\beta_n$ and an unobserved error term $\varepsilon_{nit}$ according to

$$U_{nit} = \beta_n x_{nit} + \varepsilon_{nit}. \tag{1}$$

Assuming that respondents choose the alternative that maximises utility, and further that the error term is independently and identically distributed Type I Extreme Value, the probability that respondent $n$ chooses alternative $i$ in choice occasion $t$ is given by

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

In the ECLC model, $\beta_n$ is assumed to follow a discrete distribution and belong to one class $c$ of $K$ classes. Thus, the probability conditional on class $c$ that respondent $n$ prefers alternative $i$ over alternatives $j$ is

$$P_{nit}\,|\beta_c = \frac{exp\left(\beta_c{}'x_{nit}\right)}{\sum_{j=1}^{J} exp\left(\beta_c{}'x_{njt}\right)}; \quad c \in \{1, \ldots, K\}. \tag{3}$$

The unconditional choice probability is given by

$$P_{nit} = \sum_{c=1}^{K} \frac{exp\left(z_n{}'\theta_c\right)}{\sum_{k=1}^{K} exp\left(z_n{}'\theta_k\right)} \frac{exp\left(\beta_c{}'x_{nit}\right)}{\sum_{j=1}^{J} exp\left(\beta_c{}'x_{njt}\right)} \tag{4}$$

The probability of belonging to a particular class $c$ out of all $K$ classes is given by the first ratio in (4). $\theta_c$ are $K-1$ class membership model parameters with $\theta_K$ being normalised to zero for identification. The vector $z_n$ contains individual-specific variables (such as response time) influencing class membership. What distinguishes the ECLC from a normal latent class model are the restrictions of the parameters in different classes (Scarpa et al. 2009, 2012). These are threefold: (i) if an attribute is not attended to its coefficient is zero as this attribute does not affect choice probabilities; (ii) if not zero, estimated parameters are equal across classes except for (iii) the alternative-specific constant indicating the change options which is not constrained to be equal across classes (Glenk et al. 2014). With these restrictions different attendance patterns can be created and tested against the data.

To investigate the effect of response time on scale the GMNL modelling framework is employed (Fiebig et al. 2010; Gu et al. 2013). In this model, $\beta_n$ is specified as distributed according to

$$\beta_n = \sigma_n \beta + \left[\gamma + \sigma_n (1-\gamma)\right] \eta_n \tag{5}$$

where $\sigma_n$ is a scale factor that shifts the constant coefficient vector $\beta$ up or down relative to the error component of the utility function. $\eta_n$ is a respondent-specific deviation of the individual parameters from the sample mean parameter $\beta$. The specific values that the parameters in (5) can assume determine the exact form of the individual utility function $U_{nit}$ in (1) and thus specify the type of choice model as follows.

Limiting the model's flexibility and assuming that attribute coefficients are constant for all respondents in the sample, i.e. $\beta_n = \beta$ (and $\sigma_n = 1$), yields the conditional logit model with $U_{nit} = \beta x_{nit} + \varepsilon_{nit}$. This model does not account for unobserved preference heterogeneity. Different preferences between respondents can only be detected through the inclusion of interactions of attribute levels with respondent-specific variables. To also account for preference heterogeneity not observed by means of any variables in the dataset, the elements of the coefficient vector can be modelled as random variables that may deviate from the sample mean $\beta$. To this end it is assumed that $\sigma_n = 1$ and hence $\beta_n = \beta + \eta_n$. This is the mixed logit model (Revelt and Train 1998), which assumes that attribute coefficients are respondent-specific and results in $U_{nit} = (\beta + \eta_n)x_{nit} + \varepsilon_{nit}$. Different distributional assumptions can be made with respect to $\eta_n$. Unless indicated otherwise this study uses the normal distribution for all coefficients.

More recently it has been proposed that unobserved heterogeneity might also stem from so-called scale heterogeneity (e.g. Louviere et al. 2002). Within the GMNL framework this model can be achieved in a number of ways. To simultaneously account for preference and scale heterogeneity it is assumed that the scale parameter $\sigma_n$ is random and that $\gamma = 0$ so that (5) simplifies to $\beta_n = \sigma_n(\beta + \eta_n)$. This model allows for the variation of coefficients $\beta_n$ between respondents, i.e. the variation introduced by $\eta_n$, and an overall scaling of the coefficient pattern by $\sigma_n$. According to Louviere et al. (2000) $\sigma^2$ is inversely related to the variance of the idiosyncratic error $\varepsilon_{nit}$ in (1). Consequently, it can be used to distinguish between (groups of) respondents who make more, or less, random choices. If the scale parameter increases (decreases), the deterministic part of utility is assigned a greater (smaller) weight compared to the error component. As a result, increased scale is interpreted as evidence for more determined, i.e. less random, choices. The random scale parameter is specified as

$$\sigma_n = exp\left(\bar{\sigma} + \theta z_n + \tau\right). \tag{6}$$

$\bar{\sigma} + \theta z_n$ is the mean of $\sigma_n$ with $\bar{\sigma}$ denoting a normalising constant and $z_n$ a vector of respondent-specific variables with its coefficient vector $\theta$. $\tau$ is the standard deviation of $\sigma_n$, which means the scale parameter follows a log-normal distribution, i.e. $\sigma_n \sim LN(1, \tau)$.

Hess and Rose (2012) show that random preference and scale heterogeneity cannot be identified simultaneously because everything that can be identified by means of the GMNL model is the product of $\sigma_n$ and $(\beta + \eta_n)$. In such a model it is not clear if the detected heterogeneity of coefficients stems from preference heterogeneity ($\eta_n$) or a general up- or downscaling of all (scaled) coefficients by $\sigma_n$. This criticism, however, only applies to the random, i.e. unexplained, component of preference and scale heterogeneity. If differences in scale across respondents are explained by measurable respondent characteristics (as in $z_n$), it can be detected whether these characteristics shift all (scaled) coefficients in one or the other direction. This shifting effect of all coefficients [and thus their weight relative to the error component in (1)] is the focus of this study as the scale parameter can be interpreted as a measure of choice randomness. It answers the question of how sensitive responses are to the displayed levels of the choice attributes. $z_n$ is modelled to include certain choice-related variables such as response time and choice certainty. Applications of the GMNL model with this interpretation include Czajkowski et al. (2014b, 2015), Kragt (2013), Juutinen et al. (2012) and Christie and Gibbons (2011). While the conditional logit and ECLC models can be computed using a traditional maximum likelihood approach, simulated maximum likelihood estimation with 1000 Halton draws is required for all other models.

It should be noted that the scale factor $\sigma_n$ cancels out when WTP of attribute $k$ is calculated as $WTP_k = -(\beta_k/\beta_{cost})$. Therefore, controlling for response scale should theoretically not affect estimates of mean WTP. Yet, scale does influence the size of the confidence intervals of WTP estimates. Higher scale (i.e. lower error variance) allows for the more precise estimation of attribute coefficients and consequently leads to smaller confidence intervals of mean WTP.

## 4 Results

### 4.1 Descriptive Statistics

The main survey was conducted by a market research company under close supervision of the researchers between November 2013 and January 2014. Respondents were sampled from an existing panel owned by that company.[2] The sample consists of 519 completed questionnaires. Table 2 shows descriptive statistics of response time for different sections of the questionnaire. Means of all time variables measured in seconds (hereafter 'actual

[2] Although this panel was not established through probability-based sampling and therefore representativeness of samples drawn from it is not warranted, it can be expected that with respect to the specific survey topic the selection of respondents is random.

**Table 2** Descriptive statistics of actual ($time_{sec1}$, $time_{instr}$, $time_{choice}$, $time_{total}$) and relative response time variables ($ts_{sec1}$, $ts_{instr}$, $ts_{choice}$)

| Variable | N | Unit | Mean | SD | Median | Min. | Max. |
|---|---|---|---|---|---|---|---|
| $time_{sec1}$ | 519 | Seconds | 196.43 | 98.07 | 171 | 72 | 881 |
| $time_{instr}$ | 519 | Seconds | 74.29 | 51.10 | 66 | 7 | 238 |
| $time_{choice}$ | 519 | Seconds | 184.04 | 87.40 | 166 | 45 | 576 |
| $time_{total}$ | 519 | Seconds | 454.76 | 187.55 | 412 | 128 | 1394 |
| $ts_{sec1}$ | 519 | Share | 0.441 | 0.122 | 0.420 | 0.153 | 0.862 |
| $ts_{instr}$ | 519 | Share | 0.154 | 0.078 | 0.155 | 0.016 | 0.399 |
| $ts_{choice}$ | 519 | Share | 0.404 | 0.093 | 0.410 | 0.113 | 0.667 |

response time') exceed their respective medians since distributions of these measurements are skewed to the right, and there are some positive outliers. As indicated by the minimum of $time_{instr}$ of only 7 s some respondents scarcely looked at the scenario description and choice instructions at all. The fastest respondent only needed 45 s to complete all six choice tasks ($time_{choice}$) and the respective choice certainty questions, whereas the slowest respondent did this in 9.6 min. The distributions of times corresponding with different sections, relative to total response time ($ts_{sec1}$, $ts_{instr}$, $ts_{choice}$) are not skewed as their respective means and medians are almost equal. However, the Shapiro–Wilk-test rejects the hypothesis that these variables are normally distributed ($p < 0.001$). Respondents spent between about one tenth and two thirds of their total response time on the six choice tasks and the choice certainty follow-up questions.

## 4.2 Response Time and Attribute Attendance

To shed some light on the relationship between response time and cognitive effort, Fig. 2 displays median response times for the six choice tasks ($time_{choice}$) for respondents who reported different levels of attribute attendance upon completion of the stated choice section. Medians of response time grouped by different levels of self-reported attendance reveal that respondents who state that they always considered a particular attribute when making their choices spent more time on the choice tasks than respondents who did not consider that attribute or were not sure whether they did (Fig. 2). This pattern holds with respect to self-reported attendance to all four attributes. A series of non-parametric equality-of-median tests[3] show that median response times between the 'Yes' and 'No' categories of each attribute are significantly different ($\chi^2(1) = 15.95$, $p < 0.001$ for enhanced biodiversity, $\chi^2(1) = 6.93$, $p = 0.015$ for turbine height and visibility, $\chi^2(1) = 14.95$, $p < 0.001$ for impact of electromagnetic fields and $\chi^2(1) = 3.97$, $p = 0.046$ for the cost attribute). Differences between the 'No' and 'Not sure' categories are not found to be significant except for the turbine height attribute.

In a similar way, means of the share of time to complete the six choice tasks of overall response time ($ts_{choice}$) are compared across different answers to the attribute attendance questions (Fig. 3). Respondents who state that they have considered a particular attribute when making choices have used a higher portion of overall response time for the choice task. The figure also reveals that respondents who are not sure whether they considered an

[3] These are computed based on Pearson's $\chi^2$ test.

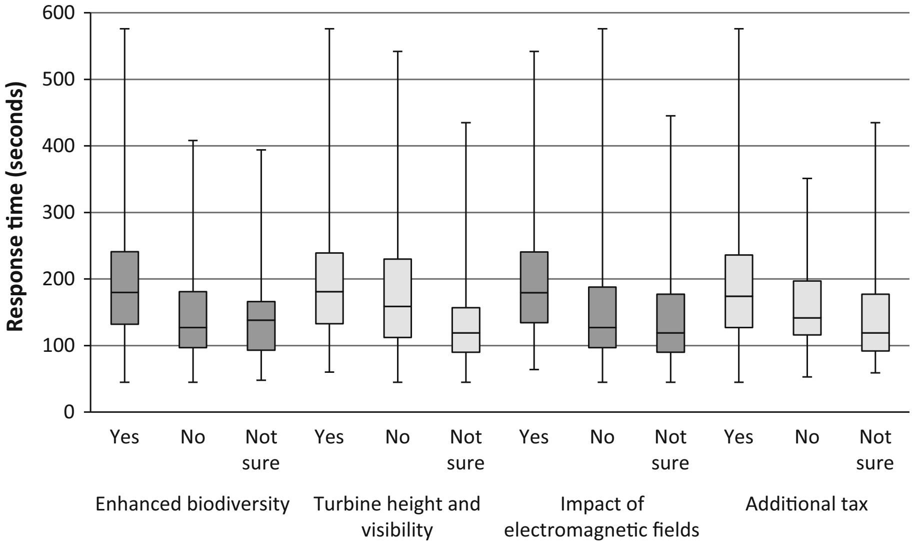


**Fig. 2** Median response times ($time_{choice}$—for all six choice tasks and the follow-up choice certainty questions) and self-reported attribute attendance. Response categories are for the question "Please indicate whether you considered each of the characteristics of the Rhiannon wind farm when choosing your preferred options"

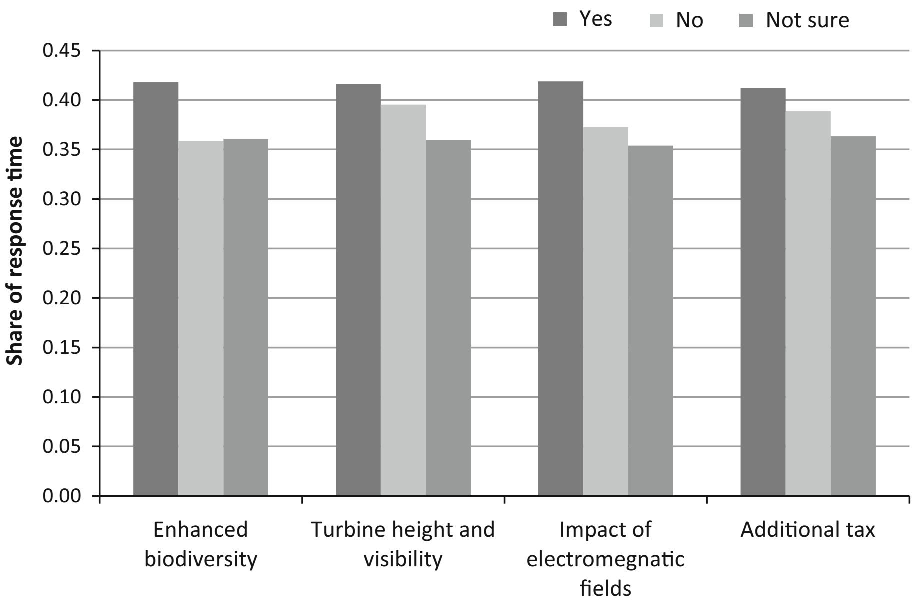


**Fig. 3** Mean share of choice task section of overall response times ($ts_{choice}$—for all six choice tasks and the follow-up choice certainty questions) and self-reported attribute attendance. Response categories are for the question "Please indicate whether you considered each of the characteristics of the Rhiannon wind farm when choosing your preferred options"

**Table 3** OLS regression of response times for the choice experiment section ($time_{choice}$)

| | Coefficient | Standard error |
|---|---|---|
| *Dependent variable* $ln(time_{choice})$ | | |
| CONSTANT | 5.083*** | (0.065) |
| MALE | −0.086** | (0.044) |
| AGE GROUP | | |
| 25–44 | 0.000 | (0.064) |
| 45–64 | 0.177** | (0.066) |
| 65 and over | 0.364*** | (0.085) |
| UNI | 0.007 | (0.045) |
| INCOME_k[a] | −0.002 | (0.010) |
| ENVORG | −0.019 | (0.058) |
| MARSEC | 0.076 | (0.116) |
| VICINITY | −0.078* | (0.044) |
| Observations | 463 | |
| $R^2$ | 0.080 | |

***, ** 1- and 5 %-level of confidence. [a] In £1000

attribute devote even less time to the choice task relative to overall time (with the exception of considering 'Enhanced biodiversity').

Applying the Mann-Whitney U-test for non-normally distributed samples shows that the means of the 'Yes' and 'No' groups are significantly different ($p < 0.001$). Together with Fig. 2 these results show that there is a positive relationship between the time a respondent needs to complete all six choice tasks and the (self-reported) level of attribute attendance. Even independent of absolute response time, those respondents who claim to have considered the choice attributes devoted a significantly greater share of the overall response time to the choice task section.

Since it is conceivable that response time is itself a function of certain respondent characteristics an ordinary least squares regression of the natural log of $time_{choice}$ is displayed in Table 3. Response time is significantly affected by the age of the respondent (AGE GROUP) and the gender of a respondent (MALE), i.e. male respondents complete the choice task faster. Respondents in higher age groups need more time to complete the choice tasks compared to the reference group between 18 and 24. Respondents living close to the Irish Sea (VICINITY) also exhibit shorter response time. All other tested variables, household income (INCOME_k), whether the respondent has a university degree (UNI), is member of an environmental organisation (ENVORG) or has previously worked in any marine sector (MARSEC) are insignificant. A non-parametric equality-of-median-test shows that median response time is lower for respondents who opt out in each choice task ($\chi^2(1) = 12.44$, $p < 0.001$). As could be expected, these respondents click through the choice experiment relatively more quickly. To test the robustness of the findings of the regression model in Table 3 an additional model excluding respondents who always opt out was run. The results of that model (n = 367) were the same.

Since it has been shown that self-reported and inferred attribute non-attendance do not always coincide (Kragt 2013), the ECLC model is used to test the influence of response time on attribute attendance as revealed by the choice data. The top section of Table 4 displays utility parameters in four classes where alternative-specific constants (ASC_CHANGE) can vary and the choice attributes are zero or constant across classes. Looking at class 1 projected increases of ten species (SPEC10) and 30 species (SPEC30) in the confines of the

**Table 4** Equality-constrained latent class (ECLC) model of attribute attendance patterns

| | Class 1 (attend all) | | Class 2 (attend only EMI) | | Class 3 (attend only COST) | | Class 4 (attend no attribute) | |
|---|---|---|---|---|---|---|---|---|
| | Coef. | SE | Coef. | SE | Coef. | SE | Coef. | SE |
| ASC_CHANGE | 1.926*** | (0.433) | 2.381*** | (0.680) | −1.185*** | (0.187) | −0.262 | (0.259) |
| SPEC10 | 0.805*** | (0.221) | 0.000 (fixed) | | 0.000 (fixed) | | 0.000 (fixed) | |
| SPEC30 | 1.125*** | (0.272) | 0.000 (fixed) | | 0.000 (fixed) | | 0.000 (fixed) | |
| HEI240 | −0.790*** | (0.276) | 0.000 (fixed) | | 0.000 (fixed) | | 0.000 (fixed) | |
| HEI300 | −0.384 | (0.243) | 0.000 (fixed) | | 0.000 (fixed) | | 0.000 (fixed) | |
| EMI | 1.254*** | (0.188) | 1.254*** | (0.188) | 0.000 (fixed) | | 0.000 (fixed) | |
| COST | −0.086*** | (0.011) | 0.000 (fixed) | | −0.086*** | (0.011) | 0.000 (fixed) | |
| *Class membership function* | | | | | | | | |
| Constant | −0.840 | (0.877) | −0.220 | (0.846) | −0.420 | (0.872) | 0.000 | |
| $time_{choice}$ | 0.010** | (0.004) | 0.009** | (0.004) | 0.008* | (0.004) | 0.000 | |
| MALE | −0.174 | (0.550) | −0.373 | (0.534) | −0.185 | (0.517) | 0.000 | |
| *AGEGROUP* | | | | | | | | |
| 25–44 | 0.503 | (0.745) | −0.087 | (0.711) | 0.707 | (0.720) | 0.000 | |
| 45–64 | 0.137 | (0.773) | −0.567 | (0.720) | 0.145 | (0.729) | 0.000 | |
| 65 and over | 0.392 | (1.105) | −0.315 | (1.101) | 0.918 | (1.045) | 0.000 | |
| VICINITY | 0.001 | (0.001) | 0.000 | (0.001) | 0.000 | (0.001) | 0.000 | |
| Class share | 0.297 | | 0.230 | | 0.364 | | 0.109 | |
| Log-likelihood | −1969 | | | | | | | |
| Observations | 8898 | | | | | | | |
| Adjusted $R^2$ | 0.393 | | | | | | | |
| BIC | 4186 | | | | | | | |

***, **, * 1-, 5- and 10 %-level of confidence windfarm have a positive effect on choices. Respondents clearly prefer an increase in species to no increase, and prefer a larger increase more strongly. Similarly, respondents have clear preferences for the prevention of any impact of the electromagnetic fields of cables as the coefficient of EMI is also significantly positive. The coefficient of the cost attribute (COST) is negative and significant as expected, which means that the higher the cost of a choice option the lower the probability that it is preferred, keeping all other attributes constant. Of the coefficients of turbine height and visibility only HEI240 is significantly different from zero. The level of 300 m high turbines (HEI300) does not affect choices on average.

The bottom half of Table 4 contains information on class membership. The positive coefficients of $time_{choice}$ for classes 1, 2 and 3 indicate that the higher response time the more likely it is for a respondent to be in any of these classes compared to reference class 4. Classes 1 to 3 are characterised by higher attribute attendance than class 4, which has no attendance to any attribute (i.e. all but the coefficient of ASC_CHANGE are assumed to be zero). Respondents taking longer for the choice experiment section of the questionnaire are therefore more likely to fall into classes with higher attribute attendance. In these classes respondents attend all attributes (class 1), only EMI (class 2) and only cost (class 3). Additional models with varying numbers of groups (3, 6, 10) were run as robustness checks which found response time consistently explained membership in classes with higher attribute attendance compared to a reference class without attendance to any attribute.

### 4.3 Assessing Determinants of Scale Heterogeneity and Choice Randomness

Turning now to the choice models involving scale heterogeneity, Table 5 displays the results of the conditional, mixed and generalised multinomial logit models. In the conditional logit model, coefficients are similar to those estimated in the ECLC model. Respondents prefer options with an increased number of species (SPEC10 and SPEC30) and the prevention of electromagnetic field impact on marine life (EMI). The cost coefficient is negative. However, none of the turbine height coefficients (HEI240 and HEI300) are significantly different from zero. This model finds a significantly positive effect on choices of $time_{choice}$ interacted with ASC_CHANGE. Higher response time leads to a higher likelihood of a respondent selecting any of the change options over the status quo. The level of self-reported certainty (CERTAINTY), however, does not influence preferences in any systematic way. The fit of the conditional logit model to the data is rather poor ($R^2 = 0.058$), but substantially improved in the mixed logit model. For this model means and standard deviations of random coefficients are reported. The significant effects of SPEC10, SPEC30, EMI and COST are also found in this model. In addition to that, this model finds the coefficient of 240 m high turbines (HEI240) to be significantly negative indicating that on average turbines of this height are not preferred by respondents compared to the status quo height of 180 m. The coefficient of 300 m high turbines (HEI300) remains insignificant. Estimates of the standard deviation of some attributes are significantly different from zero for three attribute coefficients (SPEC10, EMI and COST) and the change dummy. This is evidence of unobserved preference heterogeneity for these attributes. This model also finds that longer response time explains stronger preferences for any of the change options, but there is no effect of choice certainty.

The GMNL model in the third column confirms the signs and magnitudes of the attribute coefficients of the mixed logit model. It also finds significant standard deviations for the coefficients of almost all attributes. As in the previous models, response time is found to affect preferences whereas choice certainty does not. The estimate of scale variance $\tau$ is significantly different from zero, which means that the model finds some unobserved heterogeneity in scale.[4] However, bearing in mind the criticism in Hess and Rose (2012) that random preference and scale heterogeneity cannot be separately assessed in this modelling framework, the analysis will not dwell on this aspect any further. Whether or not there is unobserved preference or scale heterogeneity or both is irrelevant for this study. Instead we turn to analysing parameterised scale heterogeneity in the following. As shown in Table 5, the model parameter $\gamma$ is set to zero effectively resulting in the GMNL-II model characterised by a simplification of the attribute coefficient vector in Eq. (5) to $\beta_n = \sigma_n(\beta + \eta_n)$. An alternative model (not reported here) was run with an endogenous $\gamma$, which yielded both an estimate for $\gamma$ not significantly different from zero ($p = 0.403$) and slightly worse fit to the data (BIC of 4,023). Consequently, this model specification (i.e. $\gamma = 0$) is used for further GMNL models with explanatory variables of the scale parameter displayed in Table 6.

The focus of these models is the detection of systematic determinants of scale. In model 1, the coefficient of response time for the six choice tasks ($time_{choice}$) is significantly positive.[5] Longer response time thus leads to a larger individual scale parameter $\sigma_n$. This means that respondents who take longer to complete all six choices make their choices in a more determined way—determined by the attributes and their levels as displayed on the choice cards. Shorter response time, in turn, leads to responses which are more random, in which the deterministic part of utility is smaller compared to the error component. This effect is not linear as indicated by the significantly negative effect of squared response time ($time^2_{choice}$). Response time increases scale up to a certain point, but response time exceeding this turning point does in fact lead to lower scale. Given the coefficients of $time_{choice}$ and $time^2_{choice}$ in model 1 maximum scale is reached at a response time for the six choice tasks of 304 s (approximately 5.07 min). This shows that respondents taking too much time for the choice tasks state choices which are influenced more strongly by attribute levels. Introducing response time as a determinant of scale in this model leads to better model fit as compared to the basic GMNL model in Table 5, which conforms to results in Campbell et al. (2013).

The positive response time effect on scale persists when the effect of choice certainty (CERTAINTY) is also accounted for in model 2. Based on the coefficients of $time_{choice}$ and $time^2_{choice}$, a response time for the choice task section of 311 s (approximately 5.18 min) leads to maximal scale *ceteris paribus*. Choice certainty as measured as the sum of all six choice certainty questions following each choice increases scale, i.e. respondents who are more certain of their choices in total make choices which are influenced more strongly by the choice attributes and their levels.

In a third model in Table 6, choice certainty and the share of response time devoted to the choice task section ($ts_{choice}$) are included as covariates of scale. While the positive effect of certainty persists, the coefficient of $ts_{choice}$ is also significantly positive. That means that irrespective of actual response time the share of time needed to complete the choice tasks relative to total response time positively drives scale. Fit of this model to the data is similar to model 2.

Subsequently, additional models were run including alternative response time variables. The regression outputs are not displayed for the sake of brevity but relevant results are as follows. The time to complete the first section of the online questionnaire ($time_{sec1}$) does not affect scale, neither in simple nor squared form. The time to read the instructions to the choice tasks ($time_{instr}$), however, exhibits the same effects as $time_{choice}$: it positively

---

[4] This study follows advice in Fiebig et al. (2010) and only scales all attribute coefficients but not the alternative-specific constant ASC_CHANGE.

[5] This model is run after excluding all missing values of the choice certainty variable so that all models in Tables 5 and 6 are based on the same sample.

**Table 5** Conditional logit, mixed logit and generalised multinomial logit (GMNL) models

| | Conditional logit | | Mixed logit | | GMNL | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| *Mean of coefficients* | | | | | | |
| ASC_CHANGE | −1.128*** | (0.224) | −2.465* | (1.299) | −1.484 | (0.285) |
| SPEC10 | 0.376*** | (0.087) | 0.527*** | (0.135) | 0.639*** | (0.166) |
| SPEC30 | 0.477*** | (0.097) | 1.053*** | (0.171) | 1.206*** | (0.213) |
| HEI240 | −0.018 | (0.085) | −0.551*** | (0.148) | −0.614*** | (0.171) |
| HEI300 | 0.040 | (0.085) | −0.081 | (0.145) | −0.100 | (0.169) |
| EMI | 0.748*** | (0.070) | 1.544*** | (0.163) | 1.830*** | (0.231) |
| COST | −0.028*** | (0.002) | −0.081*** | (0.007) | −0.095*** | (0.011) |
| $time^{a}_{choice}$ | 0.003*** | (0.000) | 0.011*** | (0.003) | 0.010*** | (0.003) |
| CERTAINTY[a] | 0.000 | (0.007) | 0.039 | (0.047) | 0.014 | (0.053) |
| *Standard deviation of coefficients*[b] | | | | | | |
| ASC_CHANGE | | | 5.172*** | (0.389) | 5.389*** | (0.498) |
| SPEC10 | | | 0.429* | (0.375) | 0.680** | (0.338) |
| SPEC30 | | | 0.752*** | (0.303) | 1.039*** | (0.295) |
| HEI240 | | | 0.015 | (0.469) | −0.007 | (0.303) |
| HEI300 | | | 0.245 | (0.473) | −0.485 | (0.900) |
| EMI | | | 1.772*** | (0.184) | −1.972*** | (0.190) |
| COST | | | 0.071*** | (0.007) | 0.083*** | (0.009) |
| $\tau$ | | | | | 0.567*** | (0.160) |
| $\gamma$ | | | | | 0 fixed | |
| Log-likelihood | −2787 | | −1957 | | −1956 | |
| Observations | 8187 | | 8187 | | 8187 | |
| Halton draws | – | | 1000 | | 1000 | |
| Parameters | 7 | | 16 | | 17 | |
| Adjusted $R^2$ | 0.058 | | 0.335 | | 0.335 | |
| BIC | 5630 | | 4011 | | 4015 | |

***, **, * 1-, 5- and 10 %-level of confidence

[a] Interacted with ASC_CHANGE

[b] According to Gu et al. (2013) the sign of the standard deviation is irrelevant and should be interpreted as being positive. Adjusted $R^2$ is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where $LL_m$ and $LL_0$ are the log-likelihoods of the full model and the intercept-only model, respectively, and $k$ the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with $N$ denoting the number of respondents. The use of BIC was preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model

affects scale but wears off at some point as indicated by the significantly negative coefficient of $time_{instr}$ squared. It can therefore be concluded that the decisive driver of scale, when it comes to response time, is (actual and relative) time devoted to the choice experiment proper including both instructions and actual tasks. Taking a longer time to answer general and introductory questions about the survey topic does not make stated choices any more determined by the attribute levels.

To calculate WTP estimates all models in Tables 5 and 6 were rerun assuming a non-random cost parameter. Results in all choice models with respect to covariates of scale are

**Table 6** Generalised multinomial logit (GMNL) models

| | Model 1 | | Model 2 | | Model 3 | |
|---|---|---|---|---|---|---|
| | Coefficient | SE | Coefficient | SE | Coefficient | SE |
| *Mean of coefficients* | | | | | | |
| ASC_CHANGE | 0.462 | (0.350) | 0.338 | (0.360) | 0.462 | (0.337) |
| SPEC10 | 0.105** | (0.047) | 0.053* | (0.031) | 0.065** | (0.033) |
| SPEC30 | 0.173** | (0.076) | 0.085* | (0.050) | 0.112** | (0.056) |
| HEI240 | −0.075* | (0.040) | −0.037 | (0.025) | −0.051* | (0.029) |
| HEI300 | −0.001 | (0.024) | −0.002 | (0.012) | −0.007 | (0.015) |
| EMI | 0.259** | (0.109) | 0.128* | (0.073) | 0.171** | (0.081) |
| COST | −0.013** | (0.005) | −0.006* | (0.004) | −0.009** | (0.004) |
| *Standard deviation of coefficients* [a] | | | | | | |
| ASC_CHANGE | −5.372*** | (0.437) | 5.303*** | (0.459) | 5.247*** | (0.437) |
| SPEC10 | 0.044 | (0.073) | 0.008 | (0.022) | 0.058 | (0.036) |
| SPEC30 | −0.171** | (0.077) | 0.082* | (0.048) | −0.110** | (0.055) |
| HEI240 | 0.033 | (0.056) | 0.023 | (0.021) | −0.018 | (0.029) |
| HEI300 | 0.065 | (0.056) | −0.039 | (0.030) | −0.046 | (0.034) |
| EMI | −0.259** | (0.113) | 0.131* | (0.075) | 0.182** | (0.091) |
| COST | 0.012** | (0.005) | 0.006* | (0.003) | 0.008** | (0.004) |
| *Determinants of scale* | | | | | | |
| $time_{choice}$ | 0.016*** | (0.004) | 0.015*** | (0.004) | | |
| $time^2_{choice}$ | $-2.7e^{-5}$*** | (0.000) | $-2.5e^{-5}$*** | (0.000) | | |
| CERTAINTY | | | 0.034** | (0.015) | 0.034** | (0.014) |
| $ts_{choice}$ | | | | | 3.577*** | (0.763) |
| $\tau$ | −0.484*** | (0.145) | 0.488** | (0.223) | 0.331** | (0.155) |
| $\gamma$ | 0 fixed | | 0 fixed | | 0 fixed | |
| Log-likelihood | −1947 | | −1945 | | −1949 | |
| Observations | 8,187 | | 8,187 | | 8187 | |
| Halton draws | 1000 | | 1000 | | 1000 | |
| Parameters | 17 | | 18 | | 17 | |
| Adjusted $R^2$ | 0.338 | | 0.338 | | 0.338 | |
| BIC | 3997 | | 4001 | | 4002 | |

***, ** 1-, and 5 %-level of confidence

[a] According to Gu et al. (2013) the sign of the standard deviation is irrelevant and should be interpreted as being positive. Adjusted $R^2$ is computed as $R^2 = 1 - (LL_m - k)/LL_0$, where $LL_m$ and $LL_0$ are the log-likelihoods of the full model and the intercept-only model, respectively, and $k$ the number of parameters. Bayesian Information Criterion (BIC) is calculated as $BIC = -2LL_m + k \cdot \ln(N)$ with $N$ denoting the number of respondents. The use of BIC was preferred to Akaike Information Criterion because it imposes a stronger penalty on the inclusion of more parameters in the model

the same except for the coefficient of CERTAINTY, which is not significant. Regarding mean WTP, Table 7 shows little difference between the models. Most importantly, WTP is unchanged between the basic mixed logit and the GMNL models and those GMNL models controlling for the scale effect of response time.

Table 8 reports WTP estimates based on mixed logit for each half of the distributions of actual and relative response time. While there are some (but not significant) differences in

**Table 7** WTP estimates and 95 %-confidence intervals (CI) for all model specifications in Tables 5 and 6. Unit: £ GBP

| | Conditional Logit | | Mixed logit | | GMNL | |
|---|---|---|---|---|---|---|
| | WTP | 95 %-CI | WTP | 95 %-CI | WTP (£) | 95 %-CI |
| SPEC10 | 13.22 | [7.07 to 19.37] | 7.13 | [2.70 to 11.56] | 8.15 | [4.81 to 11.49] |
| SPEC30 | 16.78 | [9.91 to 23.65] | 15.87 | [10.25 to 21.50] | 13.63 | [9.58 to 17.67] |
| HEI240 | −0.64 | [−6.49 to 5.21] | −9.31 | [−13.75 to −4.86] | −8.91 | [−13.16 to −4.66] |
| HEI300 | 1.40 | [−4.56 to 7.35] | −3.45 | [−8.03 to 1.13] | −5.58 | [−10.21 to −0.95] |
| EMI | 26.32 | [21.22 to 31.42] | 25.44 | [20.34 to 30.53] | 23.90 | [19.75 to 28.06] |
| | GMNL1 | | GMNL2 | | GMNL3 | |
| | WTP | 95 %-CI | WTP | 95 %-CI | WTP (£) | 95 %-CI |
| SPEC10 | 8.85 | [5.30 to 12.40] | 9.05 | [5.64 to 12.47] | 8.98 | [5.79 to 12.18] |
| SPEC30 | 15.72 | [10.75 to 20.70] | 16.02 | [11.47 to 20.57] | 14.76 | [10.51 to 19.02] |
| HEI240 | −6.84 | [−11.16 to −2.52] | −8.04 | [−12.15 to −3.93] | −7.50 | [−11.14 to −3.86] |
| HEI300 | −1.10 | [−6.15 to 3.96] | −2.17 | [−7.05 to 2.72] | −3.56 | [−7.72 to 0.60] |
| EMI | 26.00 | [21.29 to 30.72] | 25.13 | [20.26 to 30.00] | 24.52 | [19.66 to 29.37] |

The mixed logit and all GMNL models assume a non-random cost parameter and use 1000 Halton draws for simulation. Confidence intervals are obtained using the bootstrap method in Krinsky and Robb (1986) with 1000 draws

WTP between respondents with below and above-median actual response time, splitting the sample at the median of relative response time does not have any effect on WTP estimates. What is evident, however, is the decreasing size of the confidence intervals with increasing response time. Since response time increases scale, error variance for respondents who take longer to complete the choice tasks is lower and hence WTP estimates are more precise.[6]

## 5 Discussion and Conclusions

This study investigated the link between attribute attendance and response time and the effect of the latter on the randomness of stated choices in an online choice experiment. Systematic determinants of response scale were detected in a survey dataset dealing with the valuation of ecological and amenity impacts of a proposed offshore windfarm, which are potentially remote and unfamiliar to most respondents. Results show that respondents who report stronger attribute attendance take longer to complete the choice tasks. Independent of total response time, these respondents are also found to devote significantly more time to the choice tasks relative to the time needed for the whole questionnaire. These results are supported by an equality-constrained latent class model which finds that higher actual response time explains higher attribute attendance as revealed by stated choices. These results are relevant for validating the assumption that longer response time means more cognitive effort spent on the choice task (Holmes et al. 1998). Respondents who take more time to state the choices process the information provided by the valuation scenarios and the choice cards more thoroughly than respondents who click through that section of the online questionnaire very quickly. This relationship constitutes the foundation of the main hypothesis of this study that response time, as a proxy for cognitive effort, is a factor of the level of determination of choices by the presented attribute levels.

The analysis further reveals that only respondent age (positively) and the facts that respondents are male and live close to the Irish Sea coast (negatively) influence response time. These findings partly confirm the results of Vista et al. (2009) who detect the same age effect, which is well established in more general survey methodology research (Gummer and Roßmann 2014). Apart from these influences, response time is not found to be related to any demographic characteristic. Consequently, the effects of response time on scale are in fact stemming from that variable *per se* rather than from some other (e.g. demographic) determinant of time.

The study finds a positive effect of response time on preferences for ecological and amenity impacts of an offshore windfarm. In conditional and mixed logit models that do not control for covariates of scale, respondents who take longer to complete the choice tasks are found to have stronger preferences for the change options. Estimates of mean WTP, however, are not systematically affected by response time. The main outcome of this study shows that greater (actual and relative) response time for the choice tasks increases scale (i.e. decreases error variance) and confirms earlier work by e.g. Haaijer et al. (2000), Rose and Black (2006) and Campbell et al. (2013). The positive time effect on scale means that confidence intervals of such estimates are smaller if respondents take longer to complete the choice tasks, which was demonstrated in Table 8. The response time effect is shown to persist after the influence of choice certainty is controlled for. There is further evidence that the effect of actual response time is non-linear, so that this positive effect on scale 'wears off' and turns

[6] Alternative models to estimate WTP for quartiles of the actual response time distribution not reported here show that the largest reduction in the size of the confidence intervals occurs between the first and second quartile of that variable. Evidently, imprecision in WTP estimation is most severe only among the fastest respondents.

**Table 8** WTP estimates for separate halves of the distributions of actual and relative response time based on mixed logit models. Unit: £ GBP

| | Actual response time ($time_{choice}$) | | | | Relative response time ($ts_{choice}$) | | | |
|---|---|---|---|---|---|---|---|---|
| | 1st half | | 2nd half | | 1st half | | 2nd half | |
| | WTP | 95 %-CI | WTP | 95 %-CI | WTP | 95 %-CI | WTP | 95 %-CI |
| SPEC10 | 3.27 | [−5.14 to 11.67] | 9.68 | [4.74 to 14.63] | 7.13 | [−2.16 to 16.41] | 7.14 | [2.41 to 11.86] |
| | | (16.81) | | (9.90) | | (18.57) | | (9.44) |
| SPEC30 | 13.83 | [3.31 to 24.35] | 15.67 | [9.04 to 22.30] | 17.04 | [5.99 to 28.08] | 14.89 | [8.65 to 21.12] |
| | | (21.04) | | (13.26) | | (22.09) | | (12.47) |
| HEI240 | −12.59 | [−20.96 to −4.22] | −6.46 | [−11.54 to −1.38] | −13.95 | [−22.66 to −5.24] | −6.87 | [−11.83 to −1.91] |
| | | (16.74) | | (10.16) | | (17.42) | | (9.93) |
| HEI300 | −7.09 | [−15.50 to 1.32] | −0.58 | [−5.88 to 4.73] | −3.38 | [−12.34 to 5.58] | −3.84 | [−8.98 to 1.30] |
| | | (16.83) | | (10.62) | | (17.92) | | (10.28) |
| EMI | 18.53 | [9.71 to 27.35] | 30.99 | [24.63 to 37.34] | 24.89 | [14.73to 35.05] | 26.02 | [20.45 to 31.59] |
| | | (17.64) | | (12.71) | | (20.33) | | (11.14) |
| Observations | 4092 | | 4095 | | 4026 | | 4161 | |
| Respondents | 259 | | 260 | | 259 | | 260 | |

Size of the confidence intervals in parentheses. All models assume non-random cost coefficients and use 1000 Halton draws for simulation. Confidence intervals are obtained using the bootstrap method in Krinsky and Robb (1986) with 1000 draws negative at some threshold. Choice responses of survey participants who take an extremely long time to complete this task become comparably more random. The positive effect of relative response time on scale is particularly interesting when one considers that age was found to increase actual response time. It confirms the conclusion that scale is driven by response time *per se*. While these response time effects can also be found for the webpages covering attribute description and the instructions of the choice tasks, response time for the first section of the questionnaire does not affect scale. Hence, it is in fact the cognitive effort for the choice experiment section that positively affects the level of determination of choices by the attributes.

Besides these results regarding response time, there is limited evidence of a positive effect of choice certainty on scale, supporting earlier findings (e.g. Beck et al. 2013). These authors employ the scaled multinomial logit (SMNL) model and identify self-rated choice certainty as a significant determinant of scale.

The ease with which response time data can be recorded in online surveys has been discussed and demonstrated in this study. However, this approach had no control over loading times for websites, which may have introduced bias. It might also be the case that the final sample may have already been cleaned from incomplete and other irregular responses. This might have included respondents who obviously rushed through the whole exercise as professional survey operators routinely check for speeding and discard such respondents.

It should be noted that this study did not intend to establish a cut-off point for scale in order to discard respondents whose responses are deemed too random as suggested by Bonsall and Lythgoe (2009) and who would therefore be labelled as lacking an "ability to choose" (Christie and Gibbons 2011). However, the analysis of the link between attribute non-attendance and response time can serve as a first step to establishing such a threshold. Future research could use the ECLC model to test what minimal response time is necessary in any particular survey for a respondent to be able to attend to all attributes. Such an attendance-based cut-off point would be much more defensible when excluding responses that are made very fast. Another route to specify cut-off points is to look at confidence intervals of WTP estimates. Variance of estimated WTP is driven by the relative magnitude of the random term of utility, which in turn is determined by scale (Czajkowski et al. 2014a). If for some (groups of) respondents scale is found to be so low that resulting confidence intervals of WTP estimates are too large to be meaningful, these responses could be discarded. The determination of cut-off points will also have to take into account survey mode and compare scale of responses given in more controlled (i.e. face-to-face or telephone) surveys. It will be critical for the GMNL approach to also detect any additional independent or interaction effects of respondent-specific variables on the scale parameter. Despite the ongoing debate about whether or not it is possible to separately assess unobserved preference and scale heterogeneity (Hess and Rose 2012), GMNL offers a framework to test across respondents for systematic shifts in the magnitude of coefficients compared to the error term. Response time is likely to be just one of such systematic influences on choice randomness. Familiarity of with the good to be valued (Czajkowski et al. 2015), level of education (Czajkowski et al. 2014b) or choice certainty (Beck et al. 2013) have been shown to be other potential factors. Beyond that, other variables evaluating the choice task from the perspective of the respondent, such as attribute attendance, protest attitudes or previous experience as a survey participant should be accounted for in such a model.

A limitation of the present study is that it does not account for experience in survey taking. The literature on the effects of professional respondents is still inconclusive (Hillygus et al. 2014). Yet, it is conceivable that more experience in completing surveys will decrease response time, but also increase cognitive ability as such respondents may be more familiar with typical question formats. Therefore, the effects of survey experience on scale (either directly or intermediated by response time) are not clear and are subject to future research.

Against the backdrop of the above results some of the criticism against online stated preference surveys appears to be justified. Respondents who rush through a choice experiment survey are shown to be less influenced by the specific choice attributes and their levels. This problem might be exacerbated in surveys employing pre-recruited panels of respondents owned by market research companies. The quality of the resulting data might suffer from experienced respondents who complete surveys on a regular basis and do not pay much attention to the specific survey topic. On the other hand it can be argued that analyses such as the above could be used to identify respondents who complete the survey too fast to give valid, i.e. informed, responses. While the internet offers a wide range of opportunities for survey-based environmental valuation, care has to be taken to avoid negative effects of speedy responses. Measures to avoid these detrimental effects include the better use of survey programming to require respondents to stay on one webpage (with one choice task, for instance) a minimum amount of time before being able to continue the survey. This could be done by not displaying the ‘next’ button until this minimum timespan has passed. Given the link between attribute non-attendance and short response time, it is also conceivable to display the options of each choice set with a small delay, so that respondents are prompted to assess all attributes of one option before comparing it to the next alternative. Such delaying procedures could even be extended into an online version of the time to think approach. This would involve letting respondents temporarily close the survey right after the valuation question or choice tasks and contacting them again the next day to collect their responses. The relevant literature has shown that this procedure leads to fewer preference errors and higher choice certainty (Whittington et al. 1992; Cook et al. 2007, 2012). The results of the present study also show where increased response time leads to less error variance. While it might be useful to induce respondents to take longer for the choice experiment proper (i.e. explanation and choice tasks), designing surveys with longer introduction sections including questions regarding the survey topic would have no effect on response scale. However, all these procedures might lead to increased fatigue or even respondent drop out and should therefore be carefully tested. Another less coercive way could be some form of cheap talk script that explicitly asks respondents not to rush through the questionnaire and take time to consider all options. Such cheap talk has been shown to mitigate the problem of hypothetical bias under certain circumstances (Loomis 2014) and might be able to reduce the occurrence of speedy responses in online choice experiments, as well.

**Acknowledgments** This research formed part of the programme of the UK Energy Research Centre and was supported by the UK Research Councils under Natural Environment Research Council award NE/G007748/1. The author would also like to thank Tara Hooper and Melanie Austen for contributing to survey design and implementation as well as Mikolaj Czajkowski, Nick Hanley and Danny Campbell for valuable advice regarding the estimation models. The comments and suggestions from two anonymous reviewers and the co-editor, Christian Vossler, on earlier versions of the paper are gratefully acknowledged.

## References

Alemu M, Mørkbak M, Olsen S, Jensen C (2013) Attending to the reasons for attribute non-attendance in choice experiments. Environ Resour Econ 54:333–359

Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, Dillman D, Frankel MR, Garland P, Groves RM, Kennedy C, Krosnick J, Lavrakas PJ, Lee S, Link M, Piekarski L, Rao K, Thomas RK, Zahs D (2010) Research synthesis: AAPOR report on online panels. Public Opin Q 74:711–781

Bech M, Kjaer T, Lauridsen J (2011) Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. Health Econ 20:273–286
Beck MJ, Rose JM, Hensher DA (2013) Consistently inconsistent: the role of certainty, acceptability and scale in choice. Transp Res Part E Logist Transp Rev 56:81–93
Bonsall P, Lythgoe B (2009) Factors affecting the amount of effort expended in responding to questions in behavioural choice experiments. J Choice Modell 2:216–236
Brouwer R, Dekker T, Rolfe J, Windle J (2010) Choice certainty and consistency in repeated choice experiments. Environ Resour Econ 46:93–109
Brown TC, Kingsley D, Peterson GL, Flores NE, Clarke A, Birjulin A (2008) Reliability of individual valuations of public and private goods: choice consistency, response time, and preference refinement. J Public Econ 92:1595–1606
Campbell D, Morkbak MR, Olsen SD (2013) How quick can you click? The role of response time in online stated choice experiments. Paper presented at Bioecon conference 2013, Cambridge
Celtic Array Limited (2012) Round 3 Irish Sea zone Rhiannon wind farm limited. Environmental impact assesment offshore scopign report
ChoiceMetrics (2012) Ngene 1.1.1 User manual and reference guide
Christie M, Gibbons J (2011) The effect of individual 'ability to choose' (scale heterogeneity) on the valuation of environmental goods. Ecol Econ 70:2250–2257
Cook J, Whittington D, Canh DG, Johnson FR, Nyamete A (2007) Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam. Econ Inq 45:100–114
Cook J, Jeuland M, Maskery B, Whittington D (2012) Giving stated preference respondents "time to think": results from four countries. Environ Resour Econ 51:473
Czajkowski M, Hanley N, LaRiviere J (2015) The effects of experience on preference uncertainty: theory and empirics for environmental public goods. Am J Agric Econ 97(1):333–351
Czajkowski M, Hanley N, LaRiviere J (2014a) Controlling for the effects of information in a public goods discrete choice model. Discussion papers in environmental economics 2014-04. University of St. Andrews, St. Andrews
Czajkowski M, Kądziela T, Hanley N (2014b) We want to sort! Assessing households' preferences for sorting waste. Resour Energy Econ 36:290–306
Dennis JM (2001) Are internet panels creating professional respondents? Mark Res 13:34–38
Fiebig DG, Keane MP, Louviere J, Wasi N (2010) The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. Mark Sci 29:393–421
Fleming CM, Bowden M (2009) Web-based surveys as an alternative to traditional mail methods. J Environ Manag 90:284–292
Glenk K, Martin-Ortega J, Pulido-Velazquez M, Potts J (2014) Inferring attribute non-attendance from discrete choice experiments: Implications for benefit transfer. Environ Resour Econ 1–24. doi:10.1007/s10640-014-9777-9
Gu Y, Hole AR, Knox S (2013) Estimating the generalized multinomial logit model in Stata. Stata J 13:382–397
Gummer T, Roßmann J (2014) Explaining interview duration in web surveys: a multilevel approach. Soc Sci Comput Rev. doi:10.1177/0894439314533479
Haaijer R, Kamakura W, Wedel M (2000) Response latencies in the analysis of conjoint choice experiments. J Mark Res 37:376–382
Hillygus DS, Jackson N, Young M (2014) Professional respondents in non-probability online panels. In: Callegaro M, Baker R, Bethlehem J, Göritz AS, Krosnick JA, Lavrakas PJ (eds) Online panel research: a data quality perspective. Wiley, Chichester, pp 219–237
Hess S, Rose J (2012) Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation 39:1225–1239
Hess S, Stathopoulos A (2013) Linking response quality to survey engagement: a combined random scale and latent variable approach. J Choice Model 7:1–12
Hole AR, Kolstad JR, Gyrd-Hansen D (2013) Inferred vs stated attribute non-attendance in choice experiments: a study of doctors' prescription behaviour. J Econ Behav Organ 96:21–31
Holmes T, Alger K, Zinkhan C, Mercer E (1998) The effect of response time on conjoint analysis estimates of rainforest protection values. J For Econ 4:7–28
Juutinen A, Svento R, Mitani Y, Mäntymaa E, Shojie Y, Siikamäki P (2012) Modeling observed and unobserved heterogeneity in choice experiments. Environ Econ 3:57–65
Kragt ME (2013a) Stated and inferred attribute attendance models: a comparison with environmental choice experiments. J Agric Econ 64:719–736
Kragt ME (2013b) The effects of changing cost vectors on choices and scale heterogeneity. Environ Resour Econ 54:201–221

Krinsky I, Robb AL (1986) On approximating the statistical properties of elasticities. Rev Econ Stat 68:715–719

Krueger AD, Parsons GR, Firestone J (2011) Valuing the visual disamenity of offshore wind power projects at varying distances from the shore: an application on the Delaware shoreline. Land Econ 87:268–283

Ladenburg J, Dubgaard A (2007) Willingness to pay for reduced visual disamenities from offshore wind farms in Denmark. Energy Policy 35:4059–4071

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

Lindhjem H, Navrud S (2011a) Are internet surveys an alternative to face-to-face interviews in contingent valuation? Ecol Econ 70:1628–1637

Lindhjem H, Navrud S (2011b) Using internet in stated preference surveys: a review and comparison of survey modes. Int Rev Environ Resour Econ 5:309–351

Loomis JB (2014) Strategies for overcoming hypothetical bias in stated preference surveys. J Agric Resour Econ 39:34–46

Louviere JJ, Carson RT, Ainslie A, Cameron TA, DeShazo JR, Hensher DA, Kohn R, Marley T, Street DJ (2002) Dissecting the random component of utility. Mark Lett 13:177–193

Louviere JJ, Hensher DA, Swait JD (2000) Stated choice models. Analysis and application. Cambridge University Press, Cambridge

Lundhede TH, Olsen SB, Jacobsen JB, Thorsen BJ (2009) Handling respondent uncertainty in choice experiments: evaluating recoding approaches against explicit modelling of uncertainty. J Choice Model 2:118–147

Malhotra N (2008) Completion time and response order effects in web surveys. Public Opin Q 72:914–934

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

Nielsen JS (2011) Use of the internet for willingness-to-pay surveys: a comparison of face-to-face and web-based interviews. Resour Energy Econ 33:119–129

Olsen SB (2009) Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. Environ Resour Econ 44:591–610

Revelt D, Train K (1998) Mixed logit with repeated choices: households' choices of appliances efficiency level. Rev Econ Stat 80:647–657

Rose J, Black I (2006) Means matter, but variance matter too: decomposing response latency influences on variance heterogeneity in stated preference experiments. Mark Lett 17:295–310

Scarpa R, Rose JM (2008) Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. Aust J Agric Resour Econ 52:253–282

Scarpa R, Gilbride TJ, Campbell D, Hensher DA (2009) Modelling attribute non-attendance in choice experiments for rural landscape valuation. Eur Rev Agric Econ 36:151–174

Scarpa R, Zanoli R, Bruschi V, Naspetti S (2012) Inferred and stated attribute non-attendance in food choice experiments. Am J Agric Econ 95:165–180

Schwappach DLB, Strasmann TJ (2006) "Quick and dirty numbers"?: The reliability of a stated-preference technique for the measurement of preferences for resource allocation. J Health Econ 25:432–448

Svedsäter H (2007) Ambivalent statements in contingent valuation studies: inclusive response formats and giving respondents time to think. Aust J Agric Resour Econ 51:91–107

Taylor PA, Nelson NN, Grandjean BD, Anatchkova B, Aadland D (2009) Mode effects and other potential biases in panel-based internet surveys: final report. WYSAC Technical report no. SRC-905, US EPA, Laramie

Vista AB, Rosenberger RS, Collins AR (2009) If you provide it, will they read it? Response time effects in a choice experiment. Can J Agric Econ/Revue canadienne d'agroeconomie 57:365–377

Whittington D, Smith VK, Okorafor A, Okore A, Liu JL, McPhail A (1992) Giving respondents time to think in contingent valuation studies: a developing country application. J Environ Econ Manag 22:205–225

Windle J, Rolfe J (2011) Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. Econ Anal Policy 41:83–97