Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys?

Liebe, U; Glenk, K; Oehlmann, M; Meyerhoff, J

Published in:
Journal of Choice Modelling

DOI:
10.1016/j.jocm.2015.02.002

First published: 01/01/2015

Document Version
Peer reviewed version

Does the Use of Mobile Devices (Tablets and Smartphones) Affect Survey Quality and Choice Behaviour in Web Surveys?

Abstract: Web surveys are becoming increasingly popular in survey research including stated preference surveys. Compared with face-to-face, telephone and mail surveys, web surveys may contain a different and new source of measurement error and bias: the type of device that respondents use to answer the survey questions. This is the first study that tests whether the use of mobile devices, tablets or smartphones, affects survey characteristics and stated preferences in a web-based choice experiment. The web survey on expanding renewable energy production in Germany was carried out with 3,182 respondents, of which 12 per cent used a mobile device. Propensity score matching is used to account for selection bias in the use of mobile devices for survey completion. We find that mobile device users spent more time than desktop/laptop users to answer the survey. Yet, desktop/laptop users and mobile device users do not differ in acquiescence tendency as an indicator of extreme response patterns. For mobile device users only, we find a negative correlation between screen size and interview length and a positive correlation between screen size and acquiescence tendency. In the choice experiment data, we do not find significant differences in the tendency to choose the status quo option and scale between both subsamples. However, some of the estimates of implicit prices differ, albeit not in a unidirectional fashion. Model results for mobile device users indicate a U-shaped relationship between error variance and screen size. Together, the results suggest that using mobile devices is not detrimental to survey quality.

Keywords: Acquiescence bias; choice experiment; mobile device; propensity score matching; renewable energy; sample selection bias; smartphone; survey format; survey quality

Highlights

- Propensity score matching accounts for selection bias in sample
- Effects of mobile device use: length (+); interruptions (+); acquiescence tendency (-)
- No differences in scale but differences in implicit prices (marginal WTP)
- Effects of screen size: length (-); acquiescence tendency (+); error variance (U-shaped)
- No indication that mobile device use is detrimental to survey quality

1 Introduction

Stated preference surveys are increasingly being conducted online, which can be attributed to increased internet penetration rates and the advantages online survey formats, or web surveys, offer over alternative formats (Dillman, et al. 2009, Manfreda and Vehovar 2008). Web surveys can be administered to large samples in a short period of time at a relatively low cost, and permit efficient and novel ways to convey information regarding the valuation context, for example using multi-media tools, and to efficiently control the survey flow. They also enable researchers to easily collect additional information on response conditions and behaviour (paradata) such as response times, which may be used to explain variation in choice behaviour (Rose and Black 2006; Campbell et al. 2012, 2013; Dellaert et al. 2012; Hess, S. and A. Stathopoulos 2013). Provided that the penetration of the internet and the availability of internet-based services will continue to increase, it is conceivable that web surveys will become the dominant survey format of the future.

Therefore, there is an interest in understanding how online stated preference surveys compare to other survey formats in terms of representativeness and response behaviour. Findings thus far are mixed. Compared to alternative survey formats, in terms of representativeness, web surveys may produce samples that are unbalanced towards male respondents, that are younger, more highly educated and have higher income (Kwak and Radler 2002, Lindhjem and Navrud 2011, Marta-Pedroso et al. 2007, Olsen, 2009). However, differences are study specific. In terms of response behaviour, Lindhjem and Navrud (2011), Nielsen (2011) and Marta-Pedroso, et al. (2007) find no significant differences in mean willingness to pay (WTP) in comparisons of web and face-to-face surveys applying the contingent valuation method. In a comparison of web and mail survey formats using choice experiments, both Olsen (2009) and Windle and Rolfe (2011) could not reject the hypothesis of equal WTP estimates. However, after controlling for sample frame and self-selection effects, Morrison et al. (2013) recently found that the web survey resulted in WTP estimates that were, on average, 30% lower than those derived via a mail survey.

This study differs from previous comparative studies of survey formats. Here, the focus is entirely on respondents to a web survey. In the context of preferences for expanding renewable energy production, we investigate whether there are differences in response behaviour in relation to the device used for completion. In particular, this is, to the best of our knowledge, the first time that the impact of the use of mobile devices (tablets and smartphones) is compared to using desktop computers and laptops (desktop and laptops) in completing a stated preference survey. The recent years have seen a rapid expansion of the use of internet-enabled mobile devices such as tablets and smartphones. If the internet is increasingly accessed via such devices, it can be expected that web surveys will also be increasingly be completed on tablets and smartphones (Stern et al. 2014). Research on the impacts of using mobile devices to complete surveys is still in its infancy (see, e.g., Callegaro 2010, Peytchev and Hill 2010, Buskirk and Andrus 2012, Millar and Dillman 2012 for notable exceptions). Peytchev and Hill (2010), for example, do not find differences regarding cognitive processing and use of pictures comparing mobile web surveys and other survey modes. However, they find that users of mobile phones are less likely to provide text input and show differences in response behaviour, if the survey questions extend beyond the initially visible screen. However, this study is limited by a small sample size of 92 respondents. Millar and Dillman (2012) conducted an experimental study with 600 undergraduate students, in which they tested whether the response rate of smartphone users increases, if respondents are explicitly encouraged to use the smartphone for answering the online questionnaire. This treatment group was compared to respondents, who were requested to take part in a web survey, and a third group that could choose between answering an online questionnaire or a paper copy of it. Millar and Dillman (2012) do not find that explicitly requesting to use the smartphone has an effect on the response rate.

In general, it is not easy to define and investigate the quality of survey responses (Lyberg and Biemer 2007). For example, it is not clear whether longer interviews are always ‘better’ and associated with higher quality such as greater response consistency (i.e., similar answers to similar questions). In this paper, we consider the respondents’ acquiescence tendency (Schaeffer and Presser 2003) as an indicator of survey quality. Acquiescence tendency is present if respondents agree in a survey regardless of the content of the survey questions. Such forms of extreme response patterns may bias survey results. On the one hand, acquiescence bias might be more likely when respondents complete a survey on a mobile device. The screen size of mobile devices is considerably smaller compared to most desktops and laptops, and, therefore, respondents may have to invest more time into reading the questions. This may lead to more extreme response patterns, if respondents get fatigued and want to complete the survey as quickly as possible. On the other hand, more time spent on reading the questions may result in greater deliberation, implying less extreme response patterns.

Similarly to effects on survey quality in general, the impact of using mobile devices on response behaviour in choice experiments in particular is difficult to predict, because it may depend on a large array of unobserved factors. For example, one may surmise that the use of mobile devices implies completing the survey while being mobile, for example during the daily commute to work. We would then expect that the survey is interrupted more frequently and that respondents are more distracted, which could result in a greater error variance compared to using desktop computers or a laptop. However, desktop/laptop users may equally be distracted. In the case of respondents using laptops, the circumstances may be similar, for example if laptops are used on a train or in a cafe. Regarding the use of stationary desktop computers, the use of different software and email that are competing for their attention, the radio or TV show playing in parallel or social interaction with family members or work colleagues may be examples of potential sources of distraction that could impact the accuracy of choices made.

Tablets and particularly smartphones typically have a smaller screen size compared to desktop and many laptop computers, which may require respondents to zoom in and out of choice cards frequently, or to scroll laterally to compare attribute information between different alternatives. Again, giving the apparent difficulty in accessing the whole information entailed in a choice task at once, one may conjecture that smaller screen size is associated with greater error variance. However, the difficulty in accessing the information on a mobile device may equally prompt respondents to expend more effort on taking up the information, and on making the decision, which may result in reduced error variance. Additionally, mobile device users may be very experienced in accessing information on their devices. Differences in WTP estimates may arise if respondents employ different decision rules (Hess et al. 2012), or the same rules to a differing degree. For example, it may be conceivable that non-attendance to attributes, or other decision heuristics such as always choosing the alternative in the same position within a choice set, differ between users and non-users of mobile devices. Similarly to error variance, we are not able to form any directional hypotheses regarding differences in preferences and estimates of WTP.

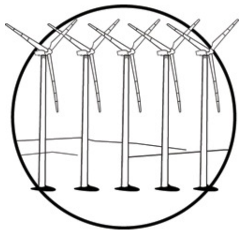
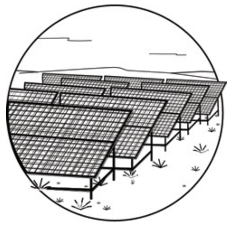
Against this backdrop, this study is largely exploratory in the sense that we test for differences in various observable survey characteristics such as interview length and acquiescence tendency as well as choice behaviour between users and non-users of mobile devices. The data was obtained in a web survey on renewable energy production in Germany, which included a choice experiment on externalities of the renewable energy production from wind, solar and biomass energy sources. Twelve per cent of the 3,182 respondents used a mobile device (tablet or smartphone) to answer the survey. We use propensity score matching to overcome selection bias and to make the subsamples of users and non-users of mobile devices comparable. In short, our findings indicate that survey quality tends to be higher for users of mobile devices compared with non-users, while choice consistency in the choice experiment does not significantly differ between subsamples. In the following, we first describe the study's background, the choice experiment, data collection and data. We proceed with presenting results regarding survey characteristics and of the choice models. The paper concludes with a discussion of our approach and findings.

2 Study and Data

Apart from the choice experiment on renewable energy expansion in Germany, the questionnaire comprises questions concerning respondents' exposure to renewables, attitudes, acceptance and fairness aspects regarding the expansion of renewable energies in Germany, and socio-demographics. Three renewable energy sources were considered: wind energy, solar energy, and biomass. At the beginning of the survey, respondents were shown pictograms and definitions of these renewables (see Table 1). It was also clarified that the survey focused on renewables in the open landscape and did not consider, for example, energy production in urban areas using solar panels.

Six focus groups in different towns spread over Germany were conducted in October 2012 to assess understanding and acceptance of the questionnaire. After discussing perceived advantages and disadvantages of renewable energies in Germany, participants completed an earlier version of the choice experiment and subsequently reported their views and impressions. Based on these comments, the choice experiment was revised. In particular, the number of choice sets and attributes was reduced. The revised questionnaire was tested in two pilot studies. The first study (N=74) was conducted with colleagues and the second (N=100) with members of the access panel provided by the survey organization that also carried out the main survey.

Table 1: Definition of renewable energy sources used in the survey

		
Wind energy refers to electricity generation with single wind turbines and wind farms exclusively onshore.	**Solar energy** refers exclusively to the production of electricity with photovoltaic systems in the open landscape, so-called solar fields.	**Biomass** refers to the production of biogas and its electricity and includes both the biogas plant as well as the cultivation of the required biomass (such as corn).

The choice experiment entails a choice among four labelled alternatives in each choice set. Three alternatives described future options for renewables expansion of wind energy, solar energy or biomass within 10-kilometers around their place of residence. The labelled renewable energy alternatives were introduced using the pictograms and definitions shown in Table 1. In addition, respondents could choose a future status quo alternative with zero cost to them. This alternative, which was labelled “no influence on renewable type”, detailed the most likely future renewables expansion scenario in the absence of any further policy intervention. The choice attributes are reported in Table 2. They relate to the minimum distance of the production sites to the edge of town (*Distance*), to the size of the production sites (*Area*) and their number (S*ite#*), to whether new high-voltage transmission lines will be built overhead or underground (*Grid*), and to the landscape area set aside for landscape protection (*Landscape*). The price attribute (*Price*) was a surcharge or rebate to the energy bill.

Figure 1 gives an example of a typical choice set. Respondents were requested to choose in each choice set their preferred alternative regarding the renewable energy future within a 10-kilometre radius of their place of residence, and their least preferred alternative[1]. As the choice refers to changes within a 10-kilometer radius from the place of residence, respondents living in big cities were asked to instead think about the landscape surrounding the city assuming that they might use the landscape for recreational purposes. The price attribute was described in terms of both monthly and annual changes in the energy bill.

[1] This study uses only the data on ‘best’ choice.

Table 2: Attributes and attribute levels

Attribute	**Alternative**	**Level**
Minimum distance to residential areas		300 / 600 / **900** / 1600 / 2500
Size of production site	Wind	small (5-10 turbines) / **medium** (18-25 turbines) / large (35-50 turbines)
	Solar	small (1-10 football fields) / **medium** (20-60 football fields) / large (100-150 football fields)
	Biomass	small (1-3 fermentation tanks) / **medium** (5-8 fermentation tanks) / large (15-25 fermentation tanks)
Number of production sites		1 / 2 / **3** / 4 / 5
Share of landscape not used for production (in %)		10 / 20 / **30** /40 / 50
Long-distance Transmission lines		**overhead** / underground
Cost in Euro (surcharge or rebate to energy bill)		-10(-120) / -5(-60) / +2(24) / +7(84) / +14(168) / +23(276)

Note: Levels of the future status quo alternative are written in bold.

In order to combine the attribute levels into choice sets, we generated a Bayesian efficient design with labelled alternatives using Ngene software. As the optimization criterion we used the C-error, which allows minimisation of the variance of the sum of the marginal WTP estimates (Scarpa and Rose 2008). The prior values were taken from model estimates based on data collected in the focus groups and in the pilot studies. The resulting design had 24 choice sets that were blocked into four blocks with 6 choice sets each. The order of appearance of choice sets was randomised. Additionally, the order of the first three non-status quo alternatives (left to right) was randomised across respondents; that is, the order of alternatives was the same for each respondent but differed across respondents.

The survey was optimized for the use with mobile devices: generally, this means that lists with response options are dynamically adjusted to the size of a mobile device. This allows users of mobile devices to more easily navigate through the survey. However, optimization has its limits concerning the display of larger survey components such as choice sets. While a choice set may be displayed in full on a mobile device screen, it may not be readable due to small screen sizes. This is more likely if the choice set is larger as in our case of sets with four alternatives and six attributes (see Figure 1). Therefore, some respondents using mobile devices will probably have had to zoom and move the choice sets to access all the information contained in the sets and also to tick the alternative they prefer[2]. This is illustrated by screen shots of a typical choice task as shown on an iPhone with a screen diagonal of 10.2cm (Appendix A).

	Electricity from wind	**Electricity from biomass**	**Electricity from solar**	**No influence on renewable type**
Minimum distance to town	600m	2500m	300m	900m
Size of production sites	large (35-50 turbines)	large (15-25 fermentation tanks)	small (1-10 football fields)	Medium
Number of production sites	4	5	5	3
Protection of landscape	20%	50%	10%	30%
Transmission lines	Underground	Underground	overhead	Overhead
Change in energy bill	+14€ (+168€)	-5€ (-60€)	+14 € (+168 €)	0 €
I choose				
…. best option	□	□	□	□
…. worst option	□	□	□	□

Figure 1: Example of choice set

The data resulted from a nationwide online survey that took place in September and October 2013. Participants were members of an access panel. A shopping voucher for an online mail order company amounting to 3.50 Euro was used as an incentive to complete the interview. In total 12,833 panel participants were invited to take part. Of these, 220 could not be admitted to the survey as quota restrictions were already fulfilled (a quota system for age and sex was applied), 4,027 persons took part in the survey, and 3,400 completed the questionnaire. After inspection of the data, 3,396 usable interviews remained. This corresponds to a response rate of 27.9% (standard RR1, AAPOR 2009).

[2] We cannot make a general claim that using mobile devices to complete the survey questions and the choice tasks required zooming and scrolling. However, we tested this with several mobile phone models and *always* found that there is a need to zoom and scroll. This was the case for only some of the tablets. Of course, the need for zooming depends on the font size that interviewees find suitable for reading the information and for completing the choice tasks.

The type of device used was determined through browser sniffing. When the respondent opens the link to the web survey, the browser used to perform this action sends a user agent string to the server hosting the web survey. The user agent string includes information relevant to optimising the display of information on the device used such as the browser used and the screen resolution. Additionally, the user agent string also contains information on the device-specific browser installation, which allows identification of the type of device used.

3 Propensity Score Matching and Econometric Approach

We first describe the original sample, the selection bias present in the original sample and the matching approach used to overcome selection bias. This is followed by an introduction of the econometric approach used to analyse the choice data.

3.1 Original Sample and Propensity Score Matching

The first two columns in Table 3 give an overview of the subsamples of respondents (desktop/laptop users and users of mobile devices, N=3,182) for which we have all information regarding the variables used in the matching procedure. The first column refers to the subsample of desktop/laptop users in the full sample, the second to the mobile-device users and the third to the matched sample of desktop/laptop users. In our study, 12% of the respondents used a mobile device to answer the web survey (378 respondents of overall 3,182 respondents), of which 53% (N=199) used a tablet and 47% (N=179) a smartphone. The type of tablets and smartphones most often used across the sample were iPad (36.5%) and iPhone (20.8%) as well as various Samsung phone models (15.7%) and Samsung tablet models (7.5%). Comparing the first and second column in Table 3, it can be seen that mobile device users are more likely to be female, younger, slightly lower educated and to have a higher income compared to those respondents who completed the survey on desktops/laptops.[3] It can further be seen that there are no remarkable differences regarding the size of the place of residence.

[3] Within the subsample of mobile device users we see further that, compared with tablet users, smartphone users are on average younger (31.60 versus 41.68 years), somewhat lower educated (13.40 versus 13.98 years of schooling) and have a lower income (1,858.59 versus 2,265,14 Euro). Based on Mann-Whitney U tests, these differences are statistically significant at the 1% level for age and income and at the 10% level for education. We do not find remarkable differences with respect to gender and place of residence (Chi2-tests).

Table 3: Overview on original and matched sample

	Unmatched sample		Matched sample		
	Desktop/laptop	Mobile device	Desktop/laptop		
	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max	Bias in %	P value of T-test
Gender (1=women)	0.45 (.50) 0/1	0.49 (.50) 0/1	0.48 (.50) 0/1	1.1	0.884
Age (years)	43.41 (14.12) 18/84	36.89 (12.20) 18/78	36.96 (12.20) 18/81	-0.5	0.945
Education (years)	13.81 (3.43) 7/18	13.72 (3.36) 8/18	13.62 (3.38) 8/18	2.8	0.698
Income (Euro)	1962.43 (929.73) 268.33/8485.28	2071.36 (960.05) 402.49/8485.28	2082.43 (936.45) 367.42/5656.85	-1.2	0.873
Small town/city	0.33 (0.47) 0/1	0.34 (0.47) 0/1	0.31 (0.46) 0/1	6.2	0.394
Large city	0.36 (0.48) 0/1	0.36 (0.48) 0/1	0.36 (0.48) 0/1	-1.1	0.880
N	2,804	378	378		

Note: Income means equivalised disposable income in Euro (monthly disposable household income divided by the square root of the number of persons living in the household). In terms of whether respondents are small town/city or large city dwellers, there is an additional category of living in a village/rural (the total of the three categories sums to one). We do not report a bias for this variable, because it is the reference variable in the logit model underlying the propensity score matching.

A logit model for use of a mobile device, which is presented in Table 4, shows that, controlling for all variables, the differences between the subsamples “desktop/laptop users” and “mobile-device users” are statistically significant with respect to respondents’ gender and education at the 10% level and age and income at the 1% level.

Table 4: Logit model for use of a mobile device

	Mobile device (1=yes, 0=no)
Gender (1=women)	0.21^{+} (1.90)
Age in years	-0.04*** (-9.21)
Education in years	-0.03^{+} (-1.86)
Income	0.0003*** (4.92)
Small town/city	0.08 (0.56)
Large city	-0.02 (-0.13)
Constant	-0.59^{+} (-1.92)
LL	-1,109.341
McFadden R^2	0.044
N	3,182

Note: z-values in parentheses; *** p<0.001, ** p<0.01, * p<0.05, $^{+}$ p<0.10. Village/rural area is the reference category for the effects of the variables small town/city and large city.

The logit model results indicate a selection bias regarding the use of a mobile device, and the socio-demographic variables such as gender, education and income may also affect stated choices on renewable energy extension. It is therefore not advised to directly compare desktop/laptop users with mobile device users. Instead, we use propensity score matching to create a control group of desktop/laptop users that is not prone to selection bias. In terms of causal analysis and with respect to matching, we are interested in the so-called treatment effect for the treated (i.e. mobile-device users, see Morgan and Winship 2007: 42): how would survey responses and stated choices look like if mobile device users had used a desktop computer or a laptop to answer the survey? The basic idea of propensity score matching is to estimate the probability of a treatment variable, here using a mobile device to complete the survey, as a function of respondents' characteristics such as, for example, gender, age, education and income (Rosenbaum and Rubin 1983, Morgan and Winship 2007). The aim is to select treatment and control cases that are equal regarding their predicted probabilities (i.e., their propensity score) and thereby correcting for confounding and selection bias in the data. We employed a nearest neighbour matching approach without replacement and conducted the matching using the Stata module psmatch2 (see Leuven and Sianesi 2003). Nearest-neighbour matching means that based on the propensity score the nearest control case (i.e., desktop/laptop user) is selected for each treatment case (i.e., mobile device user); in our case only one control case is matched with each treatment case, and the selected control case cannot be matched to another case (no replacement). It is important to stress that there is no single best matching approach. Despite the fact that we use an econometrically rather simple matching procedure, all of our results are robust when compared to using somewhat more complex procedures such as nearest-neighbour matching with replacement and caliper or coarsened exact matching (Blackwell et al. 2009, Iacus et al. 2012).

A comparison of the second and third columns in Table 3 reveals that the matched subsample of desktop/laptop users does not differ statistically from mobile device users. For the remainder of the paper, the matched data, which minimizes selection bias and confounding effects, will be used for analysis and referred to in the discussion unless explicitly stated otherwise. For completeness and comparison, we also report the results for comparisons of the mobile device user subsample with the unmatched sample of desktop/laptop users.

3.2 Econometric Approach

The modelling approach is based on the random utility theory, with a utility function U for respondent n and alternative i in choice task t characterised by price p and non-price attributes $\boldsymbol{x}$ of the experimental design, and a random error term ε:

$$U_{nit} = -\alpha_n{}' p_{nit} + \boldsymbol{\beta}_n{}' \boldsymbol{x}_{nit} + \varepsilon_{nit}, \tag{1}$$

where α and $\boldsymbol{\beta}$ are parameters to be estimated and ε is assumed to be identically and independently distributed (*iid*) and related to the choice probability with a Gumbel distributed error term.

Let the sequence of choices over T_n choice tasks for respondent n be $\boldsymbol{y}_n$, i.e. $\boldsymbol{y}_n$ = $\langle i_{n1}, i_{n2}, \dots, i_{nT_n} \rangle$. In a random parameter logit (RPL) model, heterogeneity across respondents is introduced by allowing α_n and $\boldsymbol{\beta}_n$ to deviate from the population means following a random distribution. In a RPL model, the unconditional choice probability of respondent n's sequence of choices is the integral of the logit formula over all possible values of $\boldsymbol{\eta}_{ni}$ weighed by the density of $\boldsymbol{\eta}_{ni}$:

$$\Pr(\boldsymbol{y}_n | \alpha_n, \boldsymbol{\beta}_n) = \int \prod_{t_1=1}^{T_n} \frac{\exp(-\alpha_n' p_{nit} + \boldsymbol{\beta}_n' \boldsymbol{x}_{nit})}{\sum_{j=1}^{J} \exp(-\alpha_n' p_{njt} + \boldsymbol{\beta}_n' \boldsymbol{x}_{njt})} f(\boldsymbol{\eta}_{ni} | \boldsymbol{\Omega}) d\boldsymbol{\eta}_{ni}, \tag{2}$$

where $f(\boldsymbol{\eta}_{ni}|\boldsymbol{\Omega})$ is the joint density of parameter vector for price and K non-price attributes $[\alpha_n, \beta_{n1}, \beta_{n2}, \dots, \beta_{nK}]$, $\boldsymbol{\eta}_{ni}$ is the vector comprised of the random parameters and $\boldsymbol{\Omega}$ denotes the parameters of these distributions (e.g. the mean and variance). This integral does not have a closed form and thus requires approximation through simulation (Train 2003).

All choice models are estimated in WTP space (Train and Weeks 2005; Scarpa et al. 2008), which allows the distributions of WTP to be estimated directly and hence avoids issues with calculating WTP as the ratio of two random distributions. To achieve this, the standard specification of the utility function in equation (1) can be written as follows:

$$U_{nit} = -\alpha_n' p_{nit} + (\alpha_n \boldsymbol{\omega}_n)' \boldsymbol{x}_{nit} + \varepsilon_{nit}, \tag{3}$$

where $\boldsymbol{\omega}_n = \boldsymbol{\beta}_n / \alpha_n$, that is, WTP for non-price attributes $\boldsymbol{x}$. Substituting (3) into (2) implies that $f(\boldsymbol{\eta}_{ni}|\boldsymbol{\Omega})$ now denotes the joint density of parameter vector for price and K WTP parameters $[\alpha_n, \omega_{n1}, \omega_{n2}, \dots, \omega_{nK}]$. The price attribute parameter α is assumed to follow a lognormal distribution, the WTP parameters $\boldsymbol{\omega}$ are assumed to follow a normal distribution. In all models the simulation of the log-likelihood is performed using 500 Halton draws.

The variance of the error term may differ between subgroups of respondents, in our case between users of mobile devices and respondents who used desktop/laptops to complete the survey. Relative differences in error variance can be identified by allowing scale to differ between subgroups. The unconditional choice probability then becomes:

$$\Pr(\boldsymbol{y}_n|\alpha_n, \boldsymbol{\omega}_n, \tau_n) = \int \prod_{t_1=1}^{T_n} \frac{\exp[\tau_n(-\alpha_n' p_{nit} + (\alpha_n \boldsymbol{\omega}_n)' \boldsymbol{x}_{nit})]}{\sum_{j=1}^{J} \exp[\tau_n(-\alpha_n' p_{njt} + (\alpha_n \boldsymbol{\omega}_n)' \boldsymbol{x}_{njt})]} f(\boldsymbol{\eta}_{ni}|\boldsymbol{\Omega}) d\boldsymbol{\eta}_{ni}, \qquad (4)$$

where $\tau_n = [\mu_{desk/lap}(1\text{-}I_{mob}) + \mu_{mob} I_{mob}] / \mu_{desk/lap}$ with $\mu_{desk/lap}$ and μ_{mob} being *relative* scale parameters for desktop/laptops and mobile device users, respectively, and I_{mob} is an indicator variable taking one if individual n used a mobile device to complete the survey, else zero. $\mu_{desk/lap}$ is set to one.

The error variance may also vary between individuals *within* the subgroup of mobile device users by respondent specific or observation specific characteristics. Because this paper investigates effects of mobile device usage in choice experiment surveys, we are interested in variation in error variance that can be related to device specific characteristics. Screen size might be one such characteristic that can affect choice consistency. We therefore specify a heteroskedastic logit model (Swait and Adamowicz 2001, DeShazo and Fermo 2002) that allows scale to vary between respondents n as a function of screen size, S_n; that is, $\mu_n = \mu_n(S_n)$, $\mu_n \geq 0$. This means that the error terms ε_n are independent but not identically distributed, and the estimated variances are $\sigma_n^2 = \pi^2/6\mu_n^2(S_n)$.

$\mu_n(S_n)$ can take any functional form that is appropriate. A quadratic specification as used by Swait and Adamowicz (2001) in the context of choice task complexity has appeal because of its simplicity[4]:

$$\mu_n(S_n) = \exp(\theta_1 S_n + \theta_2 S_n^2), \qquad (5)$$

with θ_1 and θ_2 being the parameters to be estimated. Assuming a quadratic relationship between screen size and scale (error variance) allows us to investigate if respondents who use small screens are less consistent in their choices (larger scale) up to a threshold of screen size, after which error variance increases (scale decreases), possibly because respondents require less effort in accessing the relevant information that characterises the alternatives. The exponential of the quadratic function ensures non-negativity of scale and excellent

[4] We tested several functional forms, including a linear specification and a Box-Cox specification. The quadratic specification was superior in terms of model fit.

convergence properties (DeShazo and Fermo, 2002). To facilitate the estimation, screen size S_n enters as a zero-centred variable, implying that at the sample mean, $\mu(\bar{S}) = \exp(0) = 1$. The unconditional choice probability of the model including heteroskedasticity in error variances is represented by:

$$\Pr(\boldsymbol{y}_n|\alpha_n, \boldsymbol{\omega}_n, \theta_1, \theta_2) \int \prod_{t_1=1}^{T_n} \frac{\exp[\mu_n(S_n)(-\alpha_n' p_{nit} + (\alpha_n \boldsymbol{\omega}_n)' \boldsymbol{x}_{nit})]}{\sum_{j=1}^{J} \exp[\mu_n(S_n)(-\alpha_n' p_{njt} + (\alpha_n \boldsymbol{\omega}_n)' \boldsymbol{x}_{njt})]} f(\boldsymbol{\eta}_{ni}|\boldsymbol{\Omega}) d\boldsymbol{\eta}_{ni}. \quad (6)$$

If both θ_1 and θ_2 are zero, equation (6) collapses to equation (2).

4 Results

This section first presents the results of comparing desktop/laptop users and mobile device users with respect to several survey characteristics. Some of these characteristics such as acquiescence tendency can serve as indicators of survey quality. We subsequently proceed with the results of the choice models investigating differences in stated preferences and WTP indicators between subsamples.

4.1 Group Comparison Regarding Survey Characteristics

Table 5 contains the descriptive statistics for several survey characteristics and response patterns for each subsample. Differences between subsamples are assessed through Mann-Whitney U tests and Chi2-tests. The mean screen size (diagonally measured) of mobile device users is 17.52. Additional analysis shows that the mean screen size of respondents' smartphones is 9.99 cm, while tablets used in the survey have a mean screen size of 24.51 cm. On average mobile device users tend to complete the survey somewhat later in the day (16.14 versus 15.46 hours past midnight) and this difference is statistically significant at the 1% level based on a Mann-Whitney U Test. A more detailed analysis reveals that most respondents, irrespective of using a desktop/laptop or mobile device, answer the survey in the evening (between 6pm and 11pm, 46%) or afternoon (between 12am and 5am, 33%), followed by morning (between 6am and 11am, 18%) and night (between 12pm and 5am, 3%). Mobile device users are more likely to answer the survey in the evening (51% versus 41%) or night (4% versus 2%) and less likely in the morning (13% versus 21%) and afternoon (31% versus 35%). All group differences, except the one for afternoon, are statistically significant (at the 1% level for morning and evening and the 10% level for night, based on Chi2-tests). Mobile device users are also more likely to interrupt the survey (8% versus 3%), and this difference is statistically significant at the 1% level. Further, mobile device users spent, on average, more time to answer the survey. The difference in mean interview length amounts to 14 minutes (45 minutes versus 31 minutes) and is statistically significant at the 1% level. In order to account for outliers and the large variance, Table 5 also reports mean interview length without the lowest and highest 5% in each subsample. While the difference in interview length between desktop/laptop and mobile device users decreases to 3 minutes, it is still highly statistically significant. The higher values for interview length might indicate a higher quality of responses, possibly because respondents read the questions and choice tasks more carefully. On the other hand, it might show that mobile device users require more time to answer the survey questions due to the smaller screen size and associated operations (zooming, scrolling, and accuracy of touch-screen entries). This tendency is confirmed if we compare, within the group of mobile device users, the interview length (excluding outliers) between tablet users and smartphone users (mean=27.88, sd=8.99 versus mean=29.93, sd=9.63, this difference is statistically significant at the 5% level based on a Mann-Whitney U Test) or directly calculate Spearman's rank correlation coefficient for screen size and interview length (excluding outliers, r=-0.100, p=0.071).

We calculated the respondents' acquiescence tendency as an indicator of response quality. The tendency to agree in a survey regardless of the content of the survey question is a well-known bias in survey research (Schaeffer and Presser 2003). It might have several causes including differences between respondents regarding cognitive skills; in our study differences in the acquiescence tendency between desktop/laptop and mobile-device users might be also interpreted as differences in respondents' effort to answer the survey question. That is, extreme response patterns such as always agreeing or disagreeing are more/less likely. Acquiescence tendency is calculated from responses to supporting questions on respondents' attitudes towards renewable energy expansion. For each respondent, we summed up the agreement answers (1=agree/completely agree) to eight questions with a four-point agreement scale and divided this sum by the number of items.[5] It follows that a value of 0 means that a respondent has never agreed (agree or completely agree) and a value of 1 that s/he has always agreed.

[5] The questions are available from the authors upon request.

Table 5: Overview on survey characteristics for the unmatched and matched sample

	Unmatched sample	Matched sample		
	Desktop/laptop	Mobile device	Desktop/laptop	Difference
	Mean (SD) Min/Max	Mean (SD) Min/Max	Mean (SD) Min/Max	significant
Screen size (diagonally measured) in cm		17.52 (7.42) 7.1/25.7		
Interview time of the day (in hours past midnight)	15.59 (4.57) 0/23	16.14 (5.20) 0/23	15.46 (4.66) 0/23	**
Interview interrupted (1=yes)	.05 0/1	.08 0/1	.03 0/1	***
Total interview length in minutes	32.10 (77.40) 5.17/2815.98	44.75 (103.67) 7.28/1390.0	31.35 (44.42) 6.07/697.98	***
Total interview length in minutes (without lowest/highest 5%)	26.79 (9.47) 14.03/59.98 N=2,462	28.84 (9.34) 14.03/55.80 N=343	25.91 (9.22) 14.18/58.97 N=320	***
Interview length of the choice-experiment part in minutes	5.65 (27.22) .37/1,366.97	7.27 (21.55) .80/284.83	5.19 (8.58) .77/104.67	***
Interview length of the choice-experiment part in minutes (without lowest/highest 5%)	4.37 (1.83) 1.78/10.20 N=2,493	4.67 (1.79) 1.82/10.15 N=347	4.21 (1.73) 1.80/9.87 N=334	***
Acquiescence tendency	.60 (.21) 0/1	.58 (.21) 0/1	.59 (.21) 0/1	n.s.
Share of status quo choices in %	10.69	10.32	10.80	n.s.
n	2,804	378	378	

Note: *** $p<0.001$, ** $p<0.01$, * $p<0.05$, $^{+}$ $p<0.10$. Significance levels for group comparison between the subsample without mobile and the subsamples with mobile are based on a Mann-Whitney U Test. Significance tests for the variable interview interrupted are based on a Chi2-Test.

Results show that mobile device users and desktop/laptop users have a similar acquiescence tendency (0.58 versus 0.59). The difference is not statistically significant. This result remains stable (0.56 versus 0.58) if we include an additional set of five survey questions into the analysis that do not directly measure agreement with a statement, but in a similar vein ask respondents to rate the fairness of considerations regarding the construction of new power plants (four-point response scales with very fair, rather fair, rather unfair, very unfair).[6] However, similar to the effects on interview length, differences in screen size might affect

[6] It is worth noting that if the unmatched subsample of desktop/laptop users had been used, the difference between mobile device users and desktop/laptop users regarding acquiescence tendency (0.58 versus 0.60) would have been found to be statistically significant at the 5% level.

acquiescence tendency. A smaller screen size may be assumed to result in a more careful reading, which in turn leads to less extreme response patterns. Within the sample of mobile device users, we tested this reasoning and indeed find a positive and statistically significant Spearman's rank correlation between screen size and acquiescence (r=0.094, p=0.075 for the eight agreement questions, and r=0.143, p=0.006 for the eight agreement plus five fairness questions). This also indicates that acquiescence bias is least likely for smartphone users (given the smaller screen size of smartphones) compared with tablet, desktop and laptop users.

4.2 Stated Choices Taking Device into Account

With regard to the choice experiment, mobile device users spent, on average, more time to answer the choice question (7 minutes versus 5 minutes, see Table 5). The difference in mean interview length is statistically significant at the 1% level, and the results are stable when we exclude the lowest and highest 5% in each subsample. This adds further evidence to the supposition that mobile device users require more time to answer the survey questions due to the smaller screen size and associated operations such as zooming, scrolling, etc. However, we do not find statistically significant differences in the mean interview length regarding the choice experiment part (excluding outliers) between tablet users and smartphone users. The tendency to choose the status quo or zero price alternative can be interpreted as an opt-out response (Kontoleon and Yabe 2003). Our data (Table 5) do not show significant differences in the share of future status quo choices between desktop/laptop users and smartphone users. The share is around 10% in each subsample.

Table 6 shows the results of the choice models. All models are statistically significant. All attribute coefficients carry the expected sign, and the alternative specific constants and most of the attribute parameters are significant at the 10% level or greater. Exceptions are renewable expansion via large areas (*Area_l*) and number of sites (*Site#*) in the mobile device subsample. Across all models, there is a tendency to move away from the future status quo. For reasons not explained by attributes, respondents of the desktop computer/laptop and mobile device subsamples prefer renewable energy expansion in their area using solar, wind and biogas over the future status quo (i.e., over not influencing the type of renewable energy and agreeing with the attribute values of the status quo alternative). Relative to medium sized areas assigned to renewable energy expansion, larger areas are associated with a disutility, while smaller areas are associated with a utility gain and therefore positive marginal WTP. A greater distance of sites dedicated for renewables, and sites being connected to the grid underground rather than above ground are preferred. Respondents also prefer to see renewables being spread to a greater number of sites. Finally, respondents have preferences for larger areas specifically assigned to landscape conservation. The standard deviation parameters are significant for several of the attributes, indicating the presence of significant heterogeneity in marginal WTP for most attributes, as well as for changes in the energy bill.

By visual inspection only, it is apparent that WTP related to the type of technology on which energy expansion focuses (i.e., WTP for the ASCs related to solar, wind and biomass) is similar for desktop/laptop and mobile device users, and that some of the estimated mean marginal WTP values for attributes differ between subsamples. WTP parameters for expansion via large production areas (Area_l) and for the number of renewable energy production sites (Site #) are not significantly different from zero in the for desktop/laptop subsample but not in the mobile device subsample. Additionally, a Poe et al. (2005) test confirms significant differences at the 5% level for one of the attributes: area set aside for landscape conservation (Landscape).

Table 6: Choice modelling results for comparisons of mobile device users and a matched subsample of desktop/laptop users

	Model 1 RPL WTP space desktop/laptop users				**Model 2** RPL WTP space mobile device users				**Model 3** RPL WTP space whole sample two scale groups				**Model 4** RPL WTP space mobile device users heteroskedastic in scale as a function of screen size			
	Mean		SD		Mean		SD		Mean		SD		Mean		SD	
ASCB	12.92	***	-		14.22	***	-		13.53	***	-		13.45	***	-	
	(3.34)				(4.84)				(2.61)				(4.81)			
ASCS	25.8	***	-		31.16	***	-		28.69	***	-		30.58	***	-	
	(4.06)				(5.29)				(3.06)				(5.2)			
ASCW	21.29	***	-		21.13	***	-		21.1	***	-		20.67	***	-	
	(3.88)				(5.2)				(2.80)				(4.87)			
Area_l	**-7.15**	***	2.78		**-2.85**		7.40		-5.14	***	4.79	*	-2.58		8.06	*
	(1.60)		(2.4)		(1.87)		(5.44)		(1.18)		(2.42)		(1.83)		(4.33)	
Area_s	3.56	**	8.92	***	5.72	***	8.17	*	4.65	***	9.32	***	5.79	***	8.28	**
	(1.29)		(2.03)		(1.37)		(4.66)		(0.92)		(2.22)		(1.57)		(4.05)	
Distance	4.12	***	0.18		3.64	***	1.66		3.85	***	0.58		3.57	***	1.07	
	(0.63)		(0.42)		(0.70)		(1.38)		(0.47)		(1.4)		(0.72)		(1.08)	
Grid	6.05	***	7.67	***	4.86	***	6.18	***	5.45	***	6.18	***	4.80	***	5.99	***
	(1.26)		(2.24)		(1.34)		(1.83)		(0.9)		(1.57)		(1.21)		(1.49)	
Landscape	**2.39**	***	1.36		**1.05**	***	0.13		1.73	***	0.5		1.05	***	0.51	
	(0.41)		(1.7)		(0.35)		(0.31)		(0.25)		(0.79)		(0.32)		(1.07)	
Site#	**1.49**	**	1.27		**0.71**		0.12		1.14	***	1.35		0.67		1.81	
	(0.55)		(2.82)		(0.59)		(1.95)		(0.38)		(2.07)		(0.57)		(1.32)	
Price	-2.98	***	0.67	***	-3	***	0.85	***	-3.03	***	0.75	***	-2.67	***	0.85	***
	(0.1)		(0.12)		(0.11)		(0.14)		(0.08)		(0.08)		(0.20)		(0.13)	
$\mu_{desk/lap}$	-		-		-		-		1 (fixed)		-		-		-	
μ_{mob}	-		-		-		-		1.09		-		-		-	
									(0.09)							
θ_1	-		-		-		-		-		-		-0.01		-	
													(0.01)			
θ_2	-		-		-		-		-		-		-0.57	*	-	
													(0.33)			
# of observations	2,268				2,268				4,536				2,268			
Rho2	0.17				0.18				0.17				0.18			
LogL	-2,623.25				-2,579.24				-5,220.56				-2,578.1			

Note: Standard errors in parentheses. *,**,*** significant at 10%,5%,1% level; in bold: differences in mean WTP between desktop/laptop and mobile device users significant at 5% level based on Poe et al. (2005) test; or not significantly different from zero in one of the subsamples**.** Some attribute levels were scaled before entering the analysis: the parameter for minimum distance to town (*Distance*) reflects WTP per 1,000 meters, the parameter for area set aside for landscape protection (*Landscape*) reflects WTP per 10%.

Table 7: Decision heuristics/rules for mobile device and desktop users

	Mobile device	Desktop/laptop
Always the same position of an alternative	N (%)	N (%)
First alternative (left to right)	16 (4.32)	11 (2.91)
Second alternative (left to right)	20 (5.29)	14 (3.70)
Third alternative (left to right)	12 (3.17)	14 (3.70)
Fourth alternative (left to right)	3 (0.79)	1 (0.26)
Always the same renewable energy source or status quo	N (%)	N (%)
Wind	7 (1.85)	12 (3.17)
Solar	36 (9.52)	20 (5.29)
Biomass	5 (1.32)	7 (1.85)
SQ (status quo)	3 (0.79)	1 (0.26)
Total	51 (13.49)	40 (10.58)
Lexicographic preferences for choice attributes	N (%)	N (%)
Maximal distance	2 (0.53)	1 (0.26)
Minimal number of sites	0 (0)	1 (0.26)
Minimal size	3 (0.79)	1 (0.26)
Maximal landscape	5 (1.32)	3 (0.79)
Grid underground	9 (2.38)	12 (3.17)
Minimal cost	22 (5.82)	17 (4.50)

Note: The position of the labelled alternatives “wind”, “solar” and “biomass” from left to right was randomized across choice sets.

One possible explanation for these differences could be that respondents’ use of information processing strategies or decision rules differs between subsamples. While a detailed empirical investigation of this possibility is beyond the scope of this paper, we descriptively compare the frequency of several decision heurists/rules in the mobile device and desktop/laptop subsamples. Table 7 summarizes the findings. We find an overall low proportion of respondents who use one of the following decision heuristics/rules: a) always choosing the alternative in the same position within a choice set; b) always choosing the same renewable energy source (labelled alternative) or status quo in a choice set; and c) lexicographic preferences for choice attributes expressed by always choosing according to the attribute level of a specific attribute that minimizes the negative external effects of power plants (e.g., maximal distance to the place of residence). We carried out Chi2-tests (but do not consider comparisons that are based on very few cases, between one and five observations). We find only one group comparison that shows statistically significant differences between mobile device and desktop/laptop users. The proportion of respondents who always choose the solar alternative in a choice set is significantly higher for mobile device users (Chi2= 4.937, p=0.026). This is reflected in a slightly higher WTP for renewable extension via solar energy

(ASCS in Table 6), which is, however, not significantly different from WTP for ASCS in the desktop/laptop subsample. Further, it can be seen in Table 7 that mobile device users have a slight tendency to choose the first or second alternative (from left to right) in a choice set more often than desktop/laptop users. This might well be due to the fact that they have to scroll more often when reading a choice set. Nevertheless, the difference is not statistically significant.

Model 3 in Table X is a pooled model of desktop/laptop and mobile device subsamples, allowing for differences in scale between respondents who completed the survey on a desktop/laptop and those who used mobile devices. Differences in scale between desktop/laptop and smartphone users are not statistically significant (p(1) = 0.36; t(1) = 0.91). Because overall there are significant differences in attribute WTP parameters and ASCs between mobile device and desktop/laptop subsamples ($\chi^2 = -2\ [\text{LogL}_{\text{Model3}} - (\text{LogL}_{\text{Model1}} - \text{LogL}_{\text{Model2}})] = 36.1$; d.f. 18; $\Pr(\chi^2 \leq 28.87) = 0.01$), the estimated scale parameter associated with mobile device users cannot strictly be interpreted as the difference in error variance between subsamples (Swait and Louviere 1993). Nevertheless, the results indicate that, on average, differences are unlikely to be large in magnitude and do not suggest that respondents using mobile devices are less consistent in their choices[7].

However, there may be differences in error variance *within* the mobile device subsample that are related to screen size. Using a likelihood ratio test, it becomes clear that model 4 represents no statistical improvement in model fit over model 2, and hence the hypothesis of H_0: $\theta_1 = \theta_2 = 0$ cannot be rejected[8]. However, it is still of interest to have a closer look at the parameter estimates of the exponentiated quadratic function of scale. θ_1 is not significantly different from zero, while θ_2 is significant at the 10% level[9]. Figure 2 illustrates the implied negative quadratic relationship between screen size (in cm) and error variance[10]. Up to a threshold value (17.45 cm), choice consistency increases as screen size increases. Beyond this point, larger screen size is associated with increasing error variance. Comparing error variance at the *mean* screen size for tablets (24.5 cm) and smartphones (10 cm), it can be easily seen that differences between the devices at sample means are small in magnitude.

---

[7] This result is consistent when comparing mobile device users with the unmatched sample of desktop/laptop users (Appendix B). In this case, the hypothesis of equal attribute parameters and ASCs cannot be rejected at the 1% level. The relative scale parameter of the mobile device subsample is insignificant and with 1.06 of a comparable magnitude to the one derived in model 4.

[8] It is worth noting that a heteroskedastic MNL model was found to outperform a basic MNL, while findings regarding θ_1 and θ_2 were very similar.

[9] Omitting the linear term θ_1 resulted in a decrease of the log likelihood function by 0.01 points with otherwise basically identical parameters including a significant mean estimate of -0.56 for θ_2.

[10] Error variance estimated as $\pi^2 / 6(\mu(S))^2$.

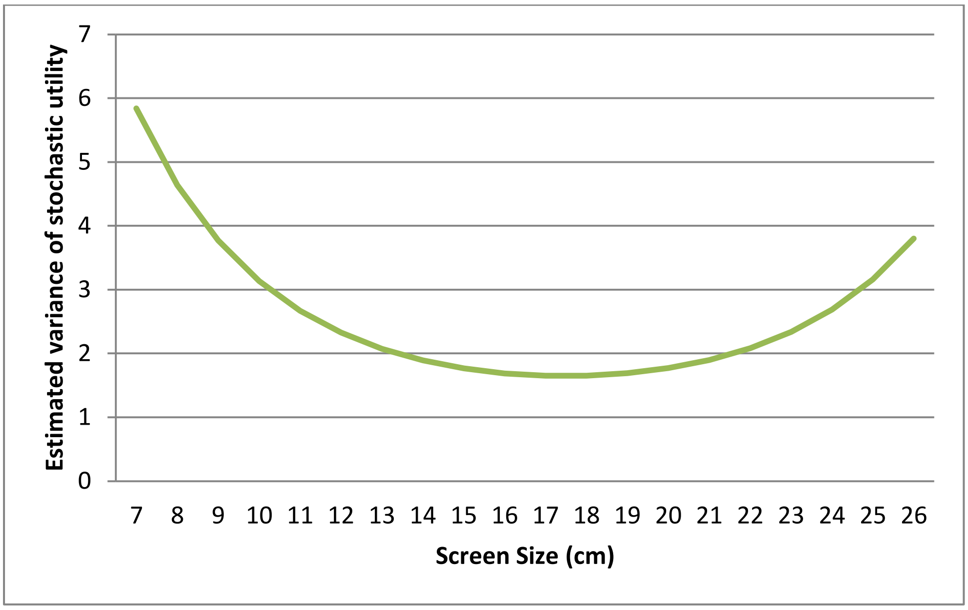


Figure 2: Estimated relationship between screen size and error variance

5 Discussion and Conclusions

To the best of our knowledge, this is the first study that tests whether the use of mobile devices affects survey characteristics and stated choices in a web-based choice experiment study. We find that some survey characteristics such as interview length are affected by the device used. Compared with respondents using a desktop/laptop, mobile device users spent more time to answer the survey. They are also slightly less likely to be prone to acquiescence, but this difference is not statistically significant. Another important finding is that within the subsample of mobile device users we see an association between screen size and interview length and acquiescence tendency, respectively. The smaller the screen size, the longer are the interviews and the *lower* is the acquiescence bias. These effects are small in magnitude, but they clearly indicate that surveys completed on smartphones are somewhat counter intuitively associated with higher survey quality.

It has to be stressed that it is important that any web survey is optimized for mobile device in order to guarantee similar visual experiences for users with and without mobile device (see, e.g., Burskirk and Andrus 2012 for approaches on how to implement smartphone surveys). In the present survey, the questionnaire has been optimized for mobile devices, and this might contribute to the similarity in terms of survey quality.

However, the usefulness of the optimization for smartphones may be limited for the choice experiment part of the survey, because the displayed choice sets may have been too small to comprehend without zooming and/or scrolling. This gives rise to investigating whether observable features of mobile devices such as screen size impact on choice behaviour. In our dataset, we cannot reject the hypothesis of equal scale by screen size. We think a main reason for this is a lack of variation given the limited sample size of mobile device users. Nonetheless, the estimated parameters of the scale function indicate a quadratic U-shaped relationship between error variance and screen size that can serve as a useful starting point for hypothesis testing in further research on this topic.

Differences in choice behaviour between desktop/laptop users and users of mobile devices can also arise from unobserved differences in accessing and processing the information contained in the choice task. We find no differences in status quo choice between the desktop/laptop and the mobile device subsamples and hence no evidence that mobile device users are more likely to exhibit a status quo bias. However, we do find differences in preferences. WTP parameters for three of the attributes differ significantly between subsamples. In this respect, it is interesting that differences are not unidirectional. WTP is larger for some of the attributes or their levels in the mobile device subsample, but smaller for others. This finding is difficult to explain. There might be characteristics additional to the ones used in the propensity score matching that affect the use of a mobile device to answer web surveys as well as responses to the choice experiment. Ideally, a sampling approach would include all relevant characteristics to ensure that we compare subsamples that have similar preferences for the environmental good at hand. However, it should be stressed that the variables included in the matching approach already cover a wide range of factors that can be expected to influence both the choice of device used for completing the survey as well as responses to the choice tasks. Further, the finding of differences in WTP for some of the attributes was consistent across variations of the matching procedure. One of the factors that has not been considered but could potentially play a role is the geographical distribution of respondents across Germany, given that there is spatial variation in the preponderance of certain renewable technologies with associated considerable differences of their impact on the landscape. This may, for example, contribute to differences between subsamples with respect to choosing the solar energy alternative. However, it is difficult to argue that there are differences in the likelihood of using a mobile device for answering the survey depending on where in Germany respondents live that are not captured by the variables already included in the matching procedure. Another potential source of differences in WTP may be related to differences in decision rules (Hess et al. 2012) and information processing strategies applied depending on the device used to complete the survey. For example, non-attendance to attributes might differ between desktop/laptop and mobile device users. An initial screening of the data for the incidence of several decision heuristics revealed only small differences between desktop/laptop and mobile device users. However, this finding should be scrutinised in future studies.

Given the large-scale nature of our survey (N=3,198) we have a large number of respondents using a mobile device in the data set (N=378 or 12% of all respondents). However, there is also much variation within the subsample of mobile device users (e.g., tablet users versus smartphone users). Therefore, an even larger sample of mobile device users is desirable to investigate the heterogeneity among mobile device users. Larger sample sizes could allow a better discrimination of effects between respondents using a smartphone versus a tablet, for example regarding the impact of screen size on error variance. At this point our results are indicative but point to the existence of intriguing effects.

In future studies investigating effects of mobile devices on responses in web surveys, the type of device should be taken into account in the sampling process. This will solve most of the problems mentioned above; however, it may be difficult to implement in choice experiment surveys where questions of survey methodology are often treated as an aside given limited research budgets. Further, respondents in our study were members of an access panel and, hence, they are experienced with answering web surveys. Differences between desktop/laptop and mobile device users might be larger if “inexperienced” respondents answer the survey. Future research on the effects of mobile device use on choice behaviour can also benefit from investigating the role of choice task complexity as indicated by choice task dimensions (i.e. number of attributes or alternatives in a choice set) or entropy as a summary measure of complexity (Swait and Adamowicz 2001). Differences between mobile device and desktop/laptop users might be less pronounced when choice task complexity decreases.

Notwithstanding limitations, this study is a first step in analysing effects of using mobile devices in web-based choice experiments and finds interesting differences between respondents using a desktop computer/laptop and mobile device. The study paves the way for more detailed studies investigating on the use of mobile devices in web surveys. Our study also adds evidence to the literature that demonstrates the usefulness of paradata to analyse the quality of survey responses (Yan and Olson 2013). Compared with other survey formats such as face-to-face and mail surveys, web surveys provide an easy way to collect paradata. There is a clear need for research that makes use of paradata to investigate sources of measurement errors with respect to survey-based experiments in general and choice experiments in particular.

References

Blackwell, Matthew, Stefano Iacus, Gary King, and Giuseppe Porro (2009). CEM: Coarsened Exact Matching in Stata. *The Stata Journal* 9: 524-546.

Buskirk, Trent D., and Charles Andrus (2012). Smart surveys for smartphone: exploringvarious approaches for conducting online mobile surveys via smartphones. *Survey Practice* (http://surveypractice.wordpress.com/2012/02/21/smart-surveys-for-smart-phones/).

Callegaro, Mario (2010). Do you know which device your respondent has used to take your online survey? *Survey Practice* (http://surveypractice.wordpress.com/2010/12/08/device-respondent-has-used/**)**.

Campbell, Danny, Morten Raun Mørkbak, and Soren Boye Olsen (2012). Response latency in stated choice experiments: Impact on preference, variance and processing heterogeneity. *Paper presented at the 19th Annual Conference of the European Association of Environmental and Resource Economists, Prague, 27 - 30 June 2012*.

Campbell, Danny, Morten Raun Mørkbak, and Soren Boye Olsen (2013). How quick can you click? Accommodating ‘quick’ responses to online stated choice questions. *Paper presented at the 15th Annual BIOECON Conference, 18-20 September 2013, Kings College, Cambridge United Kingdom*.

Dellaert, Benedict G.C., Bas Donkers, and Arthur van Soest (2012). Complexity Effects in Choice Experiment-Based Models. *Journal of Marketing Research* 49(3): 424-434.

DeShazo, J., and German Fermo (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management* 44: 123-143.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian (2009). *Internet, Mail and Mixed-Mode Surveys: The Tailored Design Method, 3rd edition*. John Wiley: Hoboken, NJ.

Hess, Stephane, and Amanda Stathopoulos (2013). Linking response quality to survey engagement: A combined random scale and latent variable approach. *Journal of Choice Modelling*, 7, 1-12.

Hess, Stephane, Stathopoulos, Amanda, and Daly, Andrew J. (2012). Allowing for heterogeneous decision rules in discrete choice models: an approach and four case studies. *Transportation*, 39, 565-591.

Iacus, Stefano M, Gary King, and Giuseppe Porro (2012). Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20(1): 1-24.

Kontoleon, Andreas, and Mitsuyasu Yabe (2003). Assessing the Impacts of Alternative ‘Opt-out’ Formats in Choice Experiment Studies. *Journal of Agricultural Policy and Resources* 5: 1-43.

Kwak, Nojin, and Barry Radler (2002). A comparison between mail and web surveys: Response pattern, respondent profile, and data quality. *Journal of Official Statistics* 18(2): 257-274.

Leuven, Edwin, and Barbara Sianesi (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Version 4.0.10 10 February 2014 (http://ideas.repec.org/c/boc/bocode/s432001.html).

Lindhjem, Henrik, and Ståle Navrud (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics* 70(9): 1628-1637.

Lyberg, Lars E., and Paul P. Biemer (2007). Quality Assurance and Quality Control in Surveys, pp. 421-441, in: Edith D.de Leeuw, Joop J. Hox, and Don A.Dillman (Eds.), *International Handbook of Survey Methodology*. Mahwah, NJ: Lawrence Erlbaum Associates / Psychology Press.

Manfreda, Katja L., and VasjaVehovar (2007). Internet Surveys, pp. 264-283, in: Edith D.de Leeuw, Joop J. Hox, and Don A. Dillman (Eds.), *International Handbook of Survey Methodology* . Mahwah, NJ: Lawrence Erlbaum Associates / Psychology Press.

Marta-Pedroso, Cristina, Helena Freitas, and Tiago Domingos (2007). Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. *Ecological Economics* 62(3): 388-398.

Millar, Morgan M., and Don A. Dillman (2012). Encouraging Survey Response via Smartphones: Effects on Respondents' Use of Mobile Devices and Survey Response Rates. *Survey Practice* 5 (No. 4).

Morgan, Stephen L., and Christopher Winship (2007). *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.

Morrison, Mark, MacDonald, Darla H., Boyle, Kevin, Rose, John, and Roderick Duncan (2013). Investigating Differences between Internet and Mail Implementation of a Stated-Preference Study While Controlling for Differences in Sample Frames and Self-Selection Effects. *Paper presented at the International Choice Modelling Conference 2013*.

Nielsen, Jytte S. (2011). Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews. *Resource and Energy Economics* 33(1): 119-129.

Olsen, Soren Boye (2009). Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and Resource Economics* 44(4): 591-610.

Peytchev, Andy, and Craig A. Hill (2010). Experiments in mobile web survey design: Similarities to other modes and unique considerations. *Social Science Computer Review* 28: 319-335.

Poe, Gregory L., Kelly L. Giraud, and John B. Loomis (2005). Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87: 353-365.

Rose, John M., and Ian R. Black (2006). Means matter, but variance matter too: Decomposing response latency influences on variance heterogeneity in stated preference experiments. *Marketing Letters*, 17, 295-310.

Rosenbaum, Paul R., and Donald B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41-55.

Scarpa, Riccardo, Thiene, Mara, and Kenneth Train (2008). Utility in willingness to pay space: A tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics* 90: 94-1010.

Schaeffer, Nora C., and Stanley Presser (2003). The Science of Asking Questions. *Annual Review of Sociology* 29: 65-88.

Stern, Michael J., Ipek Bilgen, and Don A. Dillman (2014). The State of Survey Methodology: Challenges, Dilemmas, and New Frontiers in the Era of the Tailored Design. *Field Methods* 1-18 (online first).

Swait, Joffre, and Jordan Louviere (1993). The Role of the scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research* 30: 305-314.

Swait, Joffre, and Wiktor Adamowicz (2001). Choice environment, market complexity and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes* 86: 141-167.

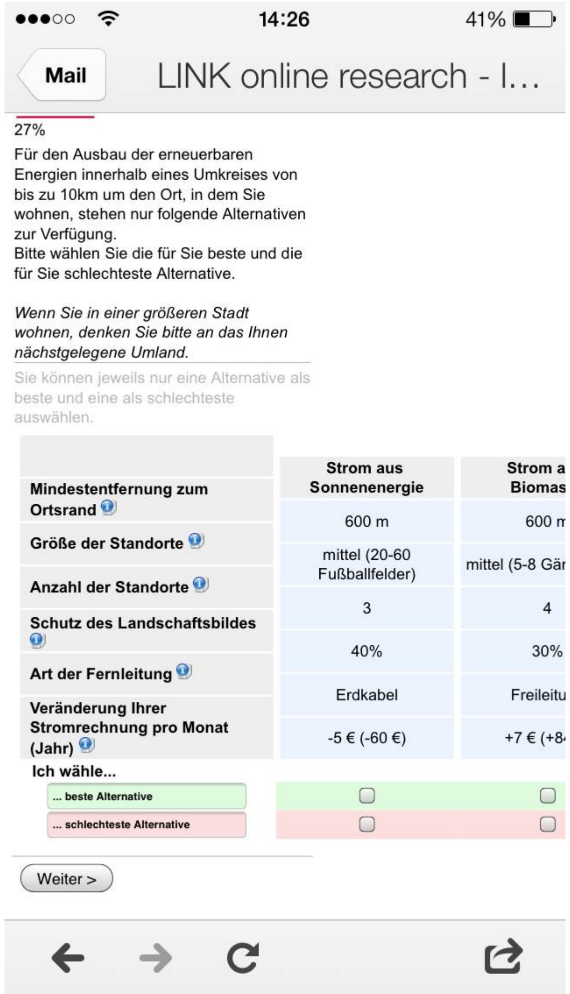
Train, Kenneth (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press

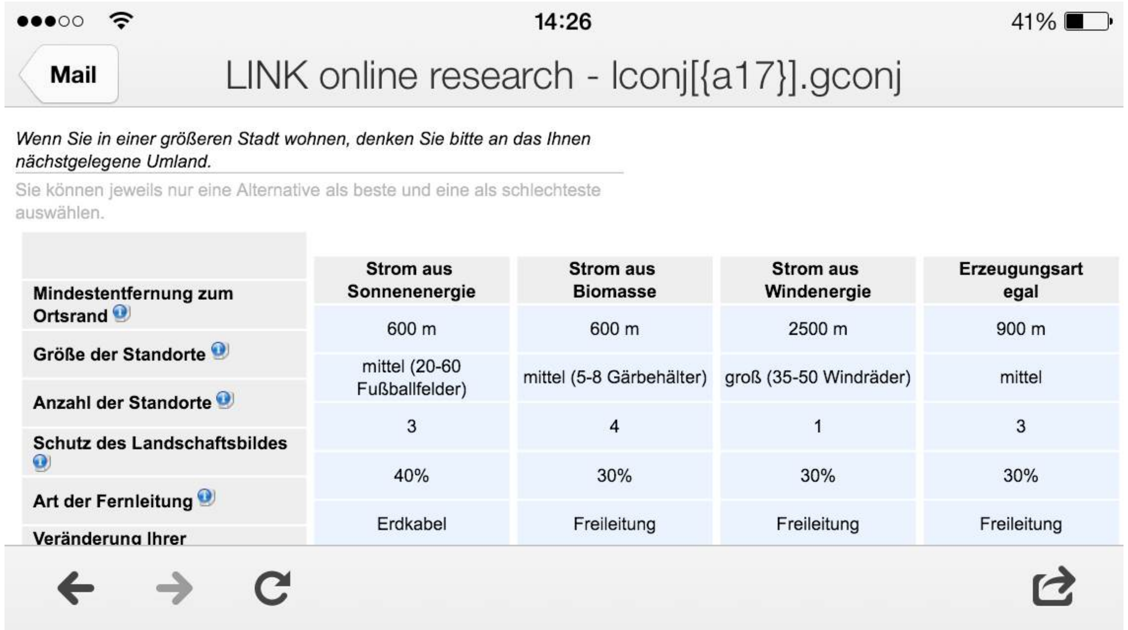
Train, Kenneth E., and Melvyn Weeks (2005). Discrete choice models in preference space and willingness-to-pay space, pp. 1-16, in: Riccardo Scarpa and Anna Alberini (Eds.), *Application of simulation methods in environmental and resource economics*. Dordrecht: Springer.

Windle, Jill, and John Rolfe (2011). Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. *Economic Analysis and Policy* 41(1): 83-97.

Yan, Ting, and Kristen Olson (2013). Analyzing Paradata to Investigate Measurement Error, pp. 73–95, in: Frauke Kreuter, *Improving Surveys with Paradata: Analytic Uses of Process Information*. Hoboken, NJ: John Wiley & Sons.

Appendix A: Two screen shots (upright format and landscape format) of a typical choice task as shown on an iPhone with a screen diagonal of 10.2cm





Appendix B: Choice modelling results for comparisons of mobile device users and the unmatched sample of desktop/laptop users

	Model A1				**Model A2**				**Model A3**			
	RPL WTP space desktop/laptop users				RPL WTP space mobile device users				RPL WTP space whole sample two scale groups			
	Mean		SD		Mean		SD		Mean		SD	
ASCB	16.22	***	-		14.22	***	-		15.62	***	-	
	(1.73)				(4.84)				(1.63)			
ASCS	34.72	***	-		31.16	***	-		33.98	***	-	
	(2.07)				(5.29)				(1.92)			
ASCW	24.93	***	-		21.13	***	-		24.19	***	-	
	(1.81)				(5.2)				(1.67)			
Area_l	**-6.41**	***	3.34		**-2.85**		7.40	**	-5.85	***	5.12	***
	(0.63)		(2.21)		(1.87)		(5.44)		(0.60)		(1.51)	
Area_s	4.82	***	12.59	***	5.72	***	8.17	***	4.87	***	12.46	***
	(0.55)		(1.05)		(1.37)		(4.66)		(0.50)		(1.05)	
Distance	4.4	***	0.8		3.64	***	1.66		4.31	***	1.09	
	(0.28)		(1.08)		(0.70)		(1.38)		(0.26)		(0.69)	
Grid	**7.39**	***	7.76	***	**4.86**	***	6.18	***	7.09	***	7.56	***
	(0.51)		(0.69)		(1.34)		(1.83)		(0.47)		(0.62)	
Landscape	**1.83**	***	1.62	**	**1.05**	***	0.13		1.73	***	1.79	***
	(0.15)		(0.65)		(0.35)		(0.31)		(0.14)		(0.62)	
Site#	**1.47**	***	0.36		**0.71**		0.12		1.3	***	0.80	
	(0.22)		(0.24)		(0.59)		(1.95)		(0.20)		(0.50)	
Price	-3.14	***	0.85	***	-3	***	0.85	***	-3.12	***	0.85	***
	(0.04)		(0.04)		(0.11)		(0.14)		(0.04)		(0.04)	
$\mu_{desk/lap}$	-		-		-		-		1		-	
									(fixed)			
μ_{mob}	-		-		-		-		1.06		-	
									(0.07)			
# of observations	16,824				2,268				19,092			
Rho2	0.17				0.18				0.17			
LogL	-19,453.38				-2,579.76				-22,037.73			

Note: Standard errors in parentheses. *,**,*** significant at 10%,5%,1% level; in bold: differences in mean WTP between desktop/laptop and mobile device users significant at 5% level based on Poe et al. (2005) test; or not significantly different from zero in one of the subsamples. Some attribute levels were scaled before entering the analysis: the parameter for minimum distance to town (*Distance*) reflects WTP per 1,000 meters, the parameter for area set aside for landscape protection (*Landscape*) reflects WTP per 10%.