



Ponderous, Proficient or Professional? Survey Experience and Smartphone Effects in Stated Preference Research

Erlend Dancke Sandorf¹ · Kristine Grimsrud² · Henrik Lindhjem^{3,4}

Accepted: 30 January 2022 / Published online: 9 March 2022
© The Author(s) 2022

Abstract

Stated preference surveys are increasingly conducted online using internet panel samples, where a fast-growing share of respondents answer on smartphones. These panel members range from novices to “professionals” in terms of previous survey experience. Little is known about the potential effects of smartphone responding and survey experience on the data quality of stated preference surveys. This paper uses a discrete choice experiment dataset on the Norwegian population’s willingness to pay to plant climate forests to explore how these two factors affect data quality. These data by type of response device, gathered using a probability-based internet panel, were combined with a unique dataset obtained from the survey company on respondents’ actual experience answering surveys on different types of devices. Our results show that differences in elicited preferences between smartphone and computer respondents are not caused by the device used, suggesting that initial concerns about smartphone responses may be exaggerated. Furthermore, more experience is associated with an increasing scale parameter (indicating lower error variance), but at a decreasing rate; and a higher propensity to choose the status quo (indicating possible simplifying strategies employed by respondents). Combined this suggest some optimal level of experience that is neither too high nor too low. We discuss the implications of our results for stated preference research and provide a few avenues for future research.

Keywords Discrete choice experiments · Internet panels · Experienced respondents · Latent class · Scale · Device effects

✉ Erlend Dancke Sandorf
erlend.dancke.sandorf@nmbu.no

¹ School of Economics and Business, Norwegian University of Life Sciences, Ås, Norway

² Research Department, Statistics Norway, Oslo, Norway

³ Menon Centre for Environmental and Resource Economics, Oslo, Norway

⁴ Norwegian Institute for Nature Research (NINA), Trondheim, Norway

1 Introduction

Research in environmental economics, and economics in general, has over the last decades become more empirical and utilise population survey data to a larger extent than in the past (Angrist et al. 2017; Kube et al. 2018). To gain useful information from surveys, more attention must be given to efforts to ensure the quality of the data (Biemer et al. 2017). While stated preference (SP) research in environmental economics traditionally has been more attentive to survey data quality than other fields of economics, not least due to the methodological discussions spurred by the NOAA panel in the early 1990s, there are two striking trends in survey research that could potentially be important and warrant more research: The increasing use of internet panels for SP surveys and the growing level of survey experience among panel members, and the rising use of smartphones to answer surveys among these respondents.

Initially, SP research focused on so-called survey mode effects and considered both measurement and sample composition, and representation effects. An early example of this effort is the comparison between interviews conducted using mail-out and (landline) telephone samples by Mannesto and Loomis (1991). From the late 2000s, the focus shifted to potential effects of conducting interviews online using samples drawn from internet panels of willing respondents recruited randomly or by opt-in and maintained by survey companies. Results from these interviews were then compared to results from more traditional methods and samples (survey modes) (Boyle et al. 2016; Lindhjem and Navrud 2011a, b; Olsen 2009; Sandorf et al. 2016, 2020). To our best knowledge, most SP research in middle- and high-income countries is currently undertaken using such internet panels due to the low cost and diminished population coverage problems (Menegaki et al. 2016; Skeie et al. 2019). Although the survey mode effects found to date typically have been small to moderate, studies are few, and results, as judged by the recent guideline on SP, are both “mixed and context specific” (Johnston et al. 2017; p. 340). Further, this guideline and current SP research have not picked up on the recent concern widely held in the survey methodology literature that the fast-growing share of internet panel respondents that answer surveys on smartphones may significantly affect survey measurement and quality (Couper et al. 2017; Hillygus et al. 2014; Peterson et al. 2017). One study estimated that around 20–30 percent of all web survey market research were conducted by mobile devices in the early 2010s, a share that is undoubtedly higher now and rising fast (Peterson et al. 2017; Poynter et al. 2014).¹ Apart from potential differences in self-selection (e.g. certain people may prefer smartphone responding) and nonresponse, the main concern is measurement effects, i.e., the gap between the ideal (true) measurement and the response obtained. If the same respondent’s answers to an equally worded and designed survey differ depending on whether a smartphone, tablet, or laptop/desktop computer (hereafter referred to as computer) is used, one can conclude that a platform or device effect is present. Early concerns regarding device effects in this literature are both related to the differences in technical attributes (e.g. screen size limitations, touch screen operations and “fat finger” problems) and the response contexts (e.g. multitasking, disturbance while “on-the-move”, presence of others) of mobile devices (smartphones and tablets) compared to computers (Couper et al. 2017). These channels may lead to interlinked effects related to how questions and visual

¹ The internet survey company used for the study in this paper reports an average of 55% of panel members responding on smartphones in general. We have found no specific estimate in the literature of the use of internet panels as a share of all online surveys (including other recruitment methods than internet panels).

stimuli are grouped and appear on the screen, which types of questions are asked (invoking more or less “socially desirable” responding), and various forms of response and scale effects (e.g., when using grids, slider bars etc.). While some earlier studies have confirmed indications of some of these effects, as summarised in, e.g., Couper et al. (2017) and de Leeuw (2018), some recent experimental studies have somewhat surprisingly found small or no device effects, generally concluding that smartphone responses are of comparable quality to computer responses (e.g. Antoun et al. 2017; Lugtig and Toepoel 2016; Schlosser and Mays 2018; Tourangeau et al. 2018; Wenz 2019). The only two studies in the SP literature in environmental and resource economics comparing devices that we are aware of have generally come to the same conclusion. However, it is a challenge in such studies to separate self-selection effects from pure measurement effects related to the device (Liebe et al. (2015) for choice experiments and Skeie et al. (2019) for contingent valuation).

However, while this initial research may look reassuring at first glance, the jury is still out on whether smartphone responding in internet panel surveys may be a potential problem in SP research; for data quality and ultimately the validity and trust in internet panel surveys and the resulting welfare estimates used in, for example, cost–benefit analysis. Research is still lacking on the subject in SP, and as pointed out by Peterson et al. (2017, p. 219): “Much of the experimental evidence on mobile web completion, including the new research [...], is based on data collected using short survey instruments and from panel members with extensive web survey experience. It is difficult to anticipate whether these findings will be replicated with longer web surveys or respondents with less web survey experience”. This point is also made by Clement et al. (2020), with reference to several experimental studies, e.g., Keusch and Yan (2017), Schlosser and Mays (2018), and Toepol and Lugtig (2014). Hence, the complexity and length of typical SP surveys indicate that the transferability of experimental survey methodology results to real-world SP survey settings may be questioned. Further, the quote brings up the issue of survey experience as a potentially critical factor in alleviating device effects over time. If that initially observed device effect may be reduced once internet panellists (and presumably the rest of the population) gain more experience handling smartphones, such device effects may only be temporary. However, when considering survey experience in internet panels more generally, other potential effects beyond the device itself are also brought into the mix. Notably, a lingering concern in the survey methodology literature has, for example, been whether internet panellists become so-called “professional”—answering surveys with little effort and primarily for money—potentially yielding lower quality responses (Zhang et al. 2020; Sandorf et al. 2020).

We hypothesise that device effects may, to a large extent, be driven by technology and experience. Surprisingly, there is no research we are aware of, even from the specialised survey methodology literature, to guide our hypotheses on the effects of survey experience.² Over time, mobile devices have become larger, and people have gained more experience using them. Website designs are now often responsive and adjust to the screen size of the device used to ensure that websites (and surveys) are displayed as intended on smaller screens. As such, we would expect that the (adverse) effects of answering on a smartphone will disappear over time and with respondent experience. If inexperience with smartphones leads to increased difficulty in answering, e.g.

² There is some research in the survey methodology literature on the effects of panelists’ previous responses in longitudinal surveys and learning effects during a SP survey for individual respondents. We consider neither of these strands of literature of relevance to our subject here.

the choice tasks, we would expect to see a higher error variance, i.e. a smaller scale parameter, for inexperienced respondents relative to the more experienced respondents (Liebe et al. 2015). In other words, inexperienced respondents would initially have a relatively more random choice process, as seen from the econometrician's point of view. However, over time, we expect this effect to disappear. Another possible outcome is that inexperienced respondents seek to simplify the task by, for example, choosing the leftmost alternative on the screen or the status quo alternative more often, as also considered by Liebe et al. (2015). The rationale for choosing the leftmost alternative is that this alternative is always the more visible to respondents, while seeing the other two alternatives could require turning the phone to landscape mode, having to "pinch to zoom", or both. This might be more difficult for inexperienced respondents and lead to a higher propensity to choose the left-most alternative. The rationale for choosing the status quo is that it is a simplifying strategy. In our case, the leftmost and status quo alternatives coincide. A final possibility is that these choice patterns are simply the result of different preferences.

We investigate the effects of survey experience and device in a nationally representative, discrete choice experiment (DCE), internet panel survey in Norway on various ecosystem service impacts of the planting of forests for carbon sequestration. We obtained data from the survey company on which device respondents used when answering the survey and, importantly, their experience with answering surveys. Experience is measured as the total (actual, not self-reported) number of surveys completed since the start of the internet panel membership. Although this is not a perfect indicator of experience of relevance to SP, since we do not have information about the types and lengths of surveys respondents have taken previously, it is still a valuable and unique dataset. Survey companies, to our knowledge, typically do not part with this kind of information and survey methodology studies usually rely on self-reported data with much lower accuracy (see, e.g., Callegaro et al. (2014)). Our results show that the respondents answering on a smartphone and answering on a computer are comparable on most socio-demographics but differ in survey experience, gender, and age. While elicited preferences differ between the groups, this is likely not caused by the device itself, as evident by insignificant interaction terms and scale parameters. We do, however, find that experience plays a substantial role. Even after controlling for unobserved heterogeneity in a latent class framework, we find that increased experience is associated with a higher propensity to choose the status quo and that the choice process, as seen from the econometrician's point of view, becomes more deterministic but at a decreasing rate. This may indicate that, beyond a certain point, respondents indeed become so-called "professional" (Sandorf et al. 2020). It is hypothesised elsewhere that these respondents may engage in further simplifying behaviour, e.g. survey satisficing, to minimise effort and maximise income from answering surveys (Sandorf 2019; Zhang et al. 2020). As such, the initial concern over lower quality responses or large device effects may be exaggerated. Still, once survey experience reaches a certain level, the effects of professional survey responding may become more important. We also discuss the implications of these results for practitioners.

The remainder of the paper is outlined as follows. Section 2 introduces the data and outlines the methods, and Sect. 3 presents and discusses the results. Section 4 concludes the article and discusses some implications for SP practitioners, including some proposed avenues for future research.

2 Data and Methods

2.1 Choice Experiment Data

We use data from a DCE survey carried out to investigate the Norwegian population's preferences for planting spruce forests on abandoned pastures where the policy goal was to increase greenhouse gas sequestration. The policy context and motivation are more thoroughly described in Iversen et al. (2019). The questionnaire contained an introductory section asking questions about preferences for policy objectives, both general and environmental. Second, the questionnaire provided text explaining the choice attributes using pictures, text, and icons, followed by the choice experiment. Third, the survey included standard follow-up questions and socio-demographic questions.

The survey topic was the management of semi-natural pastures that are gradually becoming reforested due to abandonment. One potential use of these pastures is to plant climate forests using Norway spruce, which would have to remain unlogged for at least 60 years. However, spruce planting and natural reforestation would reduce the available habitat for several endangered insects and vascular plant species dependent on semi-natural pastures. Spruce plantations and reforested areas also potentially negatively affect landscape aesthetics for some respondents, as was confirmed in qualitative testing of the survey (Grimsrud et al. 2020).

The DCE asked respondents whether to restore these semi-natural pastures through grazing, plant climate forests, or let abandoned pastures naturally reforest as mixed forest. The policy alternatives were defined as various combinations of these three land uses, compared to the status quo alternative, which was natural reforestation. Any active management choice, that is planting climate forests or restoring the semi-natural pastures through grazing would entail a cost to taxpayers while leaving the pastures for natural reforestation would be free. The cost, as explained to respondents, would have to be paid for by an annual earmarked income tax levied on all Norwegian households. Agricultural policy is paid for by everyone in Norway, so this was not expected to generate much protest. Based on focus group testing and a qualitative study conducted using the Q-methodology (see Grimsrud et al. 2020), two main attributes for the DCE, in addition to the cost, were identified: combinations of land use and biodiversity, measured using an indicator of the number of species under threat.³ The levels of the attributes are shown in Table 1.

Climate forest could, by our design, maximally be planted on 50 percent of the abandoned pastures in the country. Similarly, only maximally 50 percent of abandoned pastures could be restored to grazed pastures. The DCE design included a constraint to prevent the highest level of biodiversity to occur together with land use options where no land was utilised as pasture, e.g., a land-use where 50 percent is planted forest and 50 percent is reforested. Biodiversity levels may remain high despite some portion of abandoned pastures becoming reforested or used for climate forest if making sure to preserve the most biodiversity-rich locations. This information was given to the respondents before they were presented with the choice sets. The choice sets were optimally constructed using SAS®.

³ Before the choice cards, respondents were informed that 685 species in Norway are red listed because of land conversion of former pastures to reforested areas or to climate forests. In the choice cards, the "species under threat" attribute was described as follows for the three levels of the attribute: (1) "90 percent of the 685 threatened species will disappear from Norway", (2) "50 percent of the 685 threatened species will be preserved", (3) "685 species will still be threatened by extinction from Norway".

Table 1 Attributes and levels used in the discrete choice experiment

Attribute	Specifics	Level vector
Land use(total to sum to 100% of the area available for land use options)	Planted spruce	0% , 25%, 50%
	Pasture	0% , 25%, 50%
	Reforestation	0, 25%, 50%, 75%, 100%
Biodiversity	Species under threat	400, 550 , 700 species
Carbon sequestration	Tonnes per year(derived from land use)	0,7–1,0; 1,0–1,3 1,3–1,6 million
Cost	Additional earmarked income tax per person p.a	NOK 0 , 300, 600, 900, 1200, 1500, 1800

Reforestation is the residual of the land uses “Climate forest” and “Pasture” (so the percentages sum to 100%). The status quo levels are marked in bold

The methods and SAS-macros used are described in Kuhfeld (2007). We used a fractional factorial design with 18 choice sets.⁴

Figure 1 shows a sample choice task. Each choice task consisted of two experimentally designed alternatives and a status quo option, each described by the level of their attributes. The first attribute illustrates the percentage of the area allocated to different land uses. In the status quo option, all the abandoned pastures will become reforested. Under the experimentally designed alternatives, the percentage of land used for climate forests, grazing and reforestation varies. The second attribute captures changes in the number of endangered and threatened species due to land-use change. The third row describes the amount of carbon sequestration in the vegetation as a function of land use, e.g., more above-ground vegetation and especially spruce climate forests increase carbon sequestration. Because of the strong correlation between above-ground vegetation and carbon sequestration, the third attribute only provides information to respondents but cannot be estimated. The land-use types, the level of biodiversity and the degree to which the vegetation sequesters carbon are symbolised using icons in the choice set. Icons were included in the choice set to be recognisable to respondents from earlier information provided in the survey questionnaire. The fourth attribute, the cost of the policy, is expressed as an earmarked annual increase in personal income tax for an indefinite period

The survey and choice tasks were designed to work efficiently in a smartphone browser. However, it is hard for survey companies to make robust displays across devices, operating systems and web browsers. Therefore, the choice tasks were included as images, which loaded in actual size and not fit-to-screen on smartphones. If opened in portrait mode, respondents would see the left-most side of the choice task; and if opened in landscape mode, they would see the top row. This required respondents to use a single zoom out (or double tap on some phones) operation to see the whole choice task. This was not an issue for computer respondents because most modern computer screens are large enough

⁴ The fractional factorial design was based on the output from the MktRuns-macro. The MktEx-macro was used to design the profiles used in the choice sets. Constraints were added to the MktEx-macro to prevent unrealistic profiles. The ChoiceEff-macro optimised the combination of profiles into choice sets after the status quo alternative was added to the final output of the MktEx-macro. Finally, the Mktblock-macro was used to block the 18 choice-sets into three blocks.

Picture11A_Text
 What do you prefer?

Choose one of the three alternatives below.

Picture11A1

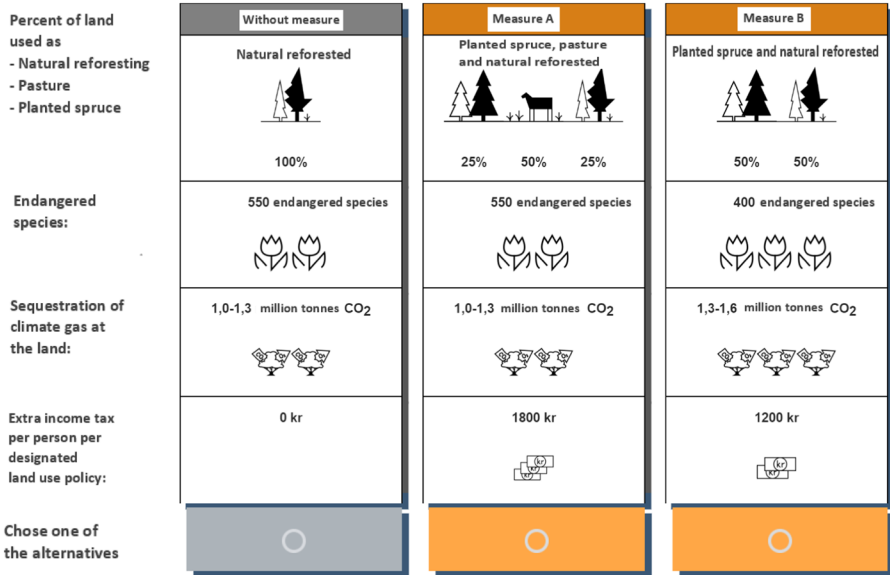


Fig. 1 A screenshot of a translated choice task used in the DCE

to view the choice task at full resolution.⁵ At the beginning of the survey, respondents were randomly allocated to receive either 6 or 12 choice tasks and encouraged to answer either on their smartphone or computer.⁶ We obtained data from the survey company on which device respondents used when answering the survey and their experience answering surveys on each device. The experience is measured as the total (actual, not self-reported) number of surveys completed on a given device since the start of the internet panel membership.⁷ This is a unique dataset, as survey companies normally do not share this

⁵ We carefully tested this on 4 types of phones and browsers. All phones performed similarly with respect to displaying the choice tasks.

⁶ The encouragement was formulated as follows: “We wish that you respond to the survey on a smartphone (not a computer or tablet)” (correspondingly for the computer encouragement). The text was placed both in the invitation email and on the front page of the survey after information about the survey length and incentive.

⁷ This could be any type of opinion or marketing survey of different degrees of complexity the respondents have carried out, most likely few if any previous SP surveys. Unfortunately, the survey company does not maintain records of the length or type of previous surveys by panel members. However, due to the low incentive paid to respondents for surveys (ca. 0.1 Euro per minute) it is reasonable to assume that a large portion of the surveys are of a length that maximises the amount of survey data collected within the time limit before response rates start dropping fast (typically around 15–20 min). Hence, although the survey experience indicator we have obtained is not a perfect match with SP surveys, it is still likely to be a useful proxy.

information. Survey methodology studies usually, therefore, rely on self-reported data with much lower accuracy (e.g., Callegaro et al. 2014; Zhang et al. 2020).⁸

The data were collected from 23rd of April 2018 to 7th of May 2018 by one of the two most reputable survey companies maintaining large, nationally representative internet survey panels in Norway.⁹ The panel consists of 80 000 Norwegian adult respondents recruited randomly (no opt-in), and the company follows strict procedures for quality control in panel management (ISO 9001 standard). The survey company invited 6929 random respondents to participate in the survey and received 1030 complete responses, a final stage response rate of 15.1 percent. Survey invitations and reminders were conducted in the same way for all samples. The survey was estimated to take 12–15 min, with a standard survey incentive of NOK 15 for a completed response. Inspection of the data revealed that 39.3 percent (405) of the respondents were encouraged to answer on a desktop computer, and 60.7 percent (635) were encouraged to answer on a smartphone. 60 percent (616) of the respondents received six choice tasks, while 40 percent (414) received twelve choice tasks. The focus in the current paper is not on these treatment groups but instead on respondents who answered on a smartphone or a desktop. We report a comparison of the treatment groups and MNL models with a full set of relative scale parameters in Online Appendix for completeness. We excluded 70 respondents who answered on a tablet device despite being encouraged not to do so, which resulted in a trimmed sample of 960 respondents. 456 of these respondents answered the choice tasks on a computer, while 504 answered on a smartphone. We note that smartphone users are somewhat less likely to complete the survey. The breakoff (drop-out) rate for desktop respondents was 14.5 percent compared to 19.5 percent for smartphone respondents, a smaller difference than typically found in the survey literature (e.g., a meta-analysis reported in Couper et al. (2017) found 2–3 times higher breakoff rates for mobile surveys). Median completion times were roughly the same at 14 min and 53 s for computer and 15 min and 25 s for smartphones.

2.2 Econometric Approach

We assume that respondents' choices can be described by a random utility model and that the utility, U , respondent n receives from alternative i in choice task s can be described by the linear additive utility function in Eq. 1:

$$U_{nis} = \beta X_{nis} + \varepsilon_{nis}, \quad (1)$$

where β is a vector of parameters to be estimated, X_{nis} a vector of attribute levels for the i th alternative in choice task s and ε_{nis} and unobserved error term commonly assumed to be Type I extreme value distributed. Under standard assumptions, the probability that alternative i , $i \in C$, is chosen can be described by the multinomial logit model (McFadden, 1974):

$$\Pr(i_{ns}) = \frac{\exp(\lambda\beta X_{nis})}{\sum_{i \in C} \exp(\lambda\beta X_{nis})} \quad (2)$$

⁸ Once exception that analyses actual data on survey experience in the stated preference literature is the study by Sandorf et al. (2020), but they do not have these data by device.

⁹ We choose to anonymise the company name, a practice that may help researchers obtain useful para- and other data from survey companies and facilitate research cooperation to improve data quality.

where λ is a scale parameter inversely related to the variance of the error term ϵ . To test for unobserved relative differences between treatment groups (Swait and Louviere 1993) and allow for the inclusion of other socio-demographic variables, we specify scale as $\lambda = \exp(\gamma Z_n)$, where γ is a vector of freely estimated parameters and Z_n a vector of individual specific variables. This functional form ensures that the scale parameter is positive and in the case of the individual-specific variables being dummy coded, will estimate relative scale parameters. The MNL model is prevalent and widely used in the literature. Still, its ability to describe preference heterogeneity is limited and rests on the Independence from Irrelevant Alternatives (IIA) assumption (Hausman and McFadden 1984). While allowing for different scales relaxes some of the strict assumptions of the MNL model, we can go further and explore unobserved preference heterogeneity. Here we allow respondents' preferences to differ using a latent class framework. In a latent class model, we assume that preferences can be described by a finite set of unique preference weights (Greene and Hensher 2003). As analysts, we do not know which set of preference weights describe which respondent, but we can estimate this up to a probability using the class probability function in Eq. 3:

$$\pi_q = \frac{\exp(\alpha_q Z_n)}{\sum_{q=1}^Q \exp(\alpha_q Z_n)}, \tag{3}$$

where α_q is a class specific vector of parameters to be estimated and Z_n is a vector of individual specific variables, including a constant, that may or may not be the same as in the scale function. The Qth vector is set to zero for identification. Under these assumptions, we can write the overall choice probability as:

$$\Pr(y_n | \pi, \beta, X) = \sum_{q=1}^Q \pi_q \prod_{s=1}^S \frac{\exp(\lambda \beta_q X_{nis})}{\sum_{i \in C} \exp(\lambda \beta_q X_{nis})}, \tag{4}$$

where β_q is now a class specific vector of parameters to be estimated. Notice that in the latent class model, we take the panel nature of the data into account by using the product over all S choice tasks answered by respondent n . The specification in Eq. 4 rests on the assumption of independence between an individual's choices. Without the independence assumption, the probability of the sequence would not simply be the product across individual choices. We believe that this assumption is not too restrictive in the current context for the following reasons: (i) the choice tasks were designed to be independent at the experimental design stage; (ii) respondents were asked to treat each choice situation independently; (iii) the choice of which policy to support in the second choice tasks is independent of the choice made in the first, i.e., there is no updating of, e.g. forest cover levels based on previous choices; and (iv) while it can be argued that a correlation across choices exists simply because respondents answer a sequence of them, we used several different versions of the survey where the order of choice tasks was different between respondents to minimise the impact ordering effects at the sample level.

To calculate the conditional willingness-to-pay estimates from the latent class model, we use the conditional, i.e., individual-specific, class membership probabilities as weights. They are obtained using the following formula:

$$\pi_{q|y_n} = \frac{\pi_q \Pr(y_n | q)}{\sum_{q=1}^Q \pi_q \Pr(y_n | q)}$$

Table 2 Summary statistics of selected variables used in the analysis

	Answered on a com- puter		Answered on a smart- phone		Significance	Test statistic
	Mean	SD	Mean	SD		
Male	0.41		0.56		***	Chi Sq. = 20.195
Age	56.50	16.40	45.40	16.20	***	t = 10.508
Income†	718,539.00	340,539.00	760,250.00	357,582.00		Chi Sq. = 0.47774
Education						Chi Sq. = 0.43855
Basic Education	0.04		0.03			
High School	0.28		0.31			
University 3 years	0.32		0.31			
University 4 years	0.19		0.17			
University > 4 years	0.14		0.14			
Other	0.04		0.04			
Surveys total	158.84	94.30	135.93	103.00	***	t = 3.6004
Surveys computer	132.82	86.30	43.90	65.90	***	t = 17.794
Surveys smartphone	26.00	47.70	92.00	82.50	***	t = - 15.339
Prefer computer	0.89	0.20	0.20	0.40	***	t = 29.383
Prefer smartphone	0.04	0.19	0.65	0.47	***	t = - 26.632
Time spent on the CTs in minutes‡	4.73	2.46	5.37	2.73	***	t = - 3.6804
N	456		504			

***-1% level, **-5% level, *-10% level, all standard errors are robust

†Variable elicited using 16 categories. Reporting the average of the midpoint means

‡Excluding 5% fastest respondents, i.e. respondents who took more than 15 min to answer the choice tasks (median completion time for survey)

where $\Pr(y_n|q)$ is the probability of the sequence of choices in class q and π_q is the unconditional probability of belonging to class q (Eq. 3).

2.3 Model Estimation

All models were coded and estimated in the statistical software R (R Core Team 2016). To reduce the risk of the latent class models converging to local rather than global optima, we generated a large number of starting values and ran the best fitting models to convergence. The models were selected based on the log-likelihood value at convergence and the Bayesian Information Criterion (BIC).

3 Results and Discussion

In Table 2, we first compare respondents who answered on a computer and a smartphone. We see that, on average, smartphone respondents are 11 years younger than computer respondents and more likely to be men. We find no significant difference in terms of

income, education, or geographic location (not shown because of the large number of locations). Overall, we see that respondents answering on a computer have answered a larger number of surveys than respondents answering on a smartphone, i.e., they have more survey experience. Furthermore, the results suggest that respondents prefer answering on a particular device. Computer respondents had on average answered 88 more surveys on a computer than on a smartphone. In contrast, smartphone respondents had answered, on average, 66 more surveys on a smartphone than on a computer. This, combined with the fact that 89 per cent of respondents answering on a computer stated that they prefer to answer on a computer and 65 percent of smartphone respondents stated that they prefer to answer on a smartphone, suggests that people will answer the survey on their preferred device (irrespective of the encouragement to use a particular device, see Online Appendix). Finally, smartphone respondents spend significantly more time answering the choice task section of the survey, with 38 s on average. A common assumption in the response latency literature is that more time spent on the choice tasks implies more effort, and consequently, better responses (Börger 2016; Campbell et al. 2017). If true, this implies that smartphone respondents would provide better responses. However, this is not necessarily the case. If the extra time spent on the choice tasks is driven by the fact that navigating the choice tasks on a small screen is more difficult, we would expect responses to be poorer. It is also possible, as is argued elsewhere (see, e.g. Sandorf et al. 2020), that response time is a function of experience. With increasing experience, it is possible to answer faster while maintaining high quality because the response format is familiar. We argue that simply considering time spent on the choice tasks across devices provides no a priori expectations about the effort expended on the choices themselves and hence response quality and that indeed the interesting exploration lies in looking at device and experience, which is explored in detail in Sect. 3.2.

3.1 Device Effects

Table 3 shows the results of our baseline MNL model and models exploring observed heterogeneity with respect to the device used and experience. From our baseline model, it is clear that respondents prefer both a 25 percent and a 50 percent increase in spruce cover (climate forests) and prefer more strongly a 25 percent and a 50 percent increase in the area used for grazing. This could reflect a preference for keeping the traditional cultural landscape in rural Norway, which historically consisted of many small-scale sheep farms dotted throughout the landscape (Iversen et al. 2019). Interestingly, there appears to be a lack of sensitivity to the size of the increase in spruce cover, suggesting that respondents have the same marginal preference for a 25 percent increase as a 50 percent increase. While respondents have a stronger preference for a 50 percent increase in grazing area than a 25 percent increase, the difference between the two is small. As expected, respondents dislike more species under threat and prefer fewer species under threat. Finally, the cost parameter is negative and significant. We estimate the full set of relative scale parameters to control for possible differences in error variance between our treatment groups and whether respondents answered on a computer or a smartphone. The scale parameters are relative to respondents who received 12 choice tasks, that were encouraged to answer using a smartphone, and that did so. None of the scale parameters are significantly different from 1, which indicates no difference in error variance between groups. While we do not consider the four treatment groups in the following analysis, it is nonetheless important to

Table 3 A baseline MNL model for comparison and baseline specifications for the effect of experience on the SQ and Scale

Utility function	Baseline		Smartphone interactions		SQ interactions		SQ interactions		
	Est	Std. Err	Est	Std. Err	Est	Std. Err	Est	Std. Err	
Tax	-0.451	0.059	***	0.085	***	0.057	***	0.076	***
25% of area covered by spruce	0.455	0.067	***	0.087	***	0.063	***	0.085	***
50% of area covered by spruce	0.439	0.069	***	0.088	***	0.063	***	0.085	***
25% of area used for grazing	0.974	0.120	***	0.131	***	0.097	***	0.149	***
50% of area used for grazing	1.132	0.140	***	0.137	***	0.109	***	0.169	***
150 fewer species under threat	0.213	0.040	***	0.053	***	0.039	***	0.048	***
150 more species under threat	-0.344	0.050	***	0.067	***	-0.375	***	-0.336	***
Status quo	0.417	0.107	***	0.142	***	0.307	0.196	0.351	0.111
Interaction terms									
Tax x smartphone			0.152	0.097					
25% of area covered by spruce x smartphone			-0.129	0.094					
50% of area covered by spruce x smartphone			-0.133	0.102					
25% of area used for grazing x smartphone			-0.113	0.124					
50% of area used for grazing x smartphone			-0.016	0.130					
150 fewer species under threat x smartphone			0.110	0.070					
150 more species under threat x smartphone			-0.036	0.080					
Status quo x smartphone			0.253	0.197					
Status quo x computer x experience					-0.988	0.533	*		
Status quo x computer x experience^2					0.744	0.370	**		
Status quo x computer x experience^3					-0.114	0.068	*		
Status quo x smartphone x experience					-0.120	0.259			
Status quo x smartphone x experience^2					0.147	0.075	**		
Scale function									
Computer x experience								0.612	**
Computer x experience^2								-0.260	**

Table 3 (continued)

Utility function	Baseline		Smartphone interactions		SQ interactions		SQ interactions	
	Est	Std. Err	Est	Std. Err	Est	Std. Err	Est	Std. Err
Smartphone x experience							0.269	0.234
Smartphone x experience ²							-0.159	0.091 *
Relative scale								
Used computer—6 CT—treatment computer	1.291	0.205						
Used smartphone—6 CT—treatment computer	1.006	0.226						
Used computer—12 CT—treatment computer	1.239	0.211						
Used smartphone—12 CT—treatment computer	1.201	0.278						
Used computer—6 CT—treatment smartphone	1.053	0.192						
Used smartphone—6 CT—treatment smartphone	0.956	0.153						
Used computer—12 CT—treatment smartphone	1.237	0.236						
Used smartphone—12 CT—treatment smartphone	1.000	(fixed)						
6 CT—Treatment computer			1.040	0.141	0.949	0.139	0.999	0.146
12 CT—Treatment computer			1.058	0.159	1.200	0.161	1.082	0.159
6 CT—Treatment smartphone			0.905	0.113	0.932	0.117	0.918	0.117
12 CT—Treatment smartphone			1.000	(fixed)	1.000	(fixed)	1.000	(fixed)
Model statistics								
LL	-8145.687		-8133.936		-8008.584		-8112.453	
Adj. R sq	0.077		0.078		0.093		0.081	
BIC	16,426.280		16,438.750		16,161.070		16,359.810	
N	8052		8052		8052		8052	
K	15		19		16		15	

***-1% level, **-5% level, *-10% level, all standard errors are robust. Tests on scale is against 1. Excluding respondents answering on a tablet

Table 4 A baseline latent class model with 3 classes for comparison

Utility function	Class 1		Class 2		Class 3		
	Est	Std. Err	Est	Std. Err	Est	Std. Err	
Skatt	-0.293	0.076	***	0.245	***	0.503	***
25% of area covered by spruce	0.577	0.078	***	0.144	**	0.209	0.391
50% of area covered by spruce	0.603	0.084	***	0.176	0.191	-0.437	0.329
25% of area used for grazing	1.362	0.144	***	0.292	0.134	1.592	0.556
50% of area used for grazing	1.606	0.170	***	0.395	0.136	1.237	0.641
150 fewer species under threat	0.221	0.049	***	0.388	0.125	0.518	0.324
150 more species under threat	-0.327	0.053	***	-0.434	0.120	-0.517	0.356
Status quo	-2.616	0.462	***	-0.800	0.269	2.012	0.761
Relative scale							
Used computer—6 CT—treatment computer	1.057	0.112					
Used smartphone—6 CT—treatment computer	0.776	0.118	*				
Used computer—12 CT—treatment computer	1.078	0.142					
Used smartphone—12 CT—treatment computer	1.080	0.162					
Used computer—6 CT—treatment smartphone	0.971	0.102					
Used smartphone—6 CT—treatment smartphone	0.874	0.086					
Used computer—12 CT—treatment smartphone	0.964	0.141					
Used smartphone—12 CT—treatment smartphone	1.000	(fixed)					
Class membership function							
Constant	1.094	0.140	***	0.030	0.133	0.000	(fixed)
Avg. Class probability	0.595			0.205		0.199	
Model statistics							
LL	-5812.780						
Adj. R sq	0.339						
BIC	11,922.350						
N	8052						

Table 4 (continued)

Utility function	Class 1		Class 2		Class 3	
	Est	Std. Err	Est	Std. Err	Est	Std. Err
K						

****-1% level, ***-5% level, **-10% level, all standard errors are robust. Tests on scale is against 1. Excluding respondents answering on a tablet

include the relative scale parameters to control for possible differences in error variances as we explore different model specifications.

In Table 4, we show the result of a latent class model that allows for unobserved preference heterogeneity. Our model search revealed that a 3-class model described the data best using a combination of model fit, information criteria and economic theory. Respondents predicted to be in Class 1 are relatively cost insensitive and while they have positive preferences for both increases in spruce cover and grazing areas, the latter dominate. Respondents in this class are also much less likely to choose the status quo as evident by the large, negative, and significant alternative specific constant. This class comprises about 60 percent of respondents and their choice behaviour is characterised by a willingness to move away from the status quo. On the other hand, respondents predicted to be in Class 2 are much more cost-sensitive, and while they also prefer increased cover of spruce and grazing areas, the strength of these preferences are smaller in magnitude and insignificant in the case of a large increase in spruce cover. This class comprises about 20 percent of respondents and appears to make real trade-offs between all alternatives, including the status quo. Lastly, respondents in Class 3 are extremely cost-sensitive and very likely to choose the status quo. They hold strong preferences for a small increase in the area used for grazing. This class also comprises around 20 percent of respondents.

Running separate models for respondents who answered using a computer or a smartphone shows that preferences follow the same pattern as for the sample as a whole and that all estimates are of the same signs and significance (results omitted for brevity). However, a log-likelihood ratio test rejects the null of jointly equal preferences. Using a set of interaction terms in a model run on the full data set (Table 3) reveals that the marginal effect of answering on a smartphone is insignificant for all attributes suggesting that the difference between respondents using a smartphone and a desktop computer is not the result of the device itself, but possibly caused by other differences between the two groups of respondents.¹⁰ From Table 2, we saw that there were differences in terms of age, gender and most notably, experience. The following analysis estimates separate experience parameters for computer and smartphone users in models run on the total sample. The benefit of such an approach is that the parameters will be directly comparable since they are subject to the same scaling. We note that there is an identification problem using the same variable in the (relative) scale function and the utility function. As such, in all models where we estimate separate parameters for computer and smartphone users, we cannot also control for the relative difference in error variance between the two.

3.2 Experience Effects

Based on the evidence above, we imagine two pathways for how experience answering surveys affect respondents' choices in the discrete choice experiment. One, if it is the case that the experience answering surveys makes it easier for a respondent to answer the survey

¹⁰ A reviewer suggested to rerun the models reported in Table 3 using samples created using propensity score matching. We created matched samples (computer vs smartphone users) to be representative with respect to device preference, experience, treatments, gender, age, education and income. We used a nearest neighbour matching algorithm, which led to a reduced sample size of 882 individuals. The overall result remained the same, however, the interaction with both spruce cover levels became marginally significant at the 10% level. In practice, we consider this to not be an indication of a device effect and that our results are robust to this sample adjustment.

and choice tasks, we would expect responses to be more reliable/deterministic. A possible way to measure this is through the scale parameter. The scale parameter is inversely related to the variance of the error term, and one interpretation of the scale parameter is as an indicator of choice determinism. A larger scale parameter indicates a more deterministic choice process, whereas a smaller scale parameter indicates a more random choice process, as seen from the analyst's point of view. Two, it is possible that experience changes the way that people respond. For example, does more experience with a given response platform affect the propensity to choose the status quo or experimentally designed alternatives? In the following, we test both pathways under various assumptions about the impact of experience using both linear and non-linear interactions and functional forms.

3.2.1 Experience and the Scale Parameter

From Table 3, we see that when we include experience in the scale function the effect is non-linear and best described by a squared function for both computer and smartphone respondents.¹¹ Briefly looking at the estimated preference parameters, we see that they are of similar magnitude, sign and significance as under the baseline model. It appears that choices become increasingly deterministic from the analyst's point of view with increasing experience but at a decreasing rate. Furthermore, this effect is stronger for respondents answering on a computer than on a smartphone. This strength of this effect is likely connected to the observation that, in general, experience levels are higher among computer respondents. We conjecture that over time, as experience among smartphone respondents increase, we will see an effect of a similar magnitude. What is possibly concerning is that it suggests a "goldilocks" level of experience where choices are as deterministic as possible. In other words, respondents should not be too inexperienced but not too experienced either. But as we will see, this relates to the propensity to choose the status quo.

From Table 5, we see that the model allowing for both unobserved preference heterogeneity through latent classes and controls for scale differences induced by different experience levels outperform the baseline latent class specification in Table 4 in terms of model fit and BIC. Furthermore, we note (without making direct comparisons) that the estimated within-class preference parameters are of comparable magnitude, sign and significance to the baseline model. As such, we do not discuss this in detail here. Interestingly, it appears that some of the scale heterogeneity attributed to experience may have been confounded with heterogeneity in preferences. While we still observe that the relationship between scale and experience follows a second-degree polynomial for computer respondents, it is linear for smartphone users. While this does suggest that perhaps more experience is unequivocally better for smartphone respondents, this is not the whole story. Estimating models using a second-degree polynomial for smartphone users (identical scale specification as the best fitting under the MNL) fit the data (marginally) better but is not supported by the BIC. That said, the estimated coefficients show a second-degree polynomial relationship similar to that observed for computer respondents. Again, we suspect that this

¹¹ The functional forms used for interactions with both the status quo and in the scale function are the result of extensive model search running separate models on both smartphone, computer and all respondents while considering linear, squared and cubic functional forms and combinations of functional forms for smartphone and computer respondents respectively. Models to include are determined on the basis of fit in combination with the BIC statistic to ensure a model that is as parsimonious as possible. Conclusions remain the same when run on the propensity score matched sample.

Table 5 A latent class model with 3 classes and experience in the scale function

Utility function	Class 1		Class 2		Class 3				
	Est	Std. Err	Est	Std. Err	Est	Std. Err			
Skatt	-0.255	0.064	***	-1.215	0.208	***	-3.126	0.479	***
25% of area covered by spruce	0.487	0.068	***	0.246	0.122	**	0.183	0.325	
50% of area covered by spruce	0.506	0.073	***	0.147	0.150		-0.363	0.254	
25% of area used for grazing	1.158	0.123	***	0.290	0.114	**	1.284	0.438	***
50% of area used for grazing	1.365	0.142	***	0.363	0.122	***	0.929	0.495	*
150 fewer species under threat	0.177	0.043	***	0.337	0.105	***	0.453	0.273	*
150 more species under threat	-0.277	0.046	***	-0.381	0.103	***	-0.388	0.303	
Status quo	-2.296	0.364	***	-0.646	0.225	***	1.656	0.611	***
Scale function									
computer x experience	0.360	0.109	***						
computer x experience ²	-0.102	0.033	***						
Smartphone x experience	0.090	0.038	**						
Relative scale									
6 CT—Treatment computer	0.925	0.101							
12 CT—Treatment computer	1.029	0.108							
6 CT—Treatment smartphone	0.915	0.073							
12 CT—Treatment smartphone	1.000	(fixed)							
Class membership function									
Constant	1.098	0.116	***	0.042	0.125		0.000	(fixed)	
Avg. Class probability	0.595			0.207			0.198		
Model statistics									
LL	-5805.012								
Adj. R sq	0.340								
BIC	11,897.820								
N	8052								
K	32								

***-1% level, **-5% level, *-10% level, all standard errors are robust. Tests on scale is against 1. Excluding respondents answering on a tablet

non-significance is likely driven by the overall lower experience level among smartphone respondents and not that their behaviour is different. Indeed, this only strengthens our conjecture that more experience is better, up to a point beyond which choices begin to appear relatively more random from the analyst's point of view. While we cannot rule out that this is potentially the result of being caught in a local optimum, we are reasonably confident that this is not the case given our extensive approach of estimating the model using a large

set of different starting values and the fact that similar results are observed with the MNL specification.

Therefore, it is reasonable to conclude that experience answering surveys does affect choice determinism and that this relationship is non-linear. This somewhat surprising result could speak to the larger issue of survey experience and why people tend to answer many surveys. There is a growing concern among practitioners that there exists a group of professional respondents that answer multiple surveys primarily for the monetary incentive (Baker et al. 2010; Hillygus et al. 2014; Sandorf et al. 2020; Zhang et al. 2020). If these respondents, who are likely to have answered many surveys also engage in survey satisficing (Krosnick 1991), then one would expect to see a pattern such as this. Said another way, as respondents gain experience, they become better at answering surveys, and we see that the quality of their answers increases, but that beyond a certain point, they become professional respondents that are in it for the money and are looking for ways to simplify and speed through the survey, for example, by making more random choices (as seen from the analyst's point of view).

3.2.2 Experience and the Status Quo Parameter

From Table 3, we see that including interaction terms between the alternative specific constant for the status quo alternative and experience do not lead to substantive changes in the parameter estimates, which remain of roughly the same magnitude, sign and significance as the baseline model. The obvious exception is the status quo constant itself. The interaction terms, as implemented here, effectively allow us to estimate separate parameters for how experience affects the propensity to choose the SQ for computer and smartphone respondents. They do not capture the marginal effect of a smartphone respondent relative to a desktop respondent. We did explore a model with the full set of interactions, but this model only fit the data marginally better and was not supported by the BIC statistic. We note that none of the relative scale parameters are significantly different from unity. Pursuing the model with status quo interactions is also in line with our *a priori* hypothesis that experience is linked to behaviour and that experienced respondents answer the survey differently than inexperienced respondents. Furthermore, the model reveals that a third-degree polynomial describes the relationship between experience and preferences for the status quo for computer respondents, and a second-degree polynomial describes this relationship for smartphone respondents. There appears to be an initial phase of experience leading to a lower propensity to choose the status quo, but this is relatively quickly dominated by the large parameter estimate on the squared term. However, it is important to note that the effect of experience now explains all heterogeneity in the model, and we observe a marked change when we allow for unobserved sources of heterogeneity.¹²

Looking at the results from Table 6, we see that when we allow for unobserved sources of heterogeneity, the relationship between the propensity to choose the status quo and experience is linear. We note a couple of points on the choice of specification. While we have run models where experience enters the class probability function, the interpretation

¹² When we use the propensity score matched sample, the patterns for how experience affects the propensity to choose the SQ remains the same, however, the level of significance changes with only the squared terms being significant at the 10% level. That said, we do not know whether this is in fact caused by changes to the sample composition or the reduced sample size itself. As we note in the paper, the strong emerging and consistent patterns but lack of significance can be the result of too little variation in the data.

Table 6 A latent class model with 3 classes and interactions for status quo

Utility function	Class 1		Class 2		Class 3				
	Est	Std. Err	Est	Std. Err	Est	Std. Err			
Skatt	-0.252	0.073	***	-1.438	0.192	***	-3.390	0.509	***
25% of area covered by spruce	0.578	0.071	***	0.268	0.123	**	-0.077	0.346	
50% of area covered by spruce	0.596	0.078	***	0.178	0.149		-0.312	0.317	
25% of area used for grazing	1.385	0.106	***	0.350	0.127	***	1.070	0.447	**
50% of area used for grazing	1.625	0.122	***	0.446	0.124	***	0.708	0.502	
150 fewer species under threat	0.204	0.050	***	0.407	0.117	***	0.395	0.273	
150 more species under threat	-0.323	0.052	***	-0.458	0.107	***	-0.371	0.335	
Status quo	-3.422	0.347	***	-1.528	0.313	***	0.917	0.605	***
Interaction terms									
Status quo x computer x experience	0.479	0.101	***						
Status quo x smartphone x experience	0.404	0.115	***						
Scale function									
6 CT—Treatment computer				0.963	0.088				
12 CT—Treatment computer				1.114	0.095				
6 CT—Treatment smartphone				0.903	0.066				
12 CT—Treatment smartphone				1.000	(fixed)				
Class membership function									
Constant	1.073	0.099	***	0.138	0.120		0.000	(fixed)	
Avg. Class probability	0.576			0.226			0.197		
Model statistics									
LL	-5798.265								
Adj. R sq	0.341								
BIC	11,875.330								
N	8052								
K	31								

***- 1% level, **- 5% level, *- 10% level, all standard errors are robust. Tests on scale is against 1. Excluding respondents answering on a tablet

of parameters becomes difficult with more than two classes. Furthermore, the interpretation would not capture behaviour in a standard latent class model but the marginal effect of one extra survey on the probability that a given preference vector (class) describes the respondent's preferences. It is also known that models with non-linear effects in the class-probability functions are notoriously hard to make converge (Hensher et al. 2005). Lastly,

we assume that the impact of experience is the same for all classes imposing the constraint that the parameter on the interaction is the same for all three classes. Allowing the impact of experience also to vary latently would imply that the role of experience is different for different people, which is not in line with the behavioural hypothesis tested here. As people become more experienced, we see that they are more likely to choose the status quo, but that depending on the class, this still means that even experienced respondents are less likely to choose the status quo.

To better understand how this ties together, we need to look at the actual choice shares, i.e., the percentage of times each alternative is chosen for each choice task. The choice shares are stable and do not vary across the choice tasks or devices. Respondents made trade-offs, and there is variation in the chosen alternative. On average, less than 30% chose the status quo (alternative 3) in any given choice task. An increased propensity to choose the status quo will bring the choice shares for experienced respondents closer to a third. This is the choice probability associated with random responses. As such, there are two effects at play here: As respondents become more experienced, they are more likely to choose the status quo. For about half of the respondents (Class 2 and Class 3), this translates into a higher propensity to choose the status quo leading to a relatively more deterministic choice process. The other half (Class 1) are extremely unlikely to choose the status quo. For these respondents, choosing the status quo more with increasing experience will bring the choice shares closer to 1/3, which is the probability associated with random responses. These two effects working together at the sample level is likely also why we observe that the relationship between behaviour and experience is best described by a non-linear relationship.

3.3 Willingness to Pay

In the following, we briefly discuss the conditional willingness-to-pay estimates derived from the baseline 3-class latent class model. Remember that scale cancels out when we calculate WTP and that in the model where we include an interaction between the SQ and experience, we capture the behaviour associated with choosing the SQ as a function of experience. Mapping the conditional WTP estimates to a respondent's experience does not reveal any strong relationship between WTP and experience, which strengthens our conclusion that experience is about behaviour in the survey and propensity to choose the SQ but is not (necessarily) a driver of WTP. These results are omitted for brevity. The willingness-to-pay estimates are the weighted sum of the class-specific WTP, where the weights are the conditional (individual-specific) class membership probabilities. For representation in the graph, the class grouping is based on which class a given respondent had the highest probability of belonging to (although all probabilities and WTPs were included in the underlying calculation).

From Fig. 2, it is clear that people in Class 1 have large and positive WTP for increases in climate forest cover and grazing areas compared to Classes 2 and 3. The latter dominates the former, and we also observe some sensitivity to scope. This is likely driven by the lower cost sensitivity among respondents predicted to be in Class 1. Respondents predicted to be in Class 2 and 3 are much more cost-sensitive and have overall weaker preferences for increases in areas used for climate forests and grazing. This leads to relatively tight willingness-to-pay distributions around 0, but substantial heterogeneity exists. This also suggests that a large segment of the population is willing to pay a considerable amount to

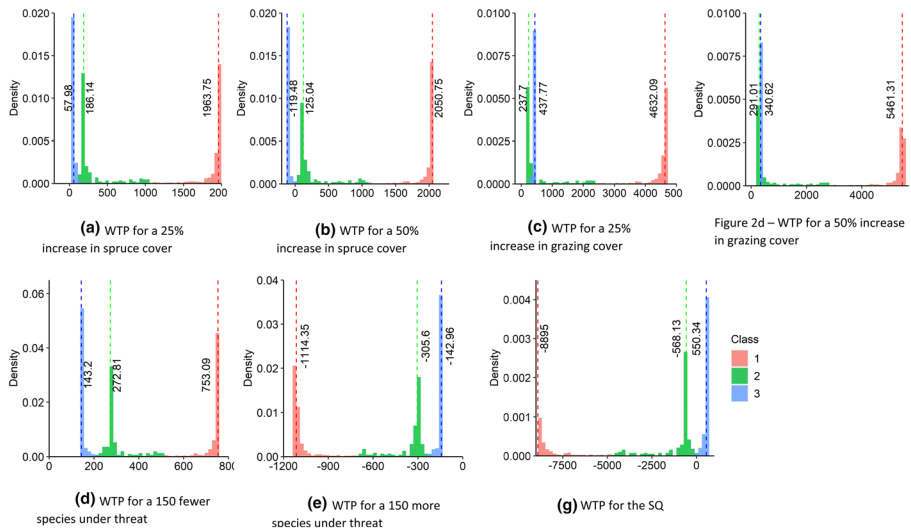


Fig. 2 Distribution of conditional willingness-to-pay from the baseline latent class model. Colored vertical lines correspond to the median for the group, with the median value printed next to the corresponding line

plant climate forests and even more for improvements in grazing areas. This same group would need to be compensated a large amount to be willing to stay with the status quo.

4 Discussion and Conclusions

In this paper, we investigated the role of survey experience and device used on choices in a stated preference survey. Survey experience was defined as the number of surveys a respondent had answered since they became members of the internet panel used for sampling.

First, we find that while our samples were comparable on most socio-demographic variables, respondents answering on a smartphone were more likely to be younger and male. Furthermore, we find that computer respondents had answered more surveys on average, but broken down by device, computer respondents had answered relatively more surveys on a computer and smartphone respondents had answered relatively more surveys on a smartphone. The results show that there are differences in preferences between computer and smartphone respondents, but that this is not caused by the device used (a result that holds on a propensity score matched sample). This is reassuring for practitioners and in line with our expectations derived from the literature.

Instead, we find that experience answering surveys affects both the scale parameter, which is an indication of choice determinism (as seen from the econometrician's point of view), and the propensity to choose the status quo. The impact of experience on scale is non-linear for both computer and smartphone respondents and is stronger for computer respondents. We argue that this difference is likely caused by the fact that computer respondents, on average, have substantially more survey experience compared to smartphone respondents. The observed relationship suggests that scale is increasing in

experience, but at a decreasing rate, ultimately suggesting a “goldilocks” level of experience where the choice process, as seen from the analyst’s point of view, is the most deterministic, and hence, that the ideal respondent has the right amount of experience: not too much and not too little. When allowing for unobserved preference heterogeneity in a latent class framework, we find that the best fitting model explains the effect of experience on scale for computer respondents as a second-degree polynomial and for smartphone respondents as a linear relationship. That said, using the second-degree polynomial also for smartphone respondents shows the same relationship between experience and scale as observed for computer respondents, although the relationship is insignificant. We hypothesise that this is simply because the overall experience level among smartphone respondents is too low to pick up this effect.

Third, we find that the relationship between experience and the propensity to choose the status quo is also non-linear in the MNL model, but that when we allow for unobserved heterogeneity, it reveals a linear relationship and that more experience is likely to lead to a higher propensity to choose the status quo. We also find that the impact of experience for computer and smartphone respondents is somewhat similar, suggesting that this is a device-independent behaviour that results from experience itself. We are, however, mindful that our experience measure is not perfect. We do not observe whether a given respondent is a member of multiple panels nor their overall survey experience if they are. We only observe their experience with answering surveys within this particular panel.

The implication of these results is that over time the potential challenge of people answering surveys on their smartphones is likely to be reduced as people gain experience doing so, at least effects that are related to technical ability and extra cognitive effort required.¹³ When combining increased experience with better devices, higher quality screens and surveys and websites being designed (“optimised”) for smartphone screens or tailor-made survey apps that are robust across devices, it leads to the tentative conclusion that the concern about lower quality responses from respondents answering on their smartphones is unwarranted. This would also explain why early papers studying device effects tended to be more concerned about respondents using smartphones (as summarised e.g. in Couper et al. 2017 and de Leuw 2018); whereas later papers, ours included, have documented relatively small device effects (e.g. Antoun et al. 2017; Tourangeau et al. 2018). However, our study is, to our knowledge, only the third investigating device effects in SP, so we would recommend more systematic experimental and practical testing of different choice experiment and contingent valuation designs and smartphone adaptations, before concluding firmly that the continuing exodus of SP survey responses to smartphones has no consequence for elicited preferences and welfare estimates.

While the observed device effects seem to be small or non-existent in this particular study, we do see that very experienced respondents tend to have a relatively more stochastic choice process as seen from the econometrician’s point of view and that they are more likely to choose the SQ, suggesting that the problem is not related to the device,

¹³ Another concern in the general survey methodology literature is related to the potentially higher level of disturbance experienced and potential influence of the presence of others when smartphone users answer surveys e.g. while travelling on public transport or in other public spaces. However, early research has shown that this concern may (also) be overblown (see e.g., Antoun et al., 2017; Wenz et al. 2019; De Bruijne and Ouedjans, 2015). Other types of effects not directly related to experience, e.g., socially desirable responding, are not likely to be different between devices; if anything, such effects may, in any case, be less of a concern in internet panel surveys compared to other types of survey modes (Lindhjem and Navrud, 2011a).

but with respondents being professional survey-takers (Sandorf et al. 2020). This issue has been a lingering concern in the survey methodology literature for some time (Zhang et al. 2020), though not thoroughly investigated in the SP literature to date. Ensuring sufficient internet panel survey data quality in SP research in general should, in our opinion, receive more attention as this is important both for validity of experimental results and the use of derived welfare estimates for cost–benefit analysis and other uses. Further investigations into which internet panel-related factors may influence stated preferences in adverse ways and by how much are important to guide SP researchers when choosing survey companies, commissioning surveys and trying to control and limit the influence of such factors. Stricter requirements should be put on the transparency and procedures of internet panel management, e. g. the recruitment processes, length of panel membership, number of surveys allowed per respondents over time, incentives for recruitment and per survey, types of devices used and interface robustness across platforms. When facing such requirements, survey companies will also have to adhere to stricter management and reporting regimes pertaining to indicators of quality in their panel management and survey data collection. This iterative process will hopefully improve the survey data quality from internet panels over time, including smartphone responding, as it seems there is no going back to in-person interviews, as in the past due to high costs and for other reasons. Further, future guidelines on SP, such as Johnston et al. (2017), should consider more thorough guidance on the data collection process and how to commission surveys from internet panels (such as e.g. Baker et al. (2010) for general surveys) moving more of this process out of the “black box” and into the control of SP researchers. Until then, SP researchers are advised to seek known well-maintained internet panels that replace respondents at regular intervals to avoid a culture of “professional responding”, ensure that the survey instrument is optimized to also be easily navigated on a smartphone, and that provide reliable paradata on e.g. types of devices used.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10640-022-00654-5>.

Acknowledgements We are grateful for funding from the Norwegian Research Council Projects no. 280393, 26842 and 235560. Erlend Dancke Sandorf acknowledges funding the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement 793163-INSPIRE.

Funding Open access funding provided by Norwegian University of Life Sciences.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Angrist J, Azoulay P, Ellison G, Hill R, Lu SF (2017) Economic research evolves: fields and styles. *Am Econ Rev* 107:293–297. <https://doi.org/10.1257/aer.p20171117>

- Antoun C, Couper MP, Conrad FG (2017) Effects of mobile versus pc web on survey response quality a crossover experiment in a probability web panel. *Public Opin Q* 81:280–306. <https://doi.org/10.1093/poq/nfw088>
- Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, Dillman D, Frankel MR, Garland P, Groves RM, Kennedy C, Krosnick JA, Lavrakas PJ, Lee S, Link M, Piekarski L, Rao K, Thomas RK, Zahs D (2010) AAPOR Report on online panels. *Public Opin Q* 74:711–781. <https://doi.org/10.2307/40927166>
- Biemer PP, de Leeuw ED, Eckman S, Edwards B, Kreuter F, Lyberg LE, Tucker NC, West BT (2017) Total survey error in practice: improving quality in the era of big data. Wiley, New Jersey
- Börger T (2016) Are fast responses more random? Testing the effect of response time on scale in an online choice experiment. *Environ Resour Econ* 65:389–413. <https://doi.org/10.1007/s10640-015-9905-1>
- Boyle KJ, Morrison M, MacDonald DH, Duncan R, Rose J (2016) Investigating internet and mail implementation of stated-preference surveys while controlling for differences in sample frames. *Environ Resour Econ* 64:401–419. <https://doi.org/10.1007/s10640-015-9876-2>
- Callegaro M, Villar A, Yeager D, Krosnick JA (2014) A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples, in: *Online Panel Research: A Data Quality Perspective*. Wiley, Hoboken, NJ, pp. 23–55
- Campbell D, Mørkbak MR, Olsen SB (2017) Response time in online stated choice experiments: the non-triviality of identifying fast and slow respondents. *J Environ Econ Policy* 6:17–35. <https://doi.org/10.1080/21606544.2016.1167632>
- Clement SL, Severin-Nielsen MK, Shamshiri-Petersen D (2020) Device effects on survey response quality. A comparison of smartphone, tablet and PC responses on a cross sectional probability sample in *Survey Methods: Insights from the Field*. Special issue: 'Advancements in Online and Mobile Survey Methods'. Retrieved from <https://surveyinsights.org/?p=13585>
- Couper MP, Antoun C, Mavletova AM (2017) Mobile web surveys: a total survey error perspective. *Total survey error in practice: improving quality in the era of big data*. Wiley, New York, NY, pp 133–153
- de Leeuw ED (2018) Mixed-mode: past present and future. *Surv Res Methods* 12:75–89
- Greene WH, Hensher DA (2003) A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp Res Part B Methodol* 37:681–698. [https://doi.org/10.1016/S0191-2615\(02\)00046-2](https://doi.org/10.1016/S0191-2615(02)00046-2)
- Grimsrud K, Graesse M, Lindhjem H (2020) Using the generalised Q method in ecological economics: a better way to capture representative values and perspectives in ecosystem service management. *Ecol Econ* 170:106588. <https://doi.org/10.1016/j.ecolecon.2019.106588>
- Hausman J, McFadden D (1984) Specification tests for the multinomial logit model. *Econometrica* 52:1219. <https://doi.org/10.2307/1910997>
- Hensher DA, Rose JM, Greene WH (2005) *Applied choice analysis: a primer*. Cambridge University Press, Cambridge
- 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, New Jersey, pp 219–238
- Iversen EK, Lindhjem H, Jacobsen JB, Grimsrud K (2019) Moving (back) to greener pastures? Social benefits and costs of climate forest planting in Norway. *Land Use Policy*. <https://doi.org/10.1016/j.landusepol.2019.104390>
- Johnston RJ, Boyle KJ, Adamowicz W, Bennett J, Brouwer R, Cameron TA, Hanemann WM, Hanley N, Ryan M, Scarpa R, Tourangeau R, Vossler CA (2017) Contemporary guidance for stated preference studies. *J Assoc Environ Resour Econ* 4:319–405. <https://doi.org/10.1086/691697>
- Keusch F, Yan T (2017) Web versus mobile web: an experimental study of device effects and self-selection effects. *Soc Sci Comput Rev* 35(6):751–769
- Krosnick J (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Appl Cogn Psychol* 5:213–236
- Kube R, Löschel A, Mertens H, Requate T (2018) Research trends in environmental and resource economics: Insights from four decades of JEEM. *J Environ Econ Manag* 92:433–464. <https://doi.org/10.1016/j.jeem.2018.08.001>
- Kuhfeld WF (2007) Marketing research methods in SAS: Experimental design, choice, conjoint, and graphical techniques, in: *SAS Document TS-694*
- Liebe U, Glenk K, Oehlmann M, Meyerhoff J (2015) Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys? *J Choice Model* 14:17–31. <https://doi.org/10.1016/j.jocm.2015.02.002>
- Lindhjem H, Navrud S (2011a) Are internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecol. Econ., Special Section - Governing the Commons: Learning from Field and Laboratory Experiments*, 70:1628–1637, <https://doi.org/10.1016/j.ecolecon.2011a.04.002>

- 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. <https://doi.org/10.1561/101.00000045>
- Lutig P, Toepool V (2016) The use of PCs, smartphones, and tablets in a probability-based panel survey: effects on survey measurement error. *Soc Sci Comput Rev* 34:78–94. <https://doi.org/10.1177/0894439315574248>
- Mannestø G, Loomis JB (1991) Evaluation of mail and in-person contingent value surveys: Results of a study of recreational boaters. *J Environ Manage* 32:177–190. [https://doi.org/10.1016/S0301-4797\(05\)80032-5](https://doi.org/10.1016/S0301-4797(05)80032-5)
- Menegaki AN, Olsen SB, Tsagarakis KP (2016) Towards a common standard—A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. *J Choice Model* 18:18–50. <https://doi.org/10.1016/j.jocm.2016.04.005>
- Olsen SB (2009) Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environ Resour Econ* 44:591–610. <https://doi.org/10.1007/s10640-009-9303-7>
- Peterson G, Griffin J, York S (2017) Smartphone participation in web surveys, in: *A Total Survey Error Perspective in Total Survey Error in Practice: Improving Quality in the Era of Big Data*. Wiley, New York, NY
- Poynter R, Williams N, York S (2014) *The handbook of mobile market research: tools and techniques for market researchers*. Wiley, New Jersey
- R Core Team (2016) R: A language and environment for statistical computing
- Sandorf ED (2019) Did you miss something? inattentive respondents in discrete choice experiments. *Environ Resour Econ* 73:1197–1235. <https://doi.org/10.1007/s10640-018-0296-y>
- Sandorf ED, Aanesen M, Navrud S (2016) Valuing unfamiliar and complex environmental goods: a comparison of valuation workshops and internet panel surveys with videos. *Ecol Econ* 129:50–61. <https://doi.org/10.1016/j.ecolecon.2016.06.008>
- Sandorf ED, Persson L, Broberg T (2020) Using an integrated choice and latent variable model to understand the impact of “professional” respondents in a stated preference survey. *Resour Energy Econ* 61:101178. <https://doi.org/10.1016/j.reseneeco.2020.101178>
- Schlosser S, Mays A (2018) Mobile and dirty: does using mobile devices affect the data quality and the response process of online surveys? *Soc Sci Comput Rev* 36:212–230. <https://doi.org/10.1177/0894439317698437>
- Skeie MA, Lindhjem H, Skjeflo S, Navrud S (2019) Smartphone and tablet effects in contingent valuation web surveys—No reason to worry? *Ecol Econ* 165:106390. <https://doi.org/10.1016/j.ecolecon.2019.106390>
- Swait J, Louviere J (1993) The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res* 30:305–314. <https://doi.org/10.2307/3172883>
- Toepol V, Lutig P (2014) What happens if you offer a mobile option to your web panel? Evidence from a probability based panel of internet users. *Soc Sci Comput Rev* 32(4):544–560
- Tourangeau R, Sun H, Yan T, Maitland A, Rivero G, Williams D (2018) Web surveys by smartphones and tablets: effects on data quality. *Soc Sci Comput Rev* 36:542–556. <https://doi.org/10.1177/089443931719438>
- Wenz A (2019) Do distractions during web survey completion affect data quality? Findings from a laboratory experiment. *Soc Sci Comput Rev*. <https://doi.org/10.1177/0894439319851503>
- Zhang C, Antoun C, Yan HY, Conrad FG (2020) Professional respondents in opt-in online panels: what do we really know? *Soc Sci Comput Rev* 38:703–719. <https://doi.org/10.1177/0894439319845102>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.