

Scotland's Rural College

A web survey application of real choice experiments

Liebe, U; Glenk, K; von Mayer-Hofer, M; Spiller, A

Published in:

Journal of Choice Modelling

DOI:

[10.1016/j.jocm.2018.07.003](https://doi.org/10.1016/j.jocm.2018.07.003)

Print publication: 01/12/2019

Document Version

Peer reviewed version

[Link to publication](#)

Citation for published version (APA):

Liebe, U., Glenk, K., von Mayer-Hofer, M., & Spiller, A. (2019). A web survey application of real choice experiments. *Journal of Choice Modelling*, 33, [100150]. <https://doi.org/10.1016/j.jocm.2018.07.003>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A Web Survey Application of Real Choice Experiments

Ulf Liebe^{a,*}, Klaus Glenk^b, Marie von Meyer-Höfer^c, Achim Spiller^c

^a University of Warwick
Department of Sociology
Social Science Building
Coventry CV4 7AL
UK
E-mail: Ulf.Liebe@warwick.ac.uk
* Corresponding author

^b Scotland's Rural College (SRUC)
Land Economy, Environment and Society Group
West Mains Road
Edinburgh EH9 3JG
UK

^c Georg-August-Universität Göttingen
Department of Agricultural Economics and Rural Development
Chair of Marketing for Food and Agricultural Products
Platz der Göttinger Sieben 5
37073 Göttingen
Germany

Abstract:

This research note presents the first study to implement a real choice experiment in a web survey. In a case study on ethical food consumption, we find statistically significant lower willingness-to-pay values for the attributes “organic production” and “fair trade” in a choice experiment involving real payments compared to a choice experiment without real payments. This holds only true for respondents who are prepared to provide their personal details in order to deliver the product (83% of the sample), providing further evidence that lack of consequentiality can be an important source of validity problems. The implementation of a real choice experiment online proves useful and can form the baseline for future tests of the effectiveness of *ex ante* approaches such as cheap talk or honesty priming as well as consequentiality scripts in web-based choice experiments.

Keywords: consequentiality; ethical food consumption; hypothetical bias; validity; stated preferences; willingness to pay

JE: C92; D12; H41; Q5; Q13

1 Introduction

Over the past decades, the use of stated choice experiments (SCEs), here defined as multifactorial survey experiments with repeated binary or multinomial choice questions, has considerably increased in economics and other social sciences. Originally developed in marketing and transportation economics (see for example: Louviere and Hensher 1982; Louviere and Woodworth 1983; Louviere 1988 and 1992), it has become a popular method in environmental, health and agricultural economics for eliciting preferences and estimating willingness-to-pay (WTP) for private and public goods that are currently not traded in markets (Louviere, Hensher and Swait 2000). This includes environmental public goods (Adamowicz, Glenk and Meyerhoff 2014), health services (Ryan, Gerard and Amaya-Amaya 2008), and novel food and consumer products (Alfnes and Rickertsen 2011). In the case of consumer products, choice experiments are often used to investigate the market potential of products that have not yet been introduced, and to estimate consumers' WTP for novel product characteristics, including health, environmental and ethical attributes (Lusk and Schroeder 2004). In the context of ethical consumption, choice experiments are also used to investigate and explain differences in preferences and WTP values for existing consumer products such as organic and fair trade coffee (Andorfer and Liebe 2013; Rousseau 2015).

Following general trends in social research, SCEs are increasingly being conducted online (Liebe et al. 2015; Menegaki, Olsen and Tsagarakis 2016), and there is an interest in investigating the differences between web-based SCEs and those using other survey modes, for example regarding sample characteristics, response rates, and preferences (Olsen 2009; Determann et al. 2017). We see a similar trend in behavioral economics where experiments are increasingly being conducted online, for example using Amazon's MTurk, and results are compared to those from laboratory experiments (Paolacci and Ipeirotis 2010; Hauser and Schwarz 2016). When compared with other research settings and survey modes, web surveys have the advantage that a larger sample representing a wider population with respect to socio-demographic characteristics such as age, education, and income, can be realized in short time and in a cost effective manner.

Irrespective of whether conducting a survey online or by another mode, the divergence of hypothetically stated and actual preferences and willingness to pay has been a major concern (e.g., List and Gallet 2001; Harrison 2006; Chang, Lusk and Norwood 2009; Hensher 2010; Fifer, Rose and Greaves 2014). Meta-analyses of the extent of hypothetical bias in stated preference methods that largely draw on contingent valuation studies find a mean ratio of hypothetical to actual willingness to pay of about 3 (List and Gallet 2001) and median ratios of 1.35 (Murphy et al. 2005) and 1.39 (Penn and Hu 2018); they further show that the hypothetical bias tends to be lower if private instead of public goods are valued, and if a choice-based format rather than a market-based elicitation mechanism (various types of auction and open-ended formats) is used to elicit WTP, possibly contributing to a persistent belief that SCEs may be less susceptible to hypothetical bias compared to other stated preference elicitation formats (Hoyos 2010), although actual findings are limited and mixed (e.g., Carlsson and Martinsson 2001; Lusk and Schroeder 2004).

In this research note we demonstrate that choice experiments can be conducted with real payments in an online survey and compare results with a purely hypothetical SCE. We argue that, complementing other approaches, web-based real choice experiments can be

used to study the validity of SCE results. By real choice experiment we refer to “[...] a straightforward extension of a stated choice experiment by including real economic incentives. As in stated choice experiments, participants are asked to make choices in a series of choice scenarios. [...] To induce real economic incentives, one of the choice scenarios is randomly drawn as binding, and the choices made in that scenario are implemented. The participants pay the price and receive the product chosen in the binding scenario” (Alfnes and Rickertsen 2011: 222). Using the terminology of Collins and Vossler (2009), this (making one choice binding) represents a typical provision rule aimed at increasing the consequentiality of the experiment.

Historically, many stated preference studies had been set up as purely hypothetical exercises. This was criticized for lacking incentive compatibility. As a consequence, it is now becoming best practice to emphasize consequentiality of SCE studies as a necessary condition of incentive compatibility (Johnston et al. 2017). Consequentiality implies that there is a non-zero probability that responses are perceived to affect (policy) outcomes and that respondents will be asked to pay for the implemented outcome. This is expected to increase the external validity of study results. The real-payment approach can be seen useful, because an actual exchange between money and the good at hand is still the best way of ensuring consequentiality in a stated preference survey: “Given equivocal evidence supporting any one (or combination) of these methods [ex ante approaches], we believe that the most promising ex ante approach remains a consequential design with a binding payment” (Johnston et al. 2017: 356).

SCEs are most useful to assess demand for private goods with novel (product) characteristics that cannot (yet) be purchased in the market place, and to elicit preferences for policies and projects with public good character. However, ensuring payment consequentiality is difficult or impossible in such contexts. Nevertheless, we argue that researchers can learn from demonstrating how results differ between incentivized (involving actual monetary costs) and non-incentivized (not involving actual monetary costs) SCEs and can use results of real choice experiments as a benchmark for testing the effectiveness of devices aimed at increasing perceived outcome consequentiality and *ex ante* procedures to enhance validity such as cheap talk (Cummings and Taylor 1999; Carlsson et al. 2005; Tonsor and Shupp 2011; Ladenburg and Olsen 2014) or honesty-based methods (de-Magistris et al. 2013; Carlsson et al. 2013; de-Magistris and Pascucci 2014; Howard et al. 2017). A contribution of this note is to demonstrate how such comparisons between incentivized and non-incentivized SCEs can be conducted online.

2 Experimental Design, Data, and Econometric Approach

Experimental Design

In order to implement a real choice experiment in a web survey, or more generally in a population survey, the product to be valued needed to have quality attributes that can be easily varied and are widely available on the market. Tea fulfills these criteria. It is easy to prepare in different product variations and unproblematic to be sent by mail. Tea is a staple product that is consumed regularly by a large part of the German population. Black tea is one of the most commonly consumed types of tea among German tea consumers (Deutscher Teeverband 2011). The Indian black tea variety Darjeeling First Flush was

chosen for its availability in different qualities in terms of production (conventional, organic, fair trade as well as organic and fair trade). A quantity of 10g was chosen for the tea products valued in the choice experiments, because it is the common sample package quantity offered at a price in specialty shops, sufficient for preparing at least five cups of tea.

The choice experiment included two generic alternatives (Tea A, Tea B) and an opt-out alternative (“none of these”). The alternatives were described by the three attributes: organic (attribute levels: no, yes), fair trade (attribute levels: no, yes), and price (attribute levels: €0.49, €0.69, €0.99, €1.19). The price levels were derived from the common tea specialty shops prices for 10g of the black tea variant Darjeeling First Flush. Figure 1 gives an example of a choice set as used in the survey.

[Figure 1 about here]

A fractional factorial main effects design was selected as an efficient design with local (fixed) priors (e.g., Scarpa and Rose 2008) using the software Ngene. First, a pretest based on an orthogonal design and a hypothetical choice experiment was conducted that resulted in 432 observations from 27 respondents (i.e., 16 choice sets per respondent) interviewed between December 2011 and January 2012. Data from this pretest was used to construct priors of parameter values as point estimates for an efficient design that minimised D_p -error for a conditional logit model. Dominant alternatives where a combined organic and fair trade tea is cheaper than an alternative tea product were excluded from the search. The design consisted of eight choice sets to be answered by each respondent. A second pretest in January 2012 based on a hypothetical choice experiment with 20 respondents resulted in statistically significant parameter estimates and hence the design was retained as the final design for the main study. Pretest observations were not used in the final analysis presented in this paper.

Participants in the hypothetical and real choice experiment received the same instructions and faced the same eight choice sets. The only difference between the two groups was that participants in the hypothetical choice experiment received the whole €2 amount for participating in the survey (which was credited to individuals’ accounts by the survey organization); in the real choice experiment, participants received their chosen product from a randomly drawn choice set and €2 minus the amount for the chosen product. We have reason to believe that respondents valued the additional endowment of €2 provided. This setup of including a seemingly low stake size is similar to experimental studies using Amazon Mechanical Turk, where it is common that a worker’s compensation is 0.50 US dollars or less (e.g., Buhrmester, Kwang and Gosling 2011; Clements et al. 2015). Those participants who chose a tea product and were willing to state a delivery address received their tea by mail within a maximum of six days after submitting the survey (see appendix for the experimental protocol). After the choice experiment, respondents were also asked to answer questions regarding, amongst others, their socio-demographic characteristics.

Data

The hypothetical and real choice experiment was implemented in a web survey that was carried out by a professional survey organization in 2012 between January 26 and February 24. Respondents were part of the organization’s online access panel and it can

therefore be assumed that they are experienced in answering survey questions online. Only members of the online access panel who were 18 years and older and who drink tea at least once a week were allowed to take part in the survey. Respondents were randomly assigned to the hypothetical or the real choice experiment.

[Table 1 about here]

We obtained 299 usable interviews in which the respondents completed all eight choice sets. Table 1 reports sample statistics for both the real and hypothetical choice experiment. Tests for group comparison (χ^2 tests, t-test, Mann-Whitney test) revealed no statistically significant differences between the two groups concerning gender composition, age, education, and income.

Econometric Approach

For analysing the choices between alternative variations of Darjeeling First Flush tea, random parameter logit models (RPL) are estimated in WTP space following Train and Weeks (2005) and Scarpa, Thiene and Train (2008). To account for the fact that respondents may treat the two tea alternatives systematically different to the opt-out alternative (Scarpa, Ferrini and Willis 2005), an additional error component (EC) was included in the utility function. Estimation of choice probabilities in the RPL models requires approximation through simulation (Train 2003), which were based on Modified Latin Hypercube Sampling with 500 draws (Hess, Train and Polak 2006).

The price attribute parameter is assumed to follow a lognormal distribution, the marginal WTP parameters of the *fair trade* and *organic* attributes are assumed to follow a normal distribution. Alternative specific constants (ASC1 and ASC2) were associated with the two alternatives and specified as fixed (i.e. non-random), after initial analyses revealed that standard deviation coefficients of normally distributed ASCs were not significant.

Differences in model results between hypothetical and real choice experiments are investigated using the procedure described by Swait and Louviere (1993). The procedure involves likelihood-ratio tests to assess if the estimated models share the same parameter vector after allowing for differences in scale between data of the two sub-samples. The assessment of differences in mean marginal WTP estimates for the attributes *organic* and *fair trade* is used to reveal the extent of potential hypothetical bias. Confidence intervals for mean marginal WTP estimates are calculated using the Krinsky and Robb (1986) bootstrapping procedure. The complete combinatorial test suggested by Poe, Giraud and Loomis (2005) is subsequently used to test for differences in mean marginal WTP for *organic* and *fair trade* between the independent samples (i.e., hypothetical versus real choice experiment).

3 Results

Table 2 presents the results of the choice models in WTP space. All models are highly statistically significant, and all coefficients are significantly different from zero. Consumers are, on average, willing to pay a premium if the tea is produced organically and is fairly traded. The statistically significant estimates for the standard deviation (SD) of the random parameters indicate the presence of taste heterogeneity for the attributes.

The coefficients of the ASCs, which are also estimated in WTP space, are positive and significant. For reasons not explained by the attributes, respondents showed a tendency to choose the tea alternatives over the opt-out alternative. At a descriptive level, 3% of the respondents in the choice experiment without real payments and 12% in the choice experiment with real payments always chose the opt-out alternative, respectively. There is thus a lower tendency to participate in the market in the case of real consequences. The significant coefficient of the error component EC indicates correlation between the stochastic parts of the utility functions for the tea alternatives. For reasons not explained through the systematic part of utility (ASCs, attributes), respondents perceive the tea alternatives as more similar relative to the opt-out alternative.

According to the Swait and Louviere (1993) test, the parameter vector between hypothetical and real choice experiments (Table 2, two leftmost models) differs significantly, providing a first indication of differences in model results [$\chi^2 = -2[\text{LogL}_{\text{Pooled/Scale}} - (\text{LogL}_{\text{Hypothetical}} + \text{LogL}_{\text{Real}})] = 24.2$; d.f. 11; $\text{Pr}(\chi^2 \leq 19.68) = 0.05$]. The differences in mean marginal WTP (mWTP) for *organic* and *fair trade* between hypothetical and real treatments are substantial and point to the presence of hypothetical bias. The estimated mean marginal WTP for organic production amounts to €0.45 in the hypothetical treatment and €0.30 in the real treatment. The corresponding values for tea produced under fair trade conditions are €0.60 and €0.43. Figure 2 shows the WTP distributions for both attributes. The Poe, Giraud and Loomis (2005) test indicates that mean marginal WTP estimates for the two attributes are significantly different at the 5% level [$\text{mWTP}_{\text{hypothetical}} \geq \text{mWTP}_{\text{real}}$; p-value (organic): 0.017; p-value (fair trade): 0.002].

[Table 2 about here]

[Figure 2 about here]

One specific characteristic of our study is that some respondents (17%) in the real treatment refused to provide their address in order to obtain the chosen tea product. This might have several reasons. Respondents did not want to provide personal information and/or doubted confidentiality. Another reason is that respondents have made their choices without considering that they would become consequential. A comparison of respondents who provided an address relative to those who did not thus reveals some evidence on the role of lack of consequentiality as an important source of hypothetical bias. In order to test for systematic differences between those respondents who provided an address in the real choice experiment and those who did not, models were estimated for both groups separately. The results (Table 2, two rightmost models) reveal that both groups are willing to pay a premium for *organic* and *fair trade*. However, those who provided an address have lower mean marginal WTP values compared to those who did not provide an address (€0.35 vs. €0.49 for organic production and €0.40 vs. €0.79 for fair trade). A Poe, Giraud and Loomis (2005) test reveals that these differences in mean marginal WTP between the sample who provided an address in the real choice experiment and those respondents who did not provide an address are statistically non-significant for organic production ($\text{mWTP}_{\text{address_no}} \geq \text{mWTP}_{\text{address_yes}}$; p=0.152) and highly significant for fair trade (p<0.001). Comparing the sample in the hypothetical treatment with those who did not provide an address, marginal WTP values are larger for the latter group (€0.45 vs. €0.49 for organic production and €0.60 vs. €0.79 for fair trade). A Poe, Giraud and Loomis (2005) test indicates a statistically non-significant difference in mean marginal WTP between the hypothetical treatment group and those who did not provide

an address for organic production ($mWTP_{\text{address_no}} \geq mWTP_{\text{hypothetical}}$; $p=0.376$) and a significant difference for fair trade ($p=0.013$).

Additional analyses based on survey items (five-point response scales) show that respondents who provided an address do not express a significantly lower/higher importance of environmental protection [$\chi^2(4)=2.13$, $p=0.711$] and fair trade [$\chi^2(4)=4.40$, $p=0.355$] when purchasing groceries than those who did not provide an address. This indicates that differences in stated preferences between those who provided an address and those who did not provide an address were not driven by differences in concern for environmental protection and fair trade.

In relative terms, i.e. $mWTP_{\text{Hypothetical}} / mWTP_{\text{Real}}$, hypothetical and real marginal WTP, diverge by a factor of 1.4 (range¹: 1.01; 1.93) for the attribute fair trade and 1.5 (range: 0.88; 2.54) for organic production. These values are considerable, and comparable to the median ratios of hypothetical and real WTP of 1.35 and 1.39 reported by Murphy et al. (2005) and Penn and Hu (2018), respectively. However, they are lower than the median ratio of 2.66 reported in Murphy et al. (2005) after excluding those observations that were using *ex-ante* methods aimed at reducing hypothetical bias.

4 Discussion and Conclusions

In this research note, we present the first study comparing the results of a hypothetical and real choice experiment using a web-based survey. Based on a hypothetical and real choice experiment on the purchase of organic and fair trade tea, lower marginal WTP values are found for the real treatment compared to the hypothetical one. Our results also provide context to studies that have investigated the effectiveness of *ex ante* approaches that aim to mitigate hypothetical bias in web-based choice experiments through survey design (e.g., Tonsor and Schupp 2011; Lin, Ortega and Caputo 2018). Such studies typically have no meaningful reference against which to compare the performance of different *ex ante* measures.

In the real treatment of our study, 83% of the respondents provided their address in order to obtain the product and 17% refused to provide their address. We informed respondents before making their choices that they will be asked to provide their address after all choices have been made and the binding choice that will result in an actual transaction has been randomly selected. Those who refused to provide their address somewhat disregarded the experimental script in the real treatment. This is supported by differences in WTP compared to the sample that provided an address, and compared to the hypothetical sample. There is mixed evidence that information regarding the consequentiality of survey results influences consequentiality beliefs and stated preferences. For example, Czajkowski et al. (2017) found in a web-survey based choice experiment that information treatments emphasizing consequentiality regarding future policies of financing theatres in Warsaw did not significantly affect stated beliefs of consequentiality and only had small effects on WTP for proposed programs for discounted theater tickets for Warsaw citizens. Lewis, Grebitus, and Nayga (2016)

¹ Based on 95% confidence intervals for attributes in both treatments.

investigated the effect of a policy consequentiality script on the likelihood of choosing a product option in an online choice experiment on U.S. consumer WTP for imported and genetically modified sugar. The script increased the belief in the consequentiality of survey responses and significantly decreased the likelihood to choose the opt-out option; yet it did not significantly affect WTP values. Oehlmann and Meyerhoff (2017) tested different policy consequentiality scripts in an online choice experiment on renewable energy expansion in Germany. They found that the scripts significantly increased perceived consequentiality, but did not significantly affect WTP. These studies show mixed evidence regarding possibilities to shape respondents' beliefs by using consequentiality scripts with small or insignificant effects on WTP. The present study varied actual consequentiality and found substantive and significant differences in WTP for one of the product's attributes. Further, in our study the frequency of opt-out choices increased when using an actual consequentiality script. It might be worthwhile to build on this finding and to compare different consequentiality scripts in a study also including real payments and a measurement of respondents' perceived consequentiality.

Testing the extent of hypothetical bias in web surveys offers a possibility of overcoming sample limitations in terms of sample size and in terms of representing the wider population with respect to socio-demographic characteristics such as age, education, and income at comparably low costs. Yet, there are also disadvantages to conducting web surveys. First, this study relies on a specific sample; that is, an online access panel. The results of any experimental study in general, and choice experiments in particular, might depend on whether the respondents have experience in answering experimental tasks (e.g., a student pool for experiments) and survey questions (e.g., online access panel) or whether they have no experience (typical for a sample drawn from the general population). In the context of food choice, results may also be sensitive to whether respondents were able to actually experience the goods on offer (e.g., MacDonald et al. 2016). The sample size in our study is also rather small. Further studies, which base real choice experiments on representative samples from the general population, are needed. This is important because many choice experiments are used to reveal monetary values and market shares for private, quasi-public or public goods, and are subsequently used to inform (policy) decision makers. Therefore, it is important that the results represent the preferences of the whole population at hand and not the preferences of a biased sample of the population; that is, for estimating the extent of hypothetical bias, studies should be based on random samples from the general public. Students and convenience samples, for example, might systematically differ from other groups in society in their valuation of specific goods (see Belot et al. 2015 for a systematic comparison of the behavior of students and non-students in economic experiments; Fréchet 2016 provides a review of experimental studies using highly demographically varied samples suggesting that age and, to a lower extent, education are relevant factors for explaining differences between student and non-student samples, albeit comparative statics, but not point estimates, are rather similar across samples).

Second, the actual tradeoff between the good at hand and money is more difficult to implement in web surveys than in other experimental settings such as laboratory and field settings. One reason for this is that respondents, as was the case in our study, have to provide a delivery address. Respondents who place low trust in survey organizations may deny doing so. Indeed, our results show that respondents who refuse to provide an address have a WTP that is similar or greater in magnitude to the one estimated from the

hypothetical choice experiment. While requiring a delivery address poses a challenge, most online surveys are based on access panels, whose members trust the survey organization which also importantly provides the monetary reimbursement for participating in the survey. This lowers the likelihood of trust issues in online experiments (in our case only 17% refused to provide an address). Another reason is that respondents will receive the product in the future and, again, not all individuals might believe that they will actually receive the product. Therefore, the presence of a projection bias is more likely (Loewenstein, O'Donoghue and Rabin 2003); participants might prefer to receive the product immediately. Again, such effects might be less severe if the respondents are members of an access panel. However, the exchange of goods and money can be implemented more easily in laboratory and field settings in which respondents do not have to provide personal information (albeit the experimenter may know the address, telephone number etc.) and receive the good immediately upon completion of the experiment.

Our study has demonstrated the potential of web surveys for studying the validity of choice experiments and offers the first benchmark for web-based choice experiments using actual transactions. It is a clear limitation that the approach cannot be easily adopted for most public goods study settings due to difficulties with implementing actual consequentiality. Our study highlights the relevance of consequentiality in a case where a real payment approach is implemented online. This can be informative for stated preference studies on private goods in areas such as environmental valuation, health, transportation, and ethical consumption. In future studies, the approach of actual consequentiality presented in this note can be compared with other scripts that are intended to induce policy and payment consequentiality. Such comparisons might also provide useful insights for public good studies. Regarding the implementation of real payments, using Amazon Mechanical Turk (see, e.g., Paolacci, Chandler and Ipeirotis 2010) or other platforms would be a comparable method to the one proposed in our study and might be a promising alternative in future studies on the validity of web-based stated choice experiments. We hope that our study paved the way for further applications in this area.

Appendix

Experimental Protocol

Respondents in the web survey were randomly assigned to the hypothetical choice experiment or the real choice experiment. Irrespective of whether they were placed in the first or in the second group, all participants were provided with the following instructions (translated from the original version in German):

“On the following pages we will present different tea products to you. We are interested in which tea you would purchase.

Many different kinds of teas are available nowadays. Here, we are concerned with loose tea, not tea bags. You can choose between different kinds of tea: tea from conventional production, organic tea, which has been produced without the use of chemical and or synthetic fertilizers and pesticides, and fair trade tea, which was obtained from trade with small-scale farmers in developing countries, who were paid enough to cover their living costs.

The choices all concern Darjeeling First Flush, a black tea from India; its aroma is described as mild and flowery.

In the following you will find 8 purchase situations in which you can choose between two teas (A and B).

The offered quantity of 10g is equivalent to approximately 5 tea bags, that is, it is sufficient for the preparation of at least 5 cups of tea. You can choose between teas that differ in growth method (organic yes/no), production (fair trade yes/no), and price (€0.49; €0.69; €0.99; €1.19).

In each of the purchase situations, we would like to know which 10g package of Darjeeling First Flush you would purchase.

Please make your choice for each of the situations independently of one another. Mark the tea product you would purchase or whether you abstain from a purchase. Tick only one tea product per page.”

Those participants who were assigned to the *hypothetical choice experiment* group additionally received the following instruction:

“At the end of the survey you will receive €2 in the form of bonus points on your opinion-people account as a thank you for your participation.

Please assume that you can spend these €2 in each of the following purchase situations.”

Those participants who were assigned to the *real choice experiment* group additionally received the following instruction:

“At the end of the survey you will receive €2 in form of bonus points on your opinion-people account as a thank you for your participation.

After you have made all 8 decisions, one of your 8 purchase decisions will be randomly selected. If, in this scenario, you have chosen one of the teas (A or B), you will get this tea by mail and its price will be deducted from your €2 bonus points.

You will actually purchase the tea! For the mail delivery, we kindly ask you to provide us with a delivery address.

All of your personal data will be treated anonymously and will be deleted directly after dispatch.

If in the randomly chosen purchase scenario you have chosen none of the teas, you will get the whole €2 accredited to your opinion-people account.”

References

Adamowicz, W., K. Glenk, and J. Meyerhoff. 2014. “Choice Modelling Research in Environmental and Resource Economics.” In S. Hess, and A. Daly, eds. *Handbook of Choice Modelling*. Cheltenham and Northampton: Edward Elgar, pp. 661–674.

Alfnes, F., and K. Rickertsen. 2011. “Non-market Valuation: Experimental Methods.” In J.L. Lusk, J. Roosen, and J.F. Shogren, eds. *The (Oxford) Handbook of the Economics of Food Consumption and Policy*. Oxford: Oxford University Press, pp. 215 – 242.

Andorfer, V.A., and U. Liebe. 2013. “Consumer behavior in moral markets On the relevance of identity, justice beliefs, social norms, status, and trust in ethical consumption.” *European Sociological Review*, 29(6): 1251–1265.

Belot, M., R. Duch, and L. Miller. 2015. “A comprehensive comparison of students and non-students in classic experimental games.” *Journal of Economic Behavior & Organization* 113: 26–33.

Buhrmester, M., T. Kwang, and S.D. Gosling. 2011. “Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data?” *Perspectives on Psychological Science* 6: 3-5.

Carlsson, F., and P. Martinsson. 2001. “Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment.” *Journal of Environmental Economics and Management* 41 (2): 179–192.

Carlsson, F., P. Frykblom, and C.J. Lagerkvist. 2005. “Using cheap-talk as a test of validity in choice experiments.” *Economics Letters* 89: 147–52.

Carlsson, F., M. Kataria, A. Krupnick, E. Lampi, A. Lofgren, P. Qin, and T. Sterner. 2013. The truth, the whole truth, and nothing but the truth: A multiple country test of an oath script. *Journal of Economic Behavior and Organization* 89: 105–21.

Chang, J.B., J. L., Lusk, and F.B. Norwood. 2009. “How closely do hypothetical surveys and laboratory experiments predict field behavior?” *American Journal of Agricultural Economics* 91 (2): 518–534.

Clements, J. M., A.M. McCright, T. Dietz, and S.T. Marquart-Pyatt. 2015. "A behavioural measure of environmental decision-making for social surveys." *Environmental Sociology* 1: 27-37.

Collins, J.P., and C.A. Vossler. 2009. Incentive Compatibility Tests of Choice Experiment Value Elicitation Questions. *Journal of Environmental Economics and Management* 58(2): 226-235.

Cummings, R.G., and L.O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89: 649–665.

Czajkowski, M., C.A. Vossler, W. Budziński, A. Wiśniewska, and E. Zawajska. 2017. "Addressing empirical challenges related to the incentive compatibility of stated preferences methods." *Journal of Economic Behavior & Organization* 142: 47-63.

de-Magistris, T., G. Azucena, and R.M. Nayga. 2013. "On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments." *American Journal of Agricultural Economics* 95(5): 1136–54.

de-Magistris, T., and S. Pascucci. 2014. "The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study." *Economics Letters* 123(2): 252–255.

Determann, Domino, M.S. Lambooi, E.W. Steyerberg, E.W. de Bekker-Grob, and G.A. de Wit GA. 2017. "Impact of Survey Administration Mode on the Results of a Health-Related Discrete Choice Experiment: Online and Paper Comparison." *Value in Health* 20(7): 953-960.

Deutscher Teeverband. 2011. *Tee als Wirtschaftsfaktor. Aktuelle Zahlen Stand Juni 2011*. http://www.teeverband.de/jahresbericht_2010.pdf. Accessed 25 May 2012.

Fifer, S., J.M. Rose, and S.P. Greaves. 2014. "Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it?" *Transportation Research Part A: Policy and Practice* 61: 164–77.

Fréchette, G.R. 2016. "Experimental economics across subject populations." In J.H. Kagel and A.E. Roth, eds. *The Handbook of Experimental Economics*, Volume 2. Princeton, NJ: Princeton University Press, pp. 435-480.

Harrison, G. W. 2006. "Experimental Evidence on Alternative Environmental Valuation Methods." *Environmental and Resource Economics* 34 (1): 125–162.

Hauser, D.J., and N. Schwarz. 2016. "Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants". *Behavior Research Methods* 48(1): 400–407.

- Hensher, D. A. 2010. "Hypothetical bias, choice experiments and willingness to pay." *Transportation Research Part 44* (6): 735–752.
- Hess, S., K. E. Train, and J. W. Polak. 2006. "On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice." *Transportation Research Part B: Methodological* 40 (2): 147–163.
- Howard, G., B.E. Roe, E.C. Nisbet, and J.F. Martin. 2017. "Hypothetical Bias Mitigation Techniques in Choice Experiments: Do Cheap Talk and Honesty Priming Effects Fade with Repeated Choices?" *Journal of the Association of Environmental and Resource Economists* 4(2), 543–573.
- Hoyos, D. 2010. "The state of the art of environmental valuation with discrete choice experiments." *Ecological Economics* 69 (8): 1595–1603.
- Johnston, R.J., K.J. Boyle, W. Adamowicz, J. Bennett, R. Brouwer, T.A. Cameron, W.M. Hanemann, N. Hanley, M. Ryan, R. Scarpa, R. Tourangeau, and C.A. Vossler. 2017. "Contemporary Guidance for Stated Preference Studies." *Journal of the Association of Environmental and Resource Economists* 4 (2): 319-405.
- Krinsky, I., and L. Robb. 1986. "On approximating the statistical properties of elasticities." *The Review of Economics and Statistics* 68 (4): 715–719.
- Ladenburg, J., and S.B. Olsen. 2014. "Augmenting Short Cheap Talk Scripts with a Repeated Opt-Out Reminder in Choice Experiment Surveys." *Resource and Energy Economics* 37: 39-63.
- Lewis, K.E., C. Grebitus, and R.M. Nayga, Jr. 2016. "U.S. consumers' preferences for imported and genetically modified sugar: Examining policy consequentiality in a choice experiment." *Journal of Behavioral and Experimental Economics* 65: 1-8.
- List, J.A., and C.A. Gallet. 2001. "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resources Economics* 20 (3): 241-254.
- Liebe, U., K. Glenk, M. Oehlmann, and J. Meyerhoff. 2015. "Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys?" *Journal of Choice Modelling* 14: 17–31.
- Lin, W., D.L. Ortega, and V. Caputo. 2018. Are Ex-Ante Hypothetical Bias Calibration Methods Context Dependent? Evidence from Online Food Shoppers in China. *The Journal of Consumer Affairs*, online first.
- Loewenstein, G., T. O'Donoghue, and M. Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4):1209–1248.
- Louviere, J.J. 1988. "Analysing Decision Making: Metric Conjoint Analysis." Sage University Paper No. 67, Newbury Park, Beverly Hills.

- Louviere, J.J. 1992. "Special issue on experimental choice analysis." *Journal of Business Research* 24 (2): 89–189.
- Louviere, J.J., and D.A. Hensher. 1982. "On the design and analysis of simulated or allocation experiments in travel choice modelling." *Transportation Research Record* 890: 11–17.
- Louviere, J.J., and G.G. Woodworth. 1983. "Design and analysis of simulated choice or allocation experiments: an approach based on aggregate data." *Journal of Marketing Research* 20 (4): 350–367.
- Louviere, J.J., D.A. Hensher, and J.D. Swait. 2000. *Stated choice methods: analysis and application*. Cambridge, UK: Cambridge University Press.
- Lusk, J.L., and T.C. Schroeder. 2004. "Are choice experiments incentive compatible? A test with quality differentiated beef steaks." *American Journal of Agricultural Economics* 86 (2): 467–482.
- MacDonald, D.H., J.M. Rose, H.J. Lease, and D.N. Cox. 2016. "Recycled wastewater and product choice: Does it make a difference if and when you taste it?" *Food Quality and Preference* 48: 283-292.
- Menegaki, A.N., S.B. Olsen, and K.P. Tsagarakis. 2016. "Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys." *Journal of Choice Modelling* 18: 18-50.
- Moore, R., R.C. Bishop, B. Provencher, and P.A. Champ. 2010. Accounting for Respondent Uncertainty to Improve Willingness-to-Pay Estimates. *Canadian Journal of Agricultural Economics* 58: 381–401.
- Murphy, J.J., P. G. Allen, T.H. Stevens, and D. Weatherhead. 2005. "A meta analysis of hypothetical bias in stated preference valuation." *Environmental Resource Economics* 30 (3): 313–325.
- Oehlmann, M., and J. Meyerhoff. 2017. "Stated Preferences Towards Renewable Energy Alternatives in Germany – Do the Consequentiality of the Survey and Trust in Institutions Matter?" *Journal of Environmental Economics and Policy* 6: 1-16.
- Olsen, S.B. 2009. "Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods." *Environmental and Resource Economics* 44: 591–610.
- Paolacci, P., J. Chandler, and P.G. Ipeirotis. 2010. "Running experiments on Amazon Mechanical Turk." *Judgment and Decision Making* 5(5): 411-419.
- Penn, J.M., and W. Hu. 2018. "Understanding Hypothetical Bias: An Enhanced Meta-Analysis." *American Journal of Agricultural Economics*, online first.

Poe G.L, K.L. Giraud, and J.B. Loomis. 2005. Computational Methods for Measuring the Difference of Empirical Distributions. *American Journal of Agricultural Economics* 87: 353-365.

Rousseau, S. 2015. "The Role of Organic and Fair Trade Labels when Choosing Chocolate." *Food Quality and Preference* 44: 92-100.

Ryan, M., K. Gerard, and M. Amaya-Amaya. 2008. *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht: Springer.

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

Scarpa, R., S. Ferrini, and K.G. Willis. 2005. "Performance of error component models for status-quo effects in choice experiments." In R. Scarpa, and A. Alberini, eds. *Applications of Simulation Methods in Environmental and Resource Economics*. Dordrecht: Springer, pp. 247–274.

Scarpa R., M. Thiene, and K. Train. 2008. "Utility willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps." *American Journal of Agricultural Economics* 90 (4): 994–1010.

Swait, J., and J. Louviere. 1993. "The role of the scale parameter in the estimation and comparison of multinomial logit models." *Journal of Marketing Research* 30: 305–314.

Tonsor, G.T., and R.S. Shupp. 2011. "Cheap talk scripts and online choice experiments: looking beyond the mean." *American Journal of Agricultural Economics* 93(4) :1015-1031.

Train, K.E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

Train K.E., and M. Weeks. 2005. "Discrete choice models in preference space and willingness-to-pay space" In R. Scarpa, and A. Alberini, eds. *Applications of Simulation Methods in Environmental and Resource Economics*. Dordrecht: Springer, pp. 1–16.

Table 1: Sample Characteristics

	Hypothetical choice experiment					Real choice experiment				
	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N
Gender (1=female)	0.54				151	0.55				139
Age in years	46.60	14.49	19	75	157	44.91	14.61	19	73	142
Education (1=higher)	0.45				157	0.52				142
Income in categories	2.81	1.16	1	5	139	2.73	1.14	1	5	124

Note: SD = standard deviation, Min = minimum, Max = maximum, N = number of respondents.
 Income in categories refers to household net income with 1 = "below 1.000 Euro," 2 = "1,000 -1,999 Euro," 3 = "2.000 - 2.999 Euro," 4 = "3.000 - 3.999 Euro," 5 = "more than 4.000 Euro."

Table 2: Model results

	Hypothetical (H)		Real (R)		Pooled		Pooled / Scale		R without address		R with address	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ASC1	0.863 *** (0.066)		0.512 *** (0.038)		0.716 *** (0.045)		0.711 *** (0.04)		0.790 *** (0.079)		0.643 *** (0.048)	
ASC2	0.860 *** (0.066)		0.545 *** (0.038)		0.734 *** (0.043)		0.731 *** (0.039)		0.930 *** (0.094)		0.663 *** (0.048)	
Organic (mWTP)	0.449 *** (0.057)	0.512 *** (0.050)	0.299 *** (0.04)	0.423 *** (0.047)	0.368 *** (0.037)	0.512 *** (0.047)	0.368 *** (0.034)	0.515 *** (0.043)	0.491 *** (0.137)	0.588 *** (0.109)	0.347 *** (0.05)	0.512 *** (0.047)
Fair Trade (mWTP)	0.598 *** (0.047)	0.456 *** (0.046)	0.428 *** (0.037)	0.401 *** (0.044)	0.542 *** (0.039)	0.452 *** (0.031)	0.541 *** (0.038)	0.452 *** (0.030)	0.788 *** (0.069)	0.651 *** (0.118)	0.400 *** (0.051)	0.345 *** (0.044)
Price	1.73 *** (0.094)	0.569 *** (0.099)	1.85 *** (0.106)	0.792 *** (0.110)	1.83 *** (0.077)	0.702 *** (0.072)	1.78 *** (0.097)	0.704 *** (0.071)	1.27 *** (0.296)	0.908 *** (0.227)	2.02 *** (0.117)	0.676 *** (0.100)
Error component	3.95 *** (0.527)		5.86 *** (0.807)		5.13 *** (0.509)		4.89 *** (0.583)		4.82 * (2.14)		5.62 *** (0.823)	
Scale _{hypothetical}							1.00					
Scale _{real}							1.11 [§] (0.156)					
Null Log- Likelihood	-1379.857		-1248.024		-2627.881		-2627.881		-210.934		-1037.090	
Log-likelihood at convergence	-933.034		-815.936		-1761.385		-1761.083		-146.649		-662.798	
Pseudo Rho2	0.324		0.346		0.330		0.330		0.305		0.361	
# Observations (# Respondents)	1256 (157)		1136 (142)		2392 (299)		2392 (299)		192 (24)		944 (118)	

Note: SD = standard deviation; mWTP = marginal willingness to pay; standard errors in parentheses; * $p[|Z|>z]<0.05$, ** $p[|Z|>z]<0.01$, *** $p[|Z|>z]<0.001$; [§] not significantly different from one.

Fig. 1 Example of a choice set

	Tea A (10g)	Tea B (10g)	
Organic	Yes	Yes	
Fair trade	Yes	No	
Price	1.19€	0.49€	
Which tea would you buy?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> None of these

Fig. 2 Marginal WTP distributions for organic and fair trade attributes estimated from the real and the hypothetical choice experiment data

