

Consequentiality and sample size in stated preferences

Pierre-Alexandre Mahieu^{*a}, Anthony Paris^b, Benoît Chèze^{b, c, d}, Pascal Gastineau^e, and Ewa Zawojńska^f

^aLEMNA, University of Nantes, Nantes, France.

^bEconomiX-CNRS, University of Paris Nanterre, Nanterre, France.

^cIFP Energies nouvelles, Rueil-Malmaison, France.

^dUMR Economie Publique, INRA-AgroParisTech, Thiverval-Grignon, France.

^eIFSTTAR, AME, EASE, Bouguenais, France.

^fFaculty of Economic Sciences, University of Warsaw, Warsaw, Poland.

March 15, 2019

Abstract

In stated preference studies, consequentiality is expected to increase the validity of the results, where consequentiality means that respondents believe that the survey outcome will have real-life consequences. Sample size has received little attention so far in this literature. In an online single-bounded dichotomous choice field study dealing with underwater turbines, we provide information on sample size to a part of participants, where the information varies across participants. We find that sample size information has no effect on subjective consequentiality, which suggests that consequential single-bounded contingent valuation studies can still ensure incentive-compatible behavior when the benefits from voting becomes very small.

Keywords: consequentiality, incentive compatibility, sample size, contingent valuation

1 Introduction

Discrete choice experiment (DCE) surveys, based on stated preferences, are employed in many fields, including environmental and health economics, transportation, to reveal the value of public goods to the society. The value estimates find

*Corresponding author, mahieu-pa@univ-nantes.fr .

29 application in various areas, such as assessment of benefits for cost-benefit analy-
30 ses of public policy projects or loss estimation in litigation processes over natural
31 damages.

32 However, there exist important concerns about the validity of stated preference
33 surveys: it has been argued that in case of many surveys, the best strategy for
34 rational participants when responding to the valuation task is not necessarily to re-
35 spond truthfully. To improve the validity of stated-preference-based value estimates,
36 Carson and Groves (2007) defined conditions necessary for truthful preference reve-
37 lation. These conditions include the use of a single binary choice format in a discrete
38 choice survey and consequentiality, which means that respondents believe that the
39 survey outcome will have real-life consequences on the probability of the public
40 good provision (often referred to as policy consequentiality) and on the probability
41 of the related payment collection (often referred to as payment consequentiality).
42 Subsequent work has developed further advancements to the conditions. Vossler
43 et al. (2012) emphasized that respondents should believe that there are chances that
44 each yes-or-no (binary) choice in a survey will increase or decrease, respectively, the
45 probability for the project and the related payment to be implemented. As a con-
46 sequence, if the sample size increases, the probability of a given response of being
47 pivotal (that is, the probability of the response changing the final survey outcome)
48 might decrease, weakening the individuals perception about own vote pivotality.

49 In the political literature, the concept of pivotality corresponds to the probability
50 of a given vote to change the outcome of the election. Intuitively, a rational individ-
51 ual will go voting if the benefits from voting are superior to the cost. In the context
52 of stated preference valuation literature, the issue of the response pivotality has been
53 paid much less attention. Mitani and Flores (2012) asked the question whether con-
54 sequential binary valuation tasks still ensure incentive-compatible behavior if the
55 expected gain from answering becomes very small. In their considerations, they
56 took into account that responding to a valuation choice task involves an effort and
57 therefore increases the cost.¹ If the cost from voting is superior to the benefit, it
58 is unclear why a rational individual would want to exert effort in a valuation task
59 and respond truthfully. Mitani and Flores (2012) conducted an induced value ex-
60 periment, in which they varied, among other things; sample sizes. The authors
61 found that varying the sample size from 1 participant to 45 participants had no

¹The authors state: "The cost of voting means any cost caused by making a voting decision, including a cognitive task of judging which alternative is better, time to make a vote, and/or participation in the voting decision".

62 impact on behavior (in their case the proportion of incentive compatible violation).
 63 Furthermore, the authors test the relationship between the group size and subjects
 64 subjective probability of being pivotal by conducting an incentivized guessing exper-
 65 iment utilizing a quadratic scoring rule. They find no relationship between sample
 66 size and subjective probability of being pivotal, which could indicate that people are
 67 not fully rationale. However, the authors are cautious in their conclusion since the
 68 number of participants is rather limited.²

69 In this paper, we test in a field experiment whether a higher sample size reduces
 70 the subjective consequentiality. We also explore the effects of the subjective conse-
 71 quentiality on willingness-to-pay and consider that the subjective consequentiality
 72 can be decomposed in at least two components as shown in Figure 1: a) the extent to
 73 which people think that their own response can influence the outcome of the study
 74 (hereafter called "*individual consequentiality*") and b) the extent to which they think
 75 the outcome of the study can influence policy makers (hereafter called "*survey con-*
 76 *sequentiality*"). Studies that assess the level of subjective consequentiality typically
 77 use a question of a type: do you think responses in this survey will influence the
 78 finally introduced outcome? This type of question may capture effects a) and b) si-
 79 multaneously. While the majority of existing studies controlling for consequentiality
 80 perceptions considers consequentiality in general, we attempt to isolate to each of
 81 the effects and check their impact on willingness to pay.

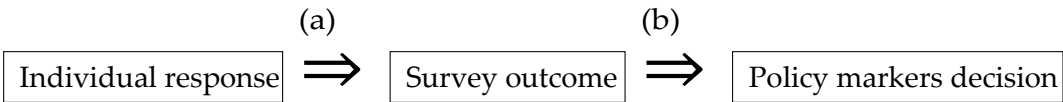


Figure 1: Decomposition of the subjective pivotality

82 On this purpose, we report the results of an online split-sample survey concern-
 83 ing a water turbines program in France. The survey was conducted in March 2018
 84 by a professional public opinion polling agency. A part of the sample is not provided
 85 with information on the sample size, like in most surveys. Another part is provided
 86 information on sample size just before the valuation choice question, where informa-
 87 tion varies across respondents in an attempt to manipulate and vary the subjective
 88 sample size. In the follow-up question stage, one question is used to elicit individual
 89 consequentiality (a) and another one for survey consequentiality (b).

²The authors state: "Thus, one might conclude that the pivotal probability does not affect a subjects truth revealing behavior in the referendum context although the biggest group size in our design is 45, which could be too small to observe the effect".

90 We find no link between sample size and subjective consequentiality, which is
91 in line with Mitani and Flores (2012), and hence suggest that consequential single-
92 bounded contingent valuation studies can still ensure incentive-compatible behavior
93 when increasing the subjective sample size. We also find that both components (a)
94 and (b) have a positive impact on willingness to pay. Implication for survey design
95 will be discussed. The remaining of the paper is structured as follows. Section 2
96 describes the survey. Section 3 presents the results. Section 4 provides a discussion
97 and Section 5 concludes.

98 **2 Survey**

99 In France, a large part of the electricity comes from nuclear power (about 75%)
100 and the share is expected to decrease to 50% by 2025 according to the French
101 energy transition for green growth act voted in 2015.³ Several programs of offshore
102 wind turbines and underwater turbines are being considered, although in March
103 2018, when our survey was performed, none of the planned offshore and underwater
104 turbines were operating.

105 The questionnaire was structured as follows. First, after a general introduction
106 inviting people to respond to the survey, it was explained that the responses will
107 be communicated to policy makers and therefore might influence policy makers.
108 Second, information was provided on renewable energy and more specifically on
109 underwater turbines. Pros and cons of underwater turbines were explained. Third,
110 participants were described with a research program consisting of constructing and
111 setting up two giant underwater turbines, of 16 meters in France. It was explained
112 that the effect of the underwater turbines on fauna and flora would be studied
113 and that the two turbines would produce electricity for about 5,000 households.
114 The location of the underwater turbines was not provided. Fourth, the following
115 valuation task was asked: "Would you be willing to pay X EUR a month during
116 a year on your electricity bill for the set-up of this program (building and testing
117 underwater turbines)?" Voluntary payment was avoided to avoid free riding. The
118 following final bid amounts were retained based on a pre-test: 0.5; 2; 5; 10 and
119 20 EUR. Fifth, different debriefing questions were asked regarding the perception
120 of the program and socio-demographic questions. Among other things, rating type
121 questions were asked, with people being asked to assess how they agreed with the

³<https://www.ecologique-solidaire.gouv.fr/loi-transition-energetique-croissance-verte>
(last consulted on 11 March 2019)

122 following statements between 1 ("I do not agree at all"), 2 ("I somewhat agree"),
123 3 ("neutral"), 4 ("I somewhat agree") and 5 ("I fully agree"): "This project is very
124 important for France"; "The outcome of this survey will influence policy makers
125 on the decision to implement or not the program" and "my valuation response can
126 affect the outcome of the survey". The two latter statement aims at measuring
127 (a) "individual consequentiality" and (b) "survey consequentiality" respectively. No
128 information was asked to elicit the so-called payment consequentiality, i.e., whether
129 people would really believe that the bill would be increased if the program was
130 implemented, to limit survey length.

131 The survey was performed by a professional company which guaranteed that
132 they would deliver a representative sample of the French population of at least 2,000
133 participants (the actual number was 2,023), where representativeness was based on
134 three socio-demographic characteristics (gender, income and age). Regarding the
135 experimental design, each individual was randomly allocated to one of four versions
136 of the questionnaires. In the baseline version of the questionnaire, hereafter called
137 V1, no information was provided on the questionnaire. In V2, a short sentence was
138 added just before the valuation question: "at least 50 participants will participate
139 to the survey". In V3, the sentence was: "at least 200 participants will participate
140 to the survey" while in V4 it was "at least 2000 participants will participate to the
141 survey". To ensure that people would read the above sentence in V2, V3 and V4
142 and would not be distracted by other information, there was little information on
143 the screen, only the sentence and a short reminder that the results of the survey
144 will provided to policy makers (see Appendix.A). Also, the sentence was worded in
145 a way to change the subjective sample size, if any, while avoiding deception which
146 is banned in economics and the participants had to wait for 5 seconds because they
147 could push on next to ensure that the information on the slide would be read.

148 Focus groups and pre-tests showed that the survey was properly worded and
149 that some persons were opposed to the program because it could potentially harm
150 fauna and flora. In the final surveys, some of the person refused to pay for the
151 program for this reason. Hence, possible negative WTP should be accounted for in
152 the econometric treatment.

153 **3 Results**

154 Table 1 provides information on socio-demographic characteristics of the respon-
155 dents, where the continuous variable income corresponds to net monthly income

156 (expressed in thousands of euro) and the binary variable education take value 1 if
 157 the individual has at least the A-level. A non-parametric Kolmogorov test is per-
 158 formed successively for each of the six possible combinations (V1 versus V2; V2
 159 versus V3, etc) and for each of the four socio-demographic variables (income, age,
 160 education, female). Results indicate that there is no difference of distribution across
 161 sub-samples at 5% statistical level for each of the four socio-demographic variables,
 162 which is not surprising given that the allocation to the different versions of the
 163 questionnaire was random.

	V1 (no info)	V2 (>50)	V3 (>200)	V4 (>2000)
Income	2.702 (1.36)	2.750 (1.523)	2.633 (1.435)	2.677 (1.475)
Age	46.111 (15.102)	47.130 (15.587)	46.238 (14.994)	45.386 (15.763)
Education	0.765 (0.424)	0.759 (0.428)	0.738 (0.44)	0.750 (0.434)
Female	0.506 (0.500)	0.504 (0.500)	0.498 (0.500)	0.513 (0.500)
<i>n</i>	468	532	520	503

Table 1: Socio-demographic characteristics

164 In the rest of the paper, we consider several ways to construct the variables
 165 related to sample size, as can be seen in Table 2. This allows testing if the results
 166 are sensitive to the variable construction.

167 3.1 Sample size and consequentiality

168 Mean comparison across sub-samples indicates that there is no link between individ-
 169 ual consequentiality and the (unobserved) subjective sample size, which is consistent
 170 with the results found by Mitani and Flores (2012). Indeed, the mean of *individ-*
 171 *ual_consequentiality* is 0.43, 0.44, 0.42 and 0.43 in V1, V2, V3 and V4 respectively.
 172 The unpaired two-sample t-tests is applied for each of the possible combinations
 173 (V1 versus V2; V2 versus V3, etc) and the null hypothesis of equal mean WTP is
 174 never rejected at conventional levels. The logit model also suggests that sample size
 175 has no impact on the subjective individual consequentiality which is robust to the
 176 inclusion of socio-demographic characteristics (see Table 3).⁴ In the logit model, the

⁴These results are also robust to the threshold retained to recode the consequential variables.

Sample size	
"Sample size information"	Binary variable that takes 1 if the respondent has been assigned the questionnaire version 1 (V1, no info), 0 if version 2 (V2, >50), version 3 (V3, >200) or version 4 (V4, >2,000).
Sample size	Continuous variable that takes value 50 if the respondent is assigned version 2 of the questionnaire (V2, >50), 200 if version 3 (V3, >200) and 2,000 if version 4 (V4, >2,000).
Sample_200	Binary variable that takes value 1 if the respondent has been assigned the version 3 (V3, >200) of the questionnaire, zero if version 2 (V2, >50) or version 4 (V4, >2,000).
Sample_2000	Binary variable that takes value 1 if the respondent has been assigned version 4 (V4, >2,000) of the questionnaire, zero if version 2 (V2, >50) or version 3 (V3, >200).
Consequentiality	
Individual_consequentiality	Binary variable that takes value 1 if the respondent reported a score of 4 ("I somewhat agree") or 5 ("I fully agree") to the statement "My valuation response can affect the outcome of the survey"; and value 0 if reporting 1 ("I do not agree at all"), 2 ("I somewhat disagree") or 3 ("neutral").
Survey_consequentiality	Binary variable that takes value 1 if the respondent reported a score of 4 ("I somewhat agree") or 5 ("I fully agree") to the statement "The outcome of this survey will influence policy makers on the decision to implement or not the program"; and value 0 if reporting 1 ("I do not agree at all"), 2 ("I somewhat disagree") or 3 ("neutral").

Table 2: Construction of variables related to sample size information and consequentiality

177 variables related to sample size (e.g., info) can be seen as instrumental variables for
178 the (unobserved) subjective sample size, as it is expected to be correlated with the
179 subjective sample size and not correlated with the explanatory variables since the
180 treatment allocation is random.

181 It is worth noting that the conclusion does not change when replacing the de-
182 pendent variable "individual consequentiality" by "survey consequentiality" or by a
183 variable which takes value one if both variables (individual consequentiality and
184 survey consequentiality) take one and zero otherwise.

185 **3.2 Willingness to pay**

186 The interval data regression model (Cameron, 1988) is employed to explore the
187 determinants of WTP. This approach relies on the maximum likelihood estimation
188 approach, which requires to assume a distribution for WTP. We retain the normal
189 distribution to account for the possible negative WTP due to the possible impact

Variables	(1)	(2)	(3)	(4)	(5)	(6)
Income				0.0682** (0.0330)	0.0760** (0.0370)	0.0767** (0.0370)
Education				0.174* (0.0970)	0.157 (0.111)	0.156 (0.111)
Female				-0.0891 (0.0910)	-0.0786 (0.104)	-0.0779 (0.104)
log(Age)				-0.00978 (0.129)	0.00948 (0.146)	0.00883 (0.146)
Sample size information	-0.00833 (0.106)			-0.00388 (0.107)		
Sample_200		-0.0842 (0.125)			-0.0753 (0.125)	
Sample_2000		-0.0424 (0.125)			-0.0337 (0.126)	
log(Sample size)			-0.00886 (0.0338)			-0.00669 (0.0340)
Constant	-0.275*** (0.0933)	-0.242*** (0.0873)	-0.234 (0.195)	-0.470 (0.504)	-0.528 (0.569)	-0.526 (0.602)
Observations	2,023	1,555	1,555	2,023	1,555	1,555

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Logistic Regression Model of the individual consequentiality

190 of the program on fauna and flora. Therefore, WTP_i is a linear function of a row
191 vector of covariates, x_i such that $WTP_i = x_i \beta + \epsilon_i$ where β is a column vector
192 of unknown parameters and ϵ_i is a normally distributed zero-mean error term with
193 standard deviation σ_i . Since the variance of the error term may depend on the
194 experimental design (i.e., the variance may differ across sample size), we allow for
195 heteroscasticity (see Vossler and Holladay (2018) and Vossler and Zawojka (2018)
196 for recent examples using the same approach). In this interval data model, "yes" and
197 "no" responses are considered as censored data, since the only information which is
198 observed is whether the WTP is above or below the assigned bid amount.

199 Table 4 displays 6 models which differ in term of independent variables. Models
200 1, 2 and 3 only includes variables related to sample size, while the other models in-
201 clude additional variables, namely sociodemographic and consequentiality. Models
202 1, 2 and 3 show that the information on sample size has no effect on willingness-
203 to-pay, regardless of the sample size variable construction. The coefficients are not
204 statistically significant from zero at conventional levels. Results also suggest that
205 the variance of the error term is also not statistically significant from zero at con-
206 ventional statistical levels, hence suggesting that increasing the sample size does

	(1)	(2)	(3)	(4)	(5)	(6)
Coefficient parameters						
Income				0.981*** (0.367)	1.148*** (0.445)	1.151*** (0.443)
Education				1.637 (1.048)	2.007 (1.279)	1.991 (1.273)
Female				-3.697*** (1.010)	-3.959*** (1.251)	-3.951*** (1.244)
log(Age)				-0.626 (1.382)	-0.0242 (1.683)	0.0465 (1.663)
Survey_consequentiality				7.489*** (1.389)	7.456*** (1.707)	7.418*** (1.699)
Individual_consequentiality				6.988*** (1.359)	7.831*** (1.707)	7.734*** (1.673)
Sample size information	0.636 (0.994)			0.693 (1.069)		
Sample_200		-1.288 (1.330)			-1.185 (1.396)	
Sample_2000		-0.816 (1.509)			-0.517 (1.474)	
log(Sample size)			-0.170 (0.372)			-0.136 (0.404)
Constant	6.696*** (0.821)	8.073*** (1.044)	8.269*** (2.118)	1.349 (5.402)	-0.519 (6.578)	-0.539 (6.894)
Standard errors						
Sample size information	0.235 (0.149)			0.177 (0.114)		
Sample_200		-0.255 (0.195)			0.133 (0.146)	
Sample_2000		0.0257 (0.219)			0.218 (0.148)	
log(Sample size)			0.0170 (0.0523)			0.0587 (0.0409)
Constant	2.603*** (0.124)	2.920*** (0.151)	2.743*** (0.299)	2.651*** (0.111)	2.746*** (0.111)	2.530*** (0.236)
Observations	2,023	1,555	1,555	2,023	1,555	1,555

Table 4: Interval data regression model

207 not lead to more random answers. The same results are obtained when including
 208 socio-demographic and questions on consequentiality: the sample size has no effect
 209 on WTP and the variance of error term. We also find that the level of income has a
 210 positive effect on WTP, which is consistent with a priori expectation. Also, people
 211 who believe that the outcome of this survey will influence policy makers on the deci-
 212 sion to implement or not the program (survey consequentiality) tend to state higher
 213 willingness-to-pay. The same pattern is observed for those who believe that their
 214 response can affect the outcome of the survey (individual consequentiality). Note
 215 that these results are robust to the threshold retained to recode the consequential
 216 variables (either 3 or 4, out of 5). We also created a series of interaction variables
 217 between the two consequential variables but results were not statistically different.

218 Table 5 displays the mean WTP and confidence interval from the interval data
 219 model without covariate and the Turnbull approach. The null hypothesis of equal
 220 mean WTP is rejected for each of the possible combinations and for both parametric
 221 and non-parametric analysis. Figure 2 shows the survival distribution for each of
 222 the treatments. The survival distributions are close, hence supporting the finding
 223 that the information on sample size has no effect on willingness to pay.

Description		V1 (no info)	V2 (>50)	V3 (>200)	V4 (>2,000)
Mean WTP (Std error)					
Parametric	Mean WTP	6.696 (0.821)	8.073 (1.044)	6.785 (0.825)	7.257 (1.090)
	Confidence interval	[5.086;8.304]	[6.027;10.119]	[5.169;8.401]	[5.121;9.394]
Non-parametric	Turnbull Mean WTP	6.525 (0.497)	7.864 (0.548)	6.712 (0.490)	7.532 (0.543)
	Confidence interval	[5.551;7.499]	[6.790;8.938]	[5.752;7.672]	[6.468;8.596]

Table 5: Parametric and non-parametric mean WTP and confidence intervals

224 We now turn to the analysis of response time, where the response time cor-
 225 responds to the number of seconds taken by participants to answer the following
 226 valuation question: "Would you be willing to pay X EUR a month during a year on
 227 your electricity bill for the set-up of this program (building and testing underwater
 228 turbines)?" A full slide was devoted to the valuation question, with no additional
 229 script/information being displayed on it except the "yes" or "no" answer. Table 6

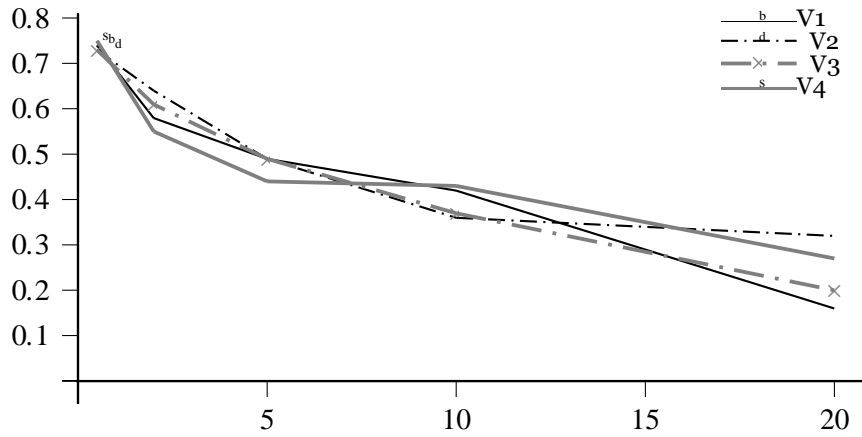


Figure 2: Survival distribution

230 shows that there is little difference across treatments. For instance, the median is
 231 identical (11 seconds). The non-parametric Kolmogorov test is employed to compare
 232 the distributions and the null hypothesis of identical distribution is never rejected,
 233 hence suggesting that the level of effort invested in the valuation task does not differ
 234 across treatments.

Percentile	V1	V2	V3	V4
1%	4	4	4	3
5%	5	6	5	5
10%	7	6	6	6
25%	8	8	8	8
50%	11	11	11	11
75%	14	14	15	14
90%	18	21	22	20

Table 6: Number of seconds taken to answer the valuation question

235 4 Discussion

236 One contribution of our paper to the literature is to give additional evidence that
 237 people fail to make the link between subjective sample size and subjective conse-
 238 quentiality, which is consistent with Mitani and Flores (2012) who performed an
 239 induced experiment that involved 45 participants, hence suggesting that consequen-

240 tial single bounded contingent valuation studies can still ensure incentive-compatible
241 behavior when increasing the subjective sample size. Indeed, while the actual bene-
242 fits from voting decreases with the sample size, the subjective benefits would remain
243 the same, which would explain why people do not increase the effort in the valu-
244 ation task (as shown by response time). This supports the use of single bounded
245 dichotomous choice surveys. However, this result is surprising given that political
246 literature has shown that the number of participants could impact the turnout in
247 election. However, the mechanism is a bit more complex in valuation since the link
248 between the survey and policy maker's decision is limited. If there is a majority of
249 "yes" respondents, the program may or may not be implemented in reality.

250 Another contribution of our paper regarding the literature on consequentiality
251 is to decompose the policy consequentiality into two components and check the
252 effect of each of the components on willingness to pay. We find that each of the
253 effects is highly significant and has a positive effect on willingness-to-pay. Asking
254 two questions rather than a single one may present some appeals. First, it may
255 be easier for the respondents. Consider the following question: "Do you think that
256 your response and the one from the other respondents will influence policy makers?"
257 If an individual thinks that the other responses will influence policy makers but
258 not his/her (e.g., the vast majority of the participant will favor the program but not
259 him/her), it is unclear whether she should respond "yes" or "no". Second, misleading
260 conclusion could be drawn if future research showed that the two components could
261 have an opposite effect on WTP and their effect would cancel out. This would
262 suggest that there is no need to control for consequentiality although it actually
263 impact results. A drawback of asking two questions instead of one is that it increases
264 the survey length.

265 Our study suffers from possible limitations, which we tried to overcome. First,
266 the information on sample size can affect protest answers or create a selection bias
267 (i.e., people refusing to take part of the survey if the sample is too high or too low),
268 and therefore alter the comparison across treatments. However, we did not provide
269 the information on sample size at the beginning of the survey to avoid selection
270 bias. As for the protest answers, the rate of protest answers is not statistically
271 different across treatments (the protest rate is 0.084, 0.083, 0.080 and 0.092 in V1,
272 V2, V3 and V4 respectively and the t-test mean comparison fails to be rejected
273 for each of the possible combination) and the exclusion of the protest answers does

274 not change any conclusion from the paper.⁵ Another possible limitation is that
275 providing information on sample size can potentially change the perception of the
276 good. When the sample size is big, participants may think that the good to be valued
277 is "important". If so, difference of mean willingness to pay could have been observed
278 across treatments. However, we do not find any difference of mean WTP, nor did
279 we find correlation between the variable "importance" and the variable related to
280 sample size, hence suggesting that it does not change the perception of the good
281 (correlation). Finally, one may argue that increasing the sample size has no impact
282 on subjective consequentiality because of the expected results. If the individual
283 thinks that 90% of the participants will vote yes, then the vote is not perceived
284 as consequential, even when the sample size is small. In the follow-up stage, we
285 asked the participants to predict other's people response and to estimate at which
286 level they think that a level will impact the decision. Even when controlling for the
287 condition vote; the information on sample size has no impact on consequentiality,
288 nor on willingness-to-pay, which reinforces the conclusion.

289 **5 Conclusion**

290 In this paper, we test in a single-bounded dichotomous choice survey dealing with
291 underwater turbines whether providing information on sample size has an impact on
292 willingness-to-pay. We find that there is no effect, and conclude that consequential
293 binary contingent valuation studies can still ensure incentive-compatible behavior
294 when the gain from voting becomes very small. Overall, this gives some support to
295 the use of the single bounded dichotomous choice surveys. Future studies could check
296 if the two considered components of consequentiality also affect behaviors in choice
297 experiment, or if sample size matters. Conditional voting requires an important
298 effort (i.e., predict the vote for each of the alternatives and voting among the top
299 two alternatives) and it might not be worth investing this effort when the sample
300 size is high and the probability to affect the outcome is therefore low.

⁵There is no clear consensus in the literature on whether or not to keep protest answers. In this paper, we decided to keep all the answers, but the results without protest answers, which are available upon request to the authors, remain the same.

301 **Acknowledgements**

290 Ewa Zawojcka gratefully acknowledges the support of the National Science Centre in
291 Poland (Preledium 8, grant no. 2014/15/N/HS4/01328) and the support of the
292 Foundation for Polish Science (FNP)

293 **References**

294 Cameron, T. A. (1988). A new paradigm for valuing non-market goods using ref-
295 erendum data: maximum likelihood estimation by censored logistic regression.
296 Journal of environmental economics and management, 15(3):355–379.

297 Carson, R. T. and Groves, T. (2007). Incentive and informational properties of
298 preference questions. Environmental and resource economics, 37(1):181–210.

299 Mitani, Y. and Flores, N. E. (2012). Robustness tests of incentive compati-
300 ble referenda: Consequential probability, group size, and value-cost difference.
301 Unpublished manuscript, Norwegian University of Life Sciences.

302 Vossler, C. A., Doyon, M., and Rondeau, D. (2012). Truth in consequentiality:
303 theory and field evidence on discrete choice experiments. American Economic
304 Journal: Microeconomics, 4(4):145–71.

305 Vossler, C. A. and Holladay, J. S. (2018). Alternative value elicitation formats
306 in contingent valuation: Mechanism design and convergent validity. Journal of
307 Public Economics, 165:133–145.

308 Vossler, C. A. and Zawojcka, E. (2018). Toward a better understanding of elicitation
309 effects in stated preference studies. Working Paper #2018-01, Department of
310 Economics, University of Tennessee.

311 Appendix

312 A. Slide positioned just before the valuation question

Version of the questionnaire: V1 (no information)

- The results of this survey will be communicated to policy makers

Next

Note: the button "Next" appears after 5 seconds.

Version of the questionnaire: V2 (> 50)

- At least 50 persons will participate to the survey
- The results of this survey will be communicated to policy makers

Next

Note: the button "Next" appears after 5 seconds.

Version of the questionnaire: V3 (> 200)

- At least 200 persons will participate to the survey
- The results of this survey will be communicated to policy makers

Next

Note: the button "Next" appears after 5 seconds.