Around and beyond the cheap talk script in Choice Experiments

Gianluca Grilli[1*], Sandra Notaro[1], Roberta Raffaelli[1]

University of Trento, Department of economics and management, Via Inama 5, 38122, Trento (TN), Italy

***Corresponding author: gianluca.grilli@unitn.it**

[Session type] prepared for presentation at the 2025 AAEA & WAEA Joint Annual Meeting in Denver, CO; July 27-29, 2025

Abstract: Using a sample of respondents interviewed face-to-face while accessing a natural park in Sardinia (Italy), we conduct a Discrete Choice Experiment to assess respondents' willingness to pay for improved environmental quality of the site. We assess the impact of four different strategies to mitigate hypothetical bias (soft cheap talk, honesty priming, consequentiality scripts, and solemn oath) and two elicitation methods (direct and inferred evaluation methods). Results indicate that none of the strategies were significantly effective in reducing HB. Conversely, inferred valuation led to significantly lower WTP estimates. The effect was especially large for attributes of pure public nature, while attributes that include utility from indirect use are less affected by elicitation method. Overall, the study suggests that inferred valuation is more effective than other strategy in removing social desirability of respondents.

Keywords: agricultural landscape, hypothetical bias, discrete choice experiments, inferred valuation, cheap talk, consequentiality, solemn oath, honesty priming.

1. Introduction

Data quality of discrete choice experiment (DCE) is often hampered by hypothetical bias (HB), i.e. the divergence between stated and real preferences of respondents. In the presence of HB, the reliability of willingness to pay (WTP) measures is at risk. Some meta-analyses suggest that in the presence of HB the ratio between hypothetical and real WTP measures exceeds two (Penn & Hu, 2018). This bias is especially important for public or socially desirable goods, for which WTP stated by respondents is biased upward (List 2003). Theoretical models suggest that WTP upward bias occurs due to social signaling or self-signaling, i.e. the tendency of respondents to overstate their WTP for impressing other people or to portray a certain self-image, respectively. Another explanation is the better-than-average effect, which is the systematic error of a person

assessing herself as better than the others.

The issue of HB is extremely relevant in non-market valuation. In public good analysis, there is a general tendency to consistently state WTP values larger than the true values, which undermines the validity of the subsequent cost-benefit analysis on development projects. In private good analysis, the bias may lead to sub-optimal pricing decisions for market goods, or to an overstatement of non-consumptive (socially desirable) attributes of the good. Therefore, addressing HB is extremely important in stated preferences.

Despite numerous empirical and theoretical contributions on HB, there is still a lack of consensus on ways to effectively address and reduce (or eliminate) HB in stated preference study. A thorough review of literature, conceptualization, external validity, and mitigation strategies for HB has been provided by Haghani et al. (2021a, 2021b). One of the oldest and most efficient ways to address HB is using cheap talk, which informs respondents about the existence and causes of HB in stated preference surveys and a plea to provide honest answers. The effectiveness of this method is still debated and often varies across goods evaluated but represents the standard approach to deal with HB issues. In addition to this, Haghani et al. find that several other measures have been implemented, with mixed effects that vary across goods and investigation fields: certainty scales, honesty priming, solemn oath, opt-out reminder, time-to-think method, RP assisted estimation, pivot design, perceived consequentiality, and inferred valuation.

While most mitigation strategies act on personal stated WTP, inferred valuation (IF) operates in a different way. In an IF study, respondents are asked to state their best estimate about other people's behavior. The strategy relies on the assumption that hypothetical studies encourage respondents to overstate personal WTP to appear socially desirable. When asked to predict WTP

of the average respondent, or the WTP of a household with similar characteristics. In this way, respondents are encouraged to state WTP free from social desirability.

In this manuscript, we further explore HB by testing and comparing different mitigation strategies, within the context of preferences for natural park management. We compare the inferred valuation through indirect questioning proposed by Lusk and Norwood (2009) with a series of HB treatments for direct questions, namely honesty priming, a consequentiality script, and a solemn oath that asked respondents to sign a commitment to provide honest answers. We test these treatments in a case study in Sardinia (Italy), one of the biggest islands in the heart of the Mediterranean Sea. Respondents were asked to express their preferences for improved management actions in a natural park that includes agricultural and landscape features. The attributes that were included in the study were both private and public in nature, to explore whether HB behavior differs when the good to assess has a private (consumptive) or public nature. While the use of treatments to reduce HB bias is not new, this study has the advantage to compare several treatments at the same time to assess the relative effectiveness.

2. Hypothetical Bias: the state of the art

In a stated preference survey, HB occurs when respondents' mis-state their true preferences. While Lee and Hwang (2016) suggest that there is still no consensus on the underlying causes of HB, some evidence suggests that it often occurs when public goods are at stake. When asked to express preferences for goods of public utility, it is morally appealing for respondents to overstate their level of support to look socially desirable (List, 1993). This desirability behavior reflects into upward biased WTP. Environmental economics is the field in which the earliest investigations of HB have been reported, due to the public nature of many environmental goods and services (e.g., Adamowicz et al. 1994, 1997).

When considering mitigation strategies, the scientific literature distinguishes between absolute and relative measures of HB mitigation. Absolute strategies consist in split-sample studies in which one sample is treated with a HB mitigation approach and the other is a bias-free control group. In this way, HB can be measured accurately. A relative HB strategy is undertaken without any bias-free benchmark, and the assessment of the HB bias effectiveness is measured as difference between treated and control group, assuming that the direction of the bias is known a-priori (Haghani et al., 2021b). The first method is more accurate but not always viable, especially in the evaluation of environmental goods for which bias-free measures are not available. Another useful distinction is between ex-ante and ex-post measures. Ex-ante aims are implemented at the survey administration stage, to prevent HB. Ex-post measures instead employ follow-up questions to correct HB after respondents expressed their preferences.

In this manuscript, we particularly explore five mitigation measures: 1) cheap talks, 2) honesty priming, 3) consequentiality script, 4) solemn oath, and 5) the inferred valuation method. Cheap talk scripts are by far the most common HB mitigation measures. First introduced by Cummings and Taylor (1999), they are often found effective as a debiasing strategy for WTP. A cheap talk informs respondents about the existence of HB and pledge to answer truthfully. Given that the use of cheap talk is becoming the standard in DCE, soft cheap talk is used as control in this study. Honesty priming is instead incidental exposures to words or sentences that encourage honesty but are not related to the choice tasks. The aim of honesty priming is to automatically activate the sense of honesty in respondents to give honest answers, and its effectiveness is still debated. Consequentiality scripts are designed to make respondents aware that the study is consequential, i.e. the results will be used for real policymaking, and that incorrect answers will translate into wrong real-world choices. Consequentiality is strictly linked to incentive

compatibility, i.e. the ability of a survey to elicit truthful answers from respondents, because respondents will commit to be truthful only if they perceive the study to be consequential. The fourth approach is the solemn oath. Through a solemn oath respondents make an official promise to provide honest answers. While relatively new in the stated preference literature, some studies indicate that oath significantly encourage truthful answers. The last method that we test is the inferred valuation, which involves asking respondents to predict the answers of other (average) respondents (Menapace & Raffaelli, 2020). The rationale behind inferred valuation is that, when the good includes normative motivations, respondents have the incentive to state larger WTP for themselves to make good impressions on society. At the same time, they have no reasons to overstate WTP for the average respondents, hence no upward bias for IQ.

3. Methodology

3.1 Study area

The empirical study was conducted in the Regional Park of Porto Conte (Sardinia, Italy), which is a highly valued farming area and especially relevant for sheep grazing. The regional park was created in 1999 to protect the unique flora and fauna. The park borders seaside and their beaches are famous summer destinations, with approx. 50,000 visits per month in summertime. The Park area contains many natural and cultural features: the famous Neptune Caves (Grotte del Nettuno), the ruins of a roman villa (Sant'Imbenia) and ancient stone buildings (nuraghe di Palmavera).

Sheep grazing is a traditional activity in Sardinia and in the park area currently takes place at the bottom of the hills. In its management plan the Regional Park has emphasized the importance of sheep grazing in maintaining open spaces and biodiversity on the top of the hills (Monte Doglia). Moreover, sheep grazing was used in the past to keep the fire breaks open. Along these

firebreaks we can find a particular protected habitat: the 6220 Nature 2000 habitat denominated Pseudo-steppe with grasses and annuals of the Thero-Brachypodietea. This habitat contains also rare types of orchids. There is scientific evidence that sheep grazing helps to maintain this particular protected habitat.

3.2 The survey design

The DCE data was collected with a face-to-face field survey and the questionnaire consisted of 4 parts, which were designed according to the best available practices (Johnston et al. 2017, Mariel et al. 2021). The initial section included warm-up questions, the second attitudinal question, the third the choice cards, and the last section was dedicated to socio-demographic data collection. The attributes of the DCE were decided after focus groups with local park managers and included: 1) landscape value of sheep grazing (high-low levels), 2) effect of sheep grazing on biodiversity (high-low levels), 3) biodiversity along firebreaks (high-low levels), 4) the possibility of attending a cultural-heritage activity, i.e. dawn with shepherds. The payment vehicle was an annual card for visiting the park, whose cost ranged between €5-30.

Table 1: Attribute and levels in the DCE

Attribute	Description	Levels
Landscape	Landscape Value of sheep grazing	Yes No
Biodiversity	Effect of sheep management on sheep grazing	Yes No
Firebreaks	Biodiversity on fire breaks corridors	Yes No
Cultural events	possibility to directly	Yes

	experience life and traditions of shepherds (attending the morning dawn with shepherds).	No
Cost	Payment vehicle: annual card to access the park	5, 10, 15, 20, 25, 30

The attribute levels were combined to generate an efficient design for the creation of choice cards. The full design was composed of 24 choice cards, with 3 alternatives each, divided into 4 blocks of 6 cards that were randomly assigned to respondents. Each respondent had to complete choice tasks twice for a total of 12 cards. In 6 cards, respondents had to indicate their own preferences (direct questions, DQ), while the remaining 6 asked respondents the indirect questions (IQ), i.e. to predict the proportion of other visitors that would choose each alternative (with the sum of proportions across the 3 alternatives equal to 100). To avoid order effect, the order of direct and IQs was randomized across respondents, so that half of respondents received direct questions first, while the other half IQs. The survey was administered face-to-face on site, so the sample comprised park visitors only.

The study was designed as a treatment-control experiment, with 3 treatments and 1 control group. Each respondent was assigned to one of these 4 sub-groups randomly:

1) Control Group (CG): Respondents in the control group received a "soft" cheap talk script, which asked respondents to consider the cost as if it was consequential. This is the most common approach to HB.
2) Honesty priming treatment (HPR): honesty priming is a psychological technique that

stimulates truthful answers by giving respondents honesty prompts in a non-explicit form. The priming was introduced with the following script "thank you for being honest" before the choice tasks. According to Pruckner (2011), this simple script activates the social norm of honesty.

3) Consequentiality treatment (CT): CT informs respondents that results would have been used by policymakers for managing the park and that reliable answers are thus necessary. In this specification of the CT, we followed Gneezy (2005) on the role of consequences on people's deception. Gneezy found that stressing the consequences that one's actions have on others might reduce incentives to lying. Hence, in our script we emphasized the consequences of respondents' choices on other tourists and the social welfare.
4) Solemn oath treatment (SOT): Our SOT treatment mimics courtroom procedures. Before the choice cards, respondents were invited to sign (in private) a "solemn oath" to tell the truth. Respondents were also informed that the interview would have continued regardless of the signature.

In addition to these 4 treatments, a fifth treatment involved the exploration of IQs. IQs were asked to all individuals, hence represents an intra-respondent treatment. Given that all respondents received both direct and IQs, confounding effects between the debiasing effect of IQ and the other treatments may arise. To avoid such confounding, we restricted the analysis of treatment evaluation to DQ only. The CG was composed of respondents who received DQ before IQ, which is the standard DCE question format. The subsample of the CG who received IQ before DQ was used as a treatment to assess the impact of IQ.

At the end of the choice card section, we included a question about the expected consequentiality of the study, which asked respondents to indicate their confidence that the study would be used

for actual policymaking. Respondents indicated the level of confidence on a 5-point scale, from "very unlikely" to "very likely". This question was included to test the impact of treatments on respondents' perception about the usefulness of the study.

3.2 Modelling Approach

The Analysis undertaken to test the impact of HB mitigation strategies starts with a random utility model (RUM), in which the utility U that respondents *n* derives from alternative I in the choice situation t is composed by a deterministic component V and a stochastic term ε:

$$U_{int} = V_{int} + \varepsilon_{int}$$

The deterministic component is assumed to be a linear combination between a matrix of attributes X and a vector of unknown β to estimate, so that the formula can be re-written in:

$$U_{int} = \beta_n X'_{int} + \varepsilon_{int}$$

Assumptions on the distribution of ε are required to assure a closed probability distribution for individual choices. In choice modelling it is common to assume an i.i.d. Gumbel-distributed ε, so that the model can be estimated using a Multinomial Logit Model (MNL), with the following probability distribution function:

$$Pr_{ij} = \sum \frac{e^{V_{ijk}}}{\sum_j e^{V_{ijk}}}$$

The MNL is a benchmark model, but it has the important limitation to assume that preferences are all the same for all respondents. In fact, the estimated coefficients are the same for all respondents. Preferences are highly heterogeneous across the population. Therefore, the MNL model is highly restrictive. One of the most common ways to relax the preference homogeneity

assumption is to adopt a Random Parameter Logit (RPL) model, in which coefficients are random and not fixed for all individuals. When parameters are random, there is no closed form for the joint probability function, which must be approximated through simulations. The simulated log-likelihood function for a RPL model using *R* random draws is the following:

$$Pr_{ij} = \int \frac{e^{V_{ijk}}}{\sum_j e^{V_{ijk}}} \theta(\beta|\boldsymbol{b}, \Omega) d\beta$$

While there is no obvious choice for the distribution of random parameters, standard practice is to assume a normal distribution. The cost coefficient is either fixed or log-normally distributed. The RPL model is advantageous because individual parameters can be estimated for each respondent, thus allowing for the maximum level of preference heterogeneity. After the parameter estimation, the Willingness to Pay (WTP) for each single attribute is calculated as a Marginal Rate of Substitution (MRS) between the non-monetary attributes and the cost attribute. For the $j-th$ attribute, the WTP formula for a model with a log-normal cost coefficient is:

$$WTP_j = \frac{\beta_j}{-exp(\beta_{cost})} \quad (5)$$

In this study, econometric models and WTPs were estimated using the R software (Team, 2020) through user-written maximum likelihood functions.

4. Results

4.1 Sample Descriptive Statistics

The survey reached 391 respondents but only 370 were included in the analysis. Of the 21 excluded respondents 14 were identified as protesters, while 7 provided unreliable answers due to specific choice pattern selection (Mariel and Artabe 2025).

The sample was fully balanced between man and female respondents, with a 50-50 split between genders. On average, respondents were 38 years of age and were highly educated: 39 percent of the sample achieved a university degree and 12 percent more had attained a PhD or equivalent post-lauream degree. As the reference population for the study is the set of park visitors, for which general descriptive statistics are missing, nothing can be said about the sample representativeness.

4.3 Estimation results

Before showing results, we tested whether treatments had an impact on respondent's self-reported expectations about the consequentiality of the study. Study consequentiality was assessed on a five-point Likert scale and analyzed using an ordered logit model, in which covariates are dummies for the treatment received. Results are displayed in table 2, which indicates that only the consequentiality script offered a significant and positive impact on expected study consequentiality of respondents. Therefore, the consequentiality treatment produced, on average, larger expectations about the actual use of the study for park management. The honesty priming and the solemn oath treatments were not effective in communicating consequentiality with respect to the baseline.

Table 2: Ordered logit analysis of self-reported study consequentiality.

Treatment (ref. = control)	Estimate
consequentiality	0.593**
	0.276
honesty	-0.23
	0.274
promise	-0.316

	0.274
Threshold parameters:	
Threshold 1	-3.862***
	0.397
Threshold 2	-0.579***
	0.202
Threshold 3	1.331***
	0.213
Threshold 4	3.881***
	0.397

In table 3 we first report MNL and RPL pooled models that serve as benchmarks. Overall, both models have concordant signs of the coefficients and only magnitudes differ. The log-likelihood level indicates, however, better fit for the RPL model.

The coefficient associated with the entrance ticket is negative and statistically significant at 1 percent confidence level. This result complies with the economic theory and indicates that respondents are cost sensitive, so that their utility reduces at increasing cost levels. With respect to the policy attributes, landscape improvement, biodiversity on firebreaks, and the heritage experience are all positive and significant, which indicate that these attributes are utility-increasing. Conversely, biodiversity improvement is negative, hence decreasing utility. The status quo coefficient is negative and significant, which supports the hypothesis that visitors are interested in the new management actions compared to the current situation of the park. In terms

of preference heterogeneity, the standard deviation of the attributes in the RPL model suggests large variability across respondents for most attributes. The only attribute associated with a non-significant standard deviation is biodiversity improvement.

Table 3: MNL and RPL polled models

Attributes	MNL	RPL
Ticket cost	-0.072***	-0.09***
	(0.003)	(0.004)
Landscape improvement	0.592***	0.723***
	(0.041)	(0.075)
Biodiversity improvement	-0.132**	-0.181***
	(0.055)	(0.070)
Biodiversity on Firebreaks	0.497***	0.605***
	(0.043)	(0.063)
Heritage experience	0.163**	0.19***
	(0.052)	(0.069)
Status quo	-1.308***	-2.366***
	(0.089)	(0.179)
Standard deviations of random parameters		
Landscape improvement		1.029***
		(0.073)
Biodiversity improvement		0.047
		(0.392)

Biodiversity on Firebreaks		0.645***
		(0.074)
Heritage experience		0.454***
		(0.106)
Status quo		2.027***
		(0.146)
LL	-4001.1	-3498.7
Obs	4440	4440
Respondents	370	

In table 3 we then focus on the impact of HB treatments on the utility functions. In the left-hand side of the panel, we reproduce an RPL model where HB treatments enter the utility function as mean-shifters, i.e. in interaction with the attributes. On the right-hand side of the table, we propose a model where direct questions are compared to IQs.

In the model with treatments, it can be noticed that overall interaction terms are significant only in a minority of occasions. Overall, there is not a strong pattern that suggests the effectiveness of one specific HB mitigation strategy over the others.

Table 4: RPL models with HB mitigation measures

RPL Model with Treatments					RPL model with direct and inferred questions	
Attributes	Control	Honesty	Consequentiality script	Solemn oath	Attributes	Attributes x indirect

		priming (t-shirt)		with signature		questioning
Ticket cost	-0.088***	0.002	0.003	-0.018*	-0.085***	-0.028***
	(0.005)	(0.007)	(0.008)	(0.009)	(0.005)	(0.006)
Landscape improvement	0.662***	0.02	0.114	0.137	1.054***	-0.606***
	(0.110)	(0.159)	(0.179)	(0.196)	(0.097)	(0.102)
Biodiversity improvement	-0.179*	0.05	-0.085	0.004	-0.071	-0.338**
	(0.104)	(0.151)	(0.169)	(0.187)	(0.108)	(0.142)
Biodiversity on Firebreaks	0.679***	-0.151	-0.106	0.002	0.901***	-0.552***
	(0.092)	(0.134)	(0.149)	(0.166)	(0.086)	(0.105)
Heritage experience	0.302***	0.135	-0.472***	-0.207	0.258**	-0.157
	(0.101)	(0.147)	(0.165)	(0.182)	(0.105)	(0.135)
Status quo	-	-0.098	-0.194	-0.21	-4.407***	1.968***

	2.257***					
	(0.251)	(0.356)	(0.400)	(0.432)	(0.289)	(0.251)
Standard deviations of random parameters						
Landscape improvement	1.024***				1.16***	
	(0.074)				(0.082)	
Biodiversity improvement	0.013				0.372**	
	(0.363)				(0.165)	
Biodiversity on Firebreaks	0.633***				0.754***	
	(0.074)				(0.080)	
Heritage experience	0.409***				0.599***	
	(0.112)				(0.109)	
Status quo	2.03***				2.762***	
	(0.146)				(0.193)	

LL	-3489.4	-3088.2
Obs	4440	
Respondents	370	

The model with IQs has meaningful coefficients. All interaction terms associated with management attributes are negative, which indicates that the inferred valuation questions lead to lower preferences. The status quo interaction coefficient is positive, therefore larger preference for the current situation. While not all coefficients are significant, overall IQs lead to reduced welfare measures.

Figure 1: WTP comparison across treatments and elicitation methods.

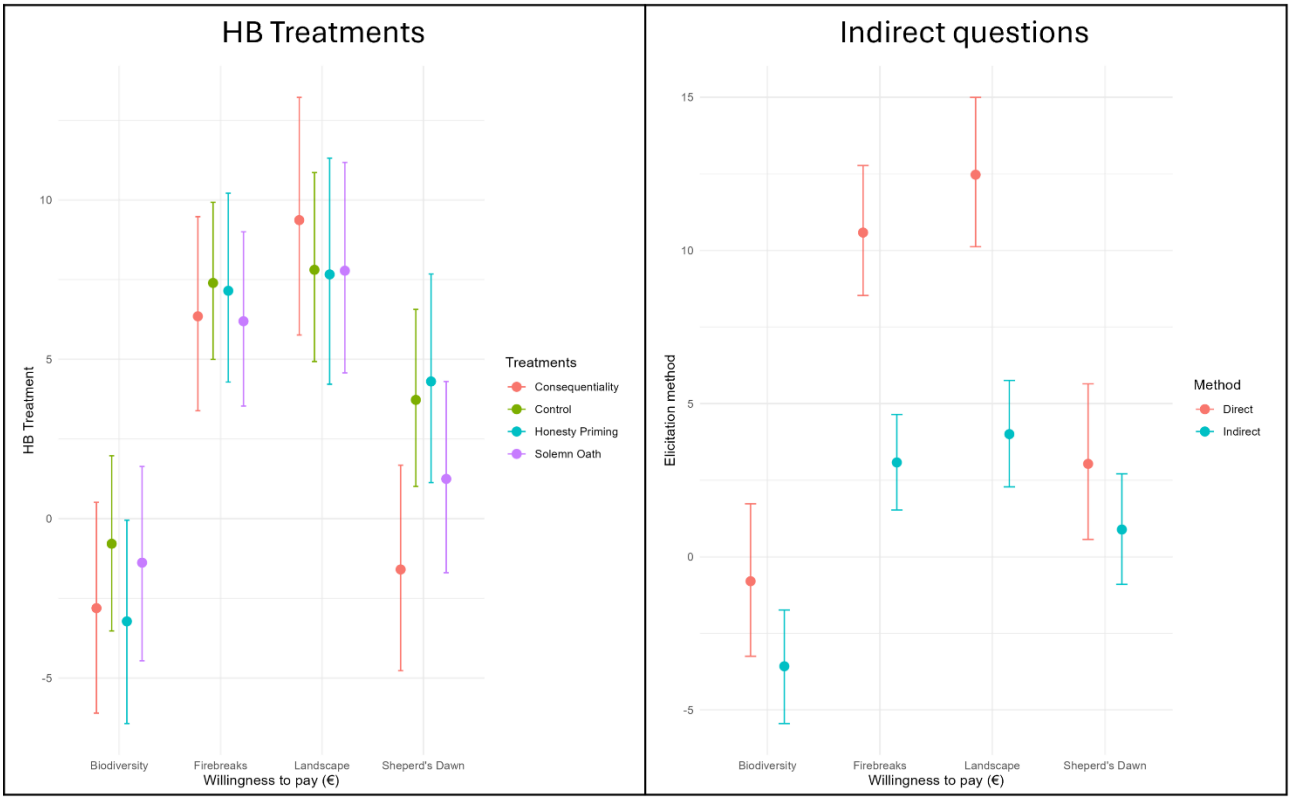

Figure 1 reports WTP estimates across treatments and elicitation methods, using the Krinski-Robb procedure with 10,000 draws to obtain confidence intervals. The left-hand side of the

figure shows that WTP did not differ significantly across treatments, in fact the mean WTP values are very close to each other for each attribute and confidence intervals overlap considerably. On the right-hand side, we can see that the inferred valuation approach reduces considerably average WTP. This is especially true for the WTP associated with biodiversity on firebreaks and landscape.

5. Discussions and Conclusions

Overall, the HB mitigation exercise provided interesting but mixed results. In terms of expected consequentiality of the study, the consequentiality treatment only was able to increase the average expectation about the actual use of the results for policymaking in the Regional Park of Porto Conte. Other treatments were not equally effective in raising consequentiality in respondents' belief.

While the consequentiality treatment was effective in the stated consequentiality, it was not equally effective in addressing upward WTP bias caused by the hypothetical nature of the study. In fact, WTP estimated in the consequentiality sample did not differ significantly from the rest of the sample. In general, none of the treatments were effective in reducing HB as WTP were statistically similar to the control group in all HB treatment groups.

Honesty priming as a HB mitigation strategy has been largely implemented in food choices and partly in environmental valuation. In the food literature, de-Magistris et al. (2013) demonstrated that honesty priming was capable to reduce marginal WTP, and that WTP in the honesty priming treatment was comparable to that of a non-hypothetical experiment. Bello and Abdual (2016) observed a significant impact of honesty priming on marginal WTP reduction and non-attendance levels when investigating preferences for organic food. On the other hand, a more recent study on organic food found no impact of honesty priming on marginal WTP

(Gschwandtner & Burton, 2020). In the environmental literature honesty priming was found ineffective when estimating preferences for a policy aimed to environmental quality preservation (Howard et al. 2017). In this mixed context, our study supports the part of the literature that suggest the limited power of honesty priming for HB mitigation.

Regarding the consequentiality treatment, our study complies with previous research that indicates that stressing consequentiality positively impact on respondents' belief about the potential real-world use of the results (e.g., Oehlmann & Meyerhoff 2017). In terms of WTP measures, there is limited evidence that consequentiality treatments significantly reduce HB, and this study is another contribution that supports previous studies.

Similarly to other mitigation measures, the empirical evidence of solemn oath effectiveness is limited, despite some studies indicate that under oath hypothetical and actual payments become indistinguishable (Stevens et al. 2013).

Overall, this work did not find support for one mitigation measure over the others. The main limitation of this analysis is that, due to the non-market nature of the study, there are no real choices that can be used to measure the extent of the bias. The study relies on the assumption that HB exists and that it causes upward bias in marginal WTP. Therefore, one explanation for our result might be that HB did not exist in first place. This possibility would explain the inefficacy of the treatments; however it is highly unlikely for at least two reasons. First, the vast majority of environmental valuation studies find upward biased WTP for socially desirable goods like environmental goods. Second, IQs provide significantly lower marginal WTP values, which is a strong indication of social desirability in direct questions for personal WTP. As a general conclusion, this study indicates that none of the treatments was an effective strategy for HB mitigation. This result, however, can at least be partially driven by the control group, which

is a soft cheap talk instead of a proper control group. Therefore, the measure of treatment effectiveness is measured with respect to the cheap talk. For this reason, another interpretation of these results might be that none of the treatments are better than soft cheap talk for measuring the upward bias of marginal WTP for public goods.

With respect to inferred valuation, we already highlighted that estimations lead to significantly lower marginal WTP. This means that respondents believe that the average visitor would support the environmental policy for the management of the park less than they do, which is strongly associated with social desirability. While IQs should be considered an alternative elicitation method rather than a proper HB mitigation strategy, the scientific literature clearly indicates that inferred valuation encourages answers free from socially desirable and moral components. Interestingly, figure 1 shows that the IQ method was least successful with the cultural attribute, namely the possibility to attend shepherd's dawn inside the park, for which the mean WTP of the IQ method falls within the confidence interval of the mean WTP estimated using DQ elicitation method. In the right-hand side of table 4, we observe that the interaction term of IQ with the cultural attribute is the only non-significant coefficient, despite the sign is negative. This is the only attribute with a private nature, as the utility derives from indirect use of the park. Other attributes are truly public, and respondents derive utility from their non-use. In the presence of private goods respondents are less subject to social desirability bias, hence less HB and lower WTP upward bias. This result suggests that the type of attribute matters and IQ is most effective with attributes of public nature.

Therefore, a recommendation that arises from the study is the use of inferred valuation as a complementary analysis of public good support measures. While IQs debias WTP, there are some drawbacks of this approach that should be considered. The main limitation of the inferred

valuation method is the lack of a direct link between WTP figures and socio-demographics of the respondents, hence no inference on observable characteristics is available. Another limitation is the projection bias, which may occur when one respondent makes inference for the others based on his own beliefs and priorities. Lastly, inferred valuation relies on the hypothesis that one respondent can accurately predict what others believe. While the "power of the crowd" has been often proven to hold, the accuracy of the prediction is probably very much linked to the sample size.

The debate on HB is far from being conclusive. This work compares multiple debiasing strategies and includes attributes that are both public (such as preferences for biodiversity) and private (heritage activity, and partly landscape view) in nature. The strategies to reduce HB can be thus assessed against the attribute nature. While standard ex-ante approaches were not successful in reducing HB, the elicitation method was. Therefore, the IQ strategy appears to be a low-cost and effective strategy for HB mitigation. This, however, is most effective when WTP values matter but the determinants of these WTP figures are less important. In fact, IQ lacks a direct link between the characteristics of the respondents and stated WTP. When respondents' personal characteristics matter for the analyst, other mitigation strategies should be explored. Considering that the current literature eon HB is still mixed and affected by the nature of the good under investigation, more research is necessary for unravelling HB in stated preference studies.

References

Bello, M., & Abdulai, A. (2016). Impact of ex-ante hypothetical bias mitigation methods on attribute non-attendance in choice experiments. *American Journal of Agricultural Economics*, *98*(5), 1486-1506.

Cummings, R. G., Taylor, L. O., Ball, H., Brown, K., Gardner, K., Murphy, J., Stevens, T., & Williams, M. (1999). Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. *American Economic Review*, *89*(3), 649–665. https://doi.org/10.1257/AER.89.3.649

De-Magistris, T., Gracia, A., & Nayga Jr, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American Journal of Agricultural Economics*, *95*(5), 1136-1154.

Fisher, R. J. (1993). Social Desirability Bias and the Validity of Indirect Questioning. *Journal of Consumer Research*, 20(2). https://www.jstor.org/stable/2489277

Gschwandtner, A., & Burton, M. (2020). Comparing treatments to reduce hypothetical bias in choice experiments regarding organic food. *European Review of Agricultural Economics*, *47*(3), 1302-1337.

Haghani, M., Bliemer, M. C. J., Rose, J. M., Oppewal, H., & Lancsar, E. (2021). Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. *Journal of Choice Modelling*, *41*. https://doi.org/10.1016/j.jocm.2021.100309

Haghani, M., Bliemer, M. C. J., Rose, J. M., Oppewal, H., & Lancsar, E. (2021b). Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. *Journal of Choice Modelling*, *41*. https://doi.org/10.1016/j.jocm.2021.100322

Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., ... & Vossler, C. A. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319–405.

Lee, J., & Hwang, U. (2016). Hypothetical Bias in Risk Preferences as a Driver of Hypothetical Bias in Willingness to Pay: Experimental Evidence. *Environmental and Resource Economics*, *65*(4), 789–811. https://doi.org/10.1007/S10640-015-9926-9/TABLES/9

List, J. A. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics*, *118*(1), 41-71.

Lusk, J. L., & Norwood, F. B. (2009). An Inferred Valuation Method. *Land Economics*, *85*(3), 500–514.

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., ... & Thiene, M. (2021). *Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis* (p. 129). Springer Nature.

Menapace, L., & Raffaelli, R. (2020). Unraveling hypothetical bias in discrete choice experiments. *Journal of Economic Behavior & Organization*, *176*, 416-430.

Oehlmann, M., & Meyerhoff, J. (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), 1-16.

Penn, J. M., & Hu, W. (2018). Understanding hypothetical bias: An enhanced meta-analysis. *American Journal of Agricultural Economics*, *100*(4), 1186-1206.

Stevens, T. H., Tabatabaei, M., & Lass, D. (2013). Oaths and hypothetical bias. *Journal of environmental management*, *127*, 135-141.