# Choice-matching for incentive-compatible elicitation of stated preferences: Field evidence for a public good

Ewa Zawojska[a,*] and Michał Krawczyk[a]

[a] Faculty of Economic Sciences, University of Warsaw; Długa 44/50, 00-241 Warsaw, Poland
* Corresponding author, ewa.zawojska@uw.edu.pl

**CRediT authorship contribution statement**:
Ewa Zawojska: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; and Writing – review and editing.
Michał Krawczyk: Conceptualization; Funding acquisition; Methodology; Resources; Software; and Writing – review and editing.

**Acknowledgements**: The authors are grateful for the helpful comments from Prof. Bradley Ruffle, participants of the Southern Economic Association 91st Annual Meeting, the 2022 Newcastle Experimental Economics Workshop, and the Behavioural Science and Policy/Environmental Policy Seminar of the University College Dublin. Ewa Zawojska gratefully acknowledges the support of the Polish National Agency for Academic Exchange within the Bekker Programme.

The published version of this study is in the *Journal of Behavioral and Experimental Economics*.

# Choice-matching for incentive-compatible elicitation of stated preferences: Field evidence for a public good

**Abstract**: Stated preferences should ideally be elicited in ways that align respondents' interests with truthful preference disclosure. Survey design conditions for such incentive compatibility typically rely on consequentiality; that is, respondents' perceptions that their survey responses matter for a final outcome. By contrast, this study aims to empirically test a novel theoretical approach that allows for incentive-compatible elicitation of preferences toward a hypothetical good. The choice-matching approach, proposed by Cvitanić et al. (2019), is applied here to design a stated-preference elicitation procedure for a public good that is intended to be incentive compatible. While choice-matching has been originally designed for a multiple-choice question, we adapt it here to an open-ended elicitation by mapping continuous responses into a finite number of value intervals. We conduct an online experiment mirroring a standard stated preference survey used for the valuation of public goods. We implement two versions of the survey questionnaire: one employing the incentive-compatible choice-matching approach and another representing a typical, unincentivized setting. We find that open-ended willingness-to-pay values are statistically significantly higher when stated under choice-matching than when expressed under unincentivized conditions. The paper discusses why the results may be regarded as supporting the use of choice-matching for stated preference elicitation.

**Keywords**: contingent valuation, choice-matching, incentive compatibility, open-ended elicitation, stated preferences

**JEL codes**: D61, D82, H43, Q51

## 1 Introduction

Stated preference (SP) surveys are among the primary tools for valuing non-market public goods (Hanley and Czajkowski 2019). Despite their widespread use (Mahieu et al. 2017), scientific debate continues on the validity of values derived from SP data. To ensure validity, contemporary guidance for SP studies suggests using incentive-compatible survey designs, which make truthful disclosure of preferences the best-response strategy (Johnston et al. 2017).

However, achieving incentive compatibility in SP surveys is often demanding. For example, a single binary-choice question is recommended for preference elicitation as the most straightforward way to encourage truthful responding, but it provides limited information about preferences per respondent. Alternative question formats may increase the statistical efficiency of value estimation but may give rise to preference misreporting (e.g., Hanemann et al. 1991; Vossler and Holladay 2018). Furthermore, incentive-compatible elicitation recommended in the SP literature requires consequentiality, meaning that respondents need to believe that their survey responses are not purely hypothetical and can matter for final policy decisions (Carson and Groves 2007; Johnston et al. 2017)—a condition that is difficult to guarantee in typical SP survey contexts querying about hypothetical policy programs.

This paper investigates an alternative approach for truthful preference elicitation in SP surveys—choice-matching, a mechanism proposed by Cvitanić et al. (2019) to elicit truthful responses even when the truth is not verifiable. This mechanism does not rely on assumptions similar to those required in the incentive-compatible approaches recommended in the SP literature, such as consequentiality (although it relies on a different set of assumptions). While choice-matching has originally been developed for eliciting truthful answers to a multiple-choice question (i.e., with discrete response options), we adapt it here to a continuous response format, such as an open-ended question, by mapping continuous willingness-to-pay (WTP) responses into a finite number of non-overlapping intervals. The application presented here is, to our knowledge, the first examination of choice-matching in a SP-based valuation of a public good.[1]

In an online experiment with two split-sample treatments, we compare the performance of choice-matching to that of a standard, unincentivized SP elicitation. The specific case study concerns the valuation of a program extending infrastructure for renewable energy production in public spaces in Warsaw, Poland. SPs are elicited with an open-ended question in an online questionnaire, which reflects a typical contingent valuation application.

[1] The note is based on a review of all papers citing the work of Cvitanić et al. (2019) listed in Google Scholar as of March 27, 2026.

We believe that this paper contributes to the literature on SP elicitation in several ways by responding to the need for tools that can support incentive-compatible elicitation of SPs. First, it provides empirical evidence on the performance of the choice-matching mechanism in the context of valuing a public good, thereby testing its ability to affect preference revelation in SP surveys. Second, the paper offers a practical illustration of how this novel, incentive-compatible mechanism can be implemented in non-market valuation surveys, including settings where the evaluated good is hypothetical. Given the widespread use of SP methods for valuation of hypothetical programs of public good provision, this application speaks directly to a broad range of empirical contexts. Third, the study shows that formats beyond the single binary choice need not be discarded on incentive-compatibility grounds. By demonstrating how an open-ended elicitation can be adapted to the choice-matching approach, the paper provides additional flexibility in applying the approach to various response formats. The possibility of obtaining truthful preference statements through an open-ended format is an appealing prospect for SP researchers, as an open-ended response provides a precise value indication (a point indication, rather than, for example, value intervals derived from binary-choice responses). Taken together, these contributions suggest that choice-matching may serve as a useful tool for incentive-compatible elicitation in SP surveys and, more broadly, that a wider set of elicitation formats can be designed to elicit valid and reliable preference information.

The remainder of the paper is structured as follows. Section 2 delineates the theoretical background. Section 3 provides details on our application of choice-matching to a SP valuation of a public good, along with a discussion of key assumptions underlying the mechanism's application. Section 4 describes the design and implementation of the survey, which provided data for the empirical study. The results of the data analysis are presented in Section 5. Section 6 discusses the validity of our findings through the lens of: (i) a critical requirement needed for the incentive compatibility of choice-matching and (ii) existing literature. Section 7 concludes.

## 2 Theoretical background

### 2.1 Incentive compatibility in SP studies

In SP literature, a single binary-choice (yes-no) question is recognized as the most straightforward approach to achieving incentive compatibility (Johnston et al. 2017). The format's simplicity and clear incentive structure underlie the recommendation for its use. Despite the long-standing recommendation (dating back at least to the report of the National Oceanic and Atmospheric Administration panel on contingent valuation; Arrow et al. 1993), various other formats continue to be applied in practice, i.a. to improve statistical efficiency or better match the decision context. While incentive compatibility is often lost with the other formats, statistical efficiency of value

estimation is gained as a single binary response reveals little information on preferences per respondent (e.g., Hanemann et al. 1991; Kanninen et al. 1993; Vossler and Holladay 2018).[2]

Disparities in incentive structures across applied SP elicitation formats appear to lead to significant variations in value estimates obtained with different formats (Vossler and Zawojska 2020). This undermines the validity of SP value measures. For instance, a well-known result from the voting literature (related to Duverger's Law, Duverger 1954) suggests that when answering a multiple-choice question, a respondent may benefit from selecting her second-best option, thus not disclosing her most preferred choice. She may be inclined to do so if she believes that her second-best option has a better chance of winning than her first-best, and there is a substantial risk that another option she considers worse may win (e.g., Cain 1978; Abramson et al. 2004). In turn, for example, in an open-ended SP question, a respondent who favors the provision of a good may be inclined to strategically overstate her true WTP for that provision if she believes that the sum of all stated amounts will determine the provision decision, or understate it if she believes that the stated amounts will be used to set the individual cost.[3] Divergences in incentive properties such as these contribute to differences in value estimates derived across various SP elicitation formats.

The influential role of differences in incentive structures across elicitation formats on value estimates emphasizes the importance of designing SP surveys to be incentive compatible. To achieve incentive compatibility with formats that are more statistically efficient than the single binary-choice question, additional conditions for the preference elicitation procedure have been developed. Vossler et al. (2012) propose conditions for a series of binary-choice questions, and Vossler and Holladay (2018) define conditions for open-ended and payment-card questions. Evidence from a lab experiment by Vossler and Zawojska (2020) suggests that when a range of common preference elicitation formats (including the single binary choice, a double-bounded binary choice, a payment card, and an open-ended question) are designed to be incentive compatible, value estimates derived from these formats are statistically indistinguishable.

Importantly, the incentive compatibility conditions proposed in SP literature—for any of the formats mentioned above—require the valuation to be non-hypothetical (e.g., Vossler et al. 2012; Carson et al. 2014; Vossler and Holladay 2018). Carson and Groves (2007) define this feature of a survey as consequentiality, meaning that respondents perceive a positive probability that their survey answers may matter for a final decision about the provision of the good considered in the preference elicitation. While respondents' beliefs in consequentiality have been shown to influence SPs and value estimates (e.g., Vossler and Watson 2013; Interis and Petrolia 2014; Zawojska et al. 2019),

[2] For example, a Monte Carlo simulation by Vossler and Holladay (2018) illustrates that in their study case, the level of precision for a mean value estimate is the same when obtained with an open-ended sample size of about 125 respondents and when based on a single binary-choice sample of 200 respondents.

[3] Response considerations related to the open-ended preference elicitation question are discussed in detail by Carson and Groves (2007) and are briefly summarized in Section 6.2.

consequentiality is not an obvious condition to satisfy in surveys. Survey scripts may fail to convince respondents of their consequentiality (e.g., Czajkowski et al. 2017; Lloyd-Smith et al. 2019); many policy projects evaluated in surveys are at hypothetical stages; and individuals' perceptions of consequentiality are not objectively measurable and are often endogenous to SPs, which makes accounting for these perceptions in SP analysis challenging (e.g., Herriges et al. 2010; Börger et al. 2021). A newly emerging line of research suggests that consequentiality might even be an undesirable survey feature, leading to preference statements that are not guided by one's personal preferences but instead driven by a sense of responsibility for a policy outcome and its impact on others (e.g., levying an obligatory tax payment on all; Comerford and Lades 2022). In light of the challenges related to consequentiality, choice-matching appears to be an attractive alternative because it can, in suitable applications, provide incentive-compatible elicitation of preferences even when the good is hypothetical.

### 2.2 Choice-matching mechanism

Choice-matching has been proposed by Cvitanić et al. (2019) as a mechanism to elicit honest responses to a multiple-choice question when the truth is not verifiable (e.g., opinions, judgments). Despite the need to elicit honest responses in settings where outcomes of interest cannot be objectively verified, these environments have received limited attention in the experimental literature (Charness et al. 2021) and empirical studies (Lehmann 2026). Incentive-compatible mechanisms for eliciting non-verifiable truth were introduced by Prelec (2004) with the Bayesian Truth Serum, followed by a peer-prediction method (Miller et al. 2005). More recent approaches in this domain include—beyond choice-matching—Bayesian markets (Baillon 2017) and top-flop betting (Baillon and Xu 2021), as well as extensions to the Bayesian Truth Serum (e.g., Witkowski and Parkes 2012; Radanovic and Faltings 2013). The choice-matching approach is viewed as standing out from other, particularly earlier-developed, methods for eliciting non-verifiable truth, as it is claimed to be sufficiently simple to communicate the benefits of truthful reporting to participants (Charness et al. 2021).[4]

In brief, choice-matching relies on the use of two linked questions: (1) one question asks about opinions, judgments, or preferences (as in our case) and is of primary interest to the researcher, as it is intended to elicit the unverifiable truth; and (2) an auxiliary question is employed to determine an outcome affecting a participant, for example, through a payoff from the study. Based on the information disclosed in question (1), participants are classified into types, for instance, those who prefer option A, those who prefer option B, and so on. A researcher then uses the information about a participant's type to determine their payoff. The payoff is based on the response of a given

[4] Applications of other approaches in the area of eliciting non-verifiable truth often forgo explaining the (complex) incentive structures to participants and rely on providing them with simple information that truthful responses may be wise or may pay off, without discussing details of the mechanism (e.g., Barrage and Lee 2010; Weaver and Prelec 2013; Loughran et al. 2014).

participant to the auxiliary question (2) and/or on responses (to that question) of other participants who are of the same type as the given participant. In the canonical version developed by Cvitanić et al. (2019), respondents answer a multiple-choice question (1) and additionally predict the distribution of others' answers to question (1) in the auxiliary task (2); compensation is a weighted sum of the respondent's own prediction score and the average prediction score of respondents who gave the same answer to question (1) (that is, those of the same type). The authors then generalize the underlying logic beyond prediction tasks. In the generalized version, the auxiliary task may take a different form, but truthful reporting is still induced by matching a respondent, on the basis of the answer to the type-revealing question (1), with respondents who gave the same answer to (1) and tying the respondent's payoff to the auxiliary-task outcomes or utilities of that matched group.

The idea underlying the incentive compatibility of choice-matching is that questions (1) and (2) need to be purposefully designed so that those who think alike in question (1) about opinions, judgments, or preferences (that is, those of the same type) are expected to give similar responses to the auxiliary question (2). In the canonical prediction-based formulation, this key assumption is called impersonal updating: respondents of the same type have the same expectations about others' type frequencies. In the generalized version, including ours, the analogous requirement is that the auxiliary question be type-separating, or approximately so: respondents of the same type should be expected to choose similar responses in the auxiliary question (2) or obtain similar auxiliary-question utilities, while respondents of different types should differ in those responses or utilities. When this condition is satisfied approximately, responding to (1) remains incentive compatible to the extent that same-type auxiliary responses are expected to be closer to one another than auxiliary responses of any other type.

Intuitively, to maximize their expected payoff, a participant benefits when the payoff is aligned with their own response to the auxiliary question (2). When the payoff is instead determined by the responses of others to that question, it is beneficial for the participant if these others are of the same type as that participant, because same-type individuals are more likely to give similar responses to the auxiliary question (which determines the payoff) than participants of different types are. This implies that it pays off to reveal one's own opinions, judgments, or preferences truthfully in question (1) in order to be assigned to the correct (true) type.

Several studies have applied the choice-matching mechanism. Myers et al. (2020) use the mechanism to study scientists' perceptions of present and future disruptions to their research related to the Covid-19 pandemic. The study by Cerroni et al. (2023) is the closest to our research context, applying choice-matching to elicit SPs for a private good (cottage pies) through a discrete choice experiment. Bauer et al. (2024) employ choice-matching in a survey to elicit investors' beliefs about the financial performance of environmental, social, and governance funds. In the study by Ceallaigh et al. (2024), the mechanism is used to help measure time inconsistency in physical activity. Oedingen et al. (2025) apply choice-matching to elicit self-reported medication adherence.

## 3 Application of choice-matching to the elicitation of stated preferences towards renewable energy

In this section, we describe our empirical application of the choice-matching approach to the elicitation of SPs for the extension of renewable energy infrastructure. Besides explaining the practical operationalization of the approach for our study, this section aims to provide an intuitive understanding of the incentive compatibility of the mechanism's application. Technical details of the implementation are discussed in the subsequent section, and a formal derivation and proof of the incentive compatibility of choice-matching are presented by Cvitanić et al. (2019).

Our application follows the logic of the example provided by Cvitanić et al. (2019, p. 189) for a possible use of choice-matching to elicit (hypothetical) WTP values for a good that cannot be offered for sale. Specifically, the authors state that the hypothetical preference elicitation question can be accompanied by an auxiliary question that asks about WTP for a different good offered for actual sale. Respondents should be informed that for the purchase transaction, "their WTP for this [i.e., non-hypothetical] good may be replaced by the WTP of a respondent who has given a similar value on the hypothetical valuation question." Cvitanić et al. (2019) claim that if the WTP values for the two goods are strongly correlated, truthful preference disclosure in the hypothetical question should be the most beneficial option for respondents. We apply a similar procedure in our study.

The good evaluated in our survey is a program to increase renewable energy use in public spaces in Warsaw by extending solar energy infrastructure. The program assumes the construction of new solar panels that would provide electricity to the city's public spaces, such as train and subway stations and parks. At the time of data collection, similar green energy solutions had been implemented, for example, on some of the city's administration buildings, but they were not extended to the public spaces considered in the survey. The program assumes a further expansion of the use of solar panels. The proposal explains to respondents that the program's implementation would require additional funds, which could be collected through taxes designated for the program's realization. The survey script states that the tax increase for this purpose would last for five years and would apply to all residents of Warsaw.

The preference elicitation question has an open-ended format and asks respondents: "How much at most would you be willing to pay annually (for 5 years) as an increased tax for the presented program of building solar panels in Warsaw?".[5] This question is of primary interest from the perspective of obtaining a value estimate of the considered program. According to the choice-matching procedure, responses to this question are used to identify a respondent's type. In our application, types are determined by the similarity of the WTP values. Specifically, based on the

[5] This and other quotations from the questionnaire used in the paper are translated from Polish.

open-ended responses, respondents are classified by the researcher into non-overlapping groups of similar WTP values, and these groups define the types. Respondents are made aware of this procedure through information in the survey script.

Distinguishing separate types of respondents is necessary to determine respondents' payoffs, which are tied to decisions in the auxiliary question (described below). In the approach originally proposed by Cvitanić et al. (2019), the question used for identifying types has a discrete response format, and the response options define the types. In our application, we translate the open-ended responses into a discrete scale to distinguish several types. In this way, we adapt the discrete-type logic of Cvitanić et al. (2019) to an open-ended setting by converting continuous WTP responses into a finite number of value intervals that serve as respondent types for payoff determination. In particular, we maintain the necessary assumptions that (i) the number of types is smaller than the number of respondents, and (ii) no type is associated with an empty set; that is, each type has representation in the sample, because types are defined by groups of similar WTP values. The survey script used to convey the grouping of similar WTP values for defining types is formulated so that it meets these assumptions.

Based on respondents' answers to the SP question, we divide them into seven types, each corresponding to one of the non-overlapping WTP intervals. This number of types is selected to satisfy the description in the survey script that separate types will be characterized by similar WTP values—this number allows us to group respondents into such types. This procedure is performed after data collection is completed, and respondents were not provided with information on the specific range of WTP values associated with their type at the time of completing the survey. This implies that the specific composition of the WTP intervals could not affect respondents' behavior and, thus, could not distort the incentive properties of the preference elicitation. Instead, respondents were provided with the rationale for why it was beneficial to be correctly grouped with other respondents with similar WTP values (as discussed below). Furthermore, the procedure of assigning types is used only to determine payoffs to individuals, and this (arbitrary) assignment of respondents to types is not used (nor needed) in the data analysis, particularly, in the examination of treatment effects. Consequently, it does not affect the conclusions of this study.

The auxiliary question is needed for the truth-inducing mechanism, and it does not provide SP data for the valuation study. In our application, the question concerns a respondent's willingness to contribute financially to Foundation "Forest Forever".[6] The foundation is a non-governmental organization in Poland that aims to restore biological balance by planting forests. In the auxiliary question, a respondent is asked: "How would you like to split 1,000 PLN [Polish zloty] from the budget of the experiment?", and the response format is: "For the new forest [planted by the foundation]: ____ PLN. For me: ____ PLN.", where the respondent must enter two amounts

[6] In Polish, Foundation "Las na Zawsze", https://lasnazawsze.org.pl/ (accessed on September 10, 2024).

summing to 1,000 PLN.[7] Respondents are informed that this question will be binding for several randomly selected individuals, meaning that they will be allocated 1,000 PLN and that some amount will be transferred to the foundation. For incentive compatibility of choice-matching, respondents are told that in the case of half of the selected individuals, the amounts allocated to the new forest and to the individual will be equal to the means of the amounts indicated by respondents of the same type (instead of using the amounts directly entered by a given individual), where the type is defined based on the preference elicitation question about renewable energy infrastructure. In the next paragraphs, we describe the logic underlying how this procedure induces truthful preference statements.

To clearly distinguish between the information elicited by the two questions described above, we henceforth refer to the answers to the preference elicitation question about the renewable energy program as *WTP values* and to the responses to the question about forest-planting support as *contribution amounts*.

The aforementioned mechanism aligns respondents' interests with truthful disclosure of their WTP values in the hypothetical SP question about the renewable energy program. Truthful value disclosure places a respondent in the WTP group whose members are expected to choose contribution amounts most similar to the respondent's own preferred money split in the auxiliary, forest-planting support question. In our implementation, if a respondent from the choice-matching treatment is selected to make a binding contribution to the foundation, the payoff-relevant contribution amount is based either on the respondent's own auxiliary answer or on the mean contribution amounts of respondents classified as the same type (that is, placed in the same WTP interval). Hence, truthful reporting of WTP is the best strategy when respondents expect that individuals with similar WTP for the renewable energy program will indicate contribution amounts for the forest that are more similar to their own preferred contribution than those indicated by individuals placed in other WTP intervals. This is the application of the key requirement for incentive compatibility of choice-matching to our study context: the auxiliary question needs to be sufficiently type-separating. In our empirical application, we treat a positive correlation between WTP values and contribution amounts as suggestive evidence consistent with that requirement, while recognizing that correlation is not identical to the formal type-separation condition. Because the SP question and the auxiliary question are thematically related, we argue that it is reasonable to expect that WTP values and contribution amounts are positively correlated (we formally investigate the correlation based on empirical data in Section 6.1). Put simply, those who like environmental programs are likely to indicate high amounts in both questions, while those who do not are likely to indicate low amounts. As a result, we argue that truthful revelation of one's type (and thus one's WTP value) in the SP question about the renewable energy program is the strategy that maximizes

[7] At the time of data collection, the exchange rate was approximately 1 EUR ≈ 4.5 PLN.

the likelihood that the contribution for the new forest, and the related payoff to the individual, will be closest to one's stated amounts in the auxiliary question.

For an illustration of the incentive structure, consider a simple example, which was also provided to the study participants. Suppose there is individual A who does not want to pay much for environmental protection. Thus, A selects a low contribution for the new forest (and, hence, a large amount for oneself) out of the 1,000 PLN from the experiment's budget. However, suppose that in the SP question about the renewable energy program, A indicates an amount substantially exceeding the true amount A wants to pay (e.g., for reasons of social self-presentation). If A is selected to make the binding contribution based on the responses of others, A will pay the mean contribution amount reported by others of the same type, that is, those who indicated a high WTP value for the renewable energy program. Because these individuals are willing to pay a lot for the renewable energy program, they likely indicated high contribution amounts for the new forest (because these are likely individuals who like supporting environmental protection). As a result, A may end up paying for the new forest a higher-than-preferred amount. An analogous example can easily be constructed to show that it does not pay off to underreport WTP for the renewable energy program (and this was also presented to study participants).

## 4 Survey design and implementation

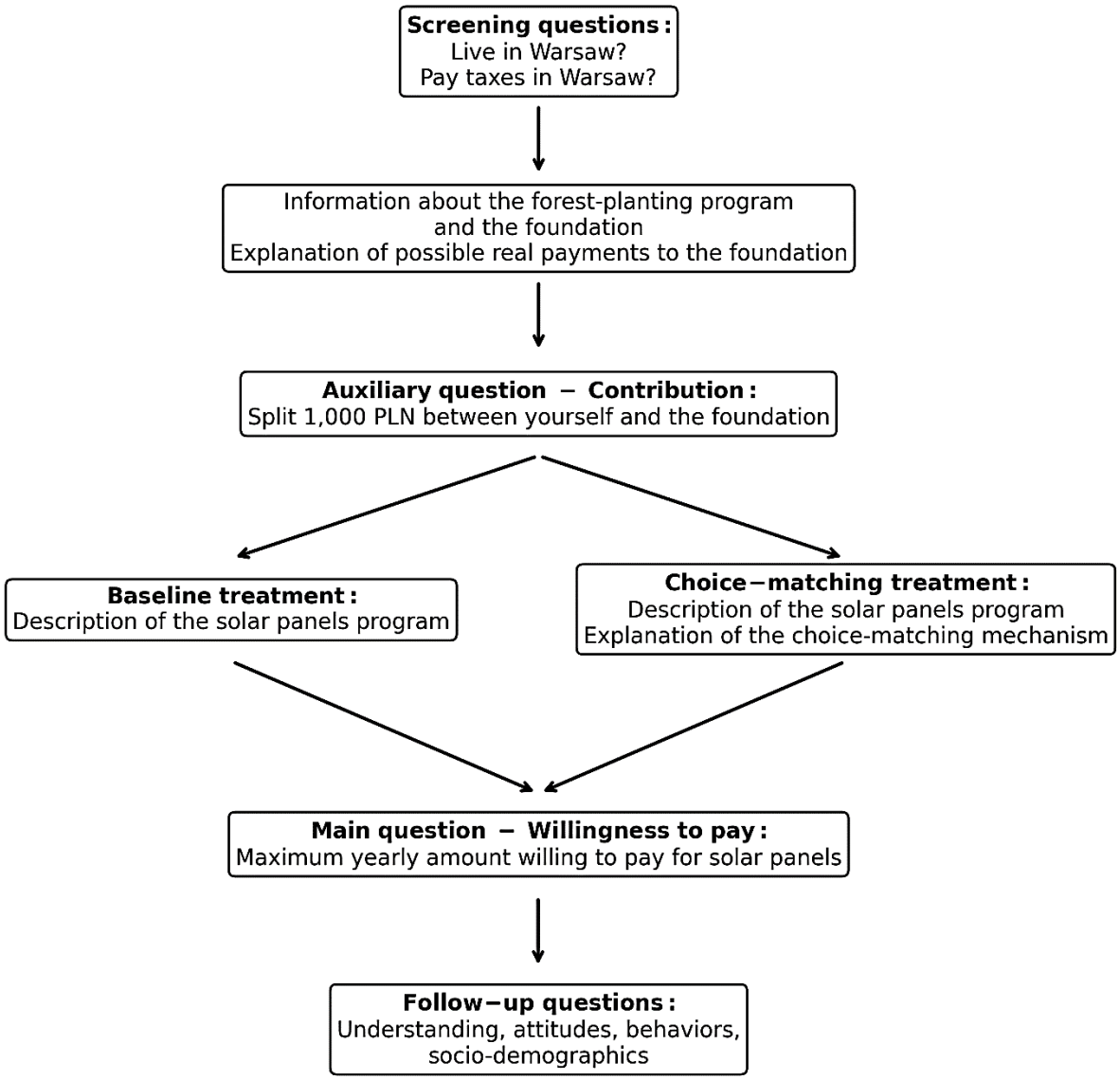
The empirical data for the study come from an online experiment mirroring a standard SP survey as used for the valuation of public goods. As described above, the evaluated good is a policy program to extend renewable energy infrastructure in Warsaw by installing solar panels in public spaces of the city. An open-ended question is employed to elicit individuals' preferences for the program. Two versions of the survey questionnaire are implemented: one employing the incentive-compatible choice-matching approach and another representing a typical, unincentivized setting. Henceforth, we refer to the former as *the choice-matching treatment* and to the latter as *the baseline treatment*. The two survey versions differ only in terms of the features necessary for implementing choice-matching. Each respondent is randomly allocated to one of the treatments and can participate in the survey only once. This study is registered in the American Economic Association's registry for randomized controlled trials (AEA RCT Registry) under the identification number [removed for review]. The study has the approval of the Ethics Committee of [removed for review] (no. 1/2020).[8]

### 4.1 Questionnaire structure

We present the structure of the questionnaire used in the experiment in Figure 1 and describe it step by step below.

[8] The survey questionnaire, both the original in Polish and the English translation, is available in the supplementary materials.

Figure 1. Timeline of the experiment. Respondents are randomly assigned to either the baseline treatment or the choice-matching treatment.




The questionnaire starts with two screening questions to ensure that the survey is completed only by individuals for whom the evaluated good is relevant—that is, individuals permanently living in Warsaw and paying taxes in the city.

The next part of the survey asks respondents (regardless of treatment) how they want to split 1,000 PLN from the experiment's budget between themselves and Foundation "Forest Forever", which will use the money for planting a new forest. This is the auxiliary question in the choice-matching mechanism. We include this question in both treatments to keep the questionnaire as similar as possible across treatments. In our application, the auxiliary question appears before the main question, that is, the one with preference elicitation, to facilitate respondents' understanding of the choice-matching mechanism and the description of procedures. Respondents are briefly informed

about the role of forests in ecosystems and about the foundation. They are also told that once data from all respondents are collected, three individuals will be randomly selected for whom responses to this question will be binding—namely, each of the selected individuals will be assigned 1,000 PLN, and the amount indicated by the individual will be paid to the foundation for the new forest, while the remaining amount will go to the individual's bank account.[9] Respondents are also informed that when the payment is received by the foundation, the contributing individual will receive a certificate confirming the area of forest planted and will later be informed by the foundation about the specific location of the planted forest. Up to this point, there are no differences across treatments. In particular, in both treatments, it is made clear that it is in respondents' best interest to respond truthfully to the auxiliary question.

The survey then proceeds with a hypothetical question eliciting the maximum amount a respondent would be willing to pay annually for extending renewable energy infrastructure with new solar panels in public spaces in Warsaw. This is the point at which the two treatments diverge. While all individuals are provided with a short description of the proposed program, respondents in the choice-matching treatment are also given information specific to the approach. Respondents in this treatment are informed that three additional respondents will be randomly selected,[10] and these respondents will make actual contributions to the foundation from the assigned 1,000 PLN from the experiment's budget. Respondents are told that in this case, the amount paid by an individual to the foundation will be equal to the mean of the contribution amounts indicated by other same-type individuals, where the type is determined based on WTP responses concerning the renewable energy program. Specifically, the script explains that after collecting answers from all respondents, we will first identify a group of survey participants who indicated a WTP value for the solar panels program similar to that of a given respondent, and then calculate the average of the contribution amounts declared by these participants for forest planting. In addition to the description of the mechanism, the script in the choice-matching treatment explains the rationale underlying the survey responses. It also provides two illustrative examples to show how truthful preference disclosure pays off when binding contributions to the foundation are based on the contribution amounts declared by other individuals. To avoid priming, the two examples are balanced: one describes a case of overstating the WTP value, while the other refers to understating the value. To avoid anchoring, neither example includes specific numerical amounts. For both treatments, this part of the survey ends with the open-ended WTP question about the program of building new solar panels in Warsaw.

[9] Selecting a subset of participants for payment, instead of paying every participant, has been shown in experimental economics to be an effective scheme allowing researchers to increase individual payments (for a review of selected studies, see Voslinsky and Azar (2021)). We follow this approach to increase the prominence of the decision in the question about the contribution amount.

[10] We note that if respondents perceive a low probability of being selected, the treatment effect may be attenuated.

Further parts of the survey collect information on respondents' understanding of the instructions, perceived importance of the considered programs (that is, afforestation and renewable energy extension with solar panels), pro-environmental behaviors, and socio-demographic characteristics. Throughout the survey, there are no back buttons, so respondents cannot revise their earlier responses.

### 4.2 Survey administration and study sample

The survey was administered online to individuals up to forty years old, randomly selected from the panel of the Experimental Economics Laboratory of the Faculty of Economic Sciences at the University of Warsaw in Poland.[11] Individuals were invited via email to participate in the study and were allowed to complete the questionnaire only once. The survey was coded in Limesurvey. Data collection took place between January and March 2021.

In total, there are 103 completed responses in the choice-matching treatment and 98 in the baseline treatment. Some observations are characterized by very short response times, which may indicate that a participant clicked through the survey without considering the questions. To address this issue, we exclude from the analysis 3% of observations in each treatment sample (equivalent to three observations per treatment) with the shortest response time to the preference elicitation question about the renewable energy program. This corresponds to omitting those who answered the question in less than 20 seconds. As a result, the data analysis is based on 100 respondents in the choice-matching treatment and 95 respondents in the baseline treatment.[12]

The median response time for the whole survey was 6.03 minutes (mean: 9.86 minutes) in the baseline treatment and 8.64 minutes (mean: 15.75 minutes) in the choice-matching treatment, which is consistent with the longer survey script in the latter.[13] From the total sample, six individuals were randomly selected to make actual contributions to planting a forest from the assigned 1,000 PLN from the experiment's budget. For three of them, each contributed exactly the amount indicated by a given individual, while for the other three, each paid the mean contribution amount of other respondents of the same type as a given individual. As a result, the binding contributions ranged from 20 PLN to 900 PLN, implying that payments to the selected individuals ranged from 100 PLN to 980 PLN.

[11] The panel involves a relatively small number of individuals aged more than forty who are registered as potential participants in economic experiments. Hence, to ensure consistency of the study sample and comparability of the experimental treatment subsamples, the age limit for participation was set at forty.
[12] The sample size was determined by budget constraints and the need to maintain a non-negligible probability of selection for actual payment. We did not conduct a formal power analysis or pre-register a pre-analysis plan, which we acknowledge as a limitation of the study.
[13] Our sample includes two observations in each treatment with total response times exceeding 100 minutes. Excluding these observations, the median and mean response times (in minutes) are 5.99 and 7.75 in the baseline treatment, and 8.62 and 11.50 in the choice-matching treatment, respectively.

The two treatment samples are described and compared in Table 1 with respect to respondents' socio-demographic characteristics and their self-reports concerning comprehension of the study instructions, decision certainty, and perceived importance of the programs. Statistical tests indicate no significant differences in the socio-demographic characteristics, as expected given the random allocation of respondents to the treatments. The mean age in the study sample is about 29, two thirds of the respondents are women, about 80% have attained an academic degree, more than half are employed full-time, and the mean net household income is slightly above 6,000 PLN per month.

We also do not observe any significant differences across treatments in respondents' views on whether the provided information was sufficient for making decisions, in their certainty about their choices, or in the perceived importance of the programs to them. In both treatments, respondents more often (definitely) agreed that they had sufficient information to indicate the amount in the forest-planting question than in the SP question about the renewable energy program. Similarly, regardless of treatment, respondents appeared to be more certain about their indicated contribution amounts in the forest-planting question than about their WTP values for the renewable energy extension. The afforestation program appeared, on average, slightly more important to respondents than the renewable energy infrastructure extension.

As shown in Table 1, the only statistically significant differences across treatments (both at the 1% level) concern self-reported comprehension of the study instructions and survey response time. While respondents in both treatments report a good understanding of the instructions, the main difference lies in the distribution of "definitely yes" and "rather yes" responses to the statement "I understood the study instructions". In the baseline treatment, 81% of respondents "definitely" understood the instructions and 18% "rather" understood them, summing to 99% who reported understanding. In the choice-matching treatment, 58% "definitely" understood the instructions and 35% "rather" understood them, amounting to 93% reporting understanding.

Table 1. Summary statistics for the treatment samples

| Variable | Definition | Means (st. dev.) for baseline | Means (st. dev.) for choice-matching | P-value for the null hypothesis of equality between the treatments |
|---|---|---|---|---|
| | *Socio-demographic characteristics* | | | |
| Age | Respondent's age in years | 28.79 (4.60) | 29.32 (4.66) | 0.3891 |
| Woman | 1 for women, 0 for men | 0.66 (0.48) | 0.66 (0.48) | 1.0000 |
| Secondary education | 1 for respondents with secondary education, 0 otherwise | 0.21 (0.41) | 0.20 (0.40) | 1.0000 |
| Academic degree | 1 for respondents with an academic degree, 0 otherwise | 0.77 (0.42) | 0.80 (0.40) | 0.7278 |

| | | | | |
|---|---|---|---|---|
| Employed full-time | 1 for respondents employed full-time, 0 otherwise | 0.54 (0.50) | 0.61 (0.49) | 0.3848 |
| Income | Respondent's household net monthly income in PLN[a] | 6,180.21 (4,878.95) | 5,971.10 (4,456.64) | 0.7814 |
| | *Comprehension, decision certainty, and importance* | | | |
| Forest: Sufficient information | "I had sufficient information to indicate well the amount in the forest-planting question" responded on a Likert scale from 1 ("definitely no") to 5 ("definitely yes") | 4.50 (0.77) | 4.21 (0.91) | 0.3155 |
| Energy: Sufficient information | "I had sufficient information to indicate well the value in the solar panels question" responded on a Likert scale from 1 ("definitely no") to 5 ("definitely yes") | 3.50 (1.31) | 3.65 (1.22) | 0.1947 |
| Forest: Certainty | Respondent's stated certainty regarding the chosen amount on a Likert scale from 1 ("totally uncertain") to 5 ("totally certain") | 4.42 (0.79) | 4.53 (0.78) | 0.5222 |
| Energy: Certainty | Respondent's stated certainty regarding the indicated value on a Likert scale from 1 ("totally uncertain") to 5 ("totally certain") | 3.73 (1.08) | 3.62 (1.11) | 0.4547 |
| Forest: Importance | "The program of forest planting is important to me" responded on a Likert scale from 1 ("definitely no") to 5 ("definitely yes") | 4.39 (0.74) | 4.38 (0.67) | 0.2462 |
| Energy: Importance | "The program of building solar panels is important to me" responded on a Likert scale from 1 ("definitely no") to 5 ("definitely yes") | 3.71 (1.06) | 3.76 (1.02) | 0.9654 |
| Comprehension | "I understood the study instructions" responded on a Likert scale from 1 ("definitely no") to 5 ("definitely yes") | 3.80 (0.43) | 3.47 (0.73) | 0.0028 |
| Choice-matching comprehension | "I understood how the amount given to Foundation 'Forest Forever' was determined in the second [choice-matching] draw" responded on a Likert scale from 1 ("definitely no") to 5 ("definitely yes") | --- | 4.32 (0.87) | --- |
| Response time | Response time for the whole survey in minutes | 9.86 (15.93) | 15.75 (32.09) | 0.0001 |
| Sample size | | 100 | 95 | |

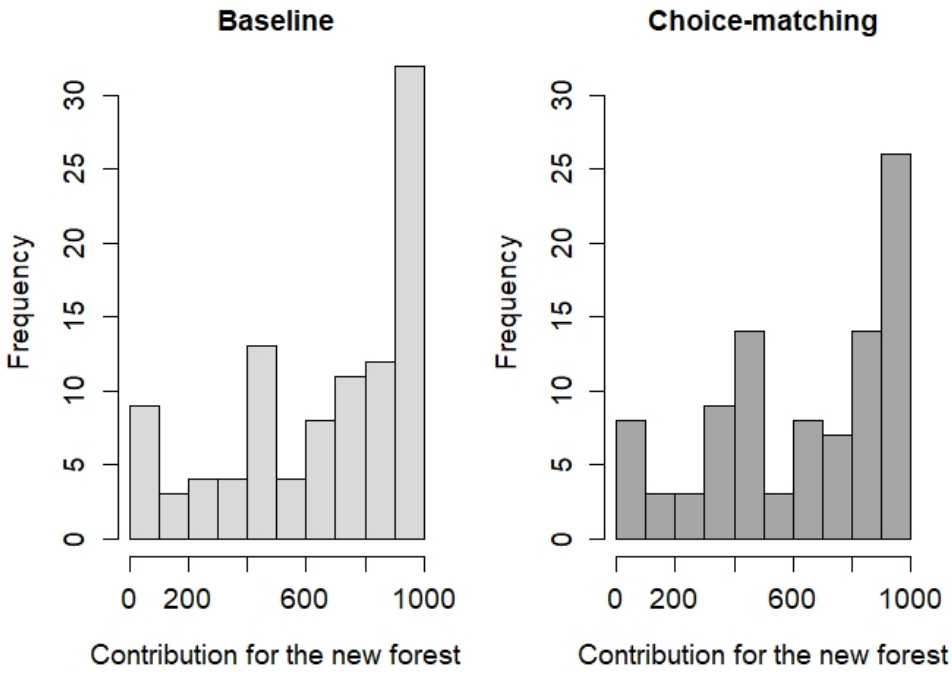
*Notes*: Means (and standard deviations in the brackets) are reported. P-values are for the null hypothesis of no difference between the treatment samples with respect to a given variable. We use Fisher's exact tests for binary indicator variables, chi-squared tests of equality of proportions for the Likert-scale statements, and Wilcoxon rank sum tests for variables measured on a continuous scale.

[a] Respondents were asked about their household income in an open-ended question. If they rejected providing an answer (67 respondents), they were directed to the same question with a discrete response scale, where the choice options represented different income intervals. 46 respondents stated their income in the closed-ended discrete-scale question. For these respondents, mid-points of the selected intervals are used here.

The reported contributions for planting a forest are similar across treatments, as expected, since the survey scripts did not differ up to this point. On average, baseline-treatment respondents selected 690 PLN as a contribution for the forest, and choice-matching-treatment respondents indicated 668 PLN. The distributions of the amounts are not statistically different across treatments according to the Wilcoxon rank sum test (p-value = 0.643 for the null hypothesis of no difference). Figure 2

presents the distributions of contribution amounts and illustrates a high degree of similarity across treatments.

Figure 2. Distributions of the contribution amounts for planting a forest in the two treatments




**5 Results**

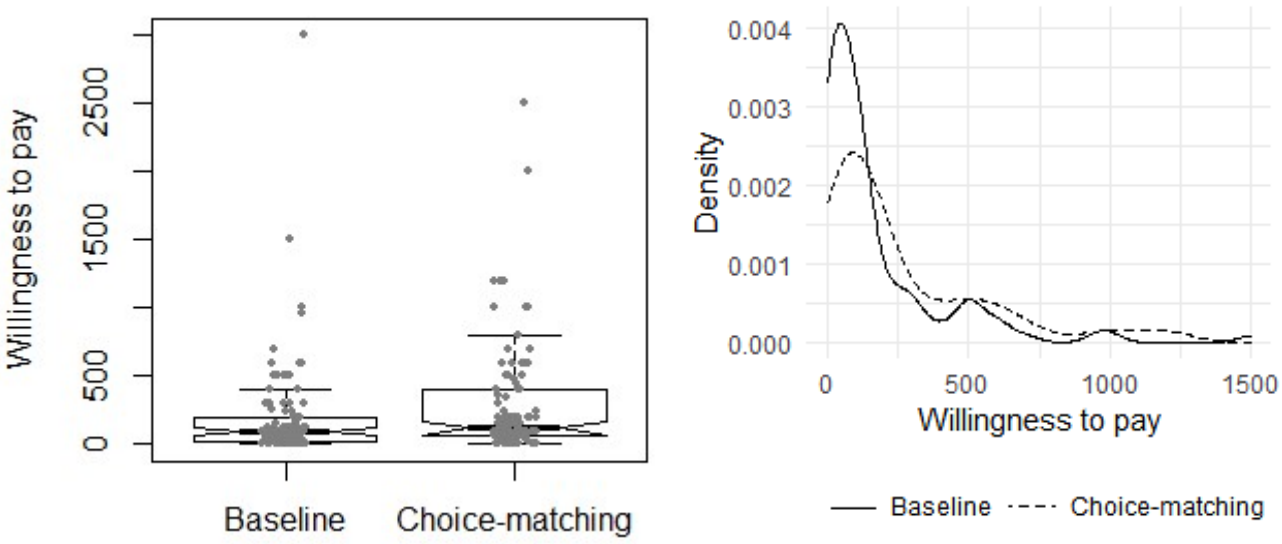
The empirical analysis focuses on whether SP responses are affected by the choice-matching treatment compared to the baseline treatment. The choice-matching treatment is designed to induce truthful preference disclosure, while the baseline treatment mirrors a typical, unincentivized application of SP methods. The SP responses come from the open-ended question about WTP for extending renewable energy infrastructure in Warsaw.

**5.1 Statistical analysis**

The WTP values range from 0 to 3,000 PLN, with a median of 100 PLN in the baseline treatment and 120 PLN in the choice-matching treatment. The mean WTP value is 198 PLN (with a standard deviation of 373) in the baseline treatment and 300 PLN (with a standard deviation of 412) in the choice-matching treatment. The middle 50% of the values (corresponding to the interquartile range) lies between 20 and 200 PLN for the baseline treatment and between 60 and 400 PLN for the choice-matching treatment. The left-hand side of Figure 3 suggests several outliers (represented by dots illustrating individual responses), in particular statements of 3,000 PLN in the baseline treatment and of 2,000 and 2,500 PLN in the choice-matching treatment, which are more than four standard

deviations from the mean values. After excluding these outliers for a clearer image, the right-hand side of Figure 3 presents kernel densities of the SP responses, separately for each treatment. The figure shows that lower WTP values are more common in the baseline treatment, while higher WTP statements appear more frequently in the choice-matching treatment. The Wilcoxon rank sum test indicates a statistically significant difference in the distributions of WTP values across treatments (p-value = 0.006 for the null hypothesis of no difference).

Figure 3. Distributions of the willingness-to-pay (WTP) values for the renewable energy program in the two treatments




The kernel density estimates presented in Figure 3 also suggest that the distribution of high WTP responses, particularly in the baseline treatment, may be characterized by concentration around prominent numbers (Albers and Albers 1983), such as 500 PLN or 1,000 PLN. Although few observations in this treatment represent such high WTP values (as shown by the dots in the boxplots in Figure 3), we examine whether respondents differ across treatments in the frequency of using prominent or round numbers for value statements.

Following the definitions in Converse and Dennis (2018), we refer to round numbers as those with one digit or divisible by 10, and to prominent numbers as "the powers of ten, their doubles, and their halves [… 5, 10, 20, 50, 100, 200…]".[14] Their findings suggest that focusing on prominent numbers may serve as a mental shortcut for individuals experiencing a high cognitive load in decision-making. As cognitive load may be higher in the choice-matching treatment, respondents might be more likely to use decision simplification strategies in this treatment.

[14] Formally, both types of numbers can be represented by $n10^i$, where (a) for round numbers, i is any non-negative integer and n is any integer {1, ..., 9}, and (b) for prominent numbers, i is any non-negative integer and n is 0.5, 1, or 2 (with the exception that for i = 0, n can be only 1 or 2).

The data analysis indicates that the shares of prominent and round numbers in WTP statements do not differ across treatments. Prominent numbers are selected by 46% of respondents in the baseline treatment and 52% in the choice-matching treatment (46 and 49 respondents, respectively). Round numbers appear in 81% of responses in the baseline treatment and 87% in the choice-matching treatment (81 and 83 respondents, respectively). Fisher's exact tests do not indicate significant differences in these proportions across treatments (p-values equal to 0.48 and 0.25, respectively). In the spirit of the conclusions of Converse and Dennis (2018), these results may contribute to the evidence that focusing on prominent (and round) numbers cannot be explained by unincentivized decision settings. This may also be interpreted as evidence that the potentially higher cognitive load in the choice-matching treatment did not lead respondents to rely more on such shortcuts in reporting WTP values, which is consistent with the evidence from Table 1 showing generally similar levels of understanding of the experimental procedures across treatments.

To estimate mean WTP values in the two treatments and to further condition them on respondents' characteristics, we now turn to econometric modelling.

**5.2 Econometric analysis**

The open-ended preference elicitation question provides a continuous measure of respondents' WTP. The data can be viewed as left-censored at zero, as only non-negative values were allowed in the questionnaire (although negative WTP is theoretically possible). In the study sample, 16 respondents indicated a value of zero, which may signal negative WTP—11 in the baseline treatment and 5 in the choice-matching treatment. To account for left censoring, we employ a censored regression—a Tobit model (Tobin 1958)—for econometric modelling of the value statements.

Formally, the model can be represented by:

$$WTP_i^* = x_i\beta + \varepsilon_i\,, \tag{1}$$

$$WTP_i = \begin{cases} 0 & \text{if } WTP_i^* \le 0 \\ WTP_i^* & \text{if } WTP_i^* > 0 \end{cases}, \tag{2}$$

where $WTP_i$ stands for the observed WTP, that is, respondent *i*'s answer to the SP question; $WTP_i^*$ is the respondent's uncensored WTP; $x_i$ denotes a row vector of explanatory variables; $\beta$ is a column vector of unknown parameters to be estimated; and $\varepsilon_i$ is a normally distributed error term with zero mean and variance $\sigma^2$.

The model is estimated with the maximum likelihood method. Setting $D_i = 1$ if a response is uncensored (that is, a positive WTP value) and $D_i = 0$ if a response is censored (that is, a zero WTP value), the log-likelihood function for the model is:

$$\log L = \sum_{i=1}^{N} \left\{ D_i \cdot \log\left[ \frac{1}{\sigma} \phi\left( \frac{WTP_i - x_i\beta}{\sigma} \right) \right] + (1 - D_i) \cdot \log\left[ 1 - \Phi\left( \frac{x_i\beta}{\sigma} \right) \right] \right\}, \quad (3)$$

where $\phi$ and $\Phi$ represent the probability density function and the cumulative distribution function of the standard normal distribution, respectively. As a result, the first term in equation (3) corresponds to the log-likelihood for the uncensored, observed data and the second term is related to the log-likelihood for the censored data.[15]

Results of the WTP Tobit regressions, together with estimated marginal effects, are provided in Table 2. Model 1 focuses on the sole treatment effect (captured by the zero-one-coded treatment indicator variable "Choice-matching"), and Model 2 includes additional explanatory variables capturing observable heterogeneity in WTP values related to respondents' characteristics and their perceptions of the SP question. "Certain about the value statement" equals 1 for respondents who are certain or totally certain about their stated WTP value, and 0 otherwise. "Sufficient information" equals 1 for respondents who agree or definitely agree with the statement "I had sufficient information to indicate well the value in the solar panels question", and 0 otherwise. Variables "Reduce car use" and "Select eco-products" equal 1 for respondents who agree or definitely agree with the respective statements, and 0 otherwise. The statements are: "I limit car/taxi use for environmental reasons" and "I choose products labeled as eco-friendly".[16] Socio-demographic variables are used as defined in Table 1.

Table 2. Results of WTP Tobit regressions

| | Model 1 | | Model 2 | |
|---|---|---|---|---|
| | Coefficient (st. err.) | Marginal effect (st. err.) | Coefficient (st. err.) | Marginal effect (st. err.) |
| Choice-matching | 119.38** (60.08) | 84.41** (42.55) | 123.74** (60.17) | 87.88** (42.81) |
| Certain about the value statement | | | -0.27 (63.23) | -0.19 (44.90) |
| Sufficient information | | | 4.50 (61.96) | 3.20 (44.01) |
| Reduce car use | | | 92.87 (65.96) | 65.96 (46.90) |
| Select eco-products | | | 117.35* (64.02) | 83.34* (45.55) |
| Age | | | -5.88 | -4.18 |

[15] While the modelling approach discussed here assumes homoscedasticity, meaning that the error variance $\sigma^2$ is constant, some studies recommend accounting for potentially different error variances when preference data from different experimental treatments are jointly used (e.g., Haab et al. 1999). Results of Tobit models allowing the error variances to vary across treatments are presented in Appendix Table A1. We do not observe the error variance to differ statistically significantly across treatments, which is why the simpler, homoscedastic model is used for the main analysis.

[16] The questionnaire includes three more statements concerning pro-environmental behaviors. These are not included in the empirical analysis as the answers are characterized by little variation, with more than 80% of respondents agreeing or definitely agreeing with the statements.

| | | | |
|---|---|---|---|
| | | (7.57) | (5.39) |
| Woman | | 19.60 | 13.92 |
| | | (64.24) | (45.61) |
| Academic degree | | 4.15 | 2.95 |
| | | (81.33) | (57.76) |
| Employed full-time | | 54.42 | 38.65 |
| | | (72.52) | (51.53) |
| Intercept | 167.76*** | 176.41 | |
| | (42.18) | (214.33) | |
| Logarithm of $\sigma$ | 6.03*** | 6.01*** | |
| | (0.05) | (0.05) | |
| Log-likelihood | -1,347.11 | -1,343.46 | |

*Notes*: Standard errors are given in the brackets. ***, **, and * imply significance at the 1%, 5%, and 10% level, respectively. Each model is based on 195 observations.

Results from both model specifications indicate that WTP statements are statistically significantly higher in the choice-matching treatment than in the baseline treatment—the coefficient on the choice-matching indicator is statistically different from zero at the 5% significance level. The WTP for the proposed extension of the renewable energy infrastructure is higher in the choice-matching treatment by 84 PLN on average according to Model 1 and by 88 PLN according to Model 2. Considering the standard errors, these estimates from the two model specifications are statistically indistinguishable. Model 2 further indicates that respondents who select ecologically labeled products are willing to pay, on average, 83 PLN more for the proposed program. The remaining explanatory variables do not appear to explain variation in WTP systematically.

To assess the robustness of our results, we consider several alternative model specifications. First, we estimate both models (Model 1 and Model 2) on a sample restricted based on response time to the preference elicitation question. Specifically, for each treatment, we exclude the 20% of respondents with the shortest response times and the 20% of respondents with the longest response times to that question. This is intended to exclude respondents who may not have paid sufficient attention to the key (preference elicitation) question. Those who answered very quickly could have skipped necessary information to make informed choices about their WTP. Those who answered very slowly could have been distracted by other tasks without paying sufficient attention to the question. Second, we estimate Model 2 excluding variables related to perceived certainty and sufficiency of information to avoid potential endogeneity concerns. Third, we estimate Model 2 including a variable measuring income. Because some respondents did not report income, this specification uses a smaller sample. Results of these supplementary regressions are reported in Appendix Table A2.

These results consistently point to the same conclusion: WTP amounts are higher in the choice-matching treatment than in the baseline treatment. The estimated effect sizes are similar, indicating increases in WTP by 76, 81, and 88 PLN in the choice-matching treatment. Only the last specification, estimated on the reduced sample with income reported, suggests a larger increase of 110 PLN on

average. Among the remaining explanatory variables, only the behavior of selecting ecologically labeled products is consistently associated with higher WTP. Thus, we conclude that the supplementary regressions support the robustness of our findings.

## 6 Discussion on the results' validity

The purpose of this section is to shed light on the validity of the obtained results. A natural question is whether the value estimates derived from the choice-matching treatment represent respondents' preferences more accurately than the estimates derived from the baseline treatment. As the true preferences are unobserved, we cannot assess this directly.[17] Instead, we discuss the results' validity by considering (1) whether the auxiliary question in our implementation plausibly separates respondent types in the sense required by the generalized version of choice-matching, and (2) whether the finding of increased WTP under the incentive-compatible mechanism is justified on the grounds of theory and existing evidence.

### 6.1 Is the auxiliary question plausibly type-separating?

For choice-matching in its generalized version to induce truth-telling, the key requirement is that the auxiliary question be sufficiently type-separating. In our application, this implies that a respondent needs to perceive that those who report WTP values for the renewable energy program similar to hers also indicate contribution amounts for the new forest that are similar to hers, and that these contribution amounts are more similar to her own than those chosen by respondents with different WTP values.[18] In our empirical analysis, we therefore treat a positive correlation between WTP values and contribution amounts as suggestive evidence consistent with this requirement, while acknowledging that it is not a direct test of the formal theoretical condition. The correlation analysis is consistent with the suggestion of Cvitanić et al. (2019). While the data support a positive correlation between the two amounts for the choice-matching treatment, the essential question is whether respondents perceive this correlation as likely—these perceptions underlie the incentive properties of the mechanism. We first discuss the correlation and then turn to the perceptions.

Spearman's rank correlation indicates a statistically significant correlation of 0.145 between WTP values and contribution amounts (p-value of 0.044), based on the full sample of 195 respondents. To assess validity of the obtained results, an important question is whether this correlation differs across treatments. Based on theory, it can be expected that the correlation is weaker in the baseline

[17] With this study, we intend to best mirror a typical environment of a SP survey. Thus, our setting relies on the elicitation of home-grown values, that is, values with which study participants come to the study, which are typically unobservable and cannot be objectively measured. An alternative design, allowing for an unambiguous assessment of the truthfulness of stated WTP values, could involve induced values, which a researcher defines and imposes on study participants.

[18] We note that this is closely related, but not identical, to impersonal updating in the canonical prediction-based model.

treatment, where unincentivized WTP responses are more likely to diverge from true preferences and, thus, be less closely related to contribution amounts. Treatment-specific results of Spearman's correlation support this expectation. In the baseline treatment, the correlation coefficient is 0.114 with p-value of 0.258, indicating no statistically significant relationship. In the choice-matching treatment, the coefficient is 0.194 with p-value of 0.060, suggesting a stronger correlation statistically significant at the 10% level. Hence, we conclude that the correlation is stronger in the choice-matching treatment, in line with theoretical expectations.

We now turn to whether individuals' perceptions support the requirement that the auxiliary question is type-separating. To examine whether perceived relationships between WTP values and contribution amounts are consistent with this requirement, we conduct a separate, brief survey. In this survey, we describe the WTP and contribution questions using, to the degree possible, identical scripts as in the original study. Respondents are told that they will not answer these questions themselves, but instead they are asked to predict how other respondents from the same panel answered them in a study that was already completed. For correctly identifying the positive relationship observed in the empirical data (and required by the type-separation logic), respondents are entered into a random draw to win 50 PLN. After this introduction, respondents are asked (boldface as in the survey):

"Please consider whether you expect any relationship between answers to the two questions. From the scenarios below, select the one that appears most likely to you. Among those who indicate the correct scenario, we will randomly draw four persons, and each of them will receive 50 PLN.

a) Those respondents who gave **more** money for the forest also gave on average **more** money for the solar panels.
b) Those respondents who gave **more** money for the forest also gave on average **less** money for the solar panels.
c) There was no relationship between answers to these two questions."

Two versions of the questionnaire are implemented. One includes the full description of the choice-matching mechanism (as in the original survey), explaining that the contribution amount paid by a selected respondent could be determined by the mean contribution amount of other respondents of the same type. The other version omits this detailed explanation and informs only that: "The study employed a special mechanism whereby it was in the best interest of respondents to provide the amount they would truly be willing to pay for the considered program of building the solar panels."

The survey was administered to 166 respondents from the same panel as the original sample in October and November 2021. It was ensured that the respondents from the main study did not participate in this follow-up survey.

In both versions of the follow-up survey, the most frequently selected answer indicates a positive correlation. This answer is chosen by 54% of respondents in the version with the abbreviated description of the choice-matching mechanism and by 50% of respondents in the version with the

full description. The opposite relationship is selected by 25% and 29% of respondents, respectively, while the remaining respondents select no relationship. These results provide suggestive evidence that, on average, respondents in the main survey are likely to expect a positive correlation. No questions for a similar assessment are included in our main survey to avoid raising doubts about the credibility of the mechanism.

Arguably, in future studies, this type of additional investigation to verify the requirement of type-separation in the auxiliary question should ideally be conducted *before* the main study, as a pretest of the stimuli to be deployed. Else, researchers run the risk of finding no treatment effect for rather uninteresting reasons or, especially if only the choice-matching treatment is conducted, of not realizing they have failed to implement the intended manipulation.

**6.2 Is it justified to observe increased WTP under choice-matching?**

Our empirical results indicate that WTP value statements are higher under choice-matching than in the unincentivized survey environment. This raises a natural question about the results' validity, particularly in light of evidence that these are typically the hypothetical settings that yield high value estimates compared to binding, real-payment settings (e.g., Penn and Hu 2018). To address this, we present arguments below demonstrating that the finding of increased open-ended-based WTP estimates upon choice-matching aligns well with the existing theoretical and empirical literature.

The main difference between our two treatments is that one (choice-matching) provides detailed information about how the collected data will be used, whereas the other one (baseline) does not. Importantly, the provision of this information does not change the hypothetical nature of the SP question, which remains hypothetical in both treatments. Therefore, standard hypothetical-bias comparisons—typically involving SP elicitation in hypothetical versus binding payment conditions—are not directly applicable to our research context. In our setting, the WTP responses neither in the choice-matching treatment nor in the baseline treatment are tied to actual payments—respondents are never required to pay their stated WTP. Furthermore, hypothetical-bias studies usually vary the hypothetical versus binding nature of choices within the same question format. Consequently, such comparisons do not take into account other biases that may arise in non-incentive-compatible mechanisms of preference elicitation (such as an open-ended SP format), which may interact with the hypothetical or binding nature of choices, for example, strategic misrepresentation of preferences. This does not apply to our study, where the preference elicitation question is hypothetical across both treatments.

Our finding of higher WTP under choice-matching applied to open-ended elicitation aligns with the results of Vossler and Holladay (2018), whose study we view as most closely related to ours. They identify theoretical conditions under which an open-ended elicitation format is incentive compatible and implement a theory-informed design in a field SP survey. They then compare WTP estimates

obtained from this incentive-compatible elicitation with those from a standard open-ended format, which lacks a mechanism to induce truthful preference disclosure. As in our study, they also compare open-ended WTP responses across incentive-compatible and unincentivized settings, where the primary difference between the two lies in the information provided about how the responses will be used. Thus, the treatments differ in whether an incentive-compatible mechanism is employed. Consistent with our findings, Vossler and Holladay (2018) report higher WTP estimates under the incentive-compatible design. These results provide empirical support for the idea that informing respondents about the data-use mechanism, which is necessary for incentive compatibility of preference elicitation, can increase stated values in open-ended SP questions. This suggests that varying the provision of such information may prompt respondents to engage in different considerations than those arising in typical hypothetical-bias studies. Finally, Vossler and Holladay (2018, p. 144) conclude "That the standard [open-ended] elicitation yields lower values [compared to the theory-informed incentive-compatible preference elicitation] is consistent with the stylized fact from the literature."

A potential concern is whether the longer survey script, needed for explaining best-response strategies in choice-matching and similarly in the theory-informed incentive-compatible mechanism of Vossler and Holladay (2018), may itself lead to higher WTP by increasing attention to or engagement with the SP elicitation procedure. While this possibility cannot be entirely ruled out, we argue that our data do not support this interpretation. If script length alone were driving the treatment effect, one might expect higher perceived importance of the evaluated program under the longer script. However, we find no statistically significant difference in perceived importance of the program of building solar panels across the baseline and choice-matching treatments with p-value of 0.9654 (cf. Table 1). This suggests that the increase in WTP under choice-matching is more likely driven by the *content* of the script, specifically the explanation of beneficial strategies to answer the preference elicitation question, rather than its length.

From a theoretical perspective, Carson and Groves (2007) discuss the beliefs that respondents in open-ended valuation settings may form regarding how their responses will be used when this is not explicitly specified—that is, when information ensuring incentive compatibility of the SP elicitation mechanism is absent. This discussion indicates that several interpretations of the possible data use mechanism conjectured by respondents can encourage understatement of the true WTP in unincentivized settings. Similarly, Carson and Hanemann (2005, p. 893) observe that "The incentive structure of open-ended CV surveys gives rise to a substantial number of 'zero' WTP responses and few positive, yet very small, responses."

Following Carson and Groves (2007), consider a respondent in a standard, unincentivized, open-ended SP survey who has some guess about the potential individual cost of providing the good in question (e.g., implementing a policy program). If the respondent perceives this cost to exceed their true WTP, the person may be inclined to underreport their true WTP and instead state a zero value.

Reporting zero requires minimal effort and may reduce the likelihood of the program being implemented or of high costs being imposed if policymakers rely on stated WTP in their decision-making. Moreover, if stated WTP values are used to determine the cost per individual, then a zero response may be viewed as a way to lower an otherwise burdensome payment. Accordingly, Carson and Groves (2007, p. 202) conclude that for individuals whose WTP is lower than the perceived cost of the program, reporting zero constitutes the best strategy "under most plausible uses of the information provided." For individuals whose WTP exceeds the perceived cost, the best strategy depends on their beliefs about how stated WTP will be used in the decision-making. If aggregate WTP is used to determine whether the good will be provided, then it may pay off for respondents to overstate their true WTP. Conversely, if stated WTP values are interpreted as signals for setting individual cost levels, understating WTP may again emerge as the strategically best response.

Taken together, both empirical evidence and theoretical considerations suggest that WTP in unincentivized, open-ended SP settings may be understated. Introducing an incentive-compatible mechanism, which encourages truthful preference disclosure, may therefore increase observed WTP values reported in the open-ended setting. Nevertheless, we acknowledge that the most direct test of validity would involve comparing elicited preferences to true preferences, for example, in an induced-value experiment.

## 7 Conclusions

We present an empirical application of a novel approach to elicit SP responses in an incentive-compatible manner, following the mechanism proposed by Cvitanić et al. (2019), which induces truthful responding when the truth is not verifiable. We illustrate the use of the mechanism in a SP survey and examine how respondents behave when facing this approach compared to behavior in a standard, unincentivized survey setting.

Our preference elicitation involves an open-ended question concerning WTP for extending renewable energy infrastructure in Warsaw with new solar panels. Both statistical and econometric analyses, including Tobit regressions, indicate that WTP value statements are statistically significantly higher under choice-matching than in unincentivized survey conditions. This finding aligns with existing empirical evidence that a standard open-ended format without a mechanism inducing truthful responding leads to understatement of true WTP (Vossler and Holladay 2018), as well as with theoretical considerations suggesting that WTP understatement may be the best response strategy in an open-ended elicitation setting that does not explain how SP data will be further used (Carson and Groves 2007; Carson and Hanemann 2005). Together with respondents' self-reports indicating that they understood the study instructions, this may suggest that respondents comprehended the choice-matching procedure in our application and thus are likely to have responded in a manner consistent with their underlying preferences. However, we acknowledge that we do not have an objective (criterion) measure of the truthfulness of the SPs.

Such a measure could be obtained, for example, through an induced-value experiment. Instead, the aim of this study is to capture, as closely as possible, the typical characteristics of SP application environments.

The use of incentive-compatible designs, which align respondents' interests with truthful responding, is recommended in SP studies (Johnston et al. 2017). One of the characteristics underlying incentive-compatible elicitation in SP literature is consequentiality, which implies that respondents' answers are not purely hypothetical but may influence final decisions. Against this background, a new line of research suggests that consequential preference statements may be biased as due to respondents experiencing responsibility utility (Comerford and Lades 2022): in consequential SP surveys, respondents may feel responsible for a final decision and thus provide answers that do not reflect the utility they would experience as passive recipients of policy outcomes (i.e., without the utility from potentially influencing the policy outcome through survey choices). Comerford and Lades (2022) argue that this creates a SP design dilemma, requiring a trade-off between biases arising from responsibility utility and those associated with unincentivized, hypothetical settings. The choice-matching approach proposed by Cvitanić et al. (2019) and empirically implemented in this study in the context of public good valuation may provide a way to address this dilemma.

Choice-matching offers a potential route to incentive-compatible elicitation of hypothetical SP responses, but the mechanism does not operate through consequentiality, unlike standard SP approaches. This is a distinguishing feature compared to existing approaches for incentive-compatible elicitation in SP surveys. The absence of direct policy consequences, combined with linking the preference elicitation question to a monetary payoff from the auxiliary question in the choice-matching procedure, may substantially weaken the role of responsibility utility. In particular, responsibility-utility considerations may be mitigated when the auxiliary question involves a decision affecting only a given individual, without consequences for others (such as additional tax burden commonly related to provision of public goods in SP surveys). Such an auxiliary question is likely the most practical way to induce truthful preference statements with choice-matching. At the same time, the choice-matching procedure provides respondents with economic reasons for truthful reporting, allowing SP researchers to mitigate potential responsibility-utility bias without resorting to unincentivized survey designs.

Naturally, the potential benefits of using choice-matching for SP elicitation come at a cost as the mechanism is not free from challenges. One is that the auxiliary question must be sufficiently type-separating for the mechanism to be incentive compatible. In our setting, this means that respondents with similar WTP for the renewable energy program should be expected to indicate similar forest contribution amounts, and that these similarities should be stronger within than across types defined by WTP groups. This requirement calls for careful survey design and thorough analysis

to understand respondents' perceptions and motivations behind answering the main and auxiliary questions in a survey.

Another challenge is the complexity of the mechanism. Although it is argued to be among the simpler methods for eliciting non-verifiable truth, it still requires extensive explanations and non-negligible attention from respondents to understand the procedures (Baillon et al. 2024). While our empirical application proposes a way to operationalize the mechanism in a SP survey, we acknowledge that further research is needed to ascertain the clarity and effectiveness of communicating the underlying rationale to respondents. In our study, comprehension is assessed based on respondents' self-reports, while incorporating objective measures, such as comprehension quizzes, would be a useful extension for future research to more rigorously verify respondents' understanding. Furthermore, implementing choice-matching increases survey length due to additional explanations and the auxiliary question, which may exacerbate fatigue effects (already documented in SP studies; e.g., Campbell et al. 2015). Consequently, a researcher considering an application of choice-matching faces important tradeoffs between the benefits of incentive-compatible elicitation of preferences for potentially hypothetical goods and the potential costs of increased complexity and respondent fatigue.

These implementation challenges also point to limitations of our study. Although we made substantial efforts to design the survey so that the auxiliary question would plausibly separate respondent types, we acknowledge that a stronger correlation between responses to the main and auxiliary questions could strengthen the mechanism's ability to elicit truthful responses. Moreover, the expected correlation between these responses relies on underlying attitudes toward environmental protection influencing responses to both questions in a similar direction. At the same time, other considerations may weaken the correlation. One of them, and likely most important in the context of our application, is a difference in attitudes toward the two payment vehicles used in the questions—an obligatory tax to all in the SP question versus a voluntary contribution in the auxiliary question. Individuals may perceive the efficiency of these two payment schemes differently. For example, some may view taxation as an efficient means of funding environmental protection, while considering voluntary payments (such as donations) an inefficient way of financing environmental protection programs. Such differences could weaken the correlation between responses to the main and auxiliary questions in our study. Nevertheless, since we observe a positive correlation in our data, it is plausible that environmental attitudes affect both responses stronger than, for instance, (diverging) attitudes towards payment schemes.

This study demonstrates the applicability of the theoretically proposed choice-matching mechanism (Cvitanić et al. 2019) to SP valuation of a public good. It also presents an adaptation of choice-matching to settings with continuous response scales (which are subsequently converted into a finite set of value intervals for the mechanism's implementation). Through this application, the paper operationalizes the technical requirements of choice-matching and illustrates how they can be

implemented in SP surveys for public good valuation. In doing so, it contributes to expanding the set of incentive-compatible designs for SP elicitation. A key feature of the choice-matching approach is that it enables incentive-compatible preference elicitation in non-consequential settings.

**Appendix**

Table A1. Results of WTP Tobit regressions allowing for the variance heteroscedasticity across the treatments

| | Model A1<br>Coefficient<br>(st. err.) | Model A2<br>Coefficient<br>(st. err.) |
|---|---|---|
| Choice-matching | 117.72*<br>(60.15) | 119.95**<br>(60.36) |
| Certain about the value statement | | -6.82<br>(63.40) |
| Sufficient information | | -9.10<br>(61.90) |
| Reduce car use | | 92.06<br>(65.83) |
| Select eco-products | | 120.91*<br>(63.83) |
| Age | | -6.11<br>(7.59) |
| Woman | | 24.56<br>(64.23) |
| Academic degree | | -1.88<br>(81.76) |
| Employed full-time | | 62.77<br>(73.37) |
| Intercept | 168.91***<br>(40.92) | 182.25<br>(213.30) |
| Standard deviation function ( $\sigma$ ) | | |
| Choice-matching | 0.06<br>(0.11) | 0.09<br>(0.11) |
| Constant | 5.99***<br>(0.08) | 5.96***<br>(0.08) |
| Log-likelihood | -1,347 | -1,343 |

*Notes*: Standard errors are given in the brackets. ***, **, and * imply significance at the 1%, 5%, and 10% level, respectively. Each model is based on 195 observations.

Table A2. Results of supplementary WTP Tobit regressions for robustness analysis

| | Model A3 | | Model A4 | | Model A5 | | Model A6 | |
|---|---|---|---|---|---|---|---|---|
| | Coeffi-cient | Marginal effect | Coeffi-cient | Marginal effect | Coeffi-cient | Marginal effect | Coeffi-cient | Marginal effect |
| Choice-matching | 105.30** | 75.86** | 111.23** | 80.60** | 124.08** | 88.12** | 152.84** | 109.61** |
| | (50.44) | (36.44) | (50.79) | (36.93) | (59.42) | (42.28) | (65.47) | (47.08) |
| Certain about the value statement | | | 12.14 | 8.8 | | | 21.11 | 15.14 |
| | | | (53.87) | (39.04) | | | (69.50) | (49.85) |
| Sufficient information | | | -35.06 | -25.4 | | | 5.18 | 3.71 |
| | | | (51.13) | (37.04) | | | (67.95) | (48.73) |
| Reduce car use | | | 59.32 | 42.99 | 92.67 | 65.81 | 99.10 | 71.07 |
| | | | (54.22) | (39.32) | (65.78) | (46.77) | (72.67) | (52.17) |
| Select eco-products | | | 80.7 | 58.48 | 117.79* | 83.65* | 119.67* | 85.82* |
| | | | (54.54) | (39.62) | (63.29) | (45.04) | (69.69) | (50.07) |
| Age | | | -5.45 | -3.95 | -5.86 | -4.16 | -4.23 | -3.04 |
| | | | (6.36) | (4.62) | (7.55) | (5.36) | (8.19) | (5.87) |
| Woman | | | 9.43 | 6.84 | 19.49 | 13.84 | 8.43 | -6.04 |
| | | | (53.05) | (38.43) | (64.22) | (45.59) | (70.23) | (50.36) |
| Academic degree | | | -54.73 | -39.66 | 3.82 | 2.71 | 15.92 | 11.42 |
| | | | (77.01) | (55.83) | (81.21) | (57.67) | (89.77) | (64.38) |
| Employed full-time | | | 33.39 | 24.2 | 54.37 | 38.61 | 63.64 | 45.64 |
| | | | (59.48) | (43.12) | (72.06) | (51.2) | (77.43) | (55.57) |
| Income | | | | | | | -0.00 | -0.00 |
| | | | | | | | (0.01) | (0.01) |
| Intercept | 105.93*** | | 224.66 | | 178.43 | | 112.19 | |
| | (35.50) | | (177.05) | | (212.46) | | (234.25) | |
| Logarithm of $\sigma$ | 5.60*** | | 5.57*** | | 6.01*** | | 6.04*** | |
| | (0.07) | | (0.07) | | (0.05) | | (0.05) | |
| Log-likelihood | -753.32 | | -750.32 | | -1,343.46 | | -1,212.35 | |
| Number of observations | 117 | | 117 | | 195 | | 174 | |

*Notes*: Standard errors are given in the brackets. ***, **, and * imply significance at the 1%, 5%, and 10% level, respectively. Models A3 and A4 are estimated on a sample restricted on a basis of the response time to the preference elicitation question—20% of respondents with the shortest response time and 20% of respondents with the longest response time to that question are excluded. Model A5 is estimated on the full sample, and Model A6 omits respondents who did not report their income level.

**References**


Abramson, P. R., Aldrich, J. H., Diamond, M., Diskin, A., Levine, R., and Scotto, T. J. (2004). Strategic abandonment or sincerely second best? The 1999 Israeli prime ministerial election. Journal of Politics, 66(3), 706-728.

Albers, W., and Albers, G. (1983). On the prominence structure of the decimal system. Advances in Psychology, 16, 271-287.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H. (1993). Report of the NOAA panel on contingent valuation. Federal Register, 58(10), 4601-4614.

Baillon, A. (2017). Bayesian markets to elicit private information. Proceedings of the National Academy of Sciences, 114(30), 7958-7962.

Baillon, A., Peker, C., and van der Zee, S. (2024). Peer prediction markets to elicit unverifiable information. Retrieved April 30, 2024, from: https://cempeker.github.io/docs/ppmarket.pdf.

Baillon, A., and Xu, Y. (2021). Simple bets to elicit private signals. Theoretical Economics, 16(3), 777-797.

Barrage, L., and Lee, M. S. (2010). A penny for your thoughts: Inducing truth-telling in stated preference elicitation. Economics Letters, 106(2), 140-142.

Bauer, R., Dong, B., and Jiao, P. (2024). Sustainability preferences: The role of beliefs. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4889330.

Börger, T., Abate, T. G., Aanesen, M., and Zawojska, E. (2021). Payment and policy consequentiality in dichotomous choice contingent valuation: Experimental design effects on self-reported perceptions. Land Economics, 97(2), 407-424.

Cain, B. E. (1978). Strategic voting in Britain. American Journal of Political Science, 22(3), 639-655.

Campbell, D., Boeri, M., Doherty, E., and Hutchinson, W. G. (2015). Learning, fatigue and preference formation in discrete choice experiments. Journal of Economic Behavior and Organization, 119, 345-363.

Carson, R. T., and Groves, T. (2007). Incentive and informational properties of preference questions. Environmental and Resource Economics, 37(1), 181-210.

Carson, R. T., Groves, T., and List, J. A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. Journal of the Association of Environmental and Resource Economists, 1(1/2), 171-207.

Carson, R.T., and Hanemann, W. M. (2005). Contingent valuation (Chapter 17). In: K.-G. Maler and J. R. Vincent (Eds.), Handbook of Environmental Economics: Valuing Environmental Changes. Amsterdam, The Netherlands: Elsevier, pp. 821-936.

Ceallaigh, D. Ó., Rohde, K. I., and Van Kippersluis, H. (2024). Skipping your workout, again? Measuring and understanding time inconsistency in physical activity. Tinbergen Institute Discussion Paper No. TI 2024-028/V.

Cerroni, S., Derbyshire, D. W., Hutchinson, W. G., and Nayga, R. M. (2023). A choice matching approach for discrete choice analysis: An experimental investigation in the laboratory. Land Economics, 99(1), 80-102.

Charness, G., Gneezy, U., and Rasocha, V. (2021). Experimental methods: Eliciting beliefs. Journal of Economic Behavior and Organization, 189, 234-256.

Comerford, D. A., and Lades, L. K. (2022). Responsibility utility and the difference between preference and desirance: Implications for welfare evaluation. Social Choice and Welfare, 58, 201-224.

Converse, B. A., and Dennis, P. J. (2018). The role of "Prominent Numbers" in open numerical judgment: Strained decision makers choose from a limited set of accessible numbers. Organizational Behavior and Human Decision Processes, 147, 94-107.

Cvitanić, J., Prelec, D., Riley, B., and Tereick, B. (2019). Honesty via choice-matching. American Economic Review: Insights, 1(2), 179-92.

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

Duverger, M. (1954). Political parties: Their organization and activity in the modern state. New York, NY: John Wiley and Sons.

Interis, M. G., and Petrolia, D. R. (2014). The effects of consequentially in binary- and multinomial-choice surveys. Journal of Agricultural and Resource Economics, 39(2), 201-216.

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., and Tourangeau, R. (2017). Contemporary guidance for stated preference studies. Journal of the Association of Environmental and Resource Economists, 4(2), 319-405.

Haab, T. C., Huang, J. C., and Whitehead, J. C. (1999). Are hypothetical referenda incentive compatible? A comment. Journal of Political Economy, 107(1), 186-196.

Hanemann, M., Loomis, J., and Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics, 73(4), 1255-1263.

Hanley, N., and Czajkowski, M. (2019). The role of stated preference valuation methods in understanding choices and informing policy. Review of Environmental Economics and Policy, 13(2), 248-266.

Herriges, J., Kling, C., Liu, C. C., and Tobias, J. (2010). What are the consequences of consequentiality? Journal of Environmental Economics and Management, 59(1), 67-81.

Kanninen, B. J. (1993). Optimal experimental design for double-bounded dichotomous choice contingent valuation. Land Economics, 69(2), 138-146.

Lehmann, N. V. (2026). Mechanisms for belief elicitation without ground truth. Journal of Economic Surveys, 40(1), 505-527.

Lloyd-Smith, P., Adamowicz, W., and Dupont, D. (2019). Incorporating stated consequentiality questions in stated preference research. Land Economics, 95(3), 293-306.

Loughran, T. A., Paternoster, R., and Thomas, K. J. (2014). Incentivizing responses to self-report questions in perceptual deterrence studies: An investigation of the validity of deterrence theory using Bayesian truth serum. Journal of Quantitative Criminology, 30(4), 677-707.

Mahieu, P. A., Andersson, H., Beaumais, O., Crastes dit Sourd, R., Hess, S., and Wolff, F. C. (2017). Stated preferences: A unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. Review of Agricultural, Food and Environmental Studies, 98, 201-220.

Miller, N., Resnick, P., and Zeckhauser, R. (2005). Eliciting informative feedback: The peer-prediction method. Management Science, 51(9), 1359-1373.

Myers, K., Lakhani, K. R., and Wang, D. (2020). Towards recovery: Scientists with better ratings of their institution's response to the COVID-19 pandemic have more optimistic forecasts about their future research. Available at SSRN 3712876, doi: 10.2139/ssrn.3712876.

Oedingen, C., van Gestel, R., Huls, S. P., Granic, G., de Bekker-Grob, E. W., and Veldwijk, J. (2025). Association of medication adherence with treatment preferences: incentivizing truthful self-reporting. The European Journal of Health Economics, 26, 1219-1232.

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

Prelec, D. (2004). A Bayesian truth serum for subjective data. Science, 306(5695), 462-466.

Radanovic, G., and Faltings, B. (2013). A robust Bayesian truth serum for non-binary signals. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica, 26(1), 24-36.

Voslinsky, A., and Azar, O. H. (2021). Incentives in experimental economics. Journal of Behavioral and Experimental Economics, 93, 101706.

Vossler, C. A., Doyon, M., and Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. American Economic Journal: Microeconomics, 4(4), 145-171.

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

Vossler, C. A., and Watson, S. B. (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. Journal of Economic Behavior and Organization, 86, 137-147.

Vossler, C. A., and Zawojska, E. (2020). Behavioral drivers or economic incentives? Toward a better understanding of elicitation effects in stated preference studies. Journal of the Association of Environmental and Resource Economists, 7(2), 279-303.

Weaver, R., and Prelec, D. (2013). Creating truth-telling incentives with the Bayesian truth serum. Journal of Marketing Research, 50(3), 289-302.

Witkowski, J., and Parkes, D. C. (2012). A robust Bayesian truth serum for small populations. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence.

Zawojska, E., Bartczak, A., and Czajkowski, M. (2019). Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. Journal of Environmental Economics and Management, 93, 63-84.