From simple to complex: A revealed preference test of discrete choice experiment designs

Christian A. Vossler
Ewa Zawojska

August 2025

WORKING PAPER #2025-03

WORKING PAPER SERIES
DEPARTMENT OF ECONOMICS
HASLAM COLLEGE OF BUSINESS
http://econ.bus.utk.edu

HASLAM COLLEGE OF BUSINESS
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

From simple to complex: A revealed preference test of discrete choice experiment designs

Christian A. Vossler*
Department of Economics, University of Tennessee, Knoxville, TN, USA
cvossler@utk.edu

and

Ewa Zawojska
Faculty of Economic Sciences, University of Warsaw, Warsaw, Poland
ewa.zawojska@uw.edu.pl

---

* Correspondence should be directed to Christian Vossler, Department of Economics, The University of Tennessee, Knoxville, TN 37996. E-mail: cvossler@utk.edu. Telephone: 865-974-1699. Fax: 865-974-4601. We thank Vic Adamowicz, Wiktor Budzinski, Mikolaj Czajkowski, Robert Johnston, and seminar participants at University of Massachusetts, University of Wyoming, Association of Environmental and Resource Economists annual conference, and European Association of Environmental and Resource Economists annual conference. Cathy Kling provided important input on the experimental design and survey instrument. Approval was obtained from the Institutional Review Board at Cornell University, as protocol IRB0145635, and the University of Tennessee, as protocol UTK IRB-23-07448-XM. The study was pre-registered as: Vossler, C., and Zawojska, E. (2023). "Effects of varying the numbers of choice options and choice questions in discrete choice experiments for valuing public goods." AEA RCT Registry. https://doi.org/10.1257/rct.11342-1.1. This article was developed under Assistance Agreement No. 83616602 awarded by the U.S. Environmental Protection Agency. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication. Ewa Zawojska acknowledges the support of the Polish National Agency for Academic Exchange under the National Component of the Bekker Programme.

From simple to complex: A revealed preference test of discrete choice experiment designs

Abstract: Researchers employing discrete choice experiments to value publicly provided goods face several experimental design decisions. Fundamental amongst them are whether to ask participants one or many choice questions to elicit valuations, and how many choice options to include in each. To provide guidance for researchers tackling these decisions, and policy makers charged with interpreting welfare estimates based on these decisions, we conducted a financially incentivized online field experiment to provide a ground truth comparison of three leading elicitation approaches. A single binary choice question and a sequence of binary choice questions yield equal willingness-to-pay estimates. A sequence of trinary choices results in lower demand estimates. The latter approach, while dominant in the stated preference literature, encourages serial status quo choices due to increased task complexity, and is prone to framing effects in that the value for one good depends on the other good included in the choice set. These behavioral effects more than offset the theoretical efficiency advantage of this elicitation approach.

Keywords: convergent validity; discrete choice experiment; mechanism design; field experiment; stated preferences; voting; elicitation effects

JEL classification: C9; D61; D82; H4; Q51

1. Introduction

For sixty years economists have used stated preference (SP) surveys to estimate the economic values people place on nonmarket goods, including potential public policies and goods currently unavailable in markets. The applications are vast, spanning the environment, health, marketing, transportation, immigration, and employment search. Moreover, many countries routinely use SP estimates in regulatory analysis. There is considerable variation in the design of these surveys and mounting evidence that valuations are sensitive to design choices; thus, fundamental design issues remain. Prominent among them is the choice of the value elicitation mechanism, and in particular the form of the question or set of questions used to elicit valuations. In this study we conduct a *revealed preference* field experiment informed by mechanism design theory to carefully examine and compare responses and welfare measures obtained from three popular elicitation approaches: a single binary choice (Single-BC), a sequence of binary choices (Seq-BC), and a sequence of trinary choices (Seq-TC).[2] Our study represents a methodological departure from prior work, which compares question formats using SP experiments; the absence of a ground truth in these SP experiments has made it challenging to translate findings into practical guidance.

A value elicitation mechanism involving a Single-BC question, a coercive payment vehicle, and advisory public referendum framing, has desirable incentive properties and has long been viewed as the benchmark method (Arrow et al. 1993; Carson and Groves 2007; Johnston et al. 2017). At present, it is instead common practice to employ an alternative form of a discrete choice experiment (DCE) that includes a sequence of valuation scenarios (i.e., choice questions), each involving three or more choice options (Rondeau and Vossler 2024). Sequential DCEs

[2] This nomenclature follows Carson and Louviere (2011) with the exception that we refer to the three-option choice set as “trinary” choice, which is more granular than “multinomial” choice.

originated in the fields of marketing and transportation economics for valuing market goods (e.g., Louviere and Hensher 1982; Louviere and Woodworth 1983). Within this context, a common feature is to include multiple alternatives to the status quo in each choice question, which helps to capture relevant market context (e.g., substitute goods). Environmental and health economists began using sequential DCEs in the 1990s for valuing public and publicly provided goods, and early applications included many choice questions and choice options (e.g., Boxall et al. 1996; Adamowicz et al. 1998). Rondeau and Vossler (2024) catalog SP studies published in five environmental economics journals over the period 2018-2022 and find that over 70 percent of studies employ sequential DCEs. Nearly half of the studies (48 percent) utilize a Seq-TC mechanism, 14 percent use Seq-BC, and 11 percent use a sequential DCE with more than three choice options per question. The Single-BC format is included in 14 percent of studies.[3]

Holding the number of respondents constant, increasing the number of choice questions or choice options is conceptually more informationally efficient and allows a researcher to obtain more precise welfare measures.[4] Increasing the number of choice questions can also foster preference learning and mechanism learning (e.g., Czajkowski, Giergiczny, and Greene 2014). Among the potential disadvantages of increasing either the number of choice questions or choice options are increased cognitive burden and, according to mechanism design theory, additional incentives to misrepresent preferences (Vossler, Doyon, and Rondeau 2012). Increasing the number of choice questions may further be associated with fatigue and order effects (e.g., stemming from anchoring on prior choice questions) (Day et al. 2012).

[3] Outside the environmental context, there is less focus on mechanism design issues, and a higher share of applications involve private goods or private provision mechanisms. For these reasons, we suspect the prevalence of sequential DCEs with three or more options to be even higher in other research domains.

[4] In the context of private goods or quasi-public goods, including recreation demand, an additional advantage of including several choice options is to capture the relevant choice context involving substitute goods (e.g., alternative recreation sites).

There is now a large SP literature that examines "elicitation effects", including whether choices are sensitive to variations in the value elicitation mechanism deployed (e.g., Welsh and Poe 1998; Champ and Bishop 2006; Stithou and Scarpa 2012). Directly relevant to our examination, Lloyd-Smith, Zawojska, and Adamowicz (2020) survey the SP literature as it relates to environmental goods, and document eight comparisons between a multinomial, sequential DCE and a Single-BC format. Based on our interpretation of the findings from these studies, Single-BC welfare measures are higher in two studies, lower in three studies, and equal in the three remaining applications. They further document two comparisons between a Single-BC and sequential binary choice elicitation, and in both cases welfare estimates are similar. Also relevant for our research, the common but not universal finding is that increasing the number of choice options per choice question in a sequential DCE leads to differences in welfare measures (see Weng et al. 2021).[5]

The studies documented in the articles referenced above have lacked indicators of actual demand, and therefore empirical differences or similarities found when comparing values elicited with different SP mechanisms do not directly imply anything about their ability to accurately elicit preferences. It is further important to acknowledge that there is no guarantee that even two incentive-compatible mechanisms will lead to equivalent estimates of demand. For example, comparisons between the Becker-DeGroot-Marschak (Becker, DeGroot, and Marschak 1964) mechanism and the second-price auction generally reveal important differences (e.g., Rutström 1998). It is therefore challenging to translate findings from prior convergent validity tests into informed guidance for SP researchers.

[5] We also note that some studies vary both the number of choice sets questions and the number of choice options within the same design (and, in some cases, other dimensions such as the number of attributes and attribute levels). Some of these report important differences in elicited preferences and WTP across elicitation formats differing along the two dimensions (e.g., Hensher 2004; Rose et al. 2009), while others observe no systematic differences in WTP (e.g., Caussade et al. 2005; Hensher 2006).

To move the literature forward, we take an alternative approach by comparing the performance of popular elicitation formats using carefully devised *revealed preference* tests of convergent validity. We construct mechanisms around these formats that, under plausible assumptions, are incentive compatible. To accomplish this, we use plurality voting rules to determine outcomes and, for the sequential formats, a random selection rule to break strategic links across choice sets.[6] Our study therefore attempts to neutralize the main confounder in SP field surveys, which is the extent to which people truthfully state their preferences. In turn, our tests reveal the extent to which the question format shapes actual preferences.

Our study, according to the literature review of Rondeau and Vossler (2024), compares the three most popular SP elicitation formats: Single-BC, Seq-BC, and Seq-TC. Consistent with most applications, the choice sets we employ include a status quo option. We use the mechanisms devised around these formats to elicit values for projects designed to improve the well-being of farmworkers and their families living in New York State.

In addition to the use of direct financial incentives and careful mechanism design, our experimental design facilitates granular comparisons that are unique to literature. We use the Single-BC mechanism to elicit values for two projects, which allows for an external scope test and multiple comparisons with the other two mechanisms. Prior literature has focused on comparing WTP for a single good or policy and, as we demonstrate, such comparisons may fail to identify important differences across single and sequential mechanisms. Respondents facing a particular sequential mechanism all answer a common set of nine choice questions designed to elicit preferences for nine farmworker assistance projects. We estimate project-specific WTP estimates using a utility model specification with project fixed effects. This avoids imposing strong

[6] Assumptions for these mechanisms to be incentive compatible are presented in more detail in Section 2.2.

functional form assumptions; importantly, the estimation avoids defining utility is a linear function of project attributes (most common in the DCE literature), which we demonstrate can mask important differences across mechanisms. Further, each project appears twice within the Seq-TC design. This allows us to test whether the value for a project depends on the other project included in the same choice set. Last, we elicit beliefs to assess the degree to which people adhere to the game form presented to them, and to gauge the extent to which Seq-TC participants may have engaged in non-truthful, strategic responses (given incentive compatibility for this mechanism depends on respondent beliefs over the expected choices of others).

Our results support the convergent validity of the Single-BC and Seq-BC mechanisms. In contrast, Seq-TC WTP estimates are generally lower than those derived from the two binary choice mechanisms. Our exploration of the underlying drivers of this result reveals that the Seq-TC mechanism induces strong framing effects, specifically that valuations for one option in the choice question can depend critically on the relative characteristics of the other included option. Further, a larger share of participants from Seq-TC compared to Seq-BC always select the status quo option. This finding is at odds with standard theory and is most likely a result of the added complexity of trinary choice questions. Evidence indicates that the above results are not an artifact of strategic Seq-TC responses; in fact, the proportion of serial status quo participants, the size of the framing effects, and WTP differences across mechanisms all increase when we drop potentially strategic participants. Last, contrary to expectations, the precision of Seq-TC welfare estimates is worse than for Seq-BC. Logically, the behavioral patterns documented above induced enough "noise" to offset the potential efficiency advantage of the Seq-TC mechanism.

2. Research methods

2.1. Overview

Experiment participants are residents of New York State and complete an online survey that is focused on agriculture and farmworkers in their state. The experiment is designed to elicit valuations, through voting questions, for a variety of projects intended to improve the well-being of farmworkers and their families. Participants are anonymously placed into groups of 50 people and vote on whether everyone in the group should pay money to fund a farmworker assistance project. Many farmworkers are immigrants that do not speak English well, are not acclimated nor prepared to live in cold climates, and do not have access to an automobile. The projects involve combinations of educational materials, winter clothing, and transportation services, and were designed in consultation with the Cornell Farmworker Program (CFP). The CFP has a developed infrastructure for identifying needy farmworkers and their families, and delivering aid to them. Partnering with the CFP thus allowed us to credibly define projects that could be implemented.

We deploy Single-BC, Seq-BC, and Seq-TC mechanisms to elicit values for assistance projects, and conduct convergent validity tests. We use the Single-BC mechanism to elicit values for two projects, which further allows us to test construct validity (scope tests) for all three mechanisms. Consequently, the experiment includes four between-subjects treatments.

2.2. Value elicitation mechanisms

A value elicitation mechanism can be described by a set of possible outcomes (i.e., goods provided along with the cost to the agent), the question(s) used to elicit preferences for the possible outcomes, and how answers to these questions influence what outcomes arise (i.e., allocation rules). Much attention has been given to differences in value elicitation question formats, as is the case in this study. However, we emphasize that when comparing question "formats", it must be recognized that other characteristics of the elicitation mechanism are important, and thus tests of elicitation effects are conditional on them. Taking the question format as given, we employ mechanisms that are incentive compatible under plausible assumptions. This helps to rule out non-truthful responses as a significant driver of elicitation effects. Our design choices thus arguably favor the finding of convergent validity.

SP valuation surveys that utilize a Single-BC question typically describe the characteristics of the good being valued in words, and the Single-BC question itself is often framed in a way that mimics an advisory public referendum. A typical, sequential DCE instead presents respondents with choice sets, where policies are presented in tabular form and described in terms of attributes and attribute levels. It is further recognized that deploying incentive compatible mechanisms in SP surveys, which in theory incentivize truthful responses, is challenging, and formats that involve a sequence of choice questions or include more than two choice options per question admit additional opportunities for non-truthful responses. The conditions for a Single-BC or a Seq-BC format to be a part of an incentive compatible mechanism in an SP survey have been established (Carson and Groves 2007; Vossler et al. 2012; Carson, Groves, and List 2014), but incentive compatibility conditions for a sequential DCE involving questions with three or more choice options have not.

All mechanisms we deploy share the following features, which helps control for other procedural or contextual variations across mechanisms: (1) choice sets include a status quo option and either one or two alternative options that define projects and their cost to the individual, and the information is presented in a tabular form; (2) choice questions that ask the respondent to select their preferred choice option are framed as referenda; (3) the cost of a project varies across participants (randomly drawn from a predefined set of values), but as in the case of SP surveys, this is not made explicit to the participants; (4) actual provision of a project is determined by a plurality vote rule; and (5) participants are informed about possible levels of the non-monetary attributes of projects prior to voting. We employ direct financial incentives and use provision rules to promote truthful preference revelation (to the extent possible). Additional details on the mechanisms are provided below.

2.2.1. Single binary choice (Single-BC)

The Single-BC mechanism involves a vote on whether to fund a project at a specific cost to the individual. The value elicitation question is presented as a selection between the project and a no project option (i.e., the status quo). The cost of the project varies across voters to permit the identification of WTP. After all participants in the voter group have submitted their vote, the votes are counted. If the majority are in favor of funding the project, the individual cost is subtracted from a participant's endowment, this money is given to the CFP, and the project is carried out. Otherwise, no money is subtracted from the endowment, and no project is implemented. A binding referendum with a majority-vote implementation rule is well known to be incentive compatible under minimal assumptions (Farquharson 1969).

2.2.2. Sequence of binary choices (Seq-BC)

The Seq-BC mechanism extends the Single-BC elicitation by presenting participants with nine separate referenda, with each involving a vote between a project and a no project option. Participants know they will be asked to vote on nine referenda, but do not know the specific projects (i.e., combinations of the attribute levels) involved in advance. To break possible strategic links between the nine votes, one referendum is randomly selected to be binding after decisions by all group members have been submitted. To promote beliefs that are consistent with incentive compatibility assumptions, each referendum is separated by a screen that encourages people to vote on each referendum without consideration of the other referenda, and to vote on each referendum as if it will be the one randomly selected.[7] The Seq-BC mechanism we deploy represents a sequence of games that, when analyzed individually, are incentive compatible. Azrieli, Chambers, and Healy (2018) prove that a mechanism that randomly selects one such game to be binding is also incentive compatible under a mild statewise monotonicity assumption.

2.2.3. Sequence of trinary choices (Seq-TC)

The Seq-TC mechanism is identical to Seq-BC with the exception that each of the nine choice sets includes two possible projects (and their costs) along with the no project option. A single choice mechanism with greater than two options and a plurality implementation rule is prone to strategic voting (McKelvey and Ordeshook 1972). Thus, although we randomly select one of the nine choice questions to be binding, this is insufficient to ensure an incentive compatible mechanism. However, unlike in a public election, there is no public discourse from which experiment participants can form strong beliefs over the likely preferences of other voters regarding the farmworker assistant projects. It is therefore plausible that participants have diffuse or uniform priors, in which case the mechanism is incentive compatible if we make this assumption along with the statewise monotonicity assumption of Azrieli, Chambers, and Healy (2018).

2.3. Choice experiment design

The survey includes valuation scenarios designed to estimate the WTP for projects providing aid to needy farmworkers and their families. Each project is defined by the amounts of the following goods provided (attribute levels in parentheses): (i) sets of education materials,

[7] This element of the design is motivated by the "independence script" of Vossler et al. (2025), which was designed to promote choice question independence in SP surveys.

where each set includes a set of English language workbooks for a farmworker and their tutor, and a set of age-appropriate children's books (1, 3, or 5); (ii) sets of winter clothing, where each set includes a pair of gloves, a winter coat, and a heavy blanket (1, 3, or 5); and (iii) transportation services, where a service is defined as a round-trip transportation to a legal service, a medical appointment, or another important engagement (1, 2, or 3). Also associated with a project is the cost to the respondent ($5, $8, $11, or $15).

Based on the three non-cost attributes, each with three levels, we defined nine projects that, in the Seq-BC context, result in a perfectly balanced and orthogonal main effects design. The nine projects are defined in Table 1. The projects can be roughly delineated by size, with three projects each characterized as "small" (S), "medium" (M), and "large" (L). We use the Single-BC mechanism to elicit valuations for one of the small projects (S3) and one of the large projects (L3). The size or "scope" of project L3 is unambiguously higher than S3 as it adds two more units to each of the three project attributes. Taking the same nine projects as given, we used a D-efficient algorithm to determine nine three-option choice sets for the Seq-TC elicitation. The Seq-TC experimental design is perfectly balanced, and each project appears twice in the design. For choice sets where one project dominates the other in terms of attribute levels, we constrained the cost of the larger project to be strictly higher to avoid asymmetrically dominated options.

We further refined the design to allow for comparisons across mechanisms that are uncontaminated by order effects. Recall that the Single-BC mechanism is used to value projects S3 and L3. We randomly assigned Seq-BC participants to first vote on S3 or L3. One of the Seq-TC choice sets includes both the S3 and L3 projects, and half of the Seq-TC participants are randomly assigned to this choice set. Other than these deliberate manipulations, the order of choice questions is randomized.

2.4. Survey development and power analysis

Four pretests were conducted in the Fall of 2022, using 115 New York State adult residents who were part of a survey panel administered by the Schlesinger Group.[8] Of potential interest, it was clear from quiz questions that many found unintuitive the idea of randomly selecting one of many votes to determine the outcome. This motivated important changes to the description of the voting procedures, including the use of repetition, the inclusion of a graphic that urged the participant to pay close attention to the instructions, and a stated justification for why only one of nine votes are ultimately binding. Pretests were used to refine the attributes and attribute levels, as well as to select which the two projects to be valued with the Single-BC mechanism.

The Seq-BC and Seq-TC surveys were piloted in December 2022 and January 2023 with a total of 100 participants registered with the Amazon Mechanical Turk (MTurk) and Prolific platforms.[9] Based on a power analysis that utilized the pilot data, we settled on sample sizes of 300 for each of the four treatments. For tests of vote proportions between the two Single-BC treatments, or when comparing Single-BC and Seq-BC for a specific project, the design can detect a treatment effect as small as 11 percentage points with 80 percent power. For a test of WTP between the two Single-BC projects, or when comparing Single-BC and Seq-BC for a specific project, the minimum detectable treatment effects range from \$1.3 to \$1.7 based on a two-sided test and a 5 percent significance level.[10] The estimated standard deviations of WTP from the Seq-

[8] Our original intent was to use Schlesinger Group for the experiment. This plan was dismissed after learning participants were paid in “points” and that there were significant delays between participation and payment. We conjecture that both factors reduce the importance of financial incentives on voting decisions.

[9] While our intent at the time was to use both platforms for the experiment, we decided against using MTurk for several reasons, including: (1) MTurk was unwilling to provide the number of active MTurk participants living in New York State; (2) although we restricted eligibility to New York State residents, many respondents indicated living elsewhere; and (3) some evidence of duplicate, fraudulent, and unreliable responses.

[10] These figures are based on Monte Carlo simulations that assumed the WTP for both groups followed normal distributions. The analysis considered a range of possible WTP standard deviation parameters, from 3.88 to 5.30, which were based on estimates derived from various interval regression models applied to the Seq-BC data. Power also depends on the relationship between the underlying WTP distributions and the range of cost levels in the value

TC pilot data are smaller than for Single-BC, indicating that WTP comparisons involving Seq-TC should be powered to detect even smaller differences.

2.4.1. Survey instrument

We coded the survey using the Qualtrics survey design platform and designed it to be completed by an online panel. The survey begins with a consent form and asks for the participant's state of residence. Only those indicating they live in New York are eligible to continue. The first section of the survey provides some information on agriculture and farmworkers in New York and includes a handful of questions to get participants thinking about these issues. The second section introduces the CFP, and describes the attributes associated with possible farmworker assistance projects, how a project would be funded, and how the CFP would implement the project. Participants are informed about possible levels of the attributes, but do not know the specific project(s) they will be asked to vote on in advance. The third section describes voting procedures, emphasizes that participants are in voting groups of 50 (people completing the survey at roughly the same time) and that real money and projects are at stake, and includes two to three quiz questions.[11] If a quiz question is answered incorrectly, additional explanations are provided. The quiz is followed by the DCE. The fourth section asks additional questions to help us better understand participants' voting choices and assess whether their beliefs are consistent with incentive compatibility assumptions. The final section collects socio-demographic information. A representative survey is included in the appendix.

elicitation choice sets. As such, in the analysis, we considered a range of true mean WTP values from \$7 to \$11 for the group with the relatively smaller WTP in the comparison.

[11] One of the three quiz question concerns the sequential value elicitation mechanism and, thus, is excluded from the Single-BC treatments.

2.5. Participants, compensation, and resolution of voting procedures

Twelve hundred participants registered on the Prolific research platform completed the experiment during the spring and summer of 2023. The median completion time was 14 minutes. The vast majority (74 percent) reported completing the survey on a desktop or laptop computer. Participants were randomized into treatment and recruited in batches to avoid long delays between participation and payment, which depended on the outcome of the group vote. Participants were guaranteed $3 for completing the survey. Participants were given an additional "starting balance" of $15 and were encouraged to treat this as they would money already in their pockets. In the case the participant's voting group did not fund a project, they were paid $15 (in addition to the guaranteed $3). Otherwise, the cost to the individual for funding the project was subtracted from the starting balance and they received the remainder (if any). Eighteen of the 24 group votes passed, and as a result over $19,000 was contributed to the CFP.

Prior to data collection, we randomly drew a sequence of random numbers, which was used to select which choice set was binding for each voter group that faced the Seq-BC or Seq-TC mechanism, and to resolve situations where two options received the same number of votes. In all cases, participants were informed of the results of the vote through a private message on Prolific. Those who helped to fund a project received confirmation from the CFP.

3. Results

3.1. Sample characteristics and beliefs

Table 2 presents socio-demographic characteristics by treatment. As expected, summary statistics are very similar across the randomized treatment groups. Using the Kruskal-Wallis test for continuous measures and Fisher's exact test for indicator variables, and a 5 percent significance level, we fail to reject the null hypothesis that the measures are statistically different across treatments. The percentage stating their race/ethnicity is White is different at the 10 percent level.

Given our use of the Prolific research platform, we make no claims that the sample is representative of the New York State population. To nevertheless provide a point of reference, relative to the New York State adult population, participants are much younger (mean age of 35 versus 49), considerably more likely to have a bachelor's degree (66 versus 39 percent), somewhat more likely to be White (61 versus 54 percent), and less likely to be married (28 versus 44 percent).

Potentially important for our application, most of the sample have lived in New York for an extensive period (sixteen or more years) and few participants describe themselves as living in a rural area (10 percent). In addition, only five participants indicated working in the "agriculture, forestry, fishing or hunting" industry. These statistics indicate that, for most respondents, the provision of farmworker assistance projects would (if anything) generate nonuse values; this is typical of many SP survey applications.

It is essential for the interpretation of results that respondents hold beliefs that are consistent with incentive compatibility assumptions. As indicators of beliefs, we asked questions before and after the voting procedures. Prior to the vote, we asked participants in all treatments a quiz question to gauge whether they understood that, if a project received the most votes, they would have to pay the stated cost, and the project would be implemented. Ninety-six percent provided the correct answer. For those in the sequential mechanism treatments, we asked an additional quiz question to assess whether they understood the procedure for determining the outcome (i.e., that we would randomly select a referendum, then apply a plurality vote rule). Eighty-nine percent answered correctly. For either question, an incorrect answer triggered additional explanations about procedures.

Following the votes, we asked respondents to self-assess whether they understood voting procedures. Summary statistics are provided in Table 3. Eighty-four percent selected "Everything was clear" and thirteen percent selected "I understood most of them". We further asked Likert-scale (1 "Agree", 2 "Neither agree nor disagree", 3 "Disagree") questions to gauge perceived payment consequences (i.e., whether they thought they would have to pay if a project received the most votes), response consequences (i.e., whether they thought the votes would be used to determine the outcome), and whether the project (if it received the most votes) would be implemented as described. The percentage selecting "Agree" to these questions are 93, 90, and 92 percent, respectively. Eighty-three percent of participants agreed with all three questions. Overall, based on both the quiz and belief questions, there is suggestive evidence that most participants held the desired/correct beliefs about the mechanism.[12]

Table 3 provides additional information from our voting follow-up questions. When asked about the influence of the DCE attributes on their choice(s), participants in all treatments indicated that winter clothing had the largest effect on voting choices. Over 90 percent of respondents answered that they were "certain" or "somewhat certain" about their voting choices.

3.2. Convergent validity tests

Table 4 presents the percentage voting in favor of projects S3 and L3 based on the Single-BC and Seq-BC mechanisms. The percentage voting in favor decreases as the project cost increases, as expected. Using Fisher's exact tests, the proportion in favor is statistically different across cost amounts in all four cases ($p < 0.01$). There is further sensitivity to scope: For both

[12] For the Seq-TC participants, we asked a follow-up question to gauge potential non-truthful, strategic voting. While we will present additional details later, responses from the overwhelming majority (91 percent) indicate that none or few of their votes were strategic.

mechanisms, the overall percentage voting in favor of the large project is 13 percentage points higher than the percentage voting for the small project. These differences are statistically significant ($p < 0.01$).[13] As can be seen from the table, vote percentages for the same project track closely between the two mechanisms. Across mechanisms, the overall percentage in favor is not statistically different for either the small project ($p = 0.74$) or the large project ($p = 0.61$), and the vote percentages differ by only two percentage points.

To estimate WTP, we deploy multinomial choice models. Let $s = 0,1, \dots ,9$ denote the projects in the experimental design, with $s = 0$ associated with the status quo or no project option. Then, assume that the utility of respondent i for option j in choice question t is given by

$$(1) \qquad V_{ijt} = \sum_{s=1}^{9} \beta_s \, 1[Project_{ijt} = s] + \alpha c_{ijt} + u_{ijt},$$

where the β_s are project-specific intercepts, c_{ijt} is the cost of the option, and u_{ijt} is an error term that is independently and identically distributed Type I extreme value.

Including project fixed effects allows identification of project-specific WTP without relying on functional form assumptions. This specification mirrors the site fixed effects utility model commonly used and recommended in the recreation demand literature (Lupi, Phaneuf, and von Haefen 2020). While it is standard in the SP literature to specify utility as a linear function of an alternative specific constant (ASC) and attribute levels, as we will demonstrate later, imposing this functional form induces averaging that masks important treatment effects.

Throughout, we estimate models separately by treatment. For the Seq-BC and Seq-TC cases, we estimate mixed logits where the project-specific intercepts are random parameters with normal distributions, and for identification purposes the coefficient on the cost variable is fixed.

[13] This result is based on Fisher's exact test in the Single-BC treatments. To account for correlated responses (i.e., Seq-BC participants vote on both projects), we use the McNemar test to compare Seq-BC vote proportions.

For the Single-BC treatments, we estimate conditional logit models, and assume all parameters are fixed; random parameters based on Single-BC data are not theoretically identified (see Revelt and Train 1998).

Table 5 reports WTP estimates for all projects valued with the three mechanisms, and Table A1 in the appendix reports coefficients and summary statistics associated with the underlying econometric models. For the Single-BC mechanism, WTP for the small and large project is \$9.78 and \$13.45, respectively. The same figures for the Seq-BC are, respectively, \$10.27 and \$13.88. The WTP point estimates are therefore very close across the two mechanisms and are not statistically different for either comparison. Consistent with the vote percentages, there is evidence of scope in WTP for both the Single-BC (diff = \$3.67; $p < 0.01$) and Seq-BC (diff = \$3.61; $p < 0.01$). The overall evidence supports convergent validity between the Single-BC and Seq-BC mechanisms.

Result 1. Comparisons of response distributions and willingness-to-pay estimates provide strong evidence of convergent validity between the single binary choice and sequential binary choice mechanisms.

Table 5 reveals important differences in WTP estimates when comparing either the Single-BC or the Seq-BC mechanisms with the Seq-TC mechanism. Seq-TC WTP estimates are lower than those generated by the other two mechanisms for all projects. The starkest differences arise for the small and medium-sized projects. When comparing the Seq-BC and Seq-TC, differences for small and medium projects range from \$2.50 (project M2) to \$6.94 (S1), and in each case the difference is statistically significant ($p < 0.01$). Differences for the large projects tend to be less pronounced, ranging from 26 cents (L2) to $1.54 (L1), and are not statistically significant. When comparing Single-BC and Seq-TC, the WTP estimates for the large project L3 are similar and not statistically different; in contrast, the Seq-TC estimate for the small project S3 is 60 percent lower than the corresponding Single-BC estimate ($p < 0.01$). Many field survey studies that compare single versus sequential discrete choice formats have relied on comparing WTP for a single policy. Had we only used the Single-BC mechanism to elicit values for L3, we would have reached a different conclusion about convergent validity.

Result 2. We reject convergent validity of willingness-to-pay estimates between the sequential trinary choice mechanism and both the single and sequential binary choice mechanisms. For most projects, the sequential trinary choice mechanism results in lower willingness-to-pay estimates.

3.3. Exploring possible drivers of elicitation effects

Why are WTP estimates from the Seq-TC mechanism generally lower than those obtained from the two binary choice mechanisms? Grounded in findings and presumptions from the SP literature, we consider four possible drivers of the observed elicitation effects: (1) framing effects; (2) task complexity; (3) order effects; and (4) strategic voting. The similarities between the Single-BC and Seq-BC mechanisms suggest that it is the addition of a choice option, rather than the inclusion of multiple valuation scenarios in the same survey, that is driving the elicitation effects for this application.

3.3.1. Framing effects

Adding another project to a choice set admits the possibility that preferences for one project become dependent on the other included project. In our experimental design, each project appears as an option twice during the sequence of trinary choice questions, and in each case is paired with a different project. This allows for a careful examination of framing effects − that is, whether the valuation of one project depends on the project with which it is paired in a choice set.

Table 6 presents choice-set dependent WTP estimates for each of the nine projects. These estimates are derived from a mixed logit model, presented in Table A2 of the appendix, that extends the prior model by allowing the WTP estimate for a given project to vary across the two choice sets wherein it appears. Of interest, across the two choice sets, a project is paired with one project from each alternative size group. For example, each small project appears once with a medium project and once with a large project. While this pairing of projects looks intentional, this is largely a result of using a D-efficient algorithm, which increases efficiency by (in part) increasing the variation in the attribute levels across options within a choice set.

The table conveys significant evidence of framing effects. We interpret this as a sign that people are forming values that depend on the relative differences of the choice options in the same choice set. While there are some exceptions, the general pattern in the WTP point estimates is that the elicited value for a project increases as the size of the paired project decreases. Said another way, adding a project to the choice set that is smaller than the other included project appears to make the relatively larger project more desirable.[14]

Medium projects are unique as they hold a relative size advantage in one choice set (when paired with a small project) and a relative size disadvantage in another choice set (when paired with a large project). In contrast, a small project always has a relative size disadvantage, and a large project always has a relative size advantage. As such, we expect framing effects to be most

---

[14] We note that this is distinct from a "decoy" effect (e.g., Bateman, Munro, and Poe 2008; Park and Kim 2005), as we assigned cost amounts (when necessary) in ways that avoid choice sets with asymmetrically dominated choice options.

pronounced for medium projects, which is consistent with the results. When a medium project is paired with a small project, WTP estimates are 33, 130, and 79 percent higher relative to when the same project is instead paired with a large project.

There is further a substantial difference in WTP for project S3, where the value of the project nearly doubles when paired with a medium instead of a large project, and this effect is statistically significant ($p < 0.01$). Again, the difference may arise due to a change of the relative size. Based on participants' statements, winter clothing was an attribute with the largest influence in the decision-making process (see Table 3). Project S3, which offers three units of winter clothing, is once paired with M3, which offers one unit of winter clothing, and again with L3, which offers five units of winter clothing. Thus, the differences in the levels of this most influential DCE attribute could be driving the observed framing effect for S3. We note that both WTP estimates for project S3 remain statistically different from those obtained with the Single-BC and Seq-BC mechanisms ($p < 0.01$). WTP point estimates for the other two small projects are similar across choice sets.

For the large projects, there is a weakly significant ($p = 0.052$) framing effect for L2. This effect works in the opposite direction from the other framing effects, although the difference is relatively small (16 percent difference). While this possible interpretation is purely speculative, participants may have effectively viewed both M2 and S2 as "small" goods relative to L2. The largest difference in the project attribute levels between L2 and S2 is two units. However, M2 only has one unit for the education attribute as opposed to five units for L2, which is the largest difference possible in the design. WTP point estimates for the other two large projects are similar across choice sets.

3.3.2. Task complexity

That a Seq-TC mechanism may introduce additional task complexity is intuitive, as one is asked to assess tradeoffs among three versus two possible options. In the face of additional complexity, one possible reaction is to select the status quo option ("no project"). The standard rational choice model predicts that adding another option to the choice set should make the decision-maker weakly better off; for some, the additional option will be preferred and increase utility. Therefore, we should expect the frequency of status quo choices to be strictly lower for Seq-TC relative to Seq-BC.

Table 7 presents the frequency of status quo choices separately for each Seq-TC choice question. Further displayed are the percentage of status quo choices that correspond with the two Seq-BC choice questions that include one of the projects contained in the comparison Seq-TC choice question. Interesting patterns emerge. In cases where the Seq-BC choice question includes a small or a medium project, the percentage of status quo choices is statistically higher in the Seq-BC relative to the Seq-TC. This is consistent with theoretical expectations and empirical findings observed in prior studies (e.g., Adamowicz et al. 2011; Rolfe and Windle 2012; Oehlmann et al. 2017). The exception to the above patterns is M1, which is the highest valued medium-size project based on Seq-BC, and for which there is a similar proportion of status quo choices for the Seq-BC question relative to when the same project is included in a Seq-TC choice set.

For the case where the Seq-BC choice question includes a large project, however, the status quo choice frequencies are very similar to those for Seq-TC, and are not statistically different. For large projects, it is as if including a smaller project (even at a lower cost) is virtually meaningless. Such a project is rarely selected, as indicated by the very low WTP estimates for small projects (when paired with a large project) reported in Table 6. At least based on Single-BC and Seq-BC, participants hold significant values for both small and large projects. Therefore, this evidence is suggestive that complexity motivates some to select the status quo as the number of choice options increases.

We further calculate the proportion of participants who select the “no project” option for all nine choice questions (i.e., serial status quo choices). For the Seq-BC mechanism, this percentage is 14.7. For the Seq-TC mechanism, the percentage increases to 21.3, which is in the *opposite* direction of theory. The difference is statistically significant ($p = 0.04$).

The leading explanations for deviations from rational choice theory that manifest as serial status quo choices are that respondents have lexicographical preferences, refuse to make tradeoffs between attributes, protest the valuation scenarios, or employ heuristics to simplify complex decision tasks (e.g., von Haefen et al. 2005; Meyerhoff, Mørkbak, and Olsen 2014; Frey and Pirscher 2019). The behavioral implications of lexicographical preferences are the same for the two mechanisms. Those who would generally refuse to tradeoff attributes in a DCE should behave the same regardless of whether facing a Seq-BC or Seq-TC mechanism. Of course, it is possible that more Seq-TC respondents refuse to make tradeoffs as the mechanism is more complex.

In our experiment, the payment vehicle, the projects, and the organization that would implement the projects are identical across the two mechanisms. As such, we would expect the frequency of protests to be similar. Protest behavior is often correlated with demographic characteristics, and these characteristics are further similar across the treatments (see Table 2). Consequently, we argue that the most plausible explanation for the larger share of serial status quo choices in our Seq-TC treatment is increased task complexity.

The estimated distributions of random parameters may provide additional evidence of differences in task complexity across mechanisms. One consequence of task complexity in preference elicitation is response error, and response errors increase variability in stated preferences (e.g., Regier et al. 2014). In a mixed logit model, this variability can be captured through the estimated standard deviations of the project fixed effects. To compare across mechanisms, we compute the coefficient of variation, which is a standardized measure defined as a ratio of the standard deviation to the mean of the distribution. For each project, the coefficient of variation is higher for Seq-TC relative to Seq-BC; on average, the coefficient of variation is 2.44 times higher for Seq-TC, and ranges between 1.39 to 6.58 across projects.

3.3.3. Order effects

The two sequential mechanisms can give rise to order effects. Various possible causes for order effects in DCE studies have been proposed in the literature, including strategic misrepresentation of preferences (driven by earlier displayed choice sets), fatigue, institutional learning, preference learning (changes in preferences), anchoring, and referencing (Day et al. 2012). The inclusion of an additional choice option could lead to differential order effects and explain differences across elicitation mechanisms.

Our experimental design allows for convergent validity testing for a subset of projects, S3 and L3, while minimizing the influence of order effects. All Seq-BC participants voted on both projects, and they were randomly assigned to vote on either project S3 or L3 in the first referendum they faced. All Seq-TC participants face a choice set that includes projects S3, L3, and the status quo, and half voted on this choice set first. Targeting the analysis to these choice sets further allows comparisons with the Single-BC treatments.

Table 8 presents WTP estimates based only on data from the choice sets described above. We estimate mixed logit models with project fixed effects and allow the fixed effects to vary according to whether the project is included in the first referendum faced by the respondent (reported in Table A3). Convergent validity tests continue to tell a similar story when the data is restricted to the first referendum faced: Seq-BC and Single-BC produce statistically equal WTP estimates for both projects, whereas there are significant differences between either of these mechanisms and Seq-TC in the case of project S3. Moreover, there is now weak evidence (p = 0.098) that Seq-BC and Seq-TC deviate for project L3 as well. Also of potential interest is that the estimated Seq-BC scope effect (S3 v. L3) based on data only from the first referendum ($3.32) is very similar to that not based on the first referendum data ($3.84). In contrast, the Seq-TC (S3 v. L3) scope effect increases by 45 percent (from $6.52 to $9.45) across the two order-based subsamples, and this difference is weakly statistically significant (p = 0.08).

We further estimate mechanism-specific mixed logit models that allow all the project-specific WTP estimates to vary linearly with choice set order (i.e., we interact with the project fixed effects an order variable coded 1 through 9). These models are reported in Table A5. Most model-specific order effects are imprecisely estimated (two are significant at the 10 percent level), and two-thirds of the order effects are negative. The lone statistical difference in order effects across mechanisms is for L2, for which the marginal WTP difference associated with order is 75 cents (p = 0.04). However, order-specific WTP estimates for L2 remain statistically equal across mechanisms.

3.3.4. Strategic voting

While we devised mechanisms around the three question formats in ways to encourage truthful responses, the Seq-TC admits the possibility that a person selects her second preferred option if she perceives her first-best choice has little to no chance of winning. After participants voted, the survey included a screen that described the well-known phenomenon of strategic voting in political candidate elections. Then, we asked then directly asked the participant how often they "selected [their] most preferred/favorite option" in the experiment. The response categories in the strategic voting question, and the percentages selecting them are as follows: "In all cases", 72 percent; "In most cases", 20 percent; "In some cases", 6 percent; and "Rarely or never", 2 percent. Therefore, the incidence of strategic voting appears low. This finding is not entirely surprising given that, in theory, the Seq-TC mechanism is incentive compatible if participants have diffuse priors over the preferences of other voters. As the welfare of New York farmworkers has not been a heavily politicized issue, we suspect that respondents had considerable uncertainty over the expected votes of others.

The percentage always selecting the status quo is considerably lower for the subsample that indicated any level of strategic voting (10.6 percent), relative to those who stated otherwise (25.6 percent), and this difference is statistically significant ($p < 0.01$). This demonstrates that the strategic voting question contains useful information as we should expect strategic voters to employ more sophisticated voting strategies.

Taking responses to the strategic voting question at face value, we re-estimated the mixed logit models while restricting the sample to those who did not engage in any strategic voting (i.e., we dropped the 85 participants who selected a response other than "In all cases"). The mixed logit models, and WTP estimates based on these models, are reported in Tables A6, A7, and A8. The subsample of possibly "non-strategic" voters are less likely to select the relatively smaller project included in choice sets, which in turn leads to more dramatic framing effects. For the three medium projects, WTP point estimates are 51, 161, and 152 percent higher when paired with a small versus a large project. Also of interest is that WTP estimates for projects S1 and S2 are near zero (and sometimes negative) and statistically insignificant. In a similar vein, WTP estimates are generally lower for medium and large projects relative to the full treatment sample, which is likely due to the higher proportion of serial status quo choosers in the subsample. While we must be careful not to interpret these findings causally, this evidence suggests that differences between Seq-TC and the other mechanisms are not driven by strategic voting—as documented here, the observed differences are even more pronounced for those least likely to have engaged in strategic voting.

Result 3. Willingness-to-pay estimates based on the sequential trinary choice mechanism are sensitive to the attribute levels of the other included option in the choice set. A higher percentage of sequential trinary participants always select the status quo relative to sequential binary choice participants, which is best explained by the fact the former mechanism increases task complexity. There is little evidence to suggest that order effects or strategic voting explain observed differences in willingness-to-pay estimates between the sequential trinary choice mechanism and the two binary choice mechanisms.

3.4. Additional analysis

3.4.1. Attribute-based utility specification

While the above analysis incorporates flexible utility specifications based on project fixed effects, it is more common for researchers to define the indirect utility function for sequential mechanisms in terms of attribute levels. It is possible that the WTP estimates from such models, which introduce considerable "averaging", may alter comparisons between the three mechanisms. Table 9 presents WTP estimates for the nine projects, separately for the Seq-BC and Seq-TC mechanisms, derived from treatment-specific mixed logit models that include an alternative-specific constant (i.e., an indicator that equals 1 for a non-status quo option), the levels of the three non-cost attributes defined as continuous variables, and a continuous cost variable. The full model results are provided in Table A9 in the appendix.

For the Seq-BC mechanism, restricting the utility specification to the attribute-based form has a minor effect on WTP, and differences in point estimates relative to the prior, flexible specification (Table 5) vary from -10.7 to 11.5 percent. The range of WTP estimates across projects decreases slightly: \$8.63 to \$14.04 (Table 9) versus \$7.74 to \$13.88 (Table 5). The effects of restricting the utility specification are much more pronounced with the Seq-TC mechanism; in particular, differences due to size are less dramatic and WTP point estimates increase for all projects. Further, WTP estimates are generally less precise. As a result, there are no longer any significant project-specific WTP differences across the two mechanisms, although the estimated utility models (and WTP functions) are statistically different ($p < 0.01$).

The attribute-based utility specification further highlights that estimated marginal WTP distributions are dramatically different across the mechanisms. For Seq-BC, marginal WTP is 57 cents, \$1.02, and 49 cents, respectively, for a one-unit increase in the education materials, winter clothing and transportation attributes. The marginal WTP estimates from the Seq-TC are much higher, at \$1.19, \$1.48, and \$1.17. Given the earlier evidence that Seq-TC choices are sensitive to the relative size of the two included projects, this logically leads to larger marginal WTP estimates.

3.4.2. Statistical precision

The statistical rationale for increasing the number of choice options is that, holding the number of respondents (and preferences) constant, this increases the precision at which one can estimate the unknown population parameters. For instance, the rule of thumb proposed by Johnson and Orme (2003) is that the sample size required for identifying the main effects in a sequential

DCE decreases proportionally with the number of choice options (i.e., that the Seq-TC sample should be 2/3 of the Seq-BC sample).

Using as a measure of precision the normalized standard error of WTP, i.e., the ratio of the standard error to the WTP estimate, and the project-specific utility specification (which coincides with Table 5), it is very clear that moving from a single binary choice to a sequence of binary choice questions represents a very cost-effective approach. Indeed, precision is very similar across the two mechanisms and, of course, from each Seq-BC respondent, we elicited values for nine projects instead of just one. Moving again from a binary to a trinary sequence of questions paints a different picture. WTP estimates derived from the Seq-BC data are generally more precise: The normalized standard error for eight of nine projects is higher for Seq-TC relative to Seq-BC, and moreover the mean error is 4.6 times larger. Perhaps the fairest comparison is for the large projects, given the similarity in WTP estimates across the two mechanisms. The mean normalized standard errors for Seq-TC are 20 percent higher on average for these three projects. Using instead the attribute-based utility specification (Table 9), even though Seq-BC and Seq-TC estimates are similar for all projects, the normalized standard errors for Seq-TC are 2.3 times larger, on average, relative to Seq-BC.

To provide some perspective on the above results, we conducted Monte Carlo simulations to assess what differences in precision we would expect if preferences were not a function of the elicitation mechanism. We use as the data generating process the project-specific utility model applied to the BC-Seq data as estimated using conditional logit. In each replication: (1) the utilities for the nine projects are drawn for a sample of 300; (2) each simulated individual faces all choice sets (separately for Seq-BC and Seq-TC), with randomly selected costs consistent with the experimental design; (3) the simulated individuals selects the choice option that yields the highest utility; and (4) mechanism-specific conditional logit models with cluster-robust standard errors are estimated for the sample.

Simulation results based on 10,000 replications are presented in Table 10. As expected, the mean WTP estimates associated with both mechanisms are virtually identical to the assumed true values. For each project, the normalized standard error based on Seq-TC mechanism is less than for the Seq-BC, indicating a gain in precision. The ratio of normalized standard errors indicates Seq-BC standard errors that are 9.8 to 28.6 percent larger than Seq-TC. On average, the Seq-BC standard errors are 20.7 percent larger. This evidence suggests that, as expected, the experimental design favored Seq-TC in terms of precision. Nevertheless, the fact remains that we find the opposite to be true.

Result 4. Willingness-to-pay estimates obtained from a sequence of trinary choices are less precise than those from a sequence of binary choices, which contrasts with expectations.

4. Discussion

With a carefully designed revealed preference field experiment grounded in mechanism design theory, we provide novel evidence on the performance of three question formats commonly used in stated preference (SP) surveys to elicit monetary valuations for nonmarket goods. The findings are univocal: the willingness-to-pay (WTP) estimates based on a single binary choice (Single-BC) and a sequence of binary choices (Seq-BC) converge, while the WTP estimates based on a sequence of trinary choices (Seq-TC) are systematically different.

The fact that the mechanism based on Seq-TC generates WTP estimates that are distinct from the other two approaches should not be taken as evidence that this format cannot recover reasonable approximations of the actual demand for nonmarket goods. Indeed, we constructed mechanisms based on the three question formats that should all (with caveats) induce truthful preference revelation as a weakly dominant strategy, and few Seq-TC respondents stated that they engaged in strategic voting. The drivers of the difference nevertheless have important implications.

First, we find that 50 percent more Seq-TC participants, relative to Seq-BC respondents, always selected the status quo option, which we attribute to the higher complexity of the Seq-TC mechanism. Adding an option to a choice set, such as when moving from a Seq-BC to a Seq-TC mechanism, should instead motivate fewer status quo choices (i.e., the added option could be better than the other two), and in our opinion, this clear violation of consumer theory is worrisome. This is particularly true when one considers our valuation scenarios. We asked people to vote on projects defined by their cost and just three non-price attributes. The attributes themselves are familiar goods (e.g., winter clothing items), and attribute levels are simple quantities. In contrast, most DCEs in the environmental and health literature ask people to value more complex and unfamiliar policies, such as those related to conserving various ecosystem services or risk reductions from a decrease in pollution.

Second, we have uncovered strong Seq-TC framing effects: Demand for one alternative to the status quo (i.e., a project) depends on the characteristics of the other alternative present in the same choice set. This finding is not entirely surprising and reflects the realities of everyday decisions. For instance, many of us are willing to pay a lot of money for a drink in a fancy restaurant when we would not fathom paying the same amount for the same drink in a causal dining establishment. And, when the application pertains to recreation choices, transportation modes, or consumer goods, there are logical arguments in favor of placing participants within the relevant decision context (e.g., selecting among a set of substitute goods). However, when SP surveys are used to inform benefit-cost analysis, the presence of framing effects makes it challenging to interpret welfare estimates. In most cases, just one of the potential policies analyzed in a survey could be implemented (i.e., a typical application asks about different possible variations of a potential policy), and so the relevant value to use in policy analysis would be one uncontaminated by alternative policies. Indeed, in this context, the purpose of the survey would be to estimate welfare changes that would occur after the introduction of the (lone) new policy.

On a related note, in some cases the purpose of a sequential DCE may be to determine which of several policies maximizes net benefits. The presence of framing effects clearly has the potential to yield a different rank-ordering of policies in terms of their expected net benefits. To illustrate this point, consider the choice of whether to implement project S3 or L3 from the experiment and, to keep it simple, assume that benefits only accrue to the voting group that helps fund the project. Using the actual cost of funding the projects, based on either binary choice mechanism, the winner is clear: S3 generates positive net benefits whereas L3 does not. For the case of Seq-TC, both projects generate negative net benefits.

Dating back to at least the NOAA Panel on Contingent Valuation (Arrow et al. 1993), and emphasized more recently (Johnson et al. 2017), a single binary choice referendum format is viewed by many as the benchmark from which to evaluate the performance of alternative value elicitation mechanisms. With this in mind, and coupled with the concerns raised over the Seq-TC mechanism, our study suggests a reasonable alternative format: a sequence of binary choices. The main advantage of asking many versus one binary choice question is that it dramatically lowers cost. The evidence we provide indicates that the differential behaviors induced by Seq-TC more than offset its potential efficiency advantage relative to Seq-BC. With obvious caveats, it bears mentioning that the two comparisons between these formats identified in the SP literature review of Lloyd-Smith, Zawojska, and Adamowicz (2020) yield the same convergent validity finding. In contrast to the Seq-TC format, an SP survey deploying a Seq-BC format, with each question being a vote between a potential policy and the status quo, can be incentive compatible (Vossler, Doyon, and Rondeau 2012). Nevertheless, an important challenge to the incentive compatibility of a Seq-BC mechanism is that people are unlikely to view their responses to the different choice questions as strategically independent. As one potential approach for mitigating this issue, Vossler et al. (2025) propose an "independence script" and demonstrate through a case study that it reduces non-truthful voting.

On a final note, a revealed preference experiment such as ours provides a benchmark of what we might expect in an SP survey that uses similar elicitation mechanisms and controls respondent beliefs about the mechanism. That revealed preference and SP studies give rise to parallel patterns of behavior, such as unexpectedly large WTP-WTA gaps, order effects, effects of asymmetric dominance, and various other elicitation effects, is well documented (Poe 2016; Lades 2025). As such, we expect results from an investigation like ours to generalize to other revealed and SP settings.

References

Adamowicz, Wiktor, Peter Boxall, Michael Williams, and Jordan Louviere. 1998. Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80 (1): 64-75.

Adamowicz, Wiktor, Diane Dupont, Alan Krupnick, and Jing Zhang. 2011. Valuation of cancer and microbial disease risk reductions in municipal drinking water: An analysis of risk context using multiple valuation methods. *Journal of Environmental Economics and Management* 61 (2): 213-226.

Arrow, Kenneth, Robert Solow, Paul Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. Report of the NOAA panel on contingent valuation. *Federal Register* 58 (10): 4601-4614.

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy. 2018. Incentives in experiments: A theoretical analysis. *Journal of Political Economy* 126 (4): 1472-1503.

Bateman, Ian J., Alistair Munro, and Gregory L. Poe. 2008. Decoy effects in choice experiments and contingent valuation: Asymmetric dominance. *Land Economics* 84 (1): 115-127.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak. 1964. Measuring utility by a single-response sequential method. *Behavioral Science* 9 (3): 226-232.

Boxall, Peter C., Wiktor L. Adamowicz, Jeffrey Swait, Michael Williams, and Jordan Louviere. 1996. A comparison of stated preference methods for environmental valuation. *Ecological Economics* 18 (3), 243-253.

Carson, Richard T., and Theodore Groves. 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics* 37: 181-210.

Carson, Richard T., Theodore Groves, and John A. List. 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, Richard T., and Jordan J. Louviere. 2011. A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics* 49: 539-559.

Caussade, Sebastián, Juan de Dios Ortúzar, Luis I. Rizzi, and David A. Hensher. 2005. Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological* 39 (7): 621-640.

Champ, Patricia A., and Richard C. Bishop. 2006. Is willingness to pay for a public good sensitive to the elicitation format? *Land Economics* 82 (2): 162-173.

Czajkowski, Mikolaj, Marek Giergiczny, and William H. Greene. 2014. Learning and fatigue effects revisited: Investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land Economics* 90 (2): 324-351.

Day, Brett, Ian J. Bateman, Richard T. Carson, Diane Dupont, Jordan J. Louviere, Sanae Morimoto, Richard Scarpa, and Paul Wang. 2012. Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management* 63 (1): 73-91.

Farquharson, Robin. 1969. *Theory of Voting*. New Haven, CT: Yale University Press.

Frey, Ulrich J., and Frauke Pirscher. 2019. Distinguishing protest responses in contingent valuation: A conceptualization of motivations and attitudes behind them. *PloS One* 14 (1): e0209872.

Hensher, David A. 2004. Identifying the influence of stated choice design dimensionality on willingness to pay for travel time savings. *Journal of Transport Economics and Policy* 38 (3): 425-446.

Hensher, David A. 2006. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and Resource Economics* 34 (1): 7-44.

Johnson, Rich, and Bryan Orme. 2003. Getting the most from CBC. Sequim: Sawtooth Software Research Paper Series, Sawtooth Software.

Johnston, Robert J., Kevin J. Boyle, Wiktor Adamowicz, Jeff Bennett, Roy Brouwer, Trudy Ann Cameron, W. Michael Hanemann, et al. 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists* 4 (2): 319-405.

Lades, Leonhard K., Ewa Zawojska, Robert J. Johnston, Nick Hanley, Liam Delaney, and Mikołaj Czajkowski. 2025. Anomalies or expected behaviors? Understanding stated preferences and welfare implications in light of contemporary behavioral economics. *Review of Environmental Economics and Policy* 19 (1): 48-68.

Lloyd-Smith, Patrick, Ewa Zawojska, and Wiktor Adamowicz. 2020. Moving beyond the contingent valuation versus choice experiment debate: Presentation effects in stated preference. *Land Economics* 96 (1): 1-24.

Louviere, Jordan J., and David A. Hensher.1982. Design and analysis of simulated choice or allocation experiments in travel choice modeling. *Transportation Research Record* 890: 11-17.

Louviere, Jordan J., and George Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: An approach based on aggregate data. *Journal of Marketing Research* 20 (4): 350-367.

Lupi, Frank, Daniel J. Phaneuf, and Roger H. von Haefen. 2020. Best practices for implementing recreation demand models. *Review of Environmental Economics and Policy* 14 (2): 302-323.

McKelvey, Richard D., and Peter C. Ordeshook. 1972. A general theory of the calculus of voting. In James F. Herndon and Joseph L. Bernd (Eds.), *Mathematical applications in political science*. Charlottesville, VA: University Press of Virginia.

Meyerhoff, Jürgen, Morten Raun Mørkbak, and Soren B. Olsen. 2014. A meta-study investigating the sources of protest behaviour in stated preference surveys. *Environmental and Resource Economics*, 58: 35-57.

Oehlmann, Malte, Jürgen Meyerhoff, Petr Mariel, and Priska Weller. 2017. Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics and Management* 81: 59-73.

Park, Jongwon, and JungKeun Kim. 2005. The effects of decoys on preference shifts: The role of attractiveness and providing justification. *Journal of Consumer Psychology* 15 (2): 94-107.

Poe, Gregory L. 2016. Behavioral anomalies in contingent values and actual choices. *Agricultural and Resource Economics Review* 45 (2): 246-269.

Regier, Dean A., Verity Watson, Heather Burnett, and Wendy J. Ungar. 2014. Task complexity and response certainty in discrete choice experiments: An application to drug treatments for juvenile idiopathic arthritis. *Journal of Behavioral and Experimental Economics* 50: 40-49.

Revelt, David, and Kenneth Train. 1998. Mixed logit with repeated choices: Households' choices of appliance efficiency level. *Review of Economics and Statistics* 80 (4): 647-657.

Rolfe, John, and Jill Windle. 2012. Restricted versus unrestricted choice in labeled choice experiments: Exploring the trade-offs of expanding choice dimensions. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 60 (1): 53-70.

Rondeau, Daniel, and Christian A. Vossler. (2024). Incentive compatibility and respondent beliefs: Consequentiality and game form. Unpublished manuscript.

Rose, John M., David A. Hensher, Sebastian Caussade, Juan de Dios Ortúzar, and Ron-Chang Jou. 2009. Identifying differences in willingness to pay due to dimensionality in stated choice experiments: A cross country analysis. *Journal of Transport Geography* 17 (1): 21-29.

Rutström, E. Elisabet. 1998. Home-grown values and the design of incentive compatible auctions. *International Journal of Game Theory* 27: 427-441.

Stithou, Mavra and Riccardo Scarpa. 2012. Collective versus voluntary payment in contingent valuation for the conservation of marine biodiversity: An exploratory study from Zakynthos, Greece. *Ocean & Coastal Management* 56: 1-9.

Welsh, Michael P., and Gregory L. Poe. 1998. Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach. *Journal of Environmental Economics and Management* 36 (2): 170-185.

Weng, Weizhe, Mark D. Morrison, Kevin J. Boyle, Peter C. Boxall, and John Rose. 2021. Effects of the number of alternatives in public good discrete choice experiments. *Ecological Economics* 182: 106904.

von Haefen, Roger H., D. Matthew Massey, and Wiktor L. Adamowicz. 2005. Serial nonparticipation in repeated discrete choice models. *American Journal of Agricultural Economics* 87 (4): 1061-1076.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* 4 (4): 145-171.

Vossler, Christian A., David A. Keiser, Catherine L. Kling, and Daniel J. Phaneuf. 2025. Information scripts and the incentive compatibility of discrete choice experiments. *Journal of the Association of Environmental and Resource Economists* 12 (2): 459-492.

Zhang, Jing, and Wiktor L. Adamowicz. 2011. Unraveling the choice format effect: A context-dependent random utility model. *Land Economics* 87 (4): 730-743.

Table 1. Farmworker assistance projects included in the experimental design

Project	Educational materials (sets)	Winter clothing (sets)	Transportation services
S1	1	1	3
S2	3	1	1
S3	1	3	1
M1	3	3	2
M2	1	5	2
M3	5	1	2
L1	5	5	1
L2	5	3	3
L3	3	5	3

Notes: A set of education materials includes a set of English language workbooks for a farmworker and their tutor, and a set of age-appropriate children's books for farmworker families. A set of winter clothing includes a pair of gloves, a winter coat, and a heavy blanket. Transportation service is the round-trip transportation to a legal service, a medical appointment, or another important engagement.

Table 2. Sample means of socio-demographic characteristics

	Single-BC [project S3]	Single-BC [project L3]	Seq-BC	Seq-TC
Female	47.0%	48.7%	54.7%	50.3%
White (non-Hispanic)	57.0%	65.3%	65.0%	57.7%
Asian	13.7%	14.0%	11.7%	13.7%
Black	13.7%	10.7%	11.7%	12.7%
Hispanic	12.0%	7.0%	8.0%	9.3%
Married	29.7%	25.3%	29.0%	26.3%
Bachelor's degree or higher	66.3%	54.7%	67.7%	65.7%
Employed	74.3%	75.6%	75.0%	77.6%
Living in a rural area	11.3%	10.8%	7.9%	8.7%
Has lived in New York for sixteen years or more	79.3%	80.0%	81.0%	80.0%
Donated in the last twelve months (time or money) to a non-profit organization	62%	57.7%	59.7%	59.0%
Age (years)	36.1	34.2	34.9	34.4
Household size (people)	2.8	2.7	2.8	2.8
Household annual income (US$)	82,217	81,585	84,717	81,408

Notes: Sample size for each treatment is 300. Household size and income variables are top-coded. "Single-BC", "Seq-BC", and "Seq-TC" denote the single binary choice, sequential binary choice, and sequential trinary choice mechanisms, respectively.

Table 3. Responses to follow-up questions related to voting

	Single-BC [project S3]	Single-BC [project L3]	Seq-BC	Seq-TC
"How well did you understand the voting procedures in this study?"				
Everything was clear	88.0	85.0	82.3	80.3
I understood most of them	9.7	13.3	14.0	15.0
I understood some of them	2.3	1.7	3.3	3.7
I was confused	0.0	0.0	0.3	1.0
"If a project is implemented, I will have to pay the stated cost of the project."				
Agree	93.7	93.7	91.3	94.7
Neither agree nor disagree	5.3	3.7	5.3	3.0
Disagree	1.0	2.7	3.3	2.3
"If a project is implemented, the stated support actions will be delivered to farmworkers."				
Agree	91.3	90.7	90.3	87.0
Neither agree nor disagree	8.0	7.3	7.7	9.7
Disagree	0.7	2.0	2.0	2.7
"The votes from my group will be used to determine whether a project is implemented or not."				
Agree	94.7	93.7	90.0	89.7
Neither agree nor disagree	5.0	5.0	5.7	8.3
Disagree	0.3	1.3	4.3	2.0
"Overall, how certain are you about your voting choice(s)?"				
Certain	55.7	57.7	50.3	47.3
Somewhat certain	35.3	32.3	39.7	44.0
Somewhat uncertain	8.0	8.3	9.3	6.7
Uncertain	1.0	1.7	0.7	2.0
Influence of choice attributes on voting (percentage indicating "large" or "moderate" effect)				
Educational materials	75.7	79.3	79.0	75.7
Winter clothing	80.0	80.7	84.0	78.3
Transportation	71.3	77.3	74.7	68.0
Cost	66.7	69.0	68.0	69.3

Notes: Table entries are percentages. "Single-BC", "Seq-BC", and "Seq-TC" denote the single binary choice, sequential binary choice, and sequential trinary choice mechanisms, respectively.

Table 4. Percentage voting in favor of projects S3 and L3 for the single binary choice (Single-BC) and sequential binary choice (Seq-BC) mechanisms

	Project cost				
Mechanism [project]	**$5**	**$8**	**$11**	**$15**	**Overall**
Single-BC [S3]	68	53	49	32	50
Seq-BC [S3]	68	56	51	32	52
Single-BC [L3]	77	72	58	44	63
Seq-BC [L3]	83	69	65	44	65

Notes: Table entries are the percentage who voted in favor of funding a project at the indicated project cost. Projects S3 and L3 are defined in Table 1.

Table 5. Willingness-to-pay estimates, by project and elicitation mechanism

Project	Single-BC	Seq-BC	Seq-TC	Single-BC = Seq-BC [Single-BC = Seq-TC]	Seq-BC = Seq-TC
S1		7.74 (0.99)	0.80 (1.69)		$p < 0.01$
S2		8.16 (0.63)	3.63 (0.84)		$p < 0.01$
S3	9.78 (0.84)	10.27 (0.85)	3.89 (0.95)	$p = 0.68$ [$p < 0.01$]	$p < 0.01$
M1		12.69 (0.78)	10.03 (0.60)		$p < 0.01$
M2		11.06 (0.59)	8.56 (0.61)		$p < 0.01$
M3		10.00 (0.63)	7.33 (0.67)		$p < 0.01$
L1		12.90 (0.86)	11.36 (0.87)		$p = 0.21$
L2		12.47 (0.72)	12.21 (1.05)		$p = 0.84$
L3	13.45 (1.08)	13.88 (1.04)	12.81 (0.98)	$p = 0.77$ [$p = 0.66$]	$p = 0.45$

Notes: Willingness-to-pay (WTP) estimates are denominated in US$. Cluster-robust standard errors (clustered by participant) are in parentheses. All WTP estimates are statistically different from zero at the 1 percent level, except for project S1 (p=0.64) based on the Seq-TC mechanism. "Single-BC", "Seq-BC", and "Seq-TC" denote the single binary choice, sequential binary choice, and sequential trinary choice mechanisms, respectively. WTP estimates are derived from conditional logit models for Single-BC and mixed logit models for Seq-BC and Seq-TC, which are reported in Table A1, that define utility as a linear function of project fixed effects and the project cost. The projects are defined in Table 1.

Table 6. Project-by-choice-set willingness-to-pay estimates, sequential trinary choice

Project X	Project A	WTP for X | Project A	Project B	WTP for X | Project B	WTP_X | A = WTP_X | B
S1	M1	1.79 (1.28)	L1	1.29 (1.40)	$p = 0.64$
S2	M2	3.31 (1.26)	L2	3.69 (0.93)	$p = 0.70$
S3	M3	5.62 (0.94)	L3	3.13 (0.98)	$p < 0.01$
M1	S1	10.96 (0.63)	L1	8.18 (0.72)	$p < 0.01$
M2	S2	10.63 (0.71)	L2	4.62 (1.08)	$p < 0.01$
M3	S3	8.98 (0.72)	L3	5.01 (0.95)	$p < 0.01$
L1	S1	11.30 (0.85)	M1	10.53 (0.84)	$p = 0.33$
L2	S2	10.88 (1.11)	M2	12.65 (1.04)	$p = 0.05$
L3	S3	12.18 (0.90)	M3	12.20 (0.87)	$p = 0.99$

Notes: Willingness-to-pay (WTP) estimates are denominated in US$. Cluster-robust standard errors (clustered by participant) are in parentheses. All WTP estimates are statistically different from zero at the 1 percent level, except those for project S1. Each project denoted in the first column appears in two choice sets, once with "Project A" and once with "Project B". WTP estimates are derived from a mixed logit model, reported in Table A2, that allows WTP to vary by project and by choice set. The projects are defined in Table 1.

Table 7. Status quo ("no project") voting choices, by project and elicitation mechanism

Sequential trinary choice (Seq-TC)		Sequential binary choice (Seq-BC)			
Projects in choice set	Status quo (%)	Project in choice set	Status quo (%)	Project in choice set	Status quo (%)
S1, M1	38.33	S1	56.67***	M1	37.33
S2, M2	36.00	S2	56.67***	M2	43.67*
S3, M3	39.33	S3	48.33**	M3	49.00**
S1, L1	38.00	S1	56.67***	L1	36.67
S2, L2	39.00	S2	56.67***	L2	37.00
S3, L3	38.67	S3	48.33**	L3	34.67
M1, L1	32.67	M1	37.33	L1	36.67
M2, L2	33.00	M2	43.67***	L2	37.00
M3, L3	33.67	M3	49.00***	L3	34.67

Notes: The first column lists the projects included in a particular Seq-TC choice set, and the second column presents the percentage selecting the status quo (i.e., "no project") option for this choice set. The fourth and sixth columns display the corresponding status quo choices for Seq-BC choice sets that include one of the two projects also contained in the Seq-TC choice set. *, **, and *** indicate that the percentage of status quo Seq-BC choices are statistically different than the percentage of the status quo Seq-TC choices at the 10, 5, and 1 percent significance levels, respectively (based on Fisher's exact tests). The projects are defined in Table 1.

Table 8. Willingness-to-pay estimates for projects S3 and L3 conditional on question order

Project [Order]	Single-BC	Seq-BC	Seq-TC	Single-BC = Seq-BC [Single-BC = Seq-TC]	Seq-BC = Seq-TC
S3 [First]	9.78 (0.84)	11.55 (1.11)	5.00 (1.11)	$p = 0.20$ [$p < 0.01$]	$p < 0.01$
S3 [Not First]		8.84 (1.10)	3.47 (1.52)		$p < 0.01$
L3 [First]	13.45 (1.08)	14.87 (1.36)	11.52 (1.51)	$p = 0.41$ [$p = 0.30$]	$p = 0.10$
L3 [Not First]		12.68 (1.13)	12.92 (1.19)		$p = 0.90$

Notes: Willingness-to-pay (WTP) estimates are denominated in US$. Cluster-robust standard errors (clustered by participant) are in parentheses. "Single-BC", "Seq-BC", and "Seq-TC" denote the single binary choice, sequential binary choice, and sequential trinary choice mechanisms, respectively. WTP estimates are derived from conditional logit models for Single-BC, and mixed logit models for Seq-BC and Seq-TC, which are reported in Table A3. "Order" refers to whether the WTP estimate is derived from the subset of participants who faced the indicated project in their first choice question (First) or in a subsequent question (Not First). The projects are defined in Table 1.

Table 9. Willingness-to-pay estimates, attribute-based utility specification

Project	Sequential binary choice (Seq-BC)	Sequential trinary choice (Seq-TC)	Seq-BC = Seq-TC
S1	8.63 (0.57)	7.38 (1.15)	$p = 0.33$
S2	8.80 (0.52)	7.42 (1.09)	$p = 0.25$
S3	9.70 (0.51)	8.00 (1.10)	$p = 0.16$
M1	11.33 (0.48)	11.55 (1.23)	$p = 0.87$
M2	12.24 (0.58)	12.13 (1.28)	$p = 0.94$
M3	10.43 (0.54)	10.97 (1.24)	$p = 0.69$
L1	14.04 (0.68)	15.73 (1.43)	$p = 0.29$
L2	12.97 (0.61)	15.10 (1.44)	$p = 0.18$
L3	13.87 (0.65)	15.68 (1.47)	$p = 0.26$

Notes: Willingness-to-pay (WTP) estimates are denominated in US$. Cluster-robust standard errors (clustered by participant) are in parentheses. All WTP estimates are statistically different from zero at the 1 percent level. WTP estimates are derived from mixed logit models, reported in Table A6, that define utility as a linear function of an alternative specific constant and continuous covariates that define levels of the four attributes (education, clothing, transportation, cost). The projects are defined in Table 1.

Table 10. Simulation results: Expected precision gains from Seq-TC relative to Seq-BC

	Seq-BC DGP		Seq-BC		Seq-TC		
Project	WTP	Std. Err.	WTP	Std. Err.	WTP	Std. Err.	Std. Err. ratio
S1	8.20	0.082	8.19 (0.666)	0.083 (0.010)	8.19 (0.630)	0.076 (0.008)	1.098 (0.129)
S2	8.04	0.082	8.04 (0.689)	0.085 (0.011)	8.03 (0.604)	0.075 (0.008)	1.128 (0.139)
S3	10.13	0.067	10.12 (0.671)	0.066 (0.006)	10.13 (0.564)	0.057 (0.004)	1.168 (0.105)
M1	12.88	0.056	12.89 (0.699)	0.055 (0.003)	12.88 (0.553)	0.044 (0.003)	1.253 (0.077)
M2	11.29	0.060	11.29 (0.682)	0.060 (0.004)	11.29 (0.553)	0.049 (0.003)	1.230 (0.085)
M3	10.01	0.066	10.01 (0.665)	0.067 (0.006)	10.01 (0.579)	0.057 (0.005)	1.163 (0.096)
L1	12.95	0.056	12.96 (0.705)	0.055 (0.003)	12.95 (0.564)	0.044 (0.003)	1.255 (0.082)
L2	12.84	0.056	12.84 (0.691)	0.055 (0.003)	12.85 (0.546)	0.043 (0.002)	1.281 (0.083)
L3	13.55	0.055	13.57 (0.726)	0.053 (0.003)	13.56 (0.557)	0.042 (0.002)	1.286 (0.081)

Notes: Reported are willingness-to-pay (WTP) means and normalized standard errors (i.e., standard error / WTP) calculated from a Monte Carlo simulation with 10,000 replications. Standard deviations are reported in parentheses. "Seq-BC" and "Seq-TC" denote the sequential binary choice and sequential trinary choice mechanisms, respectively. In each replication: (1) the utilities for the nine projects are drawn for a sample of 300 using the DGP estimated from the Seq-BC data; (2) each simulated individual faces all nine choice sets (separately for Seq-BC and Seq-TC), with randomly selected costs consistent with the experimental design; (3) the simulated individuals select the choice option that yields the highest utility; and (4) mechanism-specific conditional logit models with cluster-robust standard errors are estimated for the sample. The projects are defined in Table 1.

Appendix. Estimated models

Table A1. Utility specification with project fixed effects

	Single-BC, S3	Single-BC, L3	Seq-BC	Seq-TC
Means				
S1			2.328*** (0.572)	0.280 (0.598)
S2			2.453*** (0.548)	1.277*** (0.343)
S3	1.384*** (0.342)		3.090*** (0.715)	1.369*** (0.374)
M1			3.817*** (0.793)	3.523*** (0.353)
M2			3.327*** (0.706)	3.007*** (0.322)
M3			3.009*** (0.648)	2.577*** (0.335)
L1			3.880*** (0.824)	3.993*** (0.416)
L2			3.750*** (0.771)	4.290*** (0.459)
L3		2.074*** (0.375)	4.176*** (0.937)	4.500*** (0.458)
Cost	–0.142*** (0.0329)	–0.154*** (0.0345)	–0.301*** (0.0659)	–0.351*** (0.0293)
Standard Deviations				
S1			3.152** (1.316)	2.493*** (0.711)
S2			1.685* (0.948)	2.277*** (0.381)
S3			2.961** (1.229)	2.899*** (0.493)
M1			1.830** (0.825)	2.348*** (0.346)
M2			1.391* (0.791)	2.443*** (0.339)
M3			1.863** (0.917)	2.388*** (0.341)
L1			2.044** (0.931)	3.797*** (0.532)
L2			1.522** (0.694)	4.645*** (0.606)

L3			2.391** (1.086)	4.602*** (0.646)
Respondents	300	300	300	300
Respondents × Votes	300	300	2700	2700
Log-L	–198.2248	–186.8752	–1677.6335	–2398.7186
McFadden's R^2	0.0467	0.1013	0.1036	0.1913

Notes: *** p<0.01, ** p<0.05, * p<0.1. Cluster-robust standard errors (clustered by respondent) are in parentheses. "Single-BC", "Seq-BC", and "Seq-TC" denote the single binary choice, sequential binary choice, and sequential trinary choice mechanisms, respectively. We use conditional logit models to analyze the Single-BC data, and mixed logit models for the Seq-BC and Seq-TC data. Mixed logit models are estimated using 2000 Halton random draws. "S1", "S2", etc., are project-specific indicator variables. For mixed logit models, all variables other than "Cost" are assumed to follow normal distributions. The projects are defined in Table 1.

Table A2. Project-by-choice-set utility specification, sequential trinary choice mechanism

	Means	Standard deviations
S1	0.669 (0.488)	2.652*** (0.553)
S2	1.377*** (0.382)	2.604*** (0.497)
S3	2.100*** (0.416)	2.908*** (0.474)
M1	3.054*** (0.385)	2.482*** (0.370)
M2	3.967*** (0.404)	3.301*** (0.542)
M3	1.870*** (0.412)	2.673*** (0.387)
L1	4.218*** (0.432)	3.590*** (0.442)
L2	4.723*** (0.492)	4.779*** (0.647)
L3	4.549*** (0.490)	4.268*** (0.535)
S1 × Paired with Project B	–0.188 (0.402)	
S2 × Paired with Project B	–0.140 (0.368)	
S3 × Paired with Project B	–0.930*** (0.314)	
M1 × Paired with Project B	1.038*** (0.276)	
M2 × Paired with Project B	–2.244*** (0.406)	
M3 × Paired with Project B	1.484*** (0.333)	
L1 × Paired with Project B	–0.286 (0.292)	
L2 × Paired with Project B	–0.660* (0.340)	
L3 × Paired with Project B	0.00451 (0.277)	
Cost	–0.373*** (0.0323)	
Respondents	300	
Respondents × Votes	2700	
Log-L	–2355.1364	
McFadden's R^2	0.2060	

Notes: *** p<0.01, ** p<0.05, * p<0.1. Cluster–robust standard errors (clustered by respondent) are in parentheses. Utility parameters estimated using mixed logit with 2000 Halton random draws. "S1", "S2", etc., are project-specific indicator variables with coefficients assumed to follow normal distributions. "Paired with Project B" is an indicator that equals 1 when the project being valued is in a choice set that includes "Project B" as defined in Table 5. The projects are defined in Table 1.

Table A3. Utility specification with project fixed effects, restricted to choice sets that include S3 or L3 (Seq-BC) or both S3 and L3 (Seq-TC)

	Seq-BC	Seq-TC
Means		
S3	2.007***	1.180*
	(0.449)	(0.606)
L3	2.583***	2.716***
	(0.427)	(0.860)
S3 × Not First	–0.471*	–0.361
	(0.286)	(0.323)
L3 × Not First	–0.380	0.331
	(0.257)	(0.385)
Cost	–0.174***	–0.236***
	(0.0356)	(0.0844)
Standard Deviations		
S3	0.744	0.00125
	(1.125)	(0.0164)
L3	0.0890	1.298
	(0.274)	(1.491)
Respondents	300	300
Respondents × Votes	600	600
Log–L	–376.26733	–304.93933
McFadden's R^2	0.0953	0.0748

Notes: *** p<0.01, ** p<0.05, * p<0.1. Cluster–robust standard errors (clustered by respondent) are in parentheses. Utility parameters estimated using mixed logit models with 2000 Hammersley (Seq-BC) or Halton (Seq-TC) random draws. "Not First" is an indicator that equals 1 if the project appears in any choice set other than the first one encountered by the participant. "S3" and "L3" are project-specific indicator variables with coefficients assumed to follow normal distributions. The projects are defined in Table 1.

Table A4. Utility specification with linear order effects, sequential choice mechanisms

	Sequential binary choice	Sequential trinary choice
S1	2.864*** (1.056)	1.157* (0.553)
S2	2.560*** (0.874)	0.992* (0.497)
S3	3.523*** (1.079)	1.540*** (0.474)
M1	3.625*** (1.124)	3.981*** (0.370)
M2	2.634*** (0.797)	3.188*** (0.542)
M3	3.088*** (0.901)	3.027*** (0.387)
L1	4.906*** (1.566)	4.737*** (0.442)
L2	2.751*** (0.902)	4.540*** (0.647)
L3	4.667*** (1.359)	4.465*** (0.535)
S1 × Order	–0.0927 (0.123)	–0.107 (0.402)
S2 × Order	–0.0128 (0.0828)	0.0592 (0.368)
S3 × Order	–0.135 (0.101)	–0.00970 (0.314)
M1 × Order	0.0499 (0.093)	–0.0763 (0.276)
M2 × Order	0.135 (0.0887)	–0.0213 (0.406)
M3 × Order	–0.00959 (0.0784)	–0.0848 (0.333)
L1 × Order	–0.175 (0.110)	–0.139* (0.292)
L2 × Order	0.187* (0.101)	–0.0457 (0.340)
L3 × Order	–0.148 (0.0912)	0.0337 (0.277)
Cost	–0.303*** (0.0848)	–0.354*** (0.0307)
Respondents	300	300
Respondents × Votes	2700	2700
Log-L	–1668.3917	–2394.1895
McFadden's R^2	0.1085	0.1929

Notes: *** p<0.01, ** p<0.05, * p<0.1. Cluster–robust standard errors (clustered by respondent) are in parentheses. Utility parameters estimated using mixed logit with 2000 Halton random draws. "S1", "S2", etc., are project-specific indicator variables with coefficients assumed to follow normal distributions. Standard deviations of random coefficients are omitted for convenience. "Order" denotes the order (1 to 9) in which the indicated project was an option included in the choice question voted on by the participant. The projects are defined in Table 1.

Table A5. Utility specification with project fixed effects, Seq-TC, non-strategic voters subsample

	Means	Standard deviations	WTP
S1	–0.0172	2.488^{***}	-0.05
	(0.620)	(0.716)	(1.82)
S2	0.540	2.997^{***}	1.58
	(0.477)	(0.531)	(1.33)
S3	1.209^{***}	3.045^{***}	3.54^{***}
	(0.454)	(0.586)	(1.15)
M1	3.480^{***}	2.980^{***}	10.19^{***}
	(0.501)	(0.572)	(0.93)
M2	2.880^{***}	3.185^{***}	8.44^{***}
	(0.482)	(0.495)	(0.95)
M3	2.202^{***}	2.985^{***}	6.45^{***}
	(0.440)	(0.507)	(0.97)
L1	3.682^{***}	4.843^{***}	10.79^{***}
	(0.518)	(0.918)	(1.20)
L2	4.467^{***}	6.436^{***}	13.08^{***}
	(0.759)	(1.227)	(1.87)
L3	4.520^{***}	5.834^{***}	13.24^{***}
	(0.608)	(1.074)	(1.19)
Cost	$–0.341^{***}$		
	(0.0377)		
Respondents	215		
Log–L	–1686.7062		
McFadden's R^2	0.2066		

Notes: *** p<0.01, ** p<0.05, * p<0.1. Willingness-to-pay (WTP) estimates (last column) are denominated in US$ and derived from the estimated utility model (first two estimation columns of estimates). Cluster–robust standard errors (clustered by respondent) are in parentheses. Utility parameters estimated using mixed logit models with 2000 Halton random draws. "S1", "S2", etc., are project-specific indicator variables. Parameters on all variables other than "Cost" are assumed to follow normal distributions. The projects are defined in Table 1.

Table A6. Project-by-choice-set utility specification, Seq-TC, non-strategic voters subsample

	Means	Standard deviations
S1	–0.00742 (0.685)	2.711*** (0.840)
S2	0.601 (0.503)	3.231*** (0.641)
S3	2.047*** (0.532)	3.043*** (0.588)
M1	2.736*** (0.508)	2.943*** (0.531)
M2	3.766*** (0.591)	3.881*** (0.579)
M3	1.197** (0.516)	3.219*** (0.591)
L1	4.131*** (0.590)	4.367*** (0.712)
L2	4.909*** (0.812)	6.264*** (1.141)
L3	4.541*** (0.639)	5.143*** (0.912)
S1 × Paired with Project B	0.208 (0.525)	
S2 × Paired with Project B	0.162 (0.476)	
S3 × Paired with Project B	–1.175*** (0.375)	
M1 × Paired with Project B	1.397*** (0.356)	
M2 × Paired with Project B	–2.321*** (0.497)	
M3 × Paired with Project B	1.822*** (0.448)	
L1 × Paired with Project B	–0.661* (0.362)	
L2 × Paired with Project B	–0.733 (0.491)	
L3 × Paired with Project B	–0.204 (0.349)	
Cost	–0.361*** (0.0419)	
Respondents	215	
Log–L	–1652.332	
McFadden's R^2	0.2227	

Notes: *** p<0.01, ** p<0.05, * p<0.1. Cluster–robust standard errors (clustered by respondent) are in parentheses. Utility parameters estimated using mixed logit models with 2000 Halton random draws. "S1", "S2", etc., are project-specific indicator variables. Parameters on all variables other than "Cost" are assumed to follow normal distributions. "Paired with Project B" is an indicator that equals 1 when the project being valued is in a choice set that includes Project B as defined in Table 5. "Seq-TC" refers to the sequential trinary choice mechanism. The projects are defined in Table 1.

Table A7. Project-by-choice-set willingness-to-pay estimates, Seq-TC, non-strategic voter subsample

Project X	Project A	WTP for X | Project A	Project B	WTP for X | Project B	WTP_X | A = WTP_X | B
S1	M1	-0.02 (1.90)	L1	0.55 (2.12)	$p = 0.69$
S2	M2	2.11 (1.71)	L2	1.66 (1.33)	$p = 0.73$
S3	M3	5.66*** (1.17)	L3	2.41* (1.26)	$p < 0.01$
M1	S1	11.43*** (0.92)	L1	7.57*** (1.01)	$p < 0.01$
M2	S2	10.42*** (1.10)	L2	4.00** (1.59)	$p < 0.01$
M3	S3	8.35*** (0.96)	L3	3.31** (1.32)	$p < 0.01$
L1	S1	11.43*** (1.20)	M1	9.60*** (1.15)	$p = 0.07$
L2	S2	11.55*** (1.80)	M2	13.58*** (1.74)	$p = 0.14$
L3	S3	12.56*** (1.09)	M3	12.00*** (1.15)	$p = 0.56$

Notes: Willingness-to-pay (WTP) estimates are denominated in US$. Cluster-robust standard errors (clustered by participant) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Each project denoted in the first column appears in two choice sets, once with "Project A" and once with "Project B". WTP estimates are derived from a mixed logit model, reported in Table A5, that allows WTP to vary by project and by choice set. The projects are defined in Table 1.

Table A8. Attribute-based utility specification, sequential choice mechanisms

	Sequential binary choice	Sequential trinary choice
Means		
ASC	2.886***	0.418
	(0.558)	(0.368)
Educational materials	0.298***	0.413***
	(0.0569)	(0.0374)
Winter clothing	0.532***	0.514***
	(0.0715)	(0.0417)
Transportation services	0.255***	0.405***
	(0.0968)	(0.0643)
Cost	–0.519***	–0.347***
	(0.0527)	(0.0294)
Standard deviations		
ASC	4.110***	5.106***
	(0.448)	(0.547)
Educational materials	0.359***	0.375***
	(0.0752)	(0.0522)
Winter clothing	0.470***	0.399***
	(0.0904)	(0.0455)
Transportation services	0.541***	0.668***
	(0.165)	(0.0831)
Respondents	300	300
Log–L	–1098.8095	–1679.3587
McFadden's R^2	0.4129	0.4338

Notes: Cluster–robust standard errors (clustered by respondent) are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Utility parameters estimated using mixed logit models with 2000 Halton random draws. Parameters on all variables other than "Cost" are assumed to follow normal distributions. "ASC" is an indicator that equals 1 for a non-status quo option (i.e., a project).