Is there really a difference between "contingent valuation" and "choice experiments"?

Evidence from an induced-value experiment

Patrick Lloyd-Smith
Department of Agricultural and Resource Economics,
University of Saskatchewan, Canada

Ewa Zawojska
Faculty of Economic Sciences,
University of Warsaw, Poland

Wiktor L. Adamowicz
Department of Resource Economics and Environmental Sociology,
University of Alberta, Canada

ezawojska@wne.uw.edu.pl

Selected Paper prepared for presentation at the 2018 Agricultural & Applied Economics Association Annual Meeting, Washington, D.C., August 5-August 7

Is there really a difference between "contingent valuation" and "choice experiments"?

Patrick Lloyd-Smith[*]
Department of Agricultural and Resource Economics
University of Saskatchewan, Saskatoon, Canada

Ewa Zawojska
Faculty of Economic Sciences
University of Warsaw, Warsaw, Poland

Wiktor L. Adamowicz
Department of Resource Economics and Environmental Sociology
University of Alberta, Edmonton, Canada

[*] Corresponding author, patrick.lloydsmith@usask.ca. We thank Stephane Luchini and Verity Watson for sharing the experimental instructions and the z-Tree code, which has enabled us to conduct our study. The research was supported by the University of Alberta through the Support for the Advancement of Scholarship and the National Science Foundation (Award Number: 1360391). Ewa Zawojska gratefully acknowledges the support of the National Science Centre in Poland (Preludium 8, grant no. 2014/15/N/HS4/01328; Etiuda 4, grant no. 2016/20/T/HS4/00013).

Is there really a difference between "contingent valuation" and "choice experiments"?

Abstract: "Contingent valuation" ("CV") and "choice experiments" ("CE") are generally introduced as two separate stated preference methods to estimate welfare measures, and a large literature investigates their convergent validity. We first review the literature comparing "CV" and "CE", and show that these comparisons typically differ in (1) the number of options presented per value elicitation task, (2) the number of tasks given to a single respondent, (3) the framing of tasks, (4) the set (and order) of attributes characterizing options in tasks, (5) sizes of "CV" and "CE" samples, (6) econometric models used for data analysis, and (7) the format of information presented. Despite the wide variety of applications, we argue that the main (and perhaps only) difference between "CV" and "CE" is the presentation of information in elicitation tasks: as text in "CV" and as a table in "CE". We then assess the effect of presentation of information in an induced-value experiment. We find that participants perform equally well in "CV" and "CE" tasks in terms of making payoff-maximizing choices based on the induced values, but "CV" tasks take substantially more time to answer. A significant difference between payoff-maximizing choices in "CV" and "CE" is observed when only answers to the first elicitation task are considered. This latter finding is particularly important in light of recommendations for stated preference research that suggest that valuation studies should use only one task for eliciting preferences.

Keywords: Stated preference, Contingent valuation, Choice experiment, Experimental economics

JEL Codes: Q51 D6 H4 C91 M31

1. Introduction

Stated preference (SP) methods are employed in various policy contexts including environmental valuation, transportation choice, and health assessment. The literature presents a myriad of SP methods with various nomenclatures. Categorization of these methods has been challenging. Common parlance in the profession uses terms like "contingent valuation" ("CV") and "choice experiment" ("CE") to distinguish between two major approaches in SP methods, while other categorizations employ terms like ranking, open-ended tasks, and discrete choice. Adding to the set of names and categorizations (and likely confusion) has been the use of terms like conjoint, choice modelling, attribute-based stated preference, and a host of other labels. Nevertheless, the most common categorization of SP methods refers to "CV" versus "CE". In most nonmarket valuation textbooks, "CV" and "CE" are typically introduced as separate methods (Champ, Boyle, and Brown 2017).

Carson and Louviere (2011) attempt to clarify the nomenclature for SP methods and expunge the use of the terms "contingent valuation" and "choice experiment" when referring to the value elicitation approaches. They suggest the term "discrete choice experiment" be used to describe a SP elicitation method that asks respondents to choose a single option from a given set of options.[1] They differentiate discrete choice experiment approaches from matching approaches, in which respondents are asked to provide a numerical match to their preferences in a form of a statement of a money amount, for example, by responding to an open-ended question or by circling one amount from a list of amounts on a payment card. Carson and Louviere (2011) state that variations of discrete choice experiments or matching tasks can be characterized by whether they employ a single question or a sequence of questions. Within discrete choices, they distinguish between single binary choice (choose one from a pair of options), single multinomial choice (choose one from more than two options), complete ranking of options, best-worst choice, and choice of a subset of options.

Carson and Louviere's classification is very systematic, yet the literature has not completely adopted the nomenclature. Even the recent Contemporary Guidance for Stated Preference Methods paper by Johnston et al. (2017) makes use of the terms "CV" and "CE", defining "CV" as a valuation task examining an "indivisible whole" (p. 320)

[1] Throughout the paper, by the term "option", we mean a possible outcome provided in a valuation task which a respondent is asked to express their preferences. We note that the SP literature also uses other terms such as an alternative and a variant as synonyms to an option.

and “CE” as valuation based on attributes (separate characteristics) of a considered good. Johnston et al. (2017) identify three primary considerations for researchers in choosing between “CV” and “CE”: (1) whether marginal or total values are needed for decision making, (2) how respondents view the evaluated good (as a whole or as separable into individual attributes), and (3) how the framing of the valuation task impacts a respondent’s understanding of the task. The first two considerations are really about whether “CV” or “CE” (or specifically the decision of whether to characterize the good as having attributes or not) is appropriate for a certain valuation policy context, whereas the third consideration is the focal point of a large literature comparing the two approaches.

While the distinctions between value elicitation methods may be viewed as semantics, there are at least two important reasons for additional clarity in the description and evaluation of SP approaches. First, the differences in elicitation approaches may result in differences in validity of the obtained value estimates. A consequential single binary choice question is incentive compatible in the sense of leading to truthful preference revelation (Farquharson 1969), while other forms of valuation tasks may encourage respondents to strategically misrepresent their preferences (Carson and Groves 2007, 2011; Vossler, Doyon, and Rondeau 2012). Second, there have been, and continue to be, comparisons of so-called “CV” and “CE” approaches, with the underlying rationale often being convergent validity assessment. Mixed evidence on convergence of value estimates derived from “CV” and “CE” approaches has generated concerns about the validity of SP methods (e.g., Hanley et al. 1998a; Cameron et al. 2002; Ryan 2004; Jin, Wang, and Ran 2006; Goldberg and Rosen 2007). However, the vast majority of these studies do not make comparisons of value estimates that are based on equivalent surveys (for example, the surveys differ in the number of attributes, options to choose from, and/or valuation tasks) or on equivalent econometric analyses. Consequently, many of the observed differences in estimates coming from “CV” and “CE” could be ascribed to a variety of differences in methodological techniques.

This paper reviews the large “CV”-“CE” comparison literature to identify differences in how the two approaches are employed. The review shows that applications implementations of these approaches vary within a single comparison in different ways including (1) the number of options presented per valuation task, (2) the number of tasks given per respondent, (3) the framing of valuation tasks

(e.g., respondents provide a number in an open-ended question, while they choose their most preferred option in a multinomial choice task), (4) the set (and order) of attributes that characterize options, (5) sizes of "CV" and "CE" samples, (6) econometric models used for data analysis, and (7) the format of information presented.

Despite the diversity of applications that have been called "CV" and "CE", we contend that the main difference between the two approaches is the presentation of information in valuation tasks: in "CV", the information is typically displayed as text, while in "CE", it is displayed in a table.[2] In principle, one could present the attribute-based information as text for the evaluated good(s) commonly shown in tables in "CE" (see Randall, Ives, and Eastman 1974 as an early example); a sequence of text-based valuation tasks could be asked, although sequences of table-based valuation tasks are by far more popular; the same set of attributes and the same experimental design could be employed in both text-based and table-based elicitation tasks; and the same econometric framework could be used to examine data collected through text and table formats. Yet there are few analyses of effects of information presentation on respondents' behavior, as elaborated on in the subsequent section. In addition, the studies that do focus on information presentation do not employ incentive compatible value elicitation questions, which could affect their results.

We conduct a laboratory, induced-value experiment that concentrates solely on the format of information presentation (i.e., text versus tables) in value elicitation tasks. This allows us to clearly identify the effect of the format on stated preferences within an incentive compatible setting. We implement two treatments which provide participants with the same set of value elicitation tasks and differ only in the way the information in the tasks is presented. To mirror "CV", one treatment employs a text format, and to mirror "CE", the other treatment employs a table format. Henceforth, when referring to the treatments, we use "CV" for the text format and "CE" for the table format.

Our experimental results reveal several important findings for SP methods. First, we observe that participants perform equally well in "CV" and "CE" elicitation tasks in terms of making payoff-maximizing choices based on the induced values. However, a significant difference between payoff-maximizing choices in "CV" and "CE" is

[2] We note that the SP literature provides some exceptions to the common way of information presentation in valuation tasks (see, for example, some "CV"-"CE" comparison studies reported in Table A.1 in Appendix A).

evident when only answers to the first elicitation task are considered. This finding is particularly important in light of recommendations for SP research (Johnston et al. 2017) that using only one task for eliciting preferences is preferred for incentive compatibility. Second, responding to "CV" elicitation tasks is observed to take substantially more time than responding to "CE" elicitation tasks. "CE" is also associated with a higher frequency of "clicking through" tasks that are quickly made and not payoff-maximizing, relative to "CV".

This paper contributes to the literature by identifying the important design and methodological differences between "CV" and "CE" and by providing a controlled assessment of the effect of the information presentation format on valuation responses in an incentive compatible environment. This assessment untangles the impact of the presentation format from the effects of the experimental design, use of sequences of value elicitation tasks, econometric models, and other aspects of valuation studies, which are often confounded in comparisons of "CV" versus "CE". Our contribution helps shed some light on the issue of whether there really is any difference between "CV" and "CE", or if other methodological factors generate differences in value estimates derived from these two approaches. The contribution of our empirical inquiry goes beyond a clear and controlled comparison of "CV" and "CE". Our study examines the impact of the information presentation format on people's behavior. Surprisingly, there has been little investigation into this issue. Because we examine the role of the information presentation format in a controlled setting, we are able to focus on the format question without concerns about incentive compatibility and other differences that often arise in field applications. While the laboratory setting is "unusual" and may not always translate to field settings, we feel this is an important first step in identifying the impact of the information presentation format on responses to valuation questions and on accuracy of the value elicitation. We continue with the paper by first examining the literature comparing "CV" versus "CE" and the literature investigating the effects of the information presentation format. We then outline our experimental approach, present the results, and discuss the findings.

2. Literature review

2.1 "CV" versus "CE" comparisons

We review the literature comparing "CV" and "CE" to help clarify how researchers have framed and investigated the differences between these two approaches. Table A.1 in Appendix A summarizes 26 "CV"-"CE" comparison studies in the field of environmental economics.[3] The emphasis in this literature is whether these two approaches provide similar value estimates. In other words, the literature examines the convergent validity of the approaches. The majority of studies find that "CE" produces larger value estimates than "CV" (Hanley, Wright, and Adamowicz 1998b; Stevens et al. 2000; Foster and Mourato 2003; Lehtonen et al. 2003; Mathews, Kask, and Stewart 2004; Travisi and Nijkamp 2004; Hasler et al. 2005; Madureira, Nunes, and Sanotos 2005; Christie and Azevedo 2009; Weber and Stewart 2009; Metcalfe et al. 2012). A substantial number of studies conclude that "CV" and "CE" lead to similar value estimates (Adamowicz et al. 1998; Lockwood and Carberry 1999; Abou-ali 2003; Colombo, Cavalatra-Requena, and Hanley 2006; Jin et al. 2006; Mogas, Riera, and Bennett 2006; Tuan and Navrud 2007; Adamowicz et al. 2011; McNair, Bennett, and Hensher 2011; Loomis and Santiago 2013; He, Dupras, and Poder 2017; Price, Dupont, and Adamowicz 2017). Only one study reports higher value estimates for "CV" than for "CE" (Boxall et al. 1996). The mixed evidence has only furthered the intellectual market for these types of comparisons, and has provided the starting point for the current analysis.

A closer inspection of the literature summarized in Table A.1 reveals that the "CV" and "CE" compared *within* a single study often differ along a number of dimensions. We identify seven key dimensions of these differences. First, the approaches vary in the number of options per value elicitation task. While "CV" almost always compares a status quo scenario to a single option, "CE" typically involves a status quo scenario plus two options.[4] Second, the number of valuation questions differs, with "CV" usually including only one question and "CE" asking four or more questions, sometimes even up to 16. Third, the framing of the valuation task is often

[3] We focus the literature review on environmental economics research and do recognize that there "CV" and "CE" comparisons in many other fields, particularly health economics.

[4] For open-ended CV questions, there really is only a single option, relative to the status quo, presented to respondents and they are asked to indicate the dollar amount they would pay (or receive) for that particular option.

different between the two approaches. “CV” surveys employ various framings such as a single referendum vote, a double-bounded dichotomous choice, a payment card, or an open-ended question. “CE” surveys most commonly phrase the valuation task as a choice of the preferred option from a provided set of options, where each option is related to a set cost. These differences are especially important in light of the recent emphasis on incentive compatibility of the SP survey design Carson and Groves 2007; Johnston et al. 2017; Vossler et al. 2012). Fourth, the approaches differ in the set of attributes used to describe the options. For example, in one of the first “CV”-“CE” comparisons, Boxall et al. (1996) provide 6 attributes in “CE” and only 2 attributes in “CV”.[5] Fifth, in studies that use between-subject designs, the sample sizes sometimes substantially vary. For example, Hoehn, Lupi, and Kaplowitz (2010) have three times as many respondents in the “CE” sample compared to the “CV” sample. Sixth, applications often employ different econometric tools to examine data obtained from “CV” and “CE”. For “CV”, the econometric models range from a logit or double-bounded logit model for dichotomous choice data to an interval regression model for payment card data; alternatively, summary statistics are simply reported, in particular for open-ended data. The “CE” data are commonly analyzed using conditional logit models, and some of the most recent comparisons employ random parameter logit models to incorporate unobserved heterogeneity in preferences. Seventh, information about the evaluated good is presented differently across the two approaches. “CV” typically uses continuous text descriptions, while “CE” provides information in tables. The multitude of differences makes it difficult to conduct a clear comparison of “CV” versus “CE” and gives rise to the question of which methodological factor(s) is (are) important in explaining evidenced differences in value estimates obtained from the two approaches.

Another important finding from the reviewed “CV”-“CE” comparison literature is that there is no consistent implementation of the “CV” and “CE” approaches. In fact, the application of these two approaches vary as much *across* comparisons as they do *within* comparisons. For example, some “CV” studies present multiple valuation questions (e.g., Christie and Azevedo 2009), some “CE” tasks include only two choice options (e.g., Jin et al. 2006), and some “CV” surveys use tables to present information

[5] The degree to which respondents perceived other attributes as “fixed” in the “CV” version of the survey is not clear.

(e.g., He et al. 2016). Numerous differences in "CV" and "CE" applications have given rise to many studies that test the influence of different design factors wholly within a "CV" or a "CE" approach. For example, investigations within "CV" generally focus on comparing various elicitation tasks such as open-ended, payment card, double-bounded dichotomous choice, and dichotomous choice. Within "CE", for example, Petrolia, Interis, and Hwang (2018) compare single versus repeated elicitation questions, while Meyerhoff, Oehlmann, and Weller (2015) use a Design of Designs approach to analyze the impact of systematically varying the numbers of valuation tasks, options per task, attributes, attributes' levels, and ranges of the attributes' levels on stated preferences. Viewed in this light, the "CV"-"CE" comparison literature is best interpreted as a subset of much larger literature examining various features of SP survey design.

To summarize, we see the following two major shortcomings of the "CV"-"CE" comparison literature to date: (1) the multitude of ways in which a typical "CV" and a typical "CE" differ across important design factors *within* a single comparison and (2) the lack of consistency in how the two approaches are implemented *across* comparisons. These shortcomings provide some rationale for why the literature has produced mixed evidence of convergent validity of "CV" and "CE". On an additional note, the use of "CV" and "CE" as broad descriptive terms is perhaps misguided because it masks rather than illuminates the many design differences seen in the implementation of the approaches.

As recently emphasized by Boyle (2017) in a non-market valuation practitioner's textbook, a "CV" question based on a single dichotomous choice is conceptually and analytically equivalent to a "CE" task that uses one elicitation question involving a choice between one option and a status quo. The difference remains in how the information of the good to be valued is presented. This is the starting point for the current work, where we compare text versus table presentation displays.

2.2 Information presentation format

Studies from outside the SP literature, in particular in the marketing literature, have examined the role of the information presentation format for respondents' choices in surveys. Bettman and Kakkar (1977) conduct interviews in a supermarket regarding potential choices of cereals. They find that decisions take substantially less time when the information about available options is presented as a table than as text. Bettman and Zins (1979) examine students' choices of food products in surveys employing three

disparate ways of information presentation: as a table, as text separately describing each option, and as text separately describing every attribute and how the options differ in each attribute. Their valuation task is designed to gauge the accuracy of choices for each type of the information display, by comparing the decisions made against a prescribed set of rules. They find that the shares of correct responses do not differ across the three types of information presentation, but decisions in the table format take little time compared to the other displays. Similarly, Schkade and Kleinmuntz (1994) report shorter response time for the information presented as a table than as text, when observing students' choices of possible loan applications. We note, however, that these studies have been conducted in non-incentive compatible (particularly, hypothetical) settings.

Studies by Hoehn et al. (2010) and Oviedo and Caparros (2015) are most closely related to our research as they both address the role of table and text formats of information presentation. Hoehn et al. (2010) conduct an online survey in which respondents select between a wetland scheduled to be drained and a restored wetland developed as compensation for the drained wetland. Similar to our experiment, the information about the choice options is presented either as continuous text or as tables, and the study involves a split-sample design to inquire the effects of the information presentation. They find a larger variance and greater use of heuristics in choices made in the text display than in the table display. However, the questions about the wetland preferences are not valuation tasks, and the larger variance in the treatment with the text display may be partially explained by a much smaller sample size in this treatment.

Oviedo and Caparros (2015) compare text and table formats of information presentation within two valuation studies of a reforestation project: one conducted in the field through personal interviews and the other done in a lab with an eye tracking software. Their main finding is that respondents pay more attention (measured in time) to attributes and the bid in a table display than in a text display. Besides the presentation format differences, their text format uses four double-bounded dichotomous choice questions, while their table format includes eight questions each with three options to be ranked. As noted above, the many differences between the "CV" and "CE" make it hard to disentangle the effect of the information presentation format on choices.

In contrast to Hoehn et al. (2010) and Oviedo and Caparros (2015), we conduct an incentivized induced-value laboratory experiment where the only variation between

experimental treatments is the information display. Our study also values a private good, while these two field investigations focus on public goods.

3. Experimental design

Our experiment is based on the induced-value laboratory experiment of Luchini and Watson (2014), where participants are asked to choose between tokens with differing monetary values as determined by attribute levels. Their design constitutes our treatment with the table-based presentation of information.[6] In addition, we include a treatment with the text-based presentation of information, which modifies only the information display in comparison to the table-based treatment. Namely each elicitation task in the text-based treatment is presented as continuous text instead of as a table.

To replicate the experiment of Luchini and Watson (2014), we use the same set of instructions and the same z-Tree code.[7] The only changes that we make are: replacing the currency GBP with CAD, tiny wording adjustments in the instructions, which do not affect their content, and making the font size larger in the elicitation tasks displayed in z-Tree. The full set of instructions is included in Appendix B. In the sections below, we describe induced values and choice sets, treatments, experimental procedures, and participants.

3.1 Induced values and choice sets

The experiment consists of nine rounds. Each round involves participants answering one elicitation task. The design of the tasks mimics value elicitation questions typically implemented in field valuation surveys. Every elicitation task includes a choice between two tokens, Token A and Token B, and an option "Neither Token". The tokens are described by three non-monetary attributes: size, color and shape, and by a monetary attribute, which is a cost of purchasing a given token. The "Neither Token" option cost a participant nothing.

The value of a token is determined by the token's non-monetary attributes. Small, medium, and large sizes are linked to the values of 0.5 CAD, 2.5 CAD, and 4 CAD, respectively. A red color adds 1 CAD to a token's overall value, yellow adds 1.5 CAD,

[6] Our treatment that uses the table format is identical in all respects to the "Wide-Monetary" treatment of Luchini and Watson (2014); that is, to their treatment that induces large differences in values between each attribute's levels and that makes participants' earnings from the experiment dependent on their choices.

[7] The experiment is computerized and conducted using the software z-Tree (Fischbacher, 2007).

and blue adds 2 CAD. A circle shape is valued at 1.5 CAD, a triangle shape at 3 CAD, and a square shape at 6 CAD. Finally, the cost of purchasing a token is subtracted from the token's value, and the cost can be 2 CAD, 3 CAD, or 4 CAD. Participants are informed about the values in written instructions. The values are the same for all participants and do not change throughout the experiment. In order to ease the understanding of the induced values, the instructions provide participants with an example of how to calculate a total value of a token.

The nine choice sets used in the experiment are based on a fractional factorial design, as defined by Luchini and Watson (2014). Table 1 shows payoffs from purchasing tokens in each choice set. A payoff represents a net value of a token as determined by its size, color, and shape, after subtracting its cost. The payoffs range from negative 0.5 to positive 9.5. The order in which the choice sets are displayed is randomized for each participant.

Table 1. Payoffs from tokens across choice sets

Choice set	Payoff from Token A	Payoff from Token B
A	5.5	6.5
B	2.5	9.5
C	3.5	8
D	-0.5	7
E	8	3
F	4.5	3
G	6	4
H	3	0.5
I	8	1

3.2 Treatments

Our experiment includes two treatments. In the Text treatment, the information about tokens for choice is presented as continuous text, which is the usual format for a valuation question in a "CV" survey. In the Table treatment, the information is presented as a table, which is typical for a valuation question in a "CE" survey. Figure 1 and Figure 2 show screenshots of an example elicitation task for the Text and Table treatments, respectively. The two treatments were identical with all respects, except for the information display.

Figure 1. A screenshot of an example elicitation task in the Text treatment

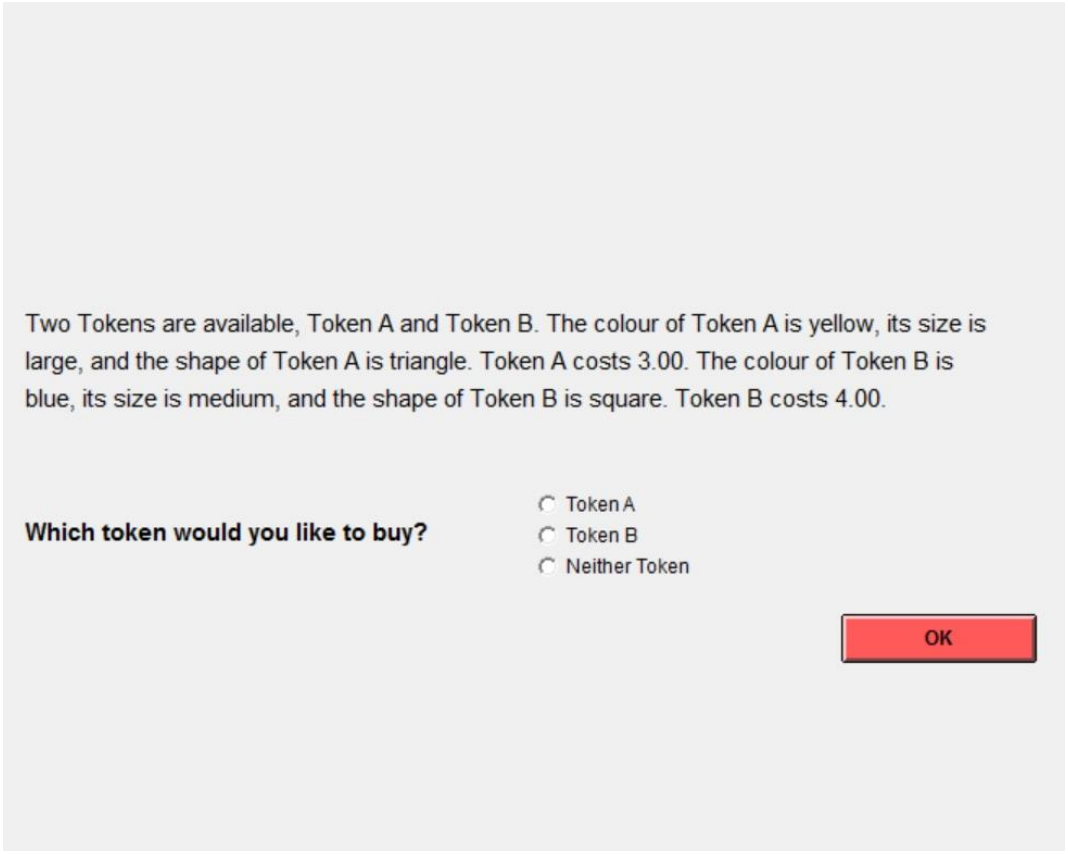
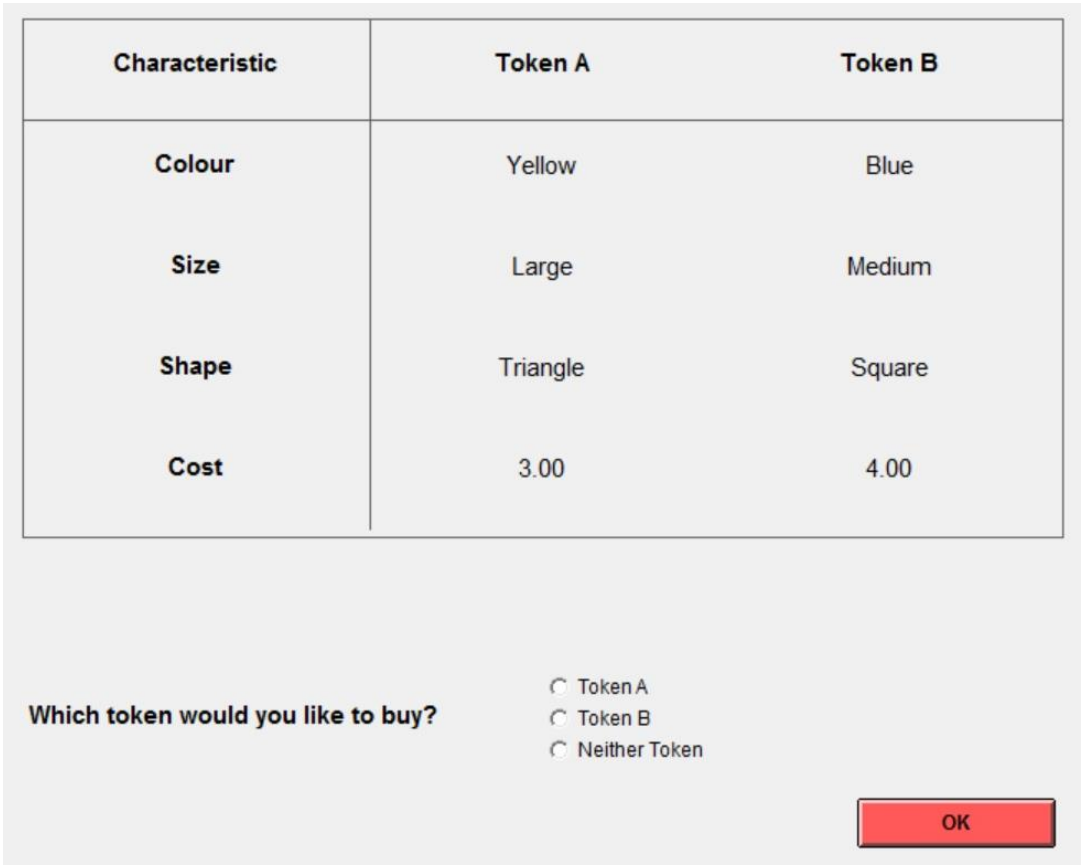


Figure 2. A screenshot of an example elicitation task in the Table treatment



3.3 Procedures

Before the experiment starts, participants receive a consent form, a written copy of the instructions, and a payment sheet. Participants sign the consent form, which is collected prior to the beginning of the experiment. Following this, the experimenter reads aloud the instructions and prompts for any clarifying questions. Participants are explicitly informed that their earnings will not be affected by the amount of time they take to make their choices, nor by choices of other participants. The instructions are identical for the two treatments.

After the instructions are read, participants make choices of tokens in a sequence of nine tasks (that is, nine experimental rounds). In every task, they select one of three options: Token A, Token B, or Neither Token. Participants have three minutes for making a choice in a task. Participants are told that after completing all nine elicitation tasks, one of the tasks will be randomly drawn and their choice from this question will determine their earnings from the experiment.

Each participant receives an initial balance of 4 CAD, which they can use to buy tokens offered in elicitation tasks. As mentioned in Section 3.1, the cost of a token ranges from 2 to 4 CAD, hence, participants can afford any of the displayed tokens. In addition to the payment determined by the participant's choice in the randomly chosen task, everybody is paid a fixed amount of 6 CAD for participation.

The experiment is followed by a short questionnaire. The questionnaire asks participants about that on what basis they selected tokens, whether they calculated monetary values of tokens, how difficult the experiment appeared to them, and whether they felt pressured by time. Finally, the questionnaire elicits basic socio-demographic information.

3.4 Participants

The experiment was conducted in a designated experimental laboratory at the University of Alberta, Canada. A total of 12 experimental sessions were organized in a three-day period from July 24 to July 26, 2017. In each session, half of the participants were randomly assigned to the Text treatment and half to the Table treatment. A single session lasted about 30 minutes. Participants earned 16.04 CAD on average (standard deviation of the earnings is equal to 2.62).

In total, 115 individuals participated in the experiment: 57 in the Text treatment and 58 in the Table treatment.[8] The participants were selected from a large pool of individuals voluntarily registered as potential participants in economic experiments at the online recruitment platform of the Department of Resource Economics and Environmental Sociology at the University of Alberta. The pool includes students, staff, and other campus community members who vary in their socio-demographic characteristics. Participants were not allowed to attend more than one session.

Table 2 presents the distribution of the experiment participants across the treatments according to their socio-demographic characteristics, along with the results of statistical tests of differences in these characteristics between the treatment samples. The samples do not differ significantly with respect to any of the characteristics. In the entire experimental sample, the average age is 29, a little more than half of the participants are females, more than half are students, and participants report that they have enough money to afford the leisure activities they like.

Table 2. Socio-demographic characteristics of the treatment samples

	Text treatment	**Table treatment**	**P-values for the null hypothesis of no difference**
Female	61%	52%	0.373
Age	29.77 (11.10)	27.91 (6.85)	0.730
Student	54%	69%	0.108
Enough money for leisure	61%	59%	0.761
Number of participants	57	58	

Notes: The p-values are for the null hypothesis of no difference between the treatment samples with respect to a given socio-demographic characteristic. Except for *Age*, the table shows the shares of participants within each treatment sample, and for these variables, chi-squared tests of equality of proportions are conducted. For *Age*, means (and standard deviations in the brackets) are reported, and the Wilcoxon signed-rank test is used for verifying the difference.

4. Results

To examine whether the information display affects respondents' behavior in preference elicitation tasks, we compare the behavior of the experiment participants across the Text and Table treatments. We assess the differences with respect to two outcomes: (1) payoff-maximizing choices and (2) time used for making a choice.

Before we proceed, we check whether the treatment samples differ in responses to the follow-up questions about participants' behavior and perception of the

[8] Given the nine elicitation tasks, this yields 1,035 observations in total. The analysis in the next section is based, however, on 1,032 observations because three participants did not give any answer in an elicitation task within the allotted time.

experiment, which are assessed in the post-experiment questionnaire. We do so because in the analysis that follows, we sometimes split participants according to their responses in the follow-up questions. Participants' answers to these questions are reported in Table A.2 in Appendix A, together with results of statistical tests of differences in distributions of the responses across the two treatments. For none of the questions do we find a statistically significant difference between treatments. The majority of participants always / precisely calculate payoffs related to purchasing tokens; perceive making choices of tokens as easy or very easy; and have enough time to make their choices.

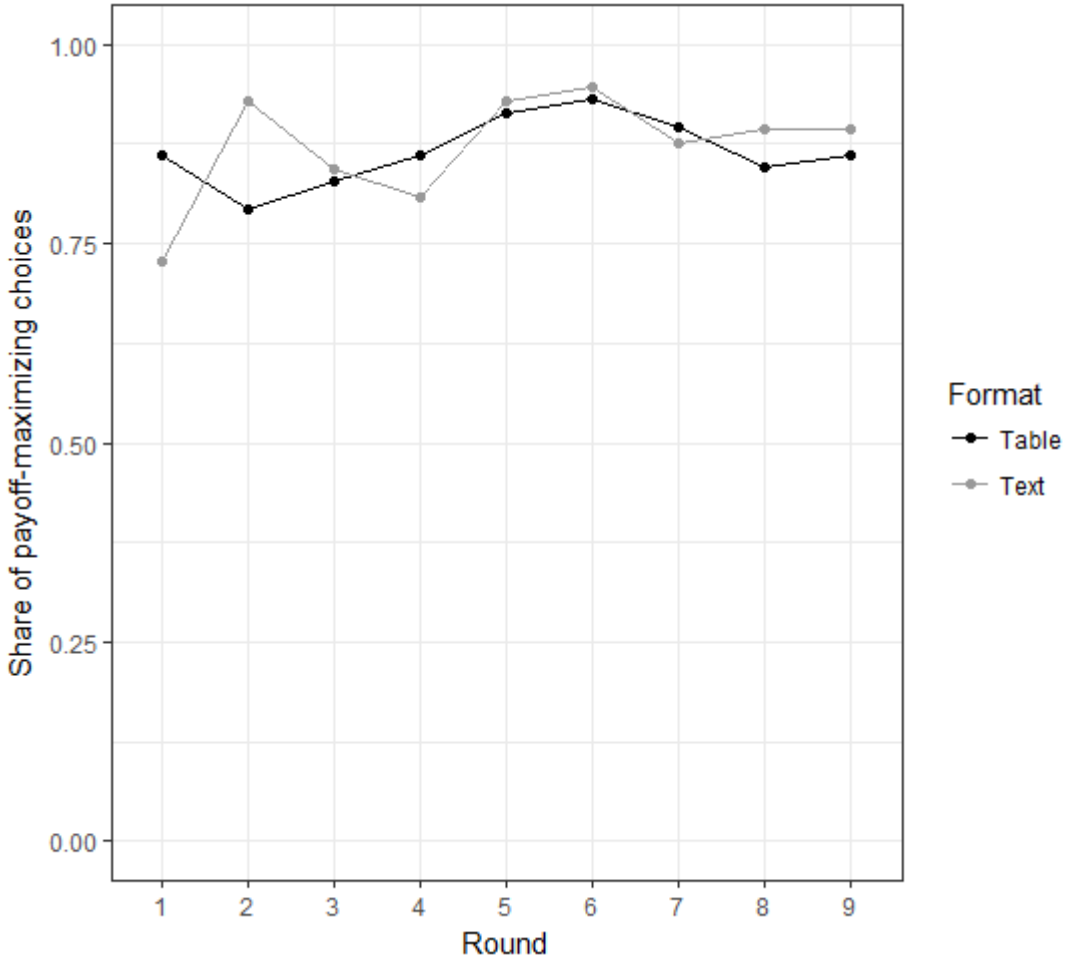
4.1 Payoff-maximizing choices

The first outcome of our examination of differences between "CV" and "CE" focuses on whether participants in the two treatments perform similarly in terms of making payoff-maximizing choices, that is, in selecting tokens with the highest net value. We find that the percentage shares of payoff-maximizing choices in all choices are almost identical across the treatments: 87.3% of choices in the Text treatment and 86.6% in the Table treatment. We also do not observe any substantial differences in these shares across choice sets, as shown in Table A.3 in Appendix A. The shares are similar across the treatments and fall into the interval from 74.1% to 94.8%. This evidence suggests that none of the choice sets is particularly difficult for participants to answer.[9]

As mentioned, the experiment consists of nine rounds, each with one elicitation task selected from the predefined choice sets. Because the order in which the choice sets are presented across rounds is randomized for each participant, we examine whether the percentage shares of payoff-maximizing choices change across rounds. This is illustrated in Figure 3. The figure does not reveal any trend in the shares of payoff-maximizing choices as participants move from a task to a task in the sequence.

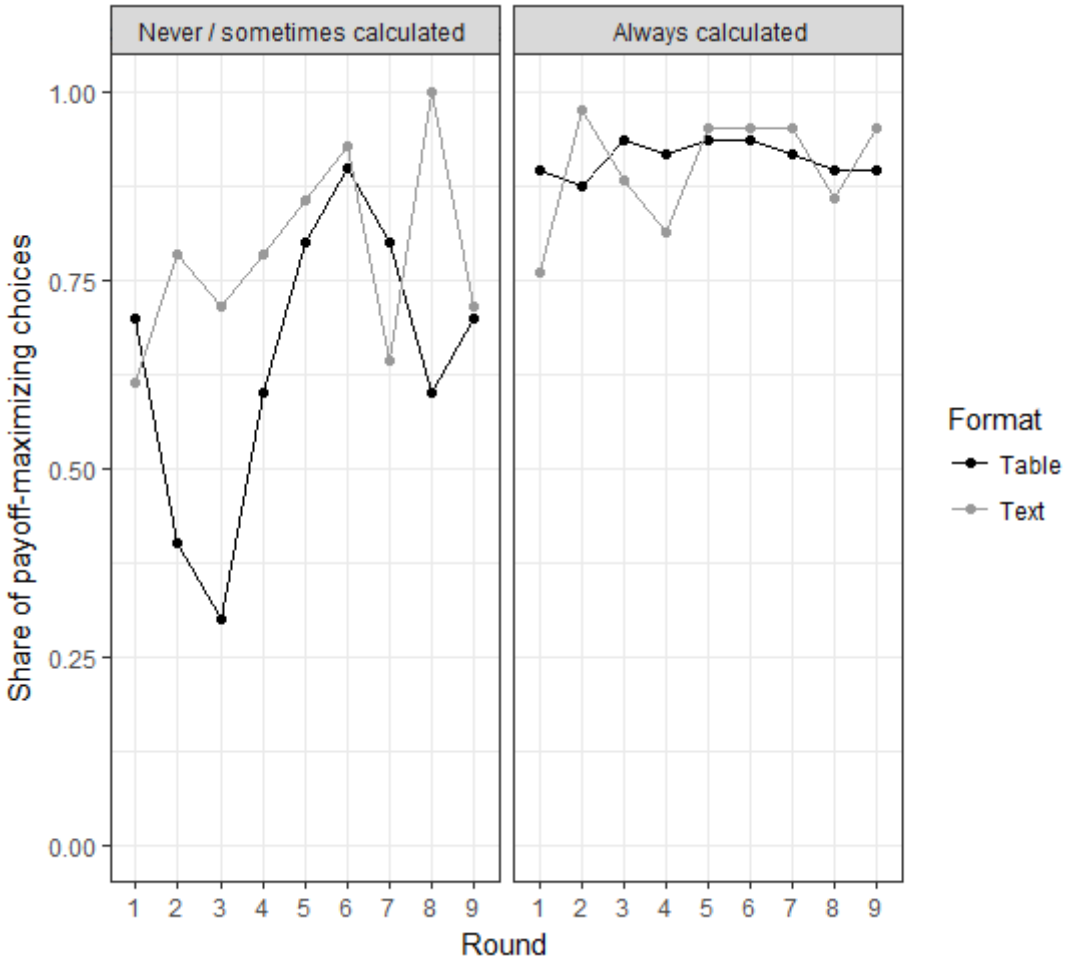
[9] Interestingly, our results diverge to some extent from the results reported by Luchini and Watson (2014). The total share of payoff-maximizing choices in their study is equal to 59.9%, and their experiment involves a sample of a similar size to the one we use, namely they have 54 experiment participants. Based on Table A.3 in Appendix A, the most pronounced differences in the shares across their study and ours are for choice sets A, B, and C.

Figure 3. Percentage shares of payoff-maximizing choices across rounds



When we split the participants according to whether they always / precisely calculate monetary values of tokens, or not, we observe some differences in how the shares of payoff-maximizing choices evolve over rounds. Figure 4 shows that for those who always / precisely calculate monetary values, the shares remain nearly constant throughout all elicitation tasks in the Table treatment, while there is some variability in the shares in the Text treatment. This may indicate more difficulty in answering the tasks in the text format.

Figure 4. Percentage shares of payoff-maximizing choices across rounds for those who always / precisely calculate monetary values of tokens, and for those who do not



We next investigate factors that affect the probability of making a payoff-maximizing choice. To examine this issue, we estimate a random-effects logit model, in which the dependent variable is binary-coded: equal to 1 for a payoff-maximizing choice and 0 otherwise. The results are displayed in Table 3. Most importantly in the context of our research question, the probability of making a payoff-maximizing choice does not differ significantly between the Text and Table treatments. This implies that how the information is displayed in an elicitation task – in a text or table format – does not affect choices made by participants. Other factors included in the model impinge on the probability of making a payoff-maximizing choice. The probability decreases for participants who make their choices quickly (in up to 20 seconds) and for those who partake in a morning session. The probability increases as a participant moves through the sequence of elicitation tasks, which may be related to a learning effect, and as the difference in payoffs from tokens gets larger.

Table 3. A random-effects logit model of the probability of making a payoff-maximizing choice of a token

Explanatory variables	**Means (Standard errors)**
Text treatment	-0.131 (0.358)
Responded in up to 20 seconds	-1.134*** (0.342)
Round	0.115*** (0.043)
Absolute difference in the payoffs from tokens in the elicitation task	0.256*** (0.049)
Morning session	-0.617* (0.356)
Constant	1.548*** (0.404)
Log-likelihood	-335.5
Log-likelihood with a constant only	-359.2
Number of observations	1,032

Notes: *** and * denote 1% and 10% significance levels, respectively.

Recent valuation literature places strong emphasis on making SP surveys incentive compatible (Johnston et al. 2017). In order to assure incentive compatibility, the literature suggests employing only one value elicitation question or task.[10] We, therefore, verify whether the information display plays a role when preferences are disclosed in a single task. We estimate a logit model using data only from the first elicitation task, with the same dependent variable as in the model above. The results are presented in Table 4. The treatment variable is a significant predictor of the probability of making a payoff-maximizing choice. It shows that, on average, participants are more likely to make a payoff-maximizing choice in the first task when the task has a table format than when it has a text format. This result may have important implications for actual applications of SP surveys. It suggests that in the surveys with one valuation task or in the first value elicitation task in a sequence, respondents may be more likely to make choices in line with their preferences when it is "CE" than when it is "CV". The table format may foster understanding and make identification of the preferred option easier. The effects of the remaining variables included in the model in Table 4 are consistent with the results from the model based on the full sample discussed above.

[10] To make a sequence of value elicitation questions incentive compatible, one needs to assure independence between the questions, which is often difficult to be reliably implemented in field valuation surveys (Vossler et al. 2012).

Table 4. A logit model of the probability of making a payoff-maximizing choice of a token in the first elicitation task

Explanatory variables	**Means (Standard errors)**
Text treatment	-1.322**
	(0.557)
Responded in up to 20 seconds	-2.135***
	(0.802)
Absolute difference in the payoffs from tokens in the elicitation task	0.215*
	(0.111)
Constant	1.456***
	(0.559)
Log-likelihood	-49.8
Log-likelihood with a constant only	-57.1
Number of observations	113

Notes: ***, **, and * denote 1%, 5%, and 10% significance levels, respectively.

4.2 Time used for making a choice

The second outcome of differences between "CV" and "CE" concerns the time used by participants to take decisions. The treatments differ with this respect: making a choice takes on average 48.2 seconds in the Text treatment and 36.5 seconds in the Table treatment. A graphical representation of the distributions of the choice-making time in the two treatments, shown in Figure 5, indicates a large similarity between the distributions, with the main difference being the Text distribution is shifted to the right compared to the Table distribution.

The table format is not only associated with quicker choices, but also with behavior consistent with "clicking through" elicitation tasks. Among the participants who self-report never or sometimes calculating monetary values of tokens, those in the Table treatment make their choices much faster and less often select payoff-maximizing tokens than those in the Text treatment. For these participants, it takes on average 19 seconds in the Table treatment and 49 seconds in the Text treatment to choose a token; and 64% of them in Table and 78% in Text make payoff-maximizing choices. These results suggest that participants who never or sometimes calculate monetary values rush through the tasks in the Table treatment, while they still devote substantial time to figure out payoff-maximizing tokens in the Text treatment. This finding could be related to a claim of Hoehn et al. (2010) that a table format may sometimes oversimplify the scenario. The oversimplification is likely to encourage more participants to rush when information is displayed as a table than when it is displayed as text.

Figure 5. Distributions of the choice-making time for the Text and Table treatments

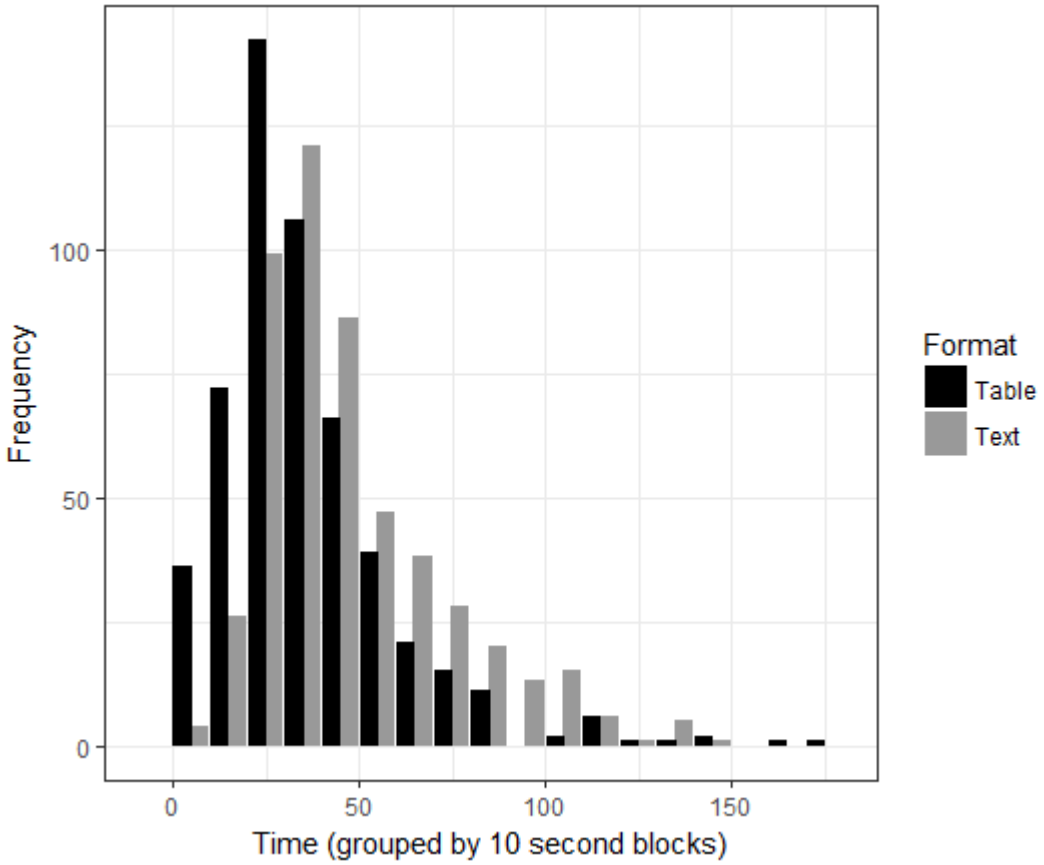
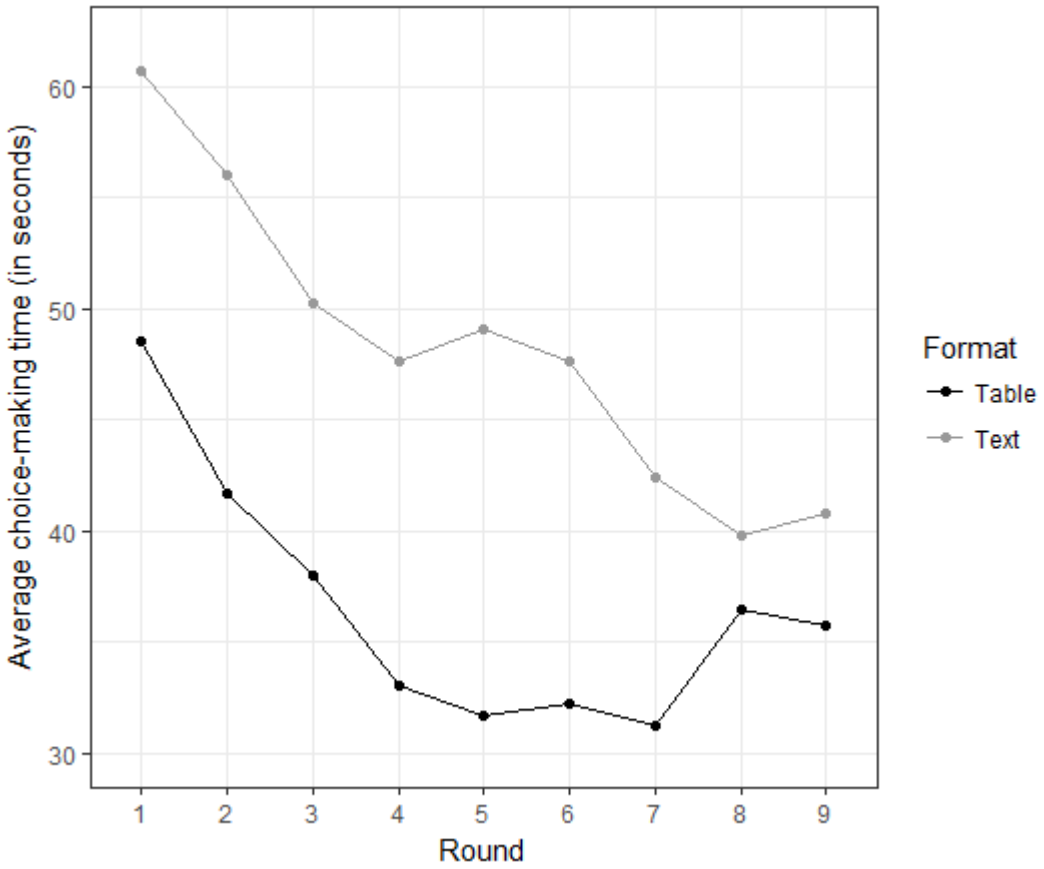


Figure 6 illustrates that the average choice-making time decreases as participants progress through the sequence of elicitation tasks. For both the Table and Text treatments, the average choice-making time rapidly decreases until the fourth round and then generally plateaus. In all rounds, the average choice-making time is higher in the Text treatment compared to the Table treatment. For the Table treatment, the relationship, however, looks non-linear, with a small increase in the average choice-making time in the two last tasks. The non-linear relationship can arise from the way the experiment is programmed: participants need to wait until all other participants in their treatment select a token before the next task is displayed. Observing that making a quick choice does not allow for proceeding faster through the tasks, participants could slow down. For all rounds, the choice-making time in the Text treatment is considerably higher, so it may cause the experience effect not to appear for this treatment.

To formally examine factors that affect the choice-making time, we estimate a random-effects linear model with the dependent variable being the number of seconds a participant uses for selecting a token. Every explanatory variable in the model is interacted with a binary-coded variable indicating the Text treatment to make possible

the identification of effects specific to the treatments. Consequently, the means can be interpreted as the effects for the Table treatment, and the interactions with the Text treatment show if the effects for Text are significantly different from the effects for Table. The results are summarized in Table 5.

Figure 6. Average choice-making time across rounds



The estimation results reported in Table 5 confirm the previously noted observations. A significant and positive coefficient of the interaction of the constant with the Text treatment demonstrates on average longer choice-making time in Text than in Table. The effect of a round on the choice-making time is non-linear in the Table treatment. Specifically, the model estimates suggest that the choice-making time decreases until about the 5th round and then starts to increase. Summing up the coefficients of the mean effect and the interaction with Text for the variable *Round squared* yields a value of 0.074, which is not statistically significantly different from zero. This means that the non-linear effect of a round on the choice-making time is not present in the Text treatment. Instead, the choice-making time in the Text treatment appears to decrease over rounds, as suggested by a negative value of the sum of the coefficients of the mean effect of the interaction with Text for the variable *Round*.

Calculating monetary values appears to considerably increase the time for making a choice in the Table format, while the effect is not found to be statistically significant for the Text format. A large absolute difference in the payoffs from the two tokens fosters quicker choice making in both treatments. The model also includes a measure of time pressure, which we approximate by the longest choice-making time observed in a given elicitation task in a session. Participants in sessions that take longer are not expected to experience as much time pressure. The results indicate that lack of or weak time pressure indeed invites participants to slow down their pace of selecting tokens, and the effect does not differ across the treatments.

Table 5. A random-effects linear model of time used for making a choice

Explanatory variables	**Means (Standard errors)**	**Interactions with Text (Standard errors)**
Constant	26.401***	22.331**
	(7.015)	(9.681)
Round	-6.255***	3.819**
	(1.338)	(1.918)
Round squared	0.556***	-0.482***
	(0.128)	(0.182)
Always / precisely calculated monetary values	20.692***	-21.564***
	(5.449)	(7.278)
Absolute difference in the payoffs from tokens in the elicitation task	-1.285***	-0.076
	(0.305)	(0.435)
Longest choice-making time in the task (No / weak time pressure)	0.139***	0.044
	(0.032)	(0.046)
Log-likelihood	-4,485.5	
Log-likelihood with a constant only	-4,597.0	
AIC	8,999.0	
BIC	9,068.1	
Number of observations	1,032	

Notes: ***, **, and * denote 1%, 5%, and 10% significance levels, respectively.

5. Conclusions

Convergent validity of contingent valuation ("CV") and choice experiments ("CE") has been examined in a large body of stated preference (SP) literature. However, the two terms used for defining different approaches within SP research have been interpreted in a variety of ways. We review the literature empirically addressing the issue of convergent validity of "CV" and "CE", and identify seven key differences in how the two approaches have been applied in practice, which include differences in design characteristics and methods of data analysis. Most "CV"-"CE" comparison studies vary across a manifold of these methodological features which makes it difficult to assess

the reasons driving the mixed evidence of convergent validity of "CV" and "CE". The use of "CV" and "CE" as broad descriptive terms is perhaps misguided because it masks rather than illuminates the many design differences seen in the implementation of the approaches. We argue researchers should shift their focus to examining impacts of specific design characteristics and methods of data analysis rather than blanket "CV" versus "CE" comparisons.

Based on our literature review, we contend that a key difference between "CV" and "CE" is the format for displaying information in valuation questions. "CV" typically presents information as text, while "CE" usually uses tables for this purpose. In other words, we do not view any of the other methodological features related to implementations of "CV" and "CE" (such as a number of value elicitation questions or tasks, a number of choice options per question, a number of attributes describing options) to be specific only to one of the two approaches. Building upon this observation and noting that the differentiation in information display has not been isolated in previous comparisons of "CV" and "CE", we empirically examine the role of the format of information presentation for respondents' behavior in valuation questions. Aside from the application to SP research, our study provides general evidence on how text-based and table-based displays of information affect people's choices.

To empirically address the problem of the role of the information presentation format, we design and implement an induced-value experiment. The laboratory setting allows us to ensure incentive compatibility of the value elicitation mechanism, which is difficult to be credibly implemented in a field context. Our main experimental findings can be summarized in the three following points. First, the vast majority of participants are observed to make payoff-maximizing choices, consistent with the induced values, and overall, the chance of selecting a payoff-maximizing option is not affected by the way information is displayed. Second, when focusing on participants' choices in the first valuation task, their behavior appears to be influenced by the format of information presentation. Specifically, selecting a payoff-maximizing option is less likely in the text display than in the table display. Third, deciding for an option is found to take substantially more time in the text format than in the table format.

These findings have important, yet mixed, implications for current practices in SP research. On one hand, the lack of difference in payoff-maximizing choices between text and table displays is positive for the generalizability of information presentation

formats and points to convergent validity of "CV" and "CE" – or, more specifically, to convergent validity of values derived from text-based and table-based tasks. On the other hand, the significant difference evident in choices made in the first elicitation task could signal difficulties with valuation that employs a single task only. This later finding is of particular importance given the current recommendation for SP research which suggests using a single value elicitation question to avoid strategic responses and, hence, to assure incentive compatibility (Johnston et al. 2017). Combined with our empirical evidence, this opens an interesting question about the recommended display in a single valuation task, so that it would help obtain a true picture of public's preferences. Although our results indicate an advantage of using a table format of information presentation over a text format for the first task, we emphasize that this finding needs verification through further analyses.

Our results correspond to previous findings related to the information presentation format, mainly derived from studies in marketing literature. Differences in information displays could be expected to lead to divergent behavior of people, because the way information is presented is claimed to influence the information processing (Bettman and Kakkar 1977; Shi et al. 2013). Some studies distinguish two separate stages of information processing which include information acquisition and information evaluation. They suggest that the information display affects to a larger degree the former rather than the latter (Schkade and Kleinmuntz 1994), and that people may accommodate differences in information presentation at the stage of information acquisition through adjusting time needed for the information analysis (Bettman and Zins 1979). This is exactly what we observe: the payoff-maximizing behavior does not differ across the information presentation formats, but participants adjust time they use for taking decisions and devote substantially more time in the text format where the information about characteristics of the evaluated good is displayed less explicitly. The short choice-making time in the table-based presentation may be related to the claim of Tversky (1969) that comparisons by attributes, as encouraged through the table-based presentation, are easier. Finally, following Ettlin and Bröder (2015), we can say that our manipulations in the display between the two formats do not induce note-worthy changes in costs of information processing and, thus, the payoff-maximizing behavior is not affected.

There are several possible extensions to our assessment of the role of the information presentation, which mainly involve relaxing the experimental laboratory

conditions to make the setting reflect closer the context of actual field valuation studies. One of such basic characteristic of our study that is in particular worth modifying is the good being evaluated. Instead of an abstract good defined by induced values, further research may consider valuing a real good. Most likely, this modification would as well increase the complexity of the valuation problem, bringing it closer to the goods evaluated in field surveys. Subsequent research could also include a good which would be less familiar for respondents to value. Given that the processes of value formation and information acquisition are likely related, we hypothesize that limited familiarity with the good being evaluated may translate into differences in the impact of information presentation on people's choices. Finally, SP methods are largely used to evaluate public goods. While our experiment focuses on a private good, an extension can use a public-good context.

Another angle of extensions to our study concerns investigating interaction effects between responses to text-based versus table-based formats of valuation questions and other aspects of SP design, such as a number of attributes describing the options, a number of attributes' levels, a survey mode (e.g., in-person versus online). Hoehn et al. (2010) discuss the role of the text versus table presentation of information in the context of complexity of the value elicitation, and note that the table display helps summarize information, reducing at the same time the complexity. They also mention, however, that the table display can possibly lead even to oversimplification of the valuation problem. Our finding of a higher frequency of "clicking through" the tasks presented in tables may provide some support to this hypothesis, although a broader investigation of motivation behind the participants' behavior is needed.

The choice of the format of presenting information in valuation questions, between text and tables, or combinations of them, is fundamental to SP research. While the recent SP literature largely uses the so-called "CE" format, in which information is typically displayed as tables, it is noteworthy that the analysis of the Deepwater Horizon oil spill (Bishop et al. 2017) employed text-based presentation of information in much of the questionnaire, even though the valuation study involved two scenarios that differed in terms of attributes and their levels. New insights into the impact of information presentation would aid SP researchers in designing valuation studies and may provide additional evidence crucial for assessing validity of SP methods.

References

Abou-Ali H (2003) Using stated preference methods to evaluate the impact of water on health: the case of metropolitan Cairo. Department of Economics Working Paper no. 113, University of Gothenberg. Available at https://gupea.ub.gu.se/bitstream/2077/2823/1/gunwpe0113.pdf. Cited 15 Mar 2018

Adamowicz WL, Boxall P, Williams M, Louviere J (1998) Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics* 80(1):64-75

Adamowicz WL, Dupont D, Krupnick A, Zhang J (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

Bettman JR, Kakkar P (1977) Effects of information presentation format on consumer information acquisition strategies. *Journal of Consumer Research* 3(4):233-240

Bettman JR, Zins MA (1979) Information format and choice task effects in decision making. *Journal of Consumer Research* 6(2):141-153

Bishop RC, Boyle KJ, Carson RT, Chapman D, Hanemann WM, Kanninen B, Kopp RJ, Krosnick JA, List J, Meade N, Paterson R, Presser S, Smith VK, Tourangeau R, Welsh M, Wooldridge JM, DeBell M, Donovan C, Konopka M, Scherer N (2017) Putting a value on injuries to natural assets: The BP oil spill. Science 356(6335):253-254

Boxall PC, Adamowicz WL, Swait J, Williams M, Louviere JJ (1996) A comparison of stated preference methods for environmental valuation. *Ecological Economics* 18(3):243-253

Boyle KJ (2017) Contingent valuation in practice. In: Champ PA, Boyle KJ, Brown TC (eds) *A Primer on Nonmarket Valuation*. Springer, Dordrecht, The Netherlands, pp 83-132

Cameron TA, Poe GL, Ethier RG, Schulze WD (2002) Alternative non-market value-elicitation methods: Are the underlying preferences the same*? Journal of Environmental Economics and Management* 44(3):391-425

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

Carson RT, Groves T (2011) Incentive and information properties of preference questions: Commentary and extensions. In: Bennett J (ed) *International Handbook of Non-Market Environmental Valuation*. Edward Elgar, Northampton, MA, pp 300-321

Carson RT, Louviere JJ (2011) A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics* 49(4):539-559

Champ PA, Boyle KJ, Brown TC (eds) (2017) *A Primer on Nonmarket Valuation*. Springer, Dordrecht, The Netherlands

Christie M, Azevedo CD (2009) Testing the consistency between standard contingent valuation, repeated contingent valuation and choice experiments. *Journal of Agricultural Economics* 60(1):154-170

Christie M, Hanley N, Warren J, Murphy K, Wright R, Hyde T (2006) Valuing the diversity of biodiversity. *Ecological Economics* 58(2):304-317

Colombo S, Calatrava-Requena J, Hanley N (2006) Analysing the social benefits of soil conservation measures using stated preference methods. *Ecological Economics* 58(4):850-861

Ettlin F, Bröder A (2015) Perceptual grouping does not affect multi-attribute decision making if no processing costs are involved. *Acta Psychologica* 157:30-43

Farquharson R (1969) *Theory of Voting*. Yale University Press, New Haven, CT

Fischbacher U (2007) Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2):171-178

Foster V, Mourato S (2003) Elicitation format and sensitivity to scope. *Environmental and Resource Economics* 24(2):141-160

Hanley N, MacMillan D, Wright RE, Bullock C, Simpson I, Parsisson D, Crabtree B (1998a) Contingent valuation versus choice experiments: Estimating the benefits of environmentally sensitive areas in Scotland. *Journal of Agricultural Economics* 49(1):1-15

Hanley N, Wright R, Adamowicz WL (1998b) Using choice experiments to value the environment. *Environmental and Research Economics* 11(3):413-428

Hasler B, Lundhede T, Martinsen L, Neye S, Schou JS (2005) Valuation of groundwater protection versus water treatment in Denmark by choice experiments and contingent valuation. National Environmental Research Institute Technical Report no. 543. Available at http://www2.dmu.dk/1_viden/2_Publikationer/3_fagrapporter/rapporter/FR543.PDF. Cited 15 Mar 2018

He J, Dupras J, Poder TG (2017) The value of wetlands in Quebec: A comparison between contingent valuation and choice experiment. *Journal of Environmental Economics and Policy* 6(1):51-78

Hoehn JP, Lupi F, Kaplowitz MD (2010) Stated choice experiments with complex ecosystem changes: The effect of information formats on estimated variances and choice parameters. *Journal of Agricultural and Resource Economics* 35(3):568-590

Jin J, Wang Z, Ran S (2006) Comparison of contingent valuation and choice experiment in solid waste management programs in Macao. *Ecological Economics* 57(3):430-441

Johnston RJ, Boyle KJ, Adamowicz WL, Bennett J, Brouwer R, Cameron TA, Hanemann WM, Hanley N, Ryan M, Scarpa R, Tourangeau R, Vossler CA (2017) Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists* 4(2):319-405

Lehtonen E, Kuuluvainen J, Pouta E, Rekola M, Li C-Z (2003) Non-market benefits of forest conservation in southern Finland. *Environmental Science and Policy* 6(3):195-204

Lockwood M, Carberry D (1999) Stated preference surveys of remnant native vegetation conservation. Paper presented at the 43rd Annual AARES Conference and the 6th Annual NZARES Conference, Christchurch Convention Centre, Christchurch, New Zealand, 20-22 January

Loomis J, Santiago L (2013) Economic valuation of beach quality improvements: Comparing incremental attribute values estimated from two stated preference valuation methods. *Coastal Management* 41(1):75-86

Luchini S, Watson V (2014) Are choice experiments reliable? Evidence from the lab. *Economics Letters* 124(1):9-13

Madureira L, Nunes LC, Sanotos JML (2005) Valuing multi-attribute environmental changes: Contingent valuation and choice experiments. Paper presented at the European Association of Environmental and Resource Economists 14th Annual Conference, International University Bremen, Bremen, Germany, 23-26 June

Mathews LG, Kask SB, Stewart S (2004) The value of the view: Valuing scenic quality using choice and contingent valuation models. Paper presented at the American Agricultural Economics Association Annual Meeting, Denver, CO, 1-4 August

McNair BJ, Bennett J, Hensher DA (2011) A comparison of responses to single and repeated discrete choice questions. *Resource and Energy Economics* 33(3):554-571

Metcalfe PJ, Baker W, Andrews K, Atkinson G, Bateman IJ, Butler S, Carson RT, East J, Gueron Y, Sheldon R, Train K (2012) An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales. *Water Resources Research*, 48(3). DOI: 10.1029/2010WR009592

Meyerhoff J, Oehlmann M, Weller P (2015) The influence of design dimensions on stated choices in an environmental context. *Environmental and Resource Economics* 61(3):385-407

Mogas J, Riera P, Bennett J (2006) A comparison of contingent valuation and choice modelling with second-order interactions. *Journal of Forest Economics* 12(1):5-30

Oviedo JL, Caparros A (2015) Information and visual attention in contingent valuation and choice modelling: Field and eye-tracking experiments applied to reforestations in Spain. *Journal of Forest Economics* 21(4):185-204

Petrolia DR, Interis MG, Hwang J (2018) Single-choice, repeated-choice, and best-worst scaling elicitation formats: Do results differ and by how much? *Environmental and Resource Economics* 69(2):365-393

Price J, Dupont D, Adamowicz WL (2017) As time goes by: Examination of temporal stability across stated preference question formats. *Environmental and Resource Economics* 68(3):643-662

Randall A, Ives B, Eastman C (1974) Bidding games for valuation of aesthetic environmental improvements. *Journal of Environmental Economics and Management* 1(2):132-149

Schkade DA, Kleinmuntz DN (1994) Information displays and choice processes: Differential effects of organization, form, and sequence. *Organizational Behavior and Human Decision Processes* 57(3):319-337

Shi SW, Wedel M, Pieters FGM (2013) Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science* 59(5):1009-1026

Stevens TH, Belkner R, Dennis D, Kittredge D, Willis C (2000) Comparison of contingent valuation and conjoint analysis in ecosystem management. *Ecological Economics* 32(1):63-74

Travisi CM, Nijkamp P (2004) Willingness to pay for agricultural environmental safety: Evidence from a survey of Milan, Italy, residents. Working Paper "Nota di Lavoro" Series no. 2004.100, Fondazione Eni Enrico Mattei. Available at https://www.feem.it/m/publications_pages/NDL2004-100.pdf. Cited 15 Mar 2018

Tuan TH, Navrud S (2007) Valuing cultural heritage in developing countries: Comparing and pooling contingent valuation and choice modelling estimates. *Environmental and Resource Economics* 38(1):51-69

Tversky A (1969) Intransitivity of preferences. *Psychological Review* 76(1):31-48

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

Weber MA, Stewart S (2009) Public values for river restoration options on the middle Rio Grande. *Restoration Ecology* 17(6):762-771

Appendix A

Table A.1. Summary of environmental valuation studies comparing "CV" and "CE"

Authors	Subject design		Valuation question framing	Number of questions / options / attributes	Sample size	Econometric model	Presentation format	Convergent validity conclusion
Abou-ali (2003)	Between	CV	DBDC with OE follow-up	3/2/3	732	Spike model	Text	CV=CE
		CE	NR	4/3/3	757	CL	Table	
Adamowicz et al. (1998)	Within	CV	Referendum vote	1/2/5	402	Logit	Text	CV=CE
		CE	Choose option	8/3/5	355	CL	Table	
Adamowicz et al. (2011)	Between	CV	DBDC	3/2/5	407	Survival model	Table/graphic	CV=CE
		CE	Choose preferred option	4/2/5 or 4/3/5	406	CL, RPL, LC	Table	
Boxall et al. (1996)	Within	CV	DC	1/2/2	271	Logit	Text	CV>CE
		CE	Choose option	16/3/6	266	CL	Table	
Cameron et al. (2002)	Between and within	CV	DC, OE, PC, polychotomous choice	1/2/2	NR	Logit, Tobit, Interval regression, Ordered logit	Text	CV=CE
		CE	NR	5/3/3	303	CL	Table	
Christie and Azevedo (2009)	Between	CV	DC	3/2/5	376	Logit	Text	CV=CE
		CE	Choose option	8/3/5	231	CL	Table	
Christie et al. (2006)	Within	CV	PC	1/2/2	741	NR	Text	Not directly compared
		CE	Choose preferred option	5/3/5	741	CL	Table	
Colombo et al. (2006)	Within	CV	OE	1/1/6	255	N/A	Text	CV=CE
		CE	Choose situation	4/3/6	201	CL, RPL	Table	

Foster and Mourato (2003)	Between	CV	DBDC	2/2/2 or 2/2/4	282	DB logit	NR	CV<CE
		CE	Choose preferred option	3/3/4	234	CL, RPL	Table	
Hanley et al. (1998a)	Between	CV	OE	1/1/3	809	N/A	Text	CV<CE
		CE	Choose option	4/3/3	256	CL	Table	
Hasler et al. (2005)	Between	CV	PC	2/2/2	663	Tobit	Text	CV<CE
		CE	Choose preferred option	6/3/3	543	CL	Table	
He et al. (2017)	Between	CV	Referendum vote	1/2/5	838	Logit	Table	CV=CE
		CE	Choose preferred program	5/3/5	858	CL, RPL	Table	
Jin et al (2006)	Between	CV	DBDC	2/2/4	252	CL, RPL	NR	CV=CE
		CE	Choose preferred option	8/2/4	241	CL	Table	
Lehtonen et al. (2003)	Between	CV	Choose preferred option	1/2/6	426-596	Logit	Table	CV<CE
		CE	Choose preferred option	8/3/6	602	NL	Table	
Lockwood and Carberry (1999)	Between	CV	DC donation	1/2/4	340-351	Logit	Text	CV=CE
		CE	Choose option	8/3/4	340-388	CL	Table	
Loomis and Santiago (2013)	Between	CV	DC	3/2/5	214	Logit	Table	CV=CE
		CE	Would you pay increased travel costs	3/2/4	213	Logit	Table	
Madureira et al. (2005)	Between	CV	DC	5/2/3	177	Logit	NR	CV<CE
		CE	Choose preferred option	5/3/3	210	CL, NL	Table	
Mathews et al. (2004)	Between and within	CV	DC with OE follow-up	1/2/2	352	Logit	Text	CV<CE
		CE	Choose option	9/3/6	152	CL	Table	
McNair et al. (2011)	Between	CV	Choose preferred option	1/2/5	1163	Logit	Table	CV=CE
		CE	Choose preferred option	4/2/5	292	Logit, RPL	Table	

Metcalfe et al. (2012)	Within	CV	DC and PC	2/2/6	1,389	Interval regression	Table	CV<CE
		CE	Choose option	7/3/6	1,389	CL	Table	
Mogas et al.(2006)	Within	CV	DC	1/2/6	1,000	Logit	NR	CV=CE
		CE	Choose preferred option	4/3/6	1,119	NL	Table	
Price et al. (2017)	Between	CV	DBDC	3/2/5	269	CL, RPL, LC	Table/graphic	CV=CE
		CE	Choose preferred option	4/2/5 or 4/3/5	273	CL, RPL, LC	Table	
Stevens et al. (2000)	Between	CV	DC	4/2/3	581	NR	Text	CV<CE
		CE	Ranking	4/1/3	692	Logit	Table	
Travisi and Nijkamp (2004)	Within	CV	DBDC	1/2/2	302	Survival model	Text	CV<CE
		CE	Choose option	4/3/4 or 5/3/4	302	CL	Table	
Tuan and Navrud (2007)	Between	CV	DC	1/2/2	484	NR	Text	CV=CE
		CE	Choose preferred option	7/2/4	446	CL	Table	
Weber and Stewart (2009)	Within	CV	PC	1/2/5	273	Logit	Table	CV<CE
		CE	Choose preferred option	4/3/5	273	CL	Table	

Notes: DC = dichotomous choice; DBDC = double-bounded dichotomous choice; PC = payment card; OE = open-ended; CL = conditional logit; RPL = random parameters logit; LC = latent class; DB logit = double-bounded logit; NL = nested logit; NR = not reported

Table A.2. Participants' responses to the follow-up questions about their behavior and perception of the experiment

	Text treatment	**Table treatment**	**P-values from chi-squared tests of equality of proportions across the treatments**
On what basis did you make your choices of tokens?			0.787
I picked the tokens randomly.	0.0%	1.7%	
I picked the tokens which seemed like giving a higher benefit.	14.0%	15.5%	
I calculated the benefit from buying each token and chose the one with a higher benefit.	84.2%	81.1%	
I don't know. / Hard to say.	1.8%	1.7%	
How difficult was it to make a choice of a token?			0.921
Very difficult	0.0%	0.0%	
Difficult	3.5%	1.7%	
Somewhat difficult	0.0%	0.0%	
Neither difficult, nor easy	12.3%	13.8%	
Somewhat easy	0.0%	0.0%	
Easy	38.6%	36.2%	
Very easy	45.6%	48.3%	
I don't know. / Hard to say.	0.0%	0.0%	
Did you have enough time to make your choices of tokens?			0.220
Yes.	96.5%	98.3%	
No.	0.0%	1.7%	
I don't know.	3.5%	0.0%	
Did you calculate the monetary values of tokens when making your choices?			0.410
Yes, I calculated the values always / precisely.	75.4%	82.7%	
Yes, I calculated the values sometimes / approximately.	21.1%	12.1%	
No.	3.5%	5.2%	

Table A.3. Percentage shares of payoff-maximizing choices and average response times across choice sets

	Percentage shares of payoff-maximizing choices			**Average response times** (in seconds)	
Choice set	**Luchini and Watson (2014)**	**Our Text treatment**	**Our Table treatment**	**Our Text treatment**	**Our Table treatment**
A	9.3	82.5	82.8	51.6	42.1
B	33.3	92.9	94.8	44.9	34.2
C	27.7	82.1	91.4	47.7	36.5
D	85.2	93.0	93.1	48.1	33.0
E	74.1	94.7	86.2	42.6	32.7
F	74.1	82.1	74.1	55.4	45.1
G	81.5	91.2	89.7	47.5	33.2
H	79.6	78.9	77.6	54.1	38.7
I	74.1	87.7	89.7	42.1	33.1

Appendix B
Instructions for the experiment

You are about to participate in an experimental study of how people make choices.

At the beginning of the experiment, you are given an account with a balance of $4, and you can use the money in this account to buy tokens that are offered for sale in this experiment.

The experiment has 9 rounds. In each round, you will be offered two tokens: token A and token B. You will be asked if you want to buy *one* of the tokens and if so, which token. The tokens have different prices and values. *You can buy at most one token in each round.*

The value of a token depends on the token's characteristics. In this experiment, the tokens have three characteristics: their size (small, medium, large); their colour (red, yellow, blue); and their shape (circle, triangle, square). The table below presents the value of each characteristic.

Size	Small Medium Large	$0.50 $2.50 $4.00
Colour	Red Yellow Blue	$1.00 $1.50 $2.00
Shape	Circle Triangle Square	$1.50 $3.00 $6.00

The total value of each token is calculated by adding up the value of each characteristic. For example: a small, yellow, triangle token has a value of: $0.50 + $1.50 + $3.00 = $5.00.

The prices of the tokens offered for sale in this experiment vary and can be equal to $2, $3, or $4.

In each choice question, you will be shown on the screen the characteristics and the price of each token.

At the end of the experiment, one round, out of the 9 rounds you participated in, will be chosen at random by the computer and displayed to you on the screen. Your account balance at the end of the experiment will depend on the choice you made in this randomly chosen round:

- If you bought a token, the price of the token will be deducted from your initial account of $4 and the total value of the token will be added to your account.
- If you did not buy a token, your account balance will be unchanged at $4.

At the end of the experiment, you will be shown on the screen: the randomly drawn round which will determine your account balance, the total value of the token you chose in this round, the price of the token you chose in this round, and your account balance. Your total earnings from the experiment will be the account balance plus $6 for participation in the experiment.

Please write down your total earnings in the payment sheet you are provided with.