methods, data, analyses | Vol. 16(2), 2022, pp. 235-272 DOI: 10.12758/mda.2022.03

The Impact of Presentation Format on Conjoint Designs: A Replication and an Extension

Sophie Cassel, Josefine Magnusson & Sebastian Lundmark
University of Gothenburg

Abstract

In recent years, conjoint experiments have been in vogue across the social sciences. A reason for the conjoint experiments' popularity is that they allow researchers to estimate the causal effects of many components of stimuli simultaneously. However, for conjoint experiments to produce valid results, respondents need to be able to process and understand the wide range of dimensions presented to them in the experiment. If the information processing is too demanding or too complicated, respondents are likely to turn to satisficing strategies, leading to poorer data quality and subsequently decreasing the researcher's ability to make accurate causal inferences. One factor that may lead to the adoption of satisficing strategies is the presentation format used for the conjoint experiment (i.e., presenting the information within a text paragraph or a table). In the present paper, a direct replication of the single conjoint presentation format experiment described in Shamon, Dülmer, and Giza's (2019) paper in *Sociological Methods & Research* is presented, and extending their work to paired conjoint experiment. The results of the direct replication showed that respondents evaluated the questionnaire more favorably when reading the table format but were, on the other hand, less likely to participate in subsequent panel waves. Albeit the number of break-offs, refusals, and non-responses did not differ between the two formats, respondents who saw the table format evaluated the scenarios with more consistency and less dimension reduction, thus favoring the table presentation format. For paired conjoint experiments, the presentation format did not affect survey evaluations or panel participation but the table format heavily outperformed the text format on every data quality measure except for dimension reduction. Conceptually, albeit not directly replicating the findings in Shamon, Dülmer, and Giza (2019), the present manuscript concludes that the table format appears preferable over the text format for conjoint experimental designs.

Keywords: conjoint experiments, satisficing, presentation format, text versus tables, replication

Conjoint experiments (also referred to as "factorial surveys") are widely used in the social sciences for measuring beliefs and preferences, and for multidimensional decision making (Shamon, Dülmer & Giza, 2019). In a typical conjoint experiment, respondents are asked to evaluate a vignette with several different dimensions (i.e., attributes of the sets/profiles to evaluate) with randomly assigned levels (i.e., the value of the dimension, for example, male/female, rich/poor).

An advantage of designs such as conjoint experiments is that the evaluation processes of the vignette resemble the judgment processes made in real life given that it involves making a single evaluation based on the information for several attributes at the same time (Hainmueller, Hopkins & Yamamoto, 2014). In comparison to the single factor experiment, the controlled variation of attribute levels in the conjoint experiment enables researchers to capture the unique impact of several dimensions at the same time (Auspurg & Jäckle, 2017; Hainmueller et al., 2014). By providing respondents with the relevant information needed to form an opinion, the conjoint experiment is thought to reduce the potentiality for respondents to change their reference frame when answering a survey (Shamon et al., 2019). Consequently, these potential benefits, combined with the growing practice of administering questionnaires via computers, have made the conjoint experiment popular across many research fields.

Despite the increasing popularity of the conjoint experiment, the number of studies examining the effect of design complexity (e.g., investigating how the vignette is presented) on respondents' cognitive burden and answer behavior has been relatively few (see Shamon et al., 2019). Except for Shamon, Dülmer, and Giza (2019), Sauer, Auspurg and Hinz (2020), and Hainmueller, Hangartner, and Yamamoto (2015), there are no studies on the impact that the presentation formats may have on the cognitive burden and answer behavior of respondents and the effect is still relatively unknown. As information intake is central to vignette and conjoint studies, knowledge of how respondents react to different presentation formats is essential for researchers' ability to draw accurate inferences from the experiments (Shamon et al., 2019).

Two presentation formats dominate the realm of online conjoint experiments: text (presenting the information in a text paragraph) and table format (a table with dimensions and levels in rows and columns). Other formats exist, such as presenting the information via video clips, illustrated cards, or pictures (see, e.g., Sato, Kubo & Namatame, 2007), but the present study focuses on text and table formats. Theoretically, using a text format may be preferable because nesting the information in stories may enhance respondents' understanding of the hypothetical situa-

Direct correspondence to
Sophie Cassel, the SOM Institute, University of Gothenburg, Sweden
E-mail: Sophie.Cassel@gu.se

tion or increase empathy with the described situation. The enhanced understanding and empathy may, in turn, increase the respondents' attention to the dimensions in the conjoint experiment (Auspurg & Hinz, 2015). Furthermore, individuals may be more accustomed to absorbing information in flowing text than from reading tables. Given respondents' likelihood of being more used to reading text than tables, presenting conjoint experiments in text format may make respondents less prone to *satisficing strategies*. Satisficing strategies are employed when respondents resort to suboptimal information processing and response techniques as a means to lower their cognitive burden of filling out a questionnaire or participating in an experiment (Krosnick, 1999).

By contrast, Shamon et al. (2019, p. 9) argued that table formats may facilitate stronger information intake in comparison to text formats because tables, generally, contain fewer words and present the relevant information at the same visual position across different scenarios, which should decrease the cognitive investment needed by respondents and especially among the respondents with lower cognitive skills (Shamon et al., 2019). However, Shamon et al.'s (2019) assumption only hold if respondents accurately comprehend the order (the rows and columns) in which the tables should be read. Such an assumption is more likely to be violated when respondents are asked to compare two different sets of characteristics (e.g., stating a preference for one of two political candidates with different dimensions), meaning that the respondents have to understand that the values in one column belong to the first set or profile (e.g., candidate A) and the values in the second column belong to the second (e.g., candidate B).

The existing empirical studies on the effect of presentation format on answering behavior paint an inconclusive picture. Whereas Sauer et al. (2020) found that table formats produced similar evaluations as text formats, Hainmueller et al. (2015) found that table format (evaluation of two sets/profiles) performed better than text formats in predicting real-life judgments. Similarly, Shamon et al. (2019) found support for the table format performing slightly better in reducing satisficing behavior compared to the text formats, although most of their measurements showed no statistically significant differences between the two formats. It appears that more data is needed to assess the role that the presentation format may have for the data quality and the respondents' experiences when participating in conjoint experiments.

To that end, the present paper presents a replication of Shamon et al.'s (2019) study on single conjoint experiments and the impact that the presentation format (text vs. table) may have on respondents' reporting behavior and subjective experience of the questionnaire is assessed. Following the advice of Sauer et al. (2020), the present study replicated Shamon et al.'s (2019) experiment in a non-student sample invited to resemble the Swedish population in terms of age, sex, and education. The sample was drawn from the Swedish Citizen Panel administered by the

Laboratory of Opinion Research (LORE) at the University of Gothenburg. Furthermore, extending Shamon et al.'s (2019) experiment, a conceptual replication of their study was performed, testing the impact of the presentation format in a paired conjoint setting (i.e., where respondents are asked to state their preferences regarding two different sets/profiles, for example by reporting their preference for one of two politicians described in the vignette). Furthermore, the importance of replicating published research has become especially acute given the many recent failures to replicate published literature (e.g., Open Science Collaboration, 2015).

This paper is organized as follows: First, a brief theoretical rationale for how text and table presentation formats may influence data quality is presented. Then, the hypotheses for the single conjoint and, thereafter, the paired conjoint experiment is introduced. The paper continues with a description of the evaluation criteria, methods and materials. The results are thereafter presented for the single conjoint and the paired conjoint experiment, separately. Next, a summary of the results and a comparison to Shamon et al.'s (2019) findings are presented. The paper ends with a discussion and some conclusions of the main takeaways of the paper.

How Does the Presentation Format Influence Data Quality and Respondent Behavior?

Single Conjoint Experiment

Survey respondents are sensitive to a range of, sometimes almost undetectable, survey design features (Schuman & Presser, 1996; Roberts et al., 2019). Previous research has suggested that the visual appearance of a questionnaire may influence respondents' satisfaction with it (although this connection is not always found – see, e.g., Mahon-Haft & Dillman, 2010). Furthermore, people have been found to interpret a task with a difficult-to-read instruction as more difficult to complete compared to when the task is described in an easy-to-read instruction (e.g., the information is presented in an easy- or difficult-to-read font) (Song & Schwarz, 2008). Hence, it stands to reason that the presentation format of a conjoint experiment that is more difficult to read, interpret, and time-consuming for the respondent, may produce less overall satisfaction with the questionnaire and may make respondents more likely to rate the questionnaire as difficult to complete.

In the presentation format of conjoint experiments, presenting the information in text format generally requires a greater number of characters and syllables, and there are more words to read compared to when the information is presented in table form. The fewer number of words and characters on the screen in a table may be interpreted as less information to process for the respondent and, therefore, as a less demanding task, potentially leading to greater respondent satisfaction com-

pared to when the information is presented as text. On the other hand, respondents may be more used to gathering information in text format than in table format, which may lead to greater survey satisfaction with the text format. Therefore, the following hypothesis will be assessed:

Hypothesis 1a (H1a): Respondents who evaluate scenarios presented in a table format may report a better respondent experience (i.e., report greater satisfaction with the questionnaire, take less time to evaluate the scenarios, and be more likely to participate in future panel waves) than respondents who evaluate the same scenarios presented as text.

According to satisficing theory, a presentation format that is difficult to understand, hard to process, or difficult to read is thought to induce stronger satisficing behavior (Krosnick & Alwin, 1987; Krosnick, 1991; Song & Schwartz, 2008). Satisficing is a decision-making process in which a person, instead of expending appropriate cognitive effort to come to an optimum decision, decides to expend only the minimum effort needed (or no effort at all) to come to a decision (Simon, 1957; Krosnick, 1991). Compared to expending the appropriate effort, satisficing strategies rarely lead to decisions or answers that best represent an individual's actual wants, needs, or attitudes. A presentation format in conjoint experiments that increases the likelihood of a respondent satisficing is expected to produce a range of negative influences on data quality.

When presenting conjoint experiments in a table format, the researcher's goal is to make the necessary information easily accessible to the respondent by only presenting the information needed for the respondent to make a decision, and to present that information through a minimum number of characters, syllables, and words, and using an easy-to-read format (Shamon et al., 2019). In line with this, Bansak and colleagues (2021) found that increasing the number of dimensions in the conjoint experiment table to as many as 18 only moderately influenced satisficing behavior, suggesting that the table format may indeed be easy for respondents to read. If the table format indeed has these intended effects, then respondents should be less likely to satisfice when reading the table presentation format than the text format, which generally uses more syllables and words, and longer sentences perhaps include more complex syntax, without further aiding the information intake. Based on satisficing theory, and replicating the predictions made in Shamon et al. (2019), the following hypotheses regarding respondent behavior will be assessed:

Hypothesis 1b (H1b): Respondents who evaluate the scenarios presented in a table format may produce data of greater quality (i.e., fewer refusals, fewer breakoffs, fewer scenario non-responses, fewer total non-responses, and less dimension reduction) than respondents who evaluate the scenarios presented in a text format.

Hypothesis 1c (H1c): The effect that the presentation format may have on the total loss of information (in terms of total non-response) may be stronger for

respondents with lower educational attainment than for respondents with higher educational attainment.

Hypothesis 1d (H1d): The effect that the presentation format may have on the total loss of information (in terms of total non-response) may be stronger for older respondents than for younger respondents.

Paired Conjoint Experiment

In paired conjoint experiments, given that more information has to be processed and that the dimensions presented in the rows and columns of a table have to be correctly attributed to the correct set/profile, the table format may not outperform the text format to the same extent as in single conjoint experiments. The cognitive process may, therefore, change in the more complex paired setting, which, consequently, changes how the presentation format affects data quality and respondent experience. However, in line with the findings that the table format did outperform the text format in Hainmueller, Hopkins, and Yamamoto (2014), we still hypothesize that table format may outperform text format while remaining open to the reported differences between the single and paired conjoint experiment.

Hypothesis 2a (H2a): Respondents who evaluate scenarios presented in a table format may report a better respondent experience (i.e., report greater satisfaction with the questionnaire, take less time to evaluate the scenarios, and be more likely to participate in future panel waves) than respondents who evaluate the scenarios presented in text format.

Hypothesis 2b (H2b): Respondents who evaluate scenarios presented in a table format may produce data of greater quality (i.e., fewer refusals, fewer break-offs, fewer scenario non-responses, fewer total non-responses, and less dimension reduction) than respondents who evaluate the scenarios presented in a text format.

Hypothesis 2c (H2c): The effect that the presentation format may have on the total loss of information (in terms of total non-response rate) may be stronger for respondents with less educational attainment than for respondents with more educational attainment.

Hypothesis 2d (H2d): The effect that the presentation format may have on the total loss of information (in terms of total non-response rate) may be stronger for older respondents than for younger respondents.

Evaluation Criteria

To evaluate the hypotheses, in the present paper, the impact of the presentation format was categorized into aspects related to the respondent experience and the data quality. The same evaluation criteria were used to investigate the impact of presentation format in the single and the paired conjoint experiment.

The impact of the presentation format on respondent experience was investigated by assessing the cost of administration (in terms of processing time), the perceived experience of the survey (survey evaluation), and the probability of participation in subsequent waves of the Swedish Citizen Panel.

The impact of presentation format on data quality was investigated by assessing the refusal to participate in the survey experiment, the probability of the respondent breaking-off from completing the questionnaire, the probability of unanswered scenario evaluations, the number of faded-out dimensions, size of coefficient for dimensions, and respondents' response inconsistency. However, some respondents may very well have valid non-attitudes, meaning that a non-response, refusal, or break-off would be the most accurate representation of their evaluation. But, in line with the argument provided by Shamon et al. (2019), omitting to make a judgment will in the present paper be perceived to be an indication of satisficing and not as a valid representation of non-attitudes. The different answer behaviors refusal, break-off, non-response, and total non-response indicate that a respondent has applied a satisficing strategy, resulting in reduced data quality. Similarly, the varying importance that the respondent assigns to different dimensions is also proposed to be a form of satisficing, as the respondent reduces the cognitive burden of completing the questionnaire by excluding dimensions or assigning varying importance to the different dimensions in the experiment. Each of these forms of satisficing will be presented in more detail below, together with how each of them is operationalized.

Respondent Experience

Cost of Administration

The impact of the presentation format was investigated in terms of the time it took the respondents to answer the scenarios (i.e., the cost of administration). Longer administration times may be an indication that the respondents are struggling with interpreting and reading the vignette. In contrast, longer administration times may also be an indication that the respondents are paying attention to the information on the screen, leading to more thoughtful responses and greater data quality. Regardless of potential benefits to data quality, longer administration times will mean less time to ask other questions, as well as having to offer higher incentives to the respondent.

Time spent on the pages with the scenarios was used to assess the cost of administration. Due to an oversight in the survey programming, the time spent on the last scenario was not recorded for the single conjoint groups. Therefore, the cost of administration analyses for the single conjoint groups includes only the time spent on the first scenarios.[1] For the paired conjoint groups, the time spent on all scenarios was recorded and analyzed.

To reduce the impact of outliers, following Tukey (1977), total response times for the scenarios that were shorter than the interquartile range (IQR) of the sample response times * 1.5, and longer than the IQR * 1.5, were excluded from the cost of administration analysis. For the scenarios in the single conjoint groups, the lower bound for the excluded outliers was 0 seconds, and the upper bound was 285.9 seconds. For the scenarios in the paired conjoint groups, the lower bound was 0 and the upper bound was 520.3 seconds.

Survey Evaluation

The impact of the presentation format was also investigated in terms of survey evaluation. Respondents reported how well designed and how difficult the questionnaire was and rated their level of annoyance and concentration while filling out the questionnaire. A more positive overall survey evaluation may be an indication that the respondent found interpreting and evaluating the scenarios less challenging, and a more positive respondent experience may lead to better data quality.

Responses to the four survey evaluation questions (well-designed, difficult, annoyed, and needed concentration) were averaged into an index and coded to range from 0 to 1, with higher values indicating a more positive overall evaluation of the questionnaire.

Participation in Subsequent Panel Waves

Taking advantage of the ability to follow each respondent's participation in the Swedish Citizen Panel, the impact that the presentation format had on participation in the subsequent waves of the panel was investigated. Panelists were randomly sampled to be invited to complete studies in the subsequent waves of the Swedish Citizen Panel which led to that not all participating respondents in this study were invited to the subsequent waves of the panel. However, the majority of the panelists were invited. Participation/non-participation in subsequent waves may have many different causes, but a between-subject comparison of the respondents who saw the text and the respondents who saw the table format may reveal whether one of the formats was particularly detrimental to respondents' willingness to complete similar future tasks and experiments. A larger dropout may be of particular interest for

1 The respondents were asked to evaluate four scenarios in total.

any sample provider attempting to estimate future costs of administering conjoint experiments.

Data Quality

Refusal

Refusing to evaluate all of the scenarios was one form of satisficing investigated. According to Shamon et al. (2019), respondents who refuse to respond decline to make judgments and thereby engage in a satisficing strategy by skipping at least one cognitive step when evaluating the scenarios. Refusal to make valid evaluations of all scenarios may indicate that the respondent found it more challenging to read or interpret the text or the table format, which has a clear negative impact on data quality.

Respondents were categorized as "refusals" if they did not evaluate any of the scenarios, answered "don't know" in all scenarios, or provided no variation in their answers. Respondents who used these answering behaviors across all scenarios were coded as 1, and 0 otherwise.

Break-offs

Another form of satisficing strategy investigated was when respondents switched to constant non-valid answering behavior at some point after evaluating the first scenario. Opting for a non-valid answering behavior after having provided valid evaluations may indicate that the respondent found it more challenging to read or interpret the vignette and reduces the data quality.

Respondents were coded as 1 if they evaluated at least the first scenario and thereafter consistently used a non-valid answering strategy. That is, they were coded as 1 if they gave a valid answer to the first scenario and then started to answer "don't know" or left a scenario evaluation unanswered, and 0 otherwise.[2]

Non-response

An alternative strategy for a respondent to decrease the cognitive burden of completing the questionnaire would be to alternate between validly evaluating scenarios and not validly evaluating scenarios (Shamon et al., 2019). Such a strategy should be considered a weaker form of satisficing than refusal or breaking-off but remains a negative influence on data quality. This evaluation criterion aims to capture the type of satisficing behavior where the respondent remains in the experiment (hence, does not refuse to answer, or break off answering the questions) but instead alternates between validly judging and not validly judging scenarios to make the survey

2 The respondents were asked to evaluate four scenarios in total.

easier to complete. Fewer invalidly judged scenarios indicate that the respondent found interpreting and reading the text or the table format less challenging and suggest better data quality.

Scenarios evaluated by respondents (which were not coded as refusal or break-off) were coded as a non-response if the respondents invalidly judged at least one scenario but not all of them. A scenario was invalidly judged if the respondent answered "don't know" or did not provide an answer. Note that this evaluation criterion was computed at the scenario level and not at the respondent level.

Total Non-response

To capture the total loss of information due to the presentation format, the total non-response was computed and captured all the scenarios that were invalidly judged, irrespective of the type of strategy used by the respondent. The criteria were computed on the scenario level, and a scenario was coded as 1 (total non-response) if it was invalidly judged by either not answering or answering "don't know," or if the respondent provided no variation in the answer across all four scenarios or broke-off their participation, and 0 otherwise.

Response Inconsistency and Partial Dimension Reduction

A presentation format that respondents have a harder time reading or understanding, or that makes it more difficult for respondents to distinguish between the dimensions may produce a weaker predictive ability of the attribute levels on the dependent variable of the conjoint experiment. As a result of increased response inconsistency, an underperforming presentation format may produce more measurement errors in the dimensions' predictions. Hence, a presentation format that yields the largest estimated parameters for the dimensions and has the lowest measurement error should be interpreted as the more valid and preferable format to use for conjoint experiments.

Furthermore, a cognitively more burdensome presentation format should have a stronger detrimental effect on both parameter estimates and measurement error as a respondent evaluates more scenarios (i.e., a partial dimension reduction). A more burdensome presentation format should increase the likelihood of the respondent putting less and less cognitive effort into distinguishing between different dimensions as the number of evaluated scenarios increases. Hence, one would expect to see a weaker and weaker predictive ability of the dimensions on the dependent variables as well as greater measurement error across scenarios (i.e., one would expect to see a reduction in the impact that the dimensions have on the dependent variables).

To investigate the partial dimension reduction and response inconsistency, the invariance in parameters and the invariance in error variance were compared across the two presentation formats by applying the structural equation modeling (SEM)

technique to predict/make a judgment on salary or party preference of respondents based on the dimensions presented in the conjoint experiment (MacDonald, 2016). All exogenous predictors were free and allowed to covary.[3]

Methods and Materials

Sample

The respondents were a pre-stratified sample of members of the Swedish Citizen Panel run by the Laboratory of Opinion Research (LORE) at the University of Gothenburg, Gothenburg, Sweden. At the time of the study, the Swedish Citizen Panel consisted of about 59,000 self-selected panelists, and members of the panel were invited to complete approximately four online omnibus questionnaires each year. The panelists were, therefore, relatively experienced, and were not paid an incentive to complete the questionnaires.

The presentation format experiment was administered to 7,000 panelists pre-stratified by sex (male, female), age (18–34, 35–49, 50–85 years), and education (low/middle education: less than 3 years of post-secondary education, high education: 3 or more years of post-secondary education) between February 24th, 2020, and March 19th, 2020. For the demographic distributions, see Table 1. Reminders were sent on March 3rd, 2020, and on March 11th, 2020, to all respondents who had not yet completed the questionnaire. Out of the 7,000 respondents invited to participate, 4,236 completed the experiment (American Association for Public Opinion Research (AAPOR) response rate 5 (RR5): 59%).

3 Parameters were estimated using the function SEM in Stata 16, with the group option and all other options set at default. By default, SEM in Stata 16 allows all exogenous predictors to covary.

Table 1 Demographic distributions of the experiment sample and the Swedish population, and the difference between the sample and Swedish population.

Variables	Sample	Population	Difference
Age, years			
18–34	18%	28%	-10%
35–49	23%	25%	-2%
50–85	59%	47%	+12%
Gender			
Male	48%	50%	-2%
Female	52%	50%	+2%
Education			
Low/middle	68%	76%	-8%
High	32%	24%	+8%

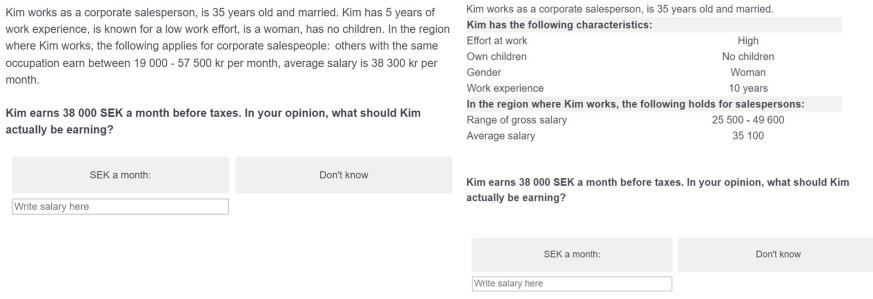
Notes. N=4,236.

Procedure

Each respondent was randomly assigned to either a single or paired conjoint experiment, and within each experiment, respondents were randomly assigned to see the conjoint in either a text or a table presentation format.

Single Conjoint Experiment

Respondents assigned to the single conjoint experiment reported the amount of salary a person deserved to earn, based on four dimensions of the person (sex, number of children, work experience, and work effort) and two contextual dimensions (average salary for others in the same region and range of salaries for people of the same occupation), approximating a direct replication of the experiment in Shamon et al. (2019). See Table S1 in the Supplementary Online Materials (SOMs) for the dimension levels. Whether the person's dimensions or the contextual dimensions were presented first was randomly determined for each participant. In addition, each level of the dimensions was determined randomly. Each respondent was presented with four scenarios to evaluate. See Figure 1 for a screenshot of the single conjoint experiment translated to English and SOM S1.1. for the full questionnaire logic.



Notes. Respondents were randomized to either the single or the paired conjoint experiment. Respondents in the single conjoint experiment group were randomized to read the information in either text or table presentation format.

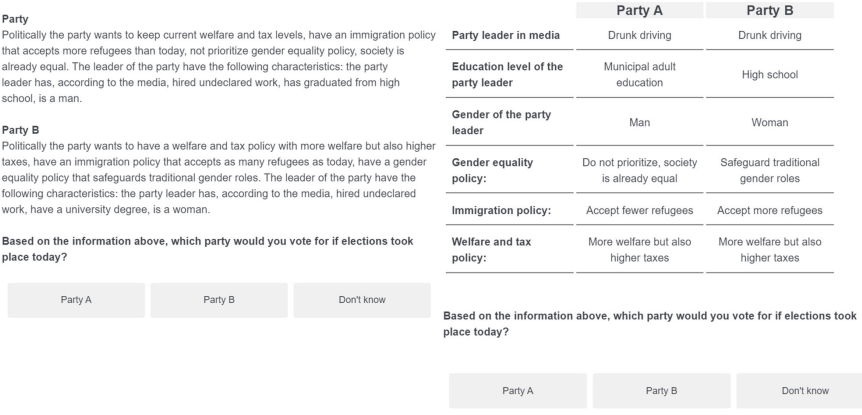
Figure 1 Screenshot of the single conjoint experiment administered on computers.

Paired Conjoint Experiment

Respondents assigned to the paired conjoint experiment reported which of two hypothetical political parties they would vote for and how likely they would be to vote for each of the parties. The paired conjoint experiment did not use the above topic (i.e., the salary of a worker) because the single conjoint experiment did not lend itself to be easily translated into a paired experiment. Instead, the paired experiment was developed to closer mimic Hainmueller et al.'s (2015) experiment where respondents evaluated two political agents (in our case, political parties). The two parties were described in terms of three political dimensions (immigration, welfare and taxes, and equality policy stances) and three dimensions of their respective party leader (sex, educational attainment, and media image). See Table S2 in the SOM for the dimension levels. Whether the leader's or the party's dimensions were presented first was randomly determined for each respondent. In addition, each level of the dimensions was determined randomly.

Each respondent was presented with four scenarios to evaluate. See Figure 2 for a screenshot of the paired conjoint experiment translated to English and SOM S1.4 for the full questionnaire logic.

After the presentation format experiment, the respondents reported their level of annoyance and concentration while filling out the questionnaire, as well as how well-designed and difficult the questionnaire was.



Notes. Respondents were randomized to either the single or the paired conjoint experiment. Respondents in the paired conjoint experiment group were randomized to read the information in either text or table presentation format.

Figure 2 Screenshot of the paired conjoint experiment administered on computers.

Differences from Shamon et al. (2019)

The procedure and sample in this paper diverge from a direct replication of Shamon et al. (2019) in three ways: Firstly, Shamon et al. administered a third condition of the text format; underlining the dimensions in the text vignette. We did not implement that condition.[4]

Secondly, Shamon et al. (2019) presented dimensions in a fixed order, whereas we randomly assigned whether the dimensions of the person or dimensions of the context were presented first for each respondent (and similarly for the order of the personal/party leader and contextual/party dimensions). The order was randomized to reduce recency and primacy effects, as well as to avoid the order effect cautioned

4 The ease of reading a paragraph (i.e., processing fluency) has been found to correlate with both actual cognitive effort and perceived effort (Reber, Schwarz, and Winkielman, 2004; Song and Schwarz, 2008). That is, a paragraph containing cursive/italicized letters has been found to be more difficult to process than a simple font such as Arial (Song and Schwarz, 2008). Similarly, underlining certain phrases likely presents respondents with yet another layer of cognitive burden compared to an easy-to-read paragraph with less clutter. This notion is supported in Shamon et al.'s (2019) findings, where the underlined text format took respondents longer to process than the other formats. Furthermore, underlining has to be accurately understood by each respondent as "the important information" in order to actually improve data quality. If the respondents interpret the underlining differently, the result may be more random measurement error.

by Auspurg and Hinz (2015) (see also Auspurg & Jäckle, 2017) but disregarded in Shamon et al. (2019).

Thirdly, Shamon et al. (2019) presented respondents with 16 scenarios that they had to evaluate, whereas we asked the respondents to evaluate four scenarios. Respondents were presented with fewer scenarios to resemble the questionnaires that they usually complete and to lower the risk of exhausting respondents. Administering fewer scenarios may contribute to weaker effects on outcomes that correlate with questionnaire fatigue, but we opted for more unique observations over having more scenarios in order to increase the variation of the type of respondents. Therefore, instead of presenting respondents with many scenarios, we included a larger sample of respondents (N=2,068 in the single conjoint experiment) compared to Shamon et al. (N=498), thus following Shamon et al.'s (2019, p. 34) suggestion to increase statistical power in order to be able to identify small effects.

However, in line with Shamon et al.'s (2019) approach, we decided to still treat respondents' answers as refusals if they provided the exact same answers for all four scenarios. Although the number of scenarios was fewer, we deem it unlikely that a respondent would validly consider all dimensions and still provide the same salary four times in a row, or in the paired scenario, always choosing Party A or Party B. Lastly, our sample included respondents older than 69 years in an attempt to better generalize to the general population compared to Shamon et al. (2019).

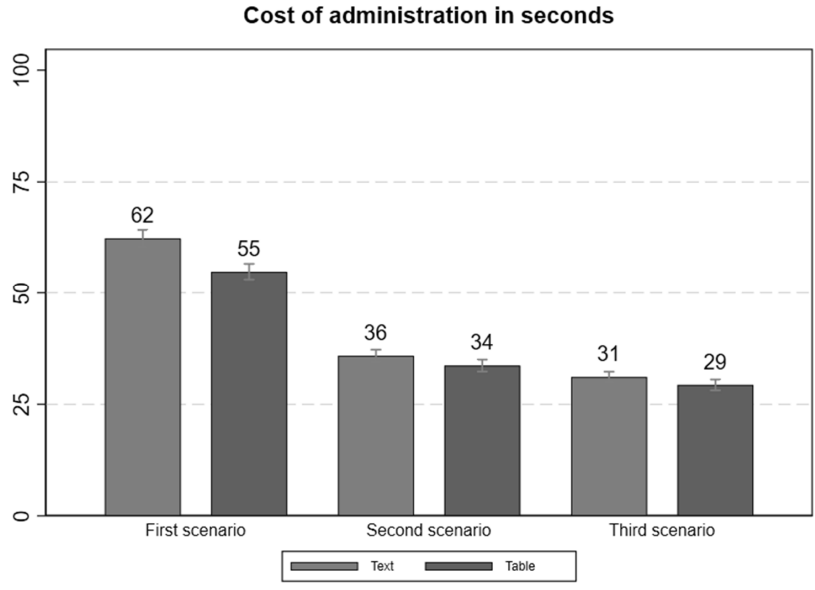
Results

In this section, the impact that the presentation format had on the single conjoint experiment will be presented first, followed by the impact it had on the paired conjoint experiment. The evaluation of the presentation format will be separated into aspects related to the participant experience (cost of administration, evaluation, participation in subsequent waves) and aspects related to data quality (refusal, break-off, non-response, total non-response, dimension reduction, and moderation of effects).

Single Conjoint Experiment

Respondent Experience

Cost of administration. Across the first three scenarios, respondents who made a judgment on the salary of the worker when reading about the dimensions in a text paragraph took statistically significantly 12 seconds longer to submit their evaluations (M = 129 seconds, standard deviation (SD) = 54) than the respondents who evaluated the salary when the dimensions were presented in a table format (M =



Notes. N=1,874. Respondents who answered the three scenarios for which time was recorded, and whose response times were not longer than 1.5 times the interquartile range (IQR) for the three scenarios, were included in the analyses (N excluded = 137).

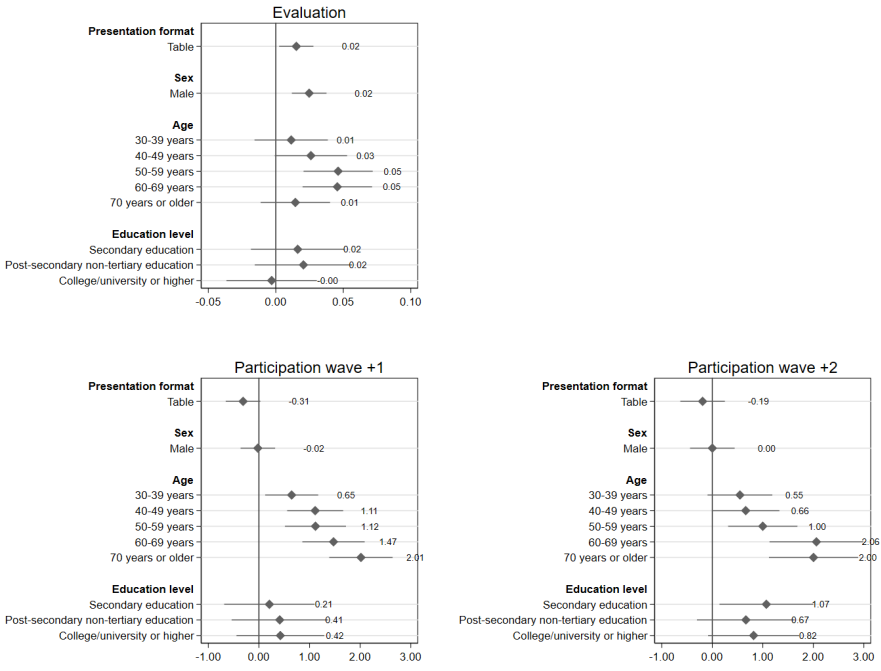
Figure 3 Cost of administration, in seconds, for the single conjoint experiment.

117 seconds, SD = 53; b = -11.72, standard error (SE) = 2.47, p<0.001), a statistically significant difference providing support for H1a.

However, the difference in administration time was greatest for the first scenario (see Figure 3). In the second and third scenario, the difference in administration time between the text and the table version statistically significantly decreased ($b_{table * scenario\ 2}$ = 5.22, SE=1.29, p<0.001; $b_{table * scenario\ 3}$ = 5.68, SE=1.30, p<0.001) (see SOM S2.1, Table S3).

Over scenarios, respondents were able to reduce the time to evaluate the specific scenario but the respondents who read the text format reduced their processing time more compared to those who read the table format (see Figure 3). However, as will be shown in the dimension reduction analyses, the stronger reduction in processing time over scenarios for respondents presented with text format seems to have stemmed from the fact that those respondents invested less and less cognitive effort in their evaluations (see section *Moderation effects*).

Evaluation. Respondents who read the table presentation format reported a more positive evaluation of the questionnaire (b=0.02, SE=0.01, p<0.01) than the respondents who read the text presentation format (see Figure 4). The significant



Notes. N=1,968 (Evaluation); N=1,484 (Participation wave +1); N=903 (Participation wave +2). The number of observations differs in the three panel waves because panelists were randomly sampled to be invited to complete the panel wave or not. Regression coefficients (gray diamonds) from one ordinary least squares regression (OLS) and two logistical regressions with their respective 95% confidence intervals (CIs) (gray solid lines). A positive value of the coefficient indicates a higher overall evaluation or a higher likelihood of participation in subsequent waves. Baseline categories were female, 18–29 years of age, and compulsory education (9 years).

Figure 4 Respondent experience in terms of overall questionnaire evaluation and participation in the subsequent waves, for the single conjoint experiment.

effect of presentation format was found in both bivariate analyses and when including controls (see Figure 4 and SOM S2.1., Table S4). Male respondents were, overall, more positive than female respondents (b=0.02, SE=0.01, p<0.001), and older respondents, in the 50–59 and 60–69-year groups, were more positive than those aged 29 years or younger (see Figure 4).

When analyzing the four separate evaluation questions used to construct the index of overall survey evaluation, the only significant effect of presentation format was found for the question asking how annoyed the respondent was when filling out the questionnaire (b=0.04, SE=0.01, p<0.01; see SOM S2.1., Table S5).

Participation in subsequent waves of the Swedish Citizen Panel. Presentation format had an initial significant effect on participation in the immediate subsequent wave of the Swedish Citizen Panel, but this effect disappeared with later waves (see Figure 4).

Respondents who were given the table presentation format were marginally less likely to participate in the following wave of the Swedish Citizen Panel, which was administered approximately 3 months after the presentation format experiment (wave +1) (b=-0.31, SE=0.17, p<0.10). However, the effect of presentation format was not statistically significant in the wave of the Swedish Citizen Panel distributed approximately 6 months after the presentation format experiment (wave +2) (see Figure 4). The age of the respondent had a significant effect on participation, both in the first and second wave following the presentation format experiment, with older respondents being more likely to participate in subsequent waves. Participation in the second wave following the experiment was significantly more likely among respondents with upper secondary education (b=1.07, SE=0.47, p<0.01).

However, on average, the results provide additional support for H1a, as respondents reported both greater satisfaction and took less time to evaluate the scenarios when receiving the table format compared to the text format. The results on participation in the subsequent waves confirm this and the immediate negative effect of the table presentation format disappeared after the first wave following the presentation format experiment.

Data Quality

Table 2 presents a descriptive summary of the sample size and answer behaviors of respondents assigned to the single conjoint experiment. Overall, 482 respondents presented with the single conjoint experiment chose to either break-off or refusal to answer the scenarios (see Table 2). Similar patterns were found between respondents presented with the text and the table presentation format. However, 26 (2.5%) respondents presented with the table format chose to stop filling out the questionnaire compared to 16 (1.5%) of those who saw the text presentation format (Table 2). The most commonly used satisficing answering behavior on the respondent level was to provide no answers or don't know answers across all of the four scenarios, 167 (16%) respondents in the text format and 161 (15.8%) respondents in the table format.

Table 3 presents a similar descriptive summary of the answering behavior at the scenario level, and the most common satisficing answering behavior was a refusal to answer any of the scenarios. The text and table presentation format yielded roughly the same amount of total loss of information (total non-response) (text: 24.4%, table: 25%) but the text presentation format had slightly fewer respondents breaking-off (1.5%) compared to the table presentation format (2.5%).

Table 2 Sample sizes and answer behavior at the respondent level, for the single conjoint experiment.

Experiment setting	Sample size	No evaluated scenarios (refusal, don't know) (1)	Invalid constant answer behavior (2)	Refusals (1) + (2)	Break-off (3)	Total (1) + (2) + (3)
Text	1047	167 (16%)	57 (5.4%)	224 (21.4%)	16 (1.5%)	240 (22.9%)
Table	1021	161 (15.8%)	55 (5.4%)	216 (21.2%)	26 (2.5%)	242 (23.7%)
Sum	2068	328 (15.9%)	112 (5.4%)	440 (21.3%)	42 (2%)	482 (23.3%)

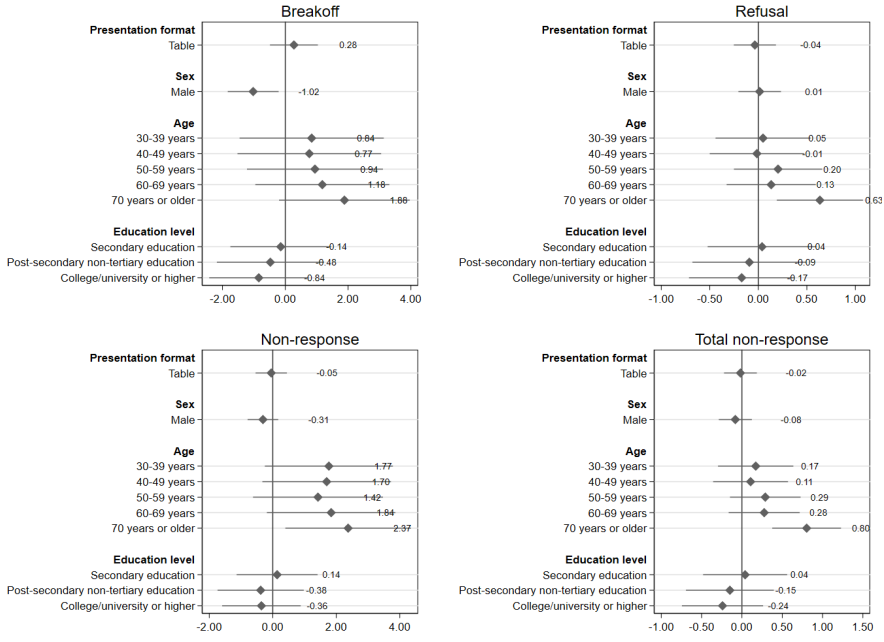
Notes. Results at the respondent level.

Table 3 Sample sizes and answer behavior at scenario level, for the single conjoint experiment.

	Gross sample size (N = 2,068)			Net sample size after excluding refusals and break-offs (N = 1,586)	
Experiment setting	Gross sample of scenarios	Refusals (1) + (2)	Break-off (3)	Non-response (4)	Total non-response (1) + (2) + (3) + (4)
Text	4,188	896 (21.4%)	64 (1.5%)	63 (1.5%)	1,023 (24.4%)
Table	4,084	864 (21.2%)	104 (2.5%)	55 (1.3%)	1,023 (25%)
Sum	8,272	1760 (21.3%)	168 (2%)	197 (2.4%)	2,046 (24.7%)

Notes. Results at scenario level. Gross sample of scenarios is calculated by multiplying the group size of an experimental setting by the set size (4 scenarios per respondent).

Refusals, break-offs, non-responses, and total non-response. The two presentation formats did not differ with regard to refusals, break-offs, non-responses, and total non-response (see Figure 5). Although the presentation format statistically significantly affected non-response (respondents who altered between judging and not judging a vignette, excluding refusals and break-offs), it did so only for the first scenario, where the text format resulted in greater levels of non-response. The effect then disappeared in the subsequent three scenarios, and on average, there was no effect of presentation format on non-responses, nor on refusals, break-offs, and total non-response. The results were not moderated by how many scenarios the respondents had answered.



Notes. N=1,480 (Break-off); N=1,965 (Refusal); N=6,067 (Non-response); N=7,866 (Total non-response). Regression coefficients (gray diamonds) from four logistical regressions with their respective 95% confidence intervals (CIs) (gray solid lines). A positive value indicates a higher likelihood of break-off, refusal, non-response, or total non-response. Baseline categories were female, 18–29 years of age, and compulsory education (9 years). Results on break-offs and refusals are at the respondent level, while results on non-response and total non-response are at the scenario level. We controlled within-participant clustering for non-response and total non-response using cluster-robust standard errors.

Figure 5 Data quality in terms of break-offs, refusals, non-responses, and total non-response, for the single conjoint experiment.

Respondents 70 years or older were more prone to refusal (b=0.63, SE=0.23, p<0.01), non-response (b=2.37, SE=1.01, p<0.05), and total non-response (b=0.80, SE=0.22, p<0.01) compared to the baseline (18–29 years old). Furthermore, women were more likely to breaking-off (b=-1.02, SE=0.41, p<0.05) than men (see SOM S2.1., Table S8).

Moderation effects. To test whether the effect of presentation format on data quality was moderated by education (H1c) and age (H1d), two new models were estimated predicting total non-response with presentation format, age, education, gender, and with an interaction between either presentation format and age or presentation format and education. All graphs on moderating effects can be found in SOM S2.1., Figures S1–S4.

Education. In contrast to the expected, education did not moderate the effect that the presentation format had on the probability of total non-response. Respondents with lower educational attainment were not more likely to adopt an invalid answer behavior due to the presentation format. Hypothesis 1c was, therefore, not supported.

Age. Similar to the effect found for educational attainment, age did not moderate the effect that the presentation format had on total non-response. Older respondents were more likely to adopt an invalid answer behavior, but there was no significant difference in the effect that the presentation format had by age of the respondents. Therefore, although older respondents found it more demanding to fill out the questionnaire, they did not find a certain presentation format more demanding compared to younger respondents, providing no support for H1d.

Response inconsistency and partial dimension reduction. In contrast to the hypothesis (H1b), in the first scenario, respondents who were presented with the table presentation format yielded statistically significantly greater response inconsistency (i.e., weaker prediction of a dimension) ($b_{\text{high effort dimension}}$ = -2,117, SE = 522, $\chi^2(1, 1{,}435) = 16.76$, $p<0.001$), albeit not with significantly more measurement error (ε = 24,141,723, SE=1,206,417), compared to those who were given the text format (ε = 22,937,796, SE=1,279,509, $\chi^2(1, 1{,}435) = 0.47$, p=0.49) (see Table 4, column 1). This underperformance in prediction strength remained for scenario 2 and 3 (see Table 4, column 2 and 3). However, by the fourth scenario, the performance difference had shifted to the table format yielding statistically significantly stronger predictions for three of the dimensions and produced statistically significantly less measurement error of the prediction ($\varepsilon_{\text{difference table versus text}}$ = -16,657,746, $\chi^2(1, 1{,}435) = 31.77$, $p<0.001$) (see Table 4, column 4).

The reversal in the outcome, from the text presentation format outperforming the table format in the first scenarios to the table format heavily outperforming the text format in the last scenario, is evidence that the text format suffered from a stronger dimension reduction than the table format. This increased partial dimension reduction across scenarios for the text format is well-illustrated by the much faster increased measurement error across scenarios among those receiving the text format compared to those receiving the table format. Respondents reading the text format evaluated the attributes with increasing measurement error over scenarios ($\varepsilon_{\text{scenario 2 – scenario 1}}$ = 4,390,654, χ^2 = 5.47, $p<0.05$; $\varepsilon_{\text{scenario 3 – scenario 2}}$ = 5,815,856, χ^2 = 6.63, $p<0.01$; $\varepsilon_{\text{scenario 4 – scenario 3}}$ = 13,945,123, χ^2 = 21.20, $p<0.01$), whereas respondents who saw the table format remained consistent in the amount of measurement error they produced ($\varepsilon_{\text{scenario 2 – scenario 1}}$ = 2,277,566, χ^2 = 1.44, p=0.23; $\varepsilon_{\text{scenario 3 – scenario 2}}$ = 4,310,024, χ^2 = 4.03, $p<0.05$; $\varepsilon_{\text{scenario 4 – scenario 3}}$ = -297,624, χ^2 = 0.02, p=0.90). Furthermore, this shift occurred even though respondents who resorted to satisficing behavior through refusal, breaking-off, or not responding were already excluded from the dimension reduction analyses. Consequently, even

Table 4 Parameter differences between text and table format predicting salary with the dimensions, for the single conjoint experiment

	Parameter differences (table – text)			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Dimensions				
Female	-798 (516)	-754 (550)	-1,148^{+} (599)	694 (668)
Two children	601 (521)	-183 (552)	226 (616)	389 (669)
10 years of experience	213 (518)	-81 (552)	-1,311* (606)	2,001** (661)
High effort	-2,117*** (522)	-1,918*** (554)	-1,547* (609)	6,611*** (667)
Medium salary	-180 (605)	434 (666)	421 (750)	-460 (785)
High salary	-610 (661)	459 (689)	328 (723)	2,596** (845)
Others earn 25,500–49,600 SEK	-707 (513)	927^{+} (553)	673 (623)	1,636* (671)
Constant	29,383*** (533)	30,776*** (558)	30,149*** (651)	37,554*** (756)
Error variance, table	24,141,723*** (1,279,509)	26,419,283*** (1,400,219)	30,729,307*** (1,628,650)	30,431,683*** (1,612,876)
Error variance, text	22,937,796*** (1,206,417)	27,328,450*** (1,437,344)	33,144,306*** (1,743,230)	47,089,429*** (2,476,676)
Error variance difference (table – text)	1,203,927	-909,167	-2,414,999	-16,657,746***
χ^2 of difference	0.47	0.21	1.03	31.77
Observations	1,435	1,435	1,435	1,435

Notes. Regression coefficients from four ordinary least squares (OLS) regression equations, standard errors (SEs) in parentheses. Positive parameters mean that the table format outperformed the text format in predicting the person's salary, whereas negative parameters mean that the text outperformed the table format. Omitted dimensions were "no children," "5 years of experience," "low effort," "low salary," "others earn 19,000–57,500 SEK." Only the respondents who answered all four scenarios were included. See SOM S2.1., Table S9, for the parameters, separately for presentation format and scenario.

$^{+}$p<0.1, *p<0.05, **p<0.01, ***p<0.001.

though the text format outperformed the table format in the first scenario, the win was short-lived and, overall, the table format produced less dimension reduction.

Hence, for the single conjoint experiment, there was no clear support for the hypothesis that the table format would outperform the text format in terms of data quality (H1b). With regard to refusals, break-offs, non-responses, and total non-response, there was no support for that hypothesis. For partial dimension reduction, the results depend on the number of scenarios the researcher wishes to include: if the respondents evaluate one scenario, the data favor text format, but when evaluating more scenarios, table format would be preferred.

Paired Conjoint Experiment

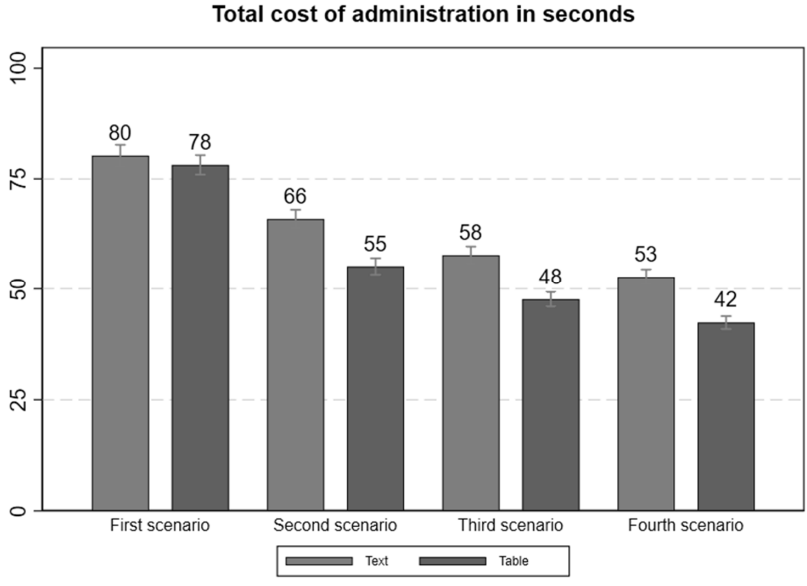
Respondent Experience

Cost of administration. Across the four scenarios, the respondents who evaluated the two political parties after reading the text format took, as hypothesized in H2a, statistically significantly 32 seconds longer to submit their evaluations (M=255 seconds, SD=94) than the respondents who evaluated the parties using the table format (M=223 seconds, SD=89; b=-32.19, SE=4.06, p<0.001).

In contrast to the single conjoint experiment, the difference in administration time between the text and the table version was the smallest for the first scenario. The differences between the formats subsequently increased as the respondents evaluated more scenarios (see Figure 6). In the second, third, and fourth scenario, the differences in administration time between the text and the table formats were statistically significantly shorter for the table format than for the text format ($b_{\text{table * scenario 2}}$ = -8.66, SE= 1.69, p<0.001; $b_{\text{table * scenario 3}}$ = -7.98, SE=1.71, p<0.001; $b_{\text{table * scenario 4}}$ = -8.31, SE=1.70, p<0.001) (see SOM S2.3., Table S10). As will be shown in the analyses of partial dimension reduction (see section *Moderating effects*), in contrast to the single conjoint experiment, the shortening of processing time for the paired conjoint experiment did not correspond to a stronger dimension reduction. Therefore, the shorter time to process the table presentation format seems to have been preferable over the, in total, longer processing time of the text presentation format.

Evaluation and participation in subsequent waves of the Swedish Citizen Panel. In contrast to the single conjoint experiment, the presentation format in the paired conjoint experiment had no significant effect on overall survey evaluation or participation in subsequent waves.

Respondents who received the table format did not evaluate the questionnaire in a significantly more positive or negative way; nor did they evaluate the separate evaluation question significantly differently than the respondents who read the text format. Male respondents evaluated the questionnaire in a significantly more positive way than female respondents (b=0.03, SE=0.01, p<0.001). In contrast to the



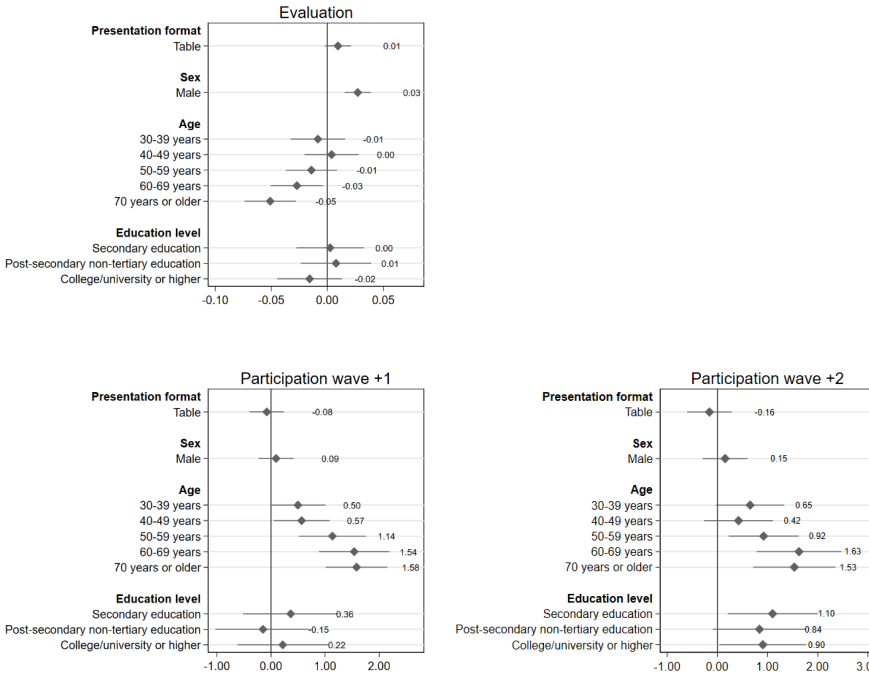
Notes. N=2,037. Respondents who answered the four scenarios and whose response times were not longer than 1.5 times the interquartile range (IQR) were included in the analyses (N excluded = 124).

Figure 6 Cost of administration, in seconds, for the paired conjoint experiment.

single conjoint experiment, in the paired conjoint experiment older respondents in the 60–69-year and over 70-year age groups gave an overall more negative survey evaluation compared to respondents under 29 years old (see Figure 7).

Respondents who read the table presentation format were not significantly more or less likely to participate in subsequent waves of the Swedish Citizen Panel following the presentation format experiment (see Figure 7). Again, the probability of participation in subsequent waves of the Swedish Citizen Panel was significantly higher among older respondents in the paired conjoint than in the single conjoint experiment. Participation in the second wave was also significantly more likely among respondents who had an upper secondary education (b=1.10, SE=0.46, p<0.01) or college/university education (b=0.90, SE=0.44, p<0.01) (see SOM S2.3., Table S12).

Therefore, the overall result provides only partial support for H2a. Respondents on average took less time to complete the survey when seeing the table compared to text, but there was no support for a difference in satisfaction with the survey or future participation depending on the presentation format.



Notes. N=2,125 (Evaluation); N=1,536 (Participation wave +1); N=965 (Participation wave +2). The number of observations differs in the three panel waves because panelists were randomly sampled to be invited to complete the panel wave or not. Regression coefficients (gray diamonds) from one ordinary least squares regression (OLS) and two logistical regressions with their respective 95% confidence intervals (CIs) (gray solid lines). A positive value indicated higher overall evaluation (Evaluation) or a higher likelihood of participation in subsequent waves (Participation wave +1 and Participation wave +2). Baseline categories were female, 18–29 years of age, and compulsory education (9 years).

Figure 7 Respondent experience in terms of overall questionnaire evaluation and participation in the subsequent waves, for the paired conjoint experiment.

Data Quality

Table 5 presents a descriptive summary of sample sizes and answer behaviors at the respondent level in the paired conjoint experiment. Descriptively, the text presentation format caused more respondents to break-off or refuse to evaluate the scenarios compared to the table presentation format, 292 (27.6%) respondents in the text format compared to 224 (19.9%) respondents in the table format (see Table 5).

The most common satisficing answering behavior for respondents presented with the text presentation format was to break-off. For respondents presented with the table presentation format, however, providing an invalid answering behavior,

such as providing no variation in their answers, was the most common satisficing answering behavior (see Table 5).

The results presented at the scenario level further show that the total loss of information (total non-response) seemed descriptively greater with the text presentation format compared to the table presentation format, 1,687 (39.8%) incorrectly judged scenarios in the text presentation format compared to 1,398 (31%) in the table format (see Table 6).

Refusals, break-offs, non-responses, and total non-response. When being presented with the table presentation format, respondents were statistically significantly less likely to adopt a refusal answering behavior, that is, to not answer, to constantly provide "don't know" answers, or to not vary their responses across the four scenarios (b=-0.29, SE=0.12, p<0.01), compared to when receiving the text

Table 5 Sample sizes and answer behavior at the respondent level, for the paired conjoint experiment.

Experiment setting	Sample size	No evaluated scenarios (refusal, don't know) (1)	Invalid constant answer behavior (2)	Refusals (1) + (2)	Break-off (3)	Total (1) + (2) + (3)
Text	1,058	98 (9.2%)	72 (6.8%)	170 (16.1%)	122 (11.5%)	292 (27.6%)
Table	1,126	60 (5.3%)	84 (7.5%)	144 (12.8%)	80 (7.1%)	224 (19.9%)
Sum	2,159	158 (7.3%)	156 (7.2%)	314 (14.5%)	202 (9.4%)	516 (23.9%)

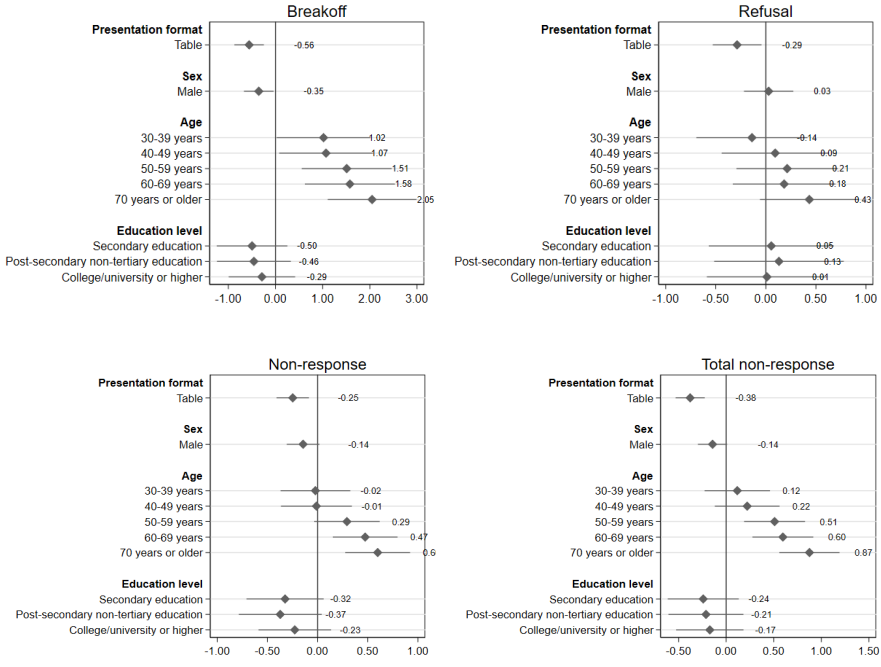
Notes. Results at the respondent level.

Table 6 Sample sizes and answer behavior at scenario level, for the paired conjoint experiment.

	Gross sample size (N = 2,159)			Net sample size after excluding refusals and break-offs (N = 1,643)	
Experiment setting	Gross sample of scenarios	Refusals (1) + (2)	Break-off (3)	Non-response (4)	Total non-response (1) + (2) + (3) + (4)
Text	4,232	680 (16.1%)	488 (11.5%)	519 (12.2%)	1,687 (39.8%)
Table	4,504	576 (12.8%)	320 (7.1%)	502 (11.1%)	1,398 (31%)
Sum	8,636	1,256 (14.5%)	808 (9.4%)	1021 (11.8%)	3,098 (35.8%)

Notes. Results at scenario level. Gross sample of scenarios is calculated by multiplying the group size of an experimental setting by the set size (4 scenarios per respondent).

presentation format. The significant effect of presentation format on refusals was found in both the bivariate analysis and when including controls (see Figure 8 and SOM S2.4., Tables S13 and S14). No other significant effects on refusals were found.



Notes. N=1,818 (Break-off); N=2,125 (Refusals); N=7,129 (Non-response); N=8,489 (Total non-response). Regression coefficients (gray diamonds) from four logistical regressions with their respective 95% confidence intervals (CIs) (gray solid lines). A positive value indicates a higher likelihood of break-off, refusal, non-response, or total non-response. Baseline categories are female, 18–29 years of age, and compulsory education (9 years). Results on break-offs and refusals are at the respondent level while results on non-response and total non-response are at the scenario level. We controlled within-participant clustering for non-response and total non-response using cluster-robust standard errors.

Figure 8 Data quality in terms of break-off, refusals, non-responses, and total non-response, for the paired conjoint experiment.

Similarly, respondents who saw the table presentation format were significantly less likely to start giving valid evaluations and then switch to a constant non-valid answer behavior (break-off) compared to respondents presented with the text format (b=-0.56, SE=0.16, p<0.001). Furthermore, male respondents were significantly less likely to breaking-off than women (b=-0.35, SE=0.16, p<0.01). The

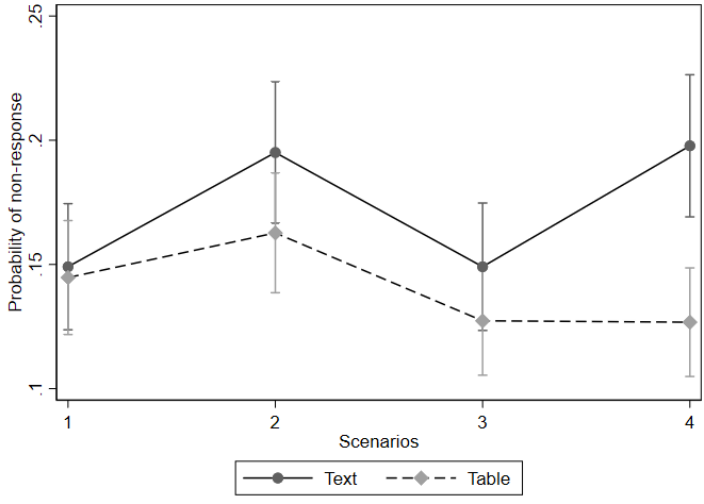
probability of a break-off increased with the age of the respondents in an almost linear fashion, where the older the respondent the higher the probability of breaking-off (see Figure 8).

The table presentation format also performed better with regard to non-response (i.e., alternating between validly judging and not validly judging scenarios). Respondents who saw the table format were less likely to resort to a non-response behavior (b=-0.25, SE=0.08, p<0.01) compared to those seeing the text format, when controlling for gender, age, and education, on non-response. Furthermore, the effect was moderated by the number of scenarios, where the adverse effect of using text format became evident only in the last scenario, where there were statistically significantly fewer non-responses among the table format respondents ($b_{\text{table * scenario 4}}$ = -0.50, SE=0.18, p<0.01).

The predictive probability of a non-response in the fourth scenario was 0.20 for the text format and 0.13 for the table format (see Figure 9). Respondents who adopted a refusal or breaking-off behavior were excluded. In line with previous results, older respondents were more likely to exhibit a non-response behavior, that is, to switch/alternate between validly answering and not validly answering a scenario.

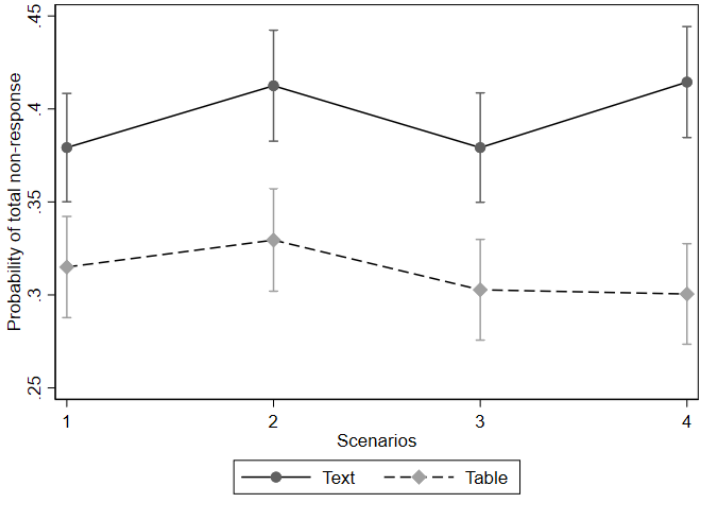
The total loss of information was, as expected, lower for the table format than for the text format (b=-0.38, SE=0.08, p<0.001). Again, the effect of presentation format on total non-response in the paired conjoint experiment was found to depend on the number of scenarios that the respondent had answered. Across all scenarios, total non-response was greater for the text format, but as the respondent evaluated more and more scenarios, the effect got stronger. Albeit the increased number of total non-response caused by the text format was not yet significant in the second and third scenario ($b_{\text{table * scenario 2}}$ = -0.07, SE=0.07, p=0.31; $b_{\text{table * scenario 3}}$ = -0.06, SE=0.08, p=0.45), by the fourth scenario this moderating effect became great enough to reach statistical significance ($b_{\text{table * scenario 4}}$ = -0.22, SE=0.08, p<0.001), see Figure 10.

As expected, there was a negative effect of age on total non-response, indicating that older respondents found the conjoint experiments more demanding and adopted satisficing strategies more often compared to younger respondents (see Figure 8).



Notes. N=6,505 (scenario level).

Figure 9 Predicted probabilities of scenario on non-response over presentation format, for the paired conjoint experiment.



Notes. N=8,489 (scenario level).

Figure 10 Predicted probabilities effect of scenario on total non-response over presentation format, for the paired conjoint experiment.

Moderating effects. Potential moderating effects of education and age were also tested in the paired setting, with total non-response as the dependent variable, age and education as moderators, and gender as a control variable (see SOM S2.4., Figures S5 and S6, for an illustration of the moderating effects).

Education. In line with the results from the single conjoint experiment, the interaction term between education and presentation format was not significant in the paired conjoint experiment. Hypothesis 2c was, therefore, not supported.

Age. The results for age as a moderator were similar to those for education. The effect that presentation format had on total non-response did not on average significantly differ between age groups. There was a small and significant effect for 60-to-69-year old's of seeing the text format on total non-response (see Figure S6), which then disappeared for respondents older than 69 years. Overall, our results did not support the hypothesis that the effect of presentation format was moderated by age (H2d).

Response inconsistency and partial dimension reduction. In contrast to the single conjoint experiment, respondents in the paired conjoint experiment who were given the vignette in a table format yielded a statistically significantly stronger prediction for one of the dimensions in the first scenario (i.e., they had less response inconsistency) ($b_{\text{education: more than high school, not university}} = -0.14$, SE=0.06, $\chi^2(1, 1{,}932) = 5.13$, $p<0.05$).[5] Furthermore, in the first scenario, the table format produced less measurement error ($\varepsilon = 0.21$, SE=0.00) than the text format ($\varepsilon = 0.23$, SE=0.00, $\chi^2(1, 1{,}932) = 6.80$, $p<0.01$) (see Table 7, column 1). In the subsequent scenarios, the parameter differences between the two presentation formats stabilized, with none of the parameters differing in scenario 2 (see Table 7, column 2), two differing in favor of the text format and one in favor of the table format in scenario 3 (see Table 7, column 3), and one favored the text and one favored the table format in scenario 4 (see Table 7, column 4).

A similar relationship was found for the error variance, where scenario 2 showed no difference in error variance, a statistically significant difference in favor of the table format in scenario 3, and no difference in scenario 4 (see Table 7, columns 2–4). Hence, the text format seems to have produced less response inconsistency in the first scenario; however, across all four scenarios, the response inconsistency analysis did not favor either of the presentation formats.

Furthermore, respondents did not evaluate the dimensions with increasingly less care over scenarios (i.e., adopted a partial dimension reduction behavior). This applied to both those who saw the questions in text format ($\varepsilon_{\text{scenario 2 – scenario 1}} = 0.01$, $\chi^2 = 0.63$, p=0.43; $\varepsilon_{\text{scenario 3 – scenario 2}} = 0.00$, $\chi^2 = 0.07$, p=0.78; $\varepsilon_{\text{scenario 4 – scenario 3}} = -0.01$, $\chi^2 = 1.31$, p=0.25) and those who saw them in table format ($\varepsilon_{\text{scenario}}$

5 In the paired conjoint experiment, negative relationships between dimensions and the dependent variable were expected. Hence, the presentation format that produced the most negative coefficient was interpreted as the better performing format.

$_{2 - \text{scenario } 1} = 0.01$, $\chi^2 = 0.13$, p=0.71; $\varepsilon_{\text{scenario } 3 - \text{scenario } 2} = 0.02$, $\chi^2 = 2.37$, p=0.12; $\varepsilon_{\text{scenario } 4 - \text{scenario } 3} = 0.01$, $\chi^2 = 0.02$, p=0.88). Therefore, in contrast to the single conjoint experiment, neither of the presentation formats produced strong evidence of dimension reduction in the paired conjoint experiment.

However, overall, H2b was supported by the majority of the evaluation criteria on data quality. The table presentation format resulted in fewer refusals, break-offs, non-responses, and total non-responses, albeit no strong evidence for a stronger partial dimension reduction or inconsistency of responses was found for either of the two formats.

Table 7 Parameter differences between text and table format when predicting party choice with the attribute dimensions, for the paired conjoint experiment.

	Parameter differences (table – text)			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Party level				
Status quo, tax and welfare	0.04	0.05	0.12*	0.05
	(0.05)	(0.06)	(0.06)	(0.06)
Less tax, less welfare	-0.02	0.02	-0.00	-0.03
	(0.05)	(0.05)	(0.05)	(0.05)
Status quo of refugees	-0.03	0.04	-0.01	-0.01
	(0.05)	(0.06)	(0.06)	(0.05)
More refugees	-0.10+	-0.01	-0.12*	-0.12*
	(0.05)	(0.05)	(0.05)	(0.05)
Status quo of gender roles	-0.03	0.07	0.08	0.09
	(0.05)	(0.06)	(0.06)	(0.05)
Traditional gender roles	-0.07	0.04	0.01	0.05
	(0.05)	(0.05)	(0.05)	(0.05)
Party leader				
Hired unreported workers	-0.01	0.00	0.05	0.06
	(0.05)	(0.05)	(0.05)	(0.05)
Drunk driving	0.01	-0.01	-0.07	-0.04
	(0.06)	(0.05)	(0.05)	(0.05)
Female party leader	-0.06	-0.02	0.05	0.06
	(0.04)	(0.04)	(0.04)	(0.04)
Education: Less than high school	-0.09	0.08 (	-0.03	0.07
	(0.06)	0.06)	(0.06)	(0.06)
Education: High school	-0.01	0.06	0.12*	0.11+
	(0.06)	(0.06)	(0.06)	(0.06)
Education: More than high school, not university	-0.14*	0.04	0.03	0.03
	(0.06)	(0.06)	(0.06)	(0.06)

	Parameter differences (table – text)			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Constant	0.76*** (0.05)	0.95*** (0.06)	0.84*** (0.06)	0.91*** (0.05)
Error variance, table format	0.21*** (0.00)	0.23*** (0.00)	0.22*** (0.00)	0.22*** (0.00)
Error variance, text format	0.23*** (0.00)	0.22*** (0.01)	0.23*** (0.00)	0.23*** (0.00)
Error variance difference (table – text)	-0.02**	0.01	-0.01*	-0.01
$\chi 2$ of difference	6.80	0.26	4.04	1.84
Observations	1,932	1,932	1,932	1,932

Notes. Regression coefficients from four ordinary least squares (OLS) regression equations, with clustered robust standard errors (SEs) in parentheses nested within respondents. Negative parameters indicate where the table format outperformed the text format in predicting the party choice, whereas positive parameters indicate where text outperformed the table format. Omitted dimensions were "more gender equality," "more welfare, higher taxes," "strive towards more gender equality," "drunk driving scandal," "male party leader," and "education: university." Only the participants who answered all four scenarios were included. See SOM S2.4., Table S15, for the parameters, separately for presentation format and scenario.
$^{+}p<0.1$, $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Summary of Results and Comparison to Shamon et al. (2019)'s Findings

In this paper, we presented a direct replication of the single conjoint presentation format experiment reported in Shamon et al. (2019), albeit with a few changes to the procedure, sample size, and analysis strategies. For a full description of differences compared to Shamon et al. (2019), see SOM S3.1–S3.2.

Our procedure makes some direct comparisons to Shamon et al. (2019) difficult. For example, our experiment evaluated only two presentation formats (text and table format), whereas Shamon et al. (2019) included an additional text format where the dimensions were underlined. We opted not to include the underlined presentation format because respondents have been reported to interpret the underlining of text in inconsistent ways (Reber, Schwarz & Winkielman, 2004), and underlining can result in more random measurement error (see Reber, Schwarz & Winkielman, 2004; Song & Schwarz, 2008).

Furthermore, presenting respondents with only four scenarios to evaluate (instead of 16, as in Shamon et al., 2019) made evaluations of consequent dimension

reduction unfeasible. However, decreasing the number of scenarios enabled us to heavily increase the statistical power of most of our other analyses (2,068 participants in the single conjoint experiment in our study compared to 498 in Shamon et al., 2019). Increased statistical power enabled us to identify whether Shamon et al.'s (2019) directional, albeit not statistically significant, effects of the text presentation format on decreased data quality were due to statistical power.

Despite some differences in design and analysis, the results in the single conjoint setting replicated several of the findings in Shamon et al. (2019). For example, despite our larger sample size, we also found no significant differences in terms of refusals or break-offs between the text and the table presentation format. Furthermore, Shamon et al. (2019) found no interaction effects between age and presentation format, which agrees with our findings.

However, some results found in this study did not replicate the findings reported by Shamon et al. (2019). For instance, in stark contrast to Shamon et al. (2019) who found no difference in partial dimension reduction and response inconsistency, we found that the table format statistically significantly outperformed the text format in less partial dimension reduction and less response inconsistency as the number of evaluated scenarios increased. Furthermore, Shamon et al. (2019) found no differences in the cost of administration, in terms of administration time, whereas our results favored the table presentation format. These differences may be due to the extra statistical power afforded by our sampling strategy.

In addition, Shamon et al. (2019) found that the text presentation format significantly showed decreased non-responses while the table format yielded fewer total non-responses, whereas we found no such differences.

The present paper, moreover, extended Shamon et al.'s (2019) work by conceptually replicating their experimental design in a paired conjoint experimental setting. Our conceptual replication produced even clearer evidence in favor of the table format. The table format outperformed the text presentation format by reducing the cost of administration and lowering refusal, break-off, non-response, and total non-response rates. The effect of presentation format on dimension reduction was, however, inconclusive.

Furthermore, the present paper extends Shamon et al.'s (2019) analyses by including two additional measurements of respondent experience, namely, respondents' evaluation of the questionnaire and participation in the waves of the Swedish Citizen Panel following the presentation format experiment. In contrast with other findings presented here, these additional measurements favored the text presentation format in terms of participation in subsequent waves but the table presentation format in terms of overall questionnaire evaluation in the single conjoint experiment. The presentation format had no significant effect on the evaluation of the questionnaire or participation in the panel waves following the experiment in the paired conjoint experiment.

Conclusion and Discussion

This paper investigated the impact that the presentation format (text or table) had on respondents' answering behavior by replicating Shamon et al.'s (2019) study on single conjoint experiments, as well as extending their work to also include paired conjoint experiments, where respondents state their preferences over two dimension sets/profiles.

Overall, the results in the present study favored the table over the text presentation format. As evidence of this, the table presentation format in both the single conjoint and the paired conjoint setting was found to statistically significantly outperform the text presentation format with regard to the cost of administration (i.e., the time it took respondents to evaluate the scenarios). However, a shorter administration time may, in fact, not be favorable if it is shorter because respondents answer faster by employing a suboptimal response process and satisficing strategies. Even though the respondents who were presented with the table format took less time to evaluate the scenarios, this shorter processing time did not clearly stem from less cognitive effort invested in the response.

Although respondents in the single conjoint setting produced stronger loadings on the dimensions in the first scenarios when reading the text instead of the table format, the respondents who read the text format suffered ever stronger partial dimension reduction (i.e., a decreasing impact of the dimensions and increased measurement error over the number of scenarios) as they evaluated more scenarios. In fact, by the fourth scenario, the table format had started producing stronger dimension loadings and significantly less measurement error than the text format. Hence, when respondents will evaluate only one scenario, the text format may be preferable, but as the number of scenarios increases, the table format seems to produce better, and more consistent, data quality. Our finding may have stemmed from the fact that respondents became fatigued more quickly by the text than by the table format, although the present study does not have the type of data that provide evidence for such a claim.

Similarly, respondents who saw the table presentation format in the paired conjoint setting evaluated the scenarios faster than respondents who read the questions in text format and did so without introducing partial dimension reduction or response inconsistency. Furthermore, in the paired conjoint setting, the table format outperformed the text format in other data quality measures, such as the number of refusals, break-offs, non-response, and total non-response. Overall, we found more distinct support for the table format in the paired conjoint setting compared to the single conjoint setting. The stronger evidence in the paired setting may be due to the presentation format having a greater impact when respondents evaluated two profiles or from the difference in topics between the single and the paired conjoint

experiments. Future studies that alternate topics on the single and paired conjoint settings to bring clarity on how sensitive the results are to the topics chosen.

The proposed theoretical benefit of the text format was that nesting the information within stories was thought to enhance respondents' understanding and empathy of the hypothetical situation. The increased understanding was, in turn, thought to increase the respondents' attention to the dimensions and increase the quality of the data. Furthermore, theoretically, respondents may be more likely to be accustomed to absorbing information in text paragraphs rather than tables. In contrast to these theories, our findings offer no support for any of these claims. Rather, respondents seem to connect to the information in the table emphatically and interpret the tables accurately, even when those tables are presented with two sets of profiles, which should have increased the complexity of the information to absorb.

Moreover, the present manuscript did not use any visual emphasis on the dimensions in the text vignettes (e.g., underlining, italicizing, or using bold fonts). Emphasizing the text that represented the dimension might have helped respondents to focus on the most relevant pieces of the vignette texts and could have made the text format perform better than what we found. However, we believe it to be unlikely that adding a visual emphasis would have negated our results because previous research has found that emphasis can make texts more difficult to read and understand (Reber, Schwarz, and Winkielman, 2004; Song and Schwarz, 2008), and emphasis can increase the time it takes respondents to evaluate conjoint vignettes (Shamon et al., 2019).

Counterintuitively, in the single conjoint setting, respondents were found to be more satisfied when receiving the table format, while not producing better data quality compared to the respondents who received the text format. By contrast, in the paired conjoint setting, respondents who were given the table format were not more satisfied with the questionnaire but produced statistically significantly better data quality compared to the respondents who were given the text format. A potential explanation for this dissimilarity may be that the most dissatisfied respondents stop filling out the questionnaire before getting to the questionnaire evaluation questions, leading to artificially greater satisfaction ratings for the worst performing presentation format (greater, because only more satisfied respondents answer the questionnaire evaluation questions). However, whereas we did find greater satisfaction among table format respondents in the single conjoint setting, we did not observe more refusals, break-offs, non-response, or total non-response for either of the presentation formats in the single conjoint.

We did observe more refusals, break-offs, non-response, and total non-response, but no differences in respondent satisfaction, in the paired conjoint setting. The only instance where break-offs, refusals, non-responses, and total non-response could artificially produce the satisfaction ratings we found would be if

respondent satisfaction among those presented with text format started at lower levels than among those receiving the table format. The artificial increase in satisfaction afforded by the break-offs, refusals, non-responses, and total non-response would then bring the mean satisfaction with text format to the same levels as satisfaction with the table format. However, random assignment of the two formats should limit such an outcome. Perhaps, rather than thinking of the findings as counterintuitive, the results of this study indicate that respondent satisfaction and data quality are two distinct phenomena, each offering different insights and advice for survey researchers. Survey researchers should be interested in both phenomena, but if forced to choose, better data quality should be preferred over respondent satisfaction, especially as respondents seem able to be unsatisfied with a questionnaire while still more likely to complete each conjoint evaluation.

Lastly, both our and Shamon et al.'s results (2019) were based on online convenience samples. Online convenience samples have been found to be more suitable for generalization of treatment effects than, for example, student samples (e.g., due to being more diverse in educational attainment, age, gender, and income of the respondents, see Berinsky, Huber & Lenz, 2012). Hence, the non-difference between text and table formats found in Sauer et al., (2020) could be due to their student sample being more accustomed to reading lengthier text paragraphs than a general population sample. In contrast, the chance remains that our self-selected panelists and those in Shamon et al. (2019) may be more literate in reading tables than the general population, potentially producing the outperforming of the table format in our experiments. Future research should attempt to replicate similar presentation format experiments among probability sampled respondents.

In the meantime, based on the results of this study and those reported in Shamon et al. (2019), we conclude that respondents simply seem less likely to resort to satisficing strategies when evaluating conjoint experiments using a table presentation format than when evaluating them in a text format. For now, we argue that a table presentation format is to be preferred when designing conjoint experiments distributed online.

References

Auspurg, K., & Hinz, T. (2015). *Factorial Survey Experiments (Quantitative Applications in the Social Sciences)*. Thousand Oaks, CA: SAGE Publications.

Auspurg, K., & Jäckle, A. (2017). First equals most important? Order effects in vignette-based Measurement. *Sociological Methods & Research*, 46(3):490–539.

Bansak, K., Hainmueller, J., Hopkins, D., & Yamamoto, T. (2021). Beyond the breaking point? Survey satisficing in conjoint experiments. *Political Science Research and Methods*, 9(1), 53–71. Doi:10.1017/psrm.2019.13

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk. *Political Analysis*, *20*(3), 351–368. https://doi.org/10.1093/pan/mpr057

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, 112(8), 2395–2400.

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, 22(1), 1–30.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. Applied cognitive psychology, 5(3), 213-236.Krosnick, J. A. (1999) Survey research. *Annual Review of Psychology, 50*(1), 537–567.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*(2), 201-219.

MacDonald, K. (2016) Group comparisons in structural equation models: Testing measurement invariance. *The Stata Blog*. Accessed December 17th, 2020, at https://blog.stata.com/2016/08/23/group-comparisons-in-structural-equation-models-testing-measurement-invariance/

Mahon-Haft, T. A., & Dillman, D. A. (2010). Does visual appeal matter? Effects of web survey aesthetics on survey quality. *Survey Research Methods*, 4(1), 43–59.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251).

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, 8(4), 364–382.

Roberts, C., Gilbert, E., Allum, N., & Eisner, L. (2019). Research synthesis: Satisficing in surveys: A systematic review of the literature. *Public Opinion Quarterly, 83*(3), 598–626.

Sato, H., Kubo, M., & Namatame, A. (2007). Video-based conjoint analysis and agent based simulation for estimating customer's behavior. *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 1102-1111). Springer, Berlin, Heidelberg.

Sauer, C., Auspurg, K., & Hinz, T. (2020). Designing Multi-Factorial Survey Experiments: Effects of Presentation Style (Text or Table), Answering Scales, and Vignette Order. *Methods*, *data*, *14*(2) Pages. https://doi.org/10.12758/MDA.2020.06

Schuman, H., & Presser, S. (1996). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Sage.

Shamon, H., Dülmer, H., & Giza, A. (2019). The Factorial Survey: The Impact of the Presentation Format of Vignettes on Answer Behavior and Processing Time. *Sociological Methods & Research*. https://doi.org/10.1177/0049124119852382

Simon, H. A. (1957). *Models of man; social and rational*. Wiley

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. *Psychological Science*, 19(10), 986–988. https://doi.org/10.1111/j.1467-9280.2008.02189.x

Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, MA.