Quadratic Survey Interface

ANONYMOUS AUTHOR(S)*

Here is the abstract that cites [50] and Posner and Weyl [50].

CCS Concepts: • **Human-centered computing** → **Empirical studies in collaborative and social computing**; **Collaborative and social computing design and evaluation methods**; *HCI design and evaluation methods.*

Additional Key Words and Phrases: Quadratic Voting; Likert scale; Empirical studies; Collective decision-making

ACM Reference Format:
Anonymous Author(s). 2024. Quadratic Survey Interface. *Proc. ACM Hum.-Comput. Interact.* 1, 1 (June 2024), 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Capturing individuals' responses, attitudes, and preferences effectively is the cornerstone of studying human subject studies, especially for the CSCW community. The effectiveness of eliciting these responses hinges upon the study protocol, survey mechanism, and design of the tool at hand [46, 14, 34]. While much research has explored the influence of the former two aspects, this research focuses on the design of a specific survey – Quadratic Surveys.

The design in any response-capturing tool significantly influences individuals' ability to express their attitudes. Political scientists have demonstrated that ballot designs alone can sway voter decisions [18], marketing and psychology researchers have examined how the presentation of questions influences responses [74, 35, 71], and Human-Computer Interaction researchers have focused on evaluating and understanding web surveys and smart interfaces for surveys [20, 78, 49]. These studies highlight the importance of studying the interface and design for survey mechanisms.

The Quadratic Mechanism is a decision mechanism where individuals express the degree of their preferences within a given budget. Quadratic Voting (QV) leverages this mechanism, allowing participants to allocate a finite amount of credits across a list of options, voting multiple times to demonstrate their strength of approval, as long as the sum of the quadratic values of their votes remains within the given credit [37]. Recent work has demonstrated that QV can gauge public opinions [53] and be transformed into Quadratic Surveys (QS) to elicit individual preferences under resource-constrained scenarios [10]. These studies suggest that more sophisticated mechanisms can elicit more truthful and carefully identified preferences in complex spaces to inform decision-making.

However, the Quadratic Mechanism is undeniably more complex than other voting and surveying mechanisms, such as the Likert scale [40], where individuals select from a few responses, and Approval Voting [5], where participants mark as many options as they approve without constraints. Responding to a QS involves expressing a numerical representation of a full set of constructed preferences. As Lichtenstein and Slovic [39] pointed out, when individuals do not have clear known

2573-0142/2024/6-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

preferences, they construct preferences in situ. This can be particularly challenging in unfamiliar contexts, when choices might conflict, and when opinions need to be translated into quantified values. Survey respondents often face these challenges, exacerbated by common constraints and budget allocation, requiring them to make difficult decisions.

This challenge emphasizes the critical importance of designing and developing suitable interfaces for Quadratic Surveys to elicit truthful and in-depth preference information from respondents. Good design is essential; without it, the quality of collected data can suffer significantly. Despite the advocacy of Quadratic Voting by Posner and Weyl [50], and its experimentation in various contexts such as the Colorado state government, the Democratic Caucus of the House of Representatives [52], government-sponsored hackathons [67], and the recent Gov4git [27], no peer-reviewed research has focused on the design perspective of such mechanisms. This increasing attention highlights the relevance and potential impact of QS. Additionally, prior research in behavioral economics and marketing has pointed out the challenge of choice overload [33] and overchoice [26]. While Quadratic Voting allows individuals to allocate resources across multiple options, the presence of many options can overwhelm participants, potentially compromising decision-making quality. It can be difficult for decision-makers to reduce the number of options present in a survey. Effective design may mitigate these overload challenges, ensuring that the Quadratic Survey mechanism fulfills its potential to capture detailed and accurate preferences. These reasons strongly motivate our main research question: *How can we design interactive interfaces to support participants in completing Quadratic Surveys?* Addressing this question fills a important gap in the literature and enhances the practical utility of QS in capturing high-quality data across various applications.

In this study, we aim to reach our goal by answering three research questions:

- RQ1a. How does the number of options on QS impact respondents' cognitive load?
- RQ1b. Across the different number of options on QS, what are the sources of cognitive load from?
- RQ2a. How does the interactive interface involving grouping and direct manipulation interface influence QS respondents' cognitive load compared to text-based interface?
- RQ2b. Across the two interfaces, what are the sources of cognitive load from?
- RQ3. What are differences in QS respondents' behaviors when coping with long lists of options?

To answer these research questions, we constructed an interactive interface informed by prior literature in questionnaire and survey response format. We iterated the interface after several rounds of pretests and pilots. Then we designed a two-by-two between-subject study where each group of participants would experience a QS with a short or long list of options, using a text-based or interactive-based interface. Study participants' cognitive load was collected using NASA-TLX, a widely used measure, followed by an interview. We recruited 41 participants from a Midwestern local community, asking members to support a wide range of societal issues in an in-lab study.

In the remainder of this paper, we focus on the related works in section 2. Then we detailed related works that informed the interactive QS interface and the design process in section 3. Experiment design follows in section **??**. Study findings and discussion are present in section 5 and sectionsec:discussion.

Contributions. In this study, we highlighted the importance of using a two-step "organize then vote" interactive interface for QS surveys with long lists of options, despite not revealing statistically significant differences in the weighted NASA-TLX scores between the groups. All four groups of participants experienced medium-to-high weighted cognitive load. Our recommendation is based on identified differences in the sources of cognitive load between participants using text-based interfaces and those using interactive interfaces when responding to QS with long lists. Supporting evidence from participants' operating behaviors showed that those using the interactive interface made more rapid and repeated updates when responding to the survey. Conversely, we observed fewer qualitative differences between the groups of participants responding to the short list. This study also provides design recommendations for deploying QS and suggests directions for future research.

2 RELATED WORKS

2.1 Quadratic Voting

A Quadratic Survey (QS) is a surveying technique transformed from Quadratic Voting (QV), sharing the same quadratic mechanism used to inform collective decision-making. This mechanism allows respondents to express their preference intensity by casting multiple votes at a quadratic cost. Made popular by Posner and Weyl [50] and Lalley and Weyl [37], it aims to mitigate the tyranny of the majority inherent in traditional one-person-one-vote systems. QV is not subject to Arrow's impossibility theorem as it does not require individuals to aggregate rankings of preferences. Quadratic Surveys adapt this mechanism for survey contexts, allowing participants to vote for or against an option, presenting two distinct choices in the same survey. This adaptation was utilized by Quarfoot et al. [53] and implemented as an open-source platform by Bassetti et al. [4]. While these studies did not explicitly label this as a 'quadratic survey', we use this term to differentiate it from the voting mechanism.

To formally define QV, in a scenario where S participants are involved, each participant is allocated a fixed quantity of voice credits, denoted as B. These credits can be distributed among various options. Importantly, each individual can cast multiple votes, either in favor of or against each option. However, this voting system incorporates a quadratic cost for voting: casting n_k votes for a particular option k incurs a cost $c(n_k)$, which is proportional to n_k^2. Consequently, the aggregate cost in voice credits for all options chosen by a participant must not exceed their allocated budget B. This necessitates that the sum of the squares of votes cast for each option ($\sum_k n_k^2$) remains within the limit of B, where n_k represents the number of votes allocated to option k. QV results determine the winner by summing up the total votes cast by all participants for each option. This design allows the marginal cost to cast one additional vote to linearly increase with the number of votes already cast on that option, inducing rational participants to vote proportionally to how much they care about an issue [50]. Quadratic Surveys extend the same mechanism but allow participants to denote positive (upvotes) or negative values (downvotes) on each option. The survey administrators compile and analyze the results by summing up the total votes and allowing cancellation between upvotes and downvotes.

Empirical studies and applications of the quadratic mechanism and quadratic voting have increased in the past few years. Several studies have explored the empirical use cases for QV, including Quarfoot et al. [53]'s study on 4500 participants' attitudes across ten public policies, highlighting differences between QV and Likert scale survey results. Cheng et al. [10] applied quadratic surveys in Human-Computer Interaction (HCI) and subsequently showed QV's effectiveness in reflecting true preferences in monetary decision tasks. Naylor et al. [44] used QV in educational research to gauge student opinions on factors affecting university success, and Cavaille and Chen [8] examined QV in polarized choice scenarios.

Another form of research focuses on the transformation and application of the quadratic mechanism into different tools. Recent work by South et al. [63] applies the quadratic mechanism as part of the management framework to support networked authority, which was later applied to Gov4git [27]. Quadratic Funding focuses on the redistribution of funds following outcomes from consensus made using the quadratic mechanism [6, 23]. Yet across these research, there is little attention focused on the user experience and interface design supporting individuals to express attitudes and opinions under the quadratic mechanism.

2.2 Survey, Voting, and QV Design

To emphasize the significance of interface design and understanding user experience, we reviewed prior literature on the influence of systems in surveying and voting. The notorious butterfly ballot [73] highlighted the impact interface design can make on election outcomes. Researchers like Engstrom and Roberts [18], Chisnell [12], and organizations like the Center for Civic Design, which publishes reports like "Designing Usable Ballots" [17], stress that democracy is a design problem. We group this literature into three main categories: designs that shifted voter decisions, designs that influenced human errors, and designs that incorporated technologies to improve usability.

Designs that shifted voter decisions: For example, states without the option for straight-party ticket voting (the option to circle an option that votes for all the candidates in the same party) exhibited higher rates of split-ticket voting [18]. Another example from the Australian ballot with an office block and no party box (having a box that clearly segments the position that the candidates are competing for) has been shown to enhance incumbency advantages.

Designs that influenced errors: Butterfly ballots increased voter errors because voters could not correctly identify the punch hole on the ballot. Splitting contestants across columns increases the chance for voters to overvote [54]. On the other hand, Everett et al. [19] showed the use of incorporating physical voting behaviors, like lever voting, into GUI interfaces.

Designs that incorporated technologies: Other projects like the Caltech-MIT Voting Technology Project have sparked research to address accessibility challenges, resulting in innovations like EZ Ballot [38], Anywhere Ballot [65], and Prime III [16]. In addition, Gilbert et al. [25] investigated optimal touchpoints on voting interfaces, and Conrad et al. [13] examined zoomable voting interfaces.

These related works highlight the importance of design and how it influences elicited individual attitudes. Research in the marketing and research community studying survey and questionnaire design, usability, and interaction finds similar trends. The term 'Response Format' is often used to describe the style and presentation of a question presented on a survey. Various studies have shown that different designs of response formats can influence outcomes. For example, Weijters et al. [75] demonstrated that horizontal distances between options are more influential than vertical distances, with the latter recommended for reduced bias. Slider bars, which operate on a drag-and-drop principle, show lower mean scores and higher nonresponse rates compared to buttons, indicating they are more prone to bias and difficult to use. In contrast, visual analogue scales that operate on a point-and-click principle perform better [70].

While both fields have deep insights into understanding design's influence on attitude elicitation, QS's unique capability of supporting both ranking and rating [10] makes designing an interface important and challenging. Subsequently, this research aims to understand how this interface influences an individual's QS response behavior. Requiring the distribution of budgets following the quadratic mechanism introduces new and complex decisions.

2.3 Cognitive Challenges and Choice Overload

Quadratic Surveys contain a series of complex decisions. QS respondents are asked to express preferences while budgeting across multiple options. Lichtenstein and Slovic [39] laid out the three key elements that make decisions difficult. They include people making decisions within an unfamiliar context, people forced to make tradeoffs due to conflicts among choices, and people quantifying values for their opinions. QS fits into the description of all three elements, as participants can face options placed by the decision maker which they have never seen before. Participants are bounded by budgets that force them to make tradeoffs, and the final votes are presented in values. Hence, we believe that QS introduces high cognitive load.

Previous studies have demonstrated that cognitive overload can adversely affect performance, for instance, causing individuals to rely more on heuristics rather than engaging in deliberate and logical decision-making [15]. In addition, some researchers believe that preferences are constructed in situ just as memories are. Thus, when too much information is presented to an individual, they can 'satisfice' their decisions [62, 48, 72]. This overload can happen because of the presence of too many options. Subsequently, too many options can lead to individuals feeling overloaded, leading to decision paralysis, demotivation, and dissatisfaction [33].

Additionally, Alwin and Krosnick [3] highlighted that the use of ranking techniques in surveys can be time-consuming and potentially more costly to administer. These challenges are compounded when there are numerous items to rank, requiring substantial cognitive sophistication and concentration from survey respondents [21].

However, in several notable applications of Quadratic Voting in society, there can be hundreds of options within a single QV question. For instance, the 2019 Colorado House of Representatives considered 107 bills [1], and the 2019 Taiwan Presidential Hackathon featured 136 proposals [51]. These psychological and behavioral research highlighted the importance of understanding how individuals navigate and can potentially benefit from interfaces under long-list QS conditions.

3 INTERFACE DESIGN

In this study, we developed an interactive interface for QS based on prior literature for the experiment condition. As there no standardlized QV interface, we constructed a text based QS interface that includes most of what existing interfaces offers and make it comparable with the interactive interface. In the following subsections, we describe these interface design decisions and supporting literature.

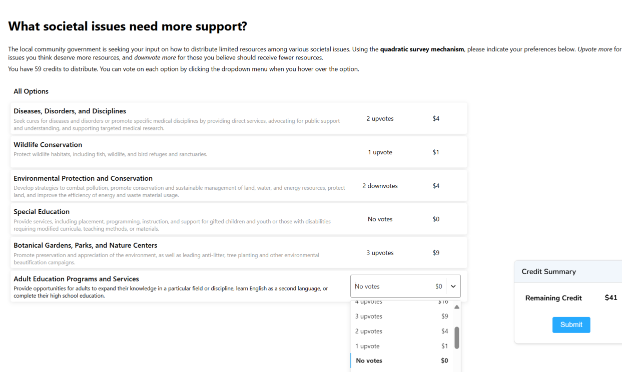
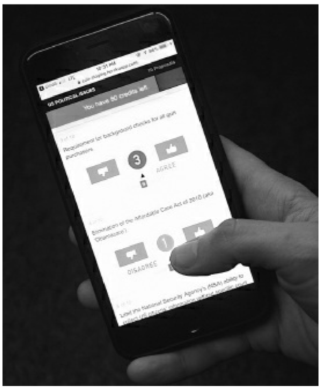


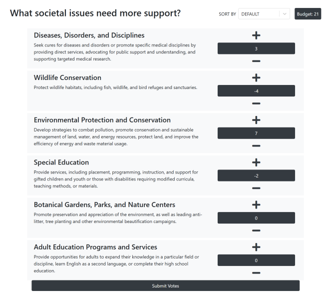
Fig. 2. The text-based interface

3.1 Text-based Interface (Text Condition)

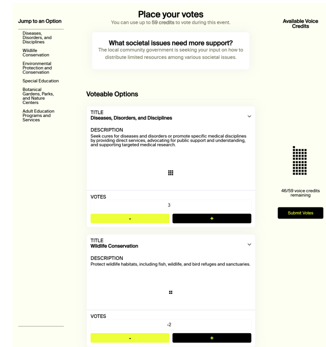
First, we surveyed the current implementation of QV interfaces to understand the development of such tools. We present a selection in Figure 1. All five interfaces retain and present the following components:



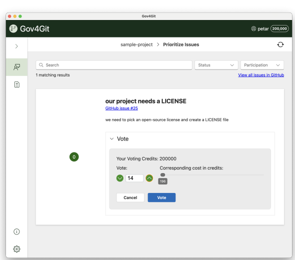
(a) Software designed by WeDesign used in [53]. Image taken from [50].



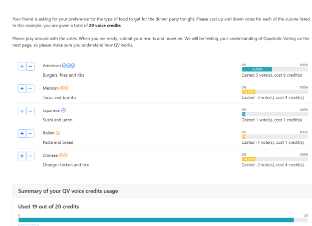
(b) An open-source QV interface [79] with a publicly available service.



(c) An open-source QV interface [55] forked from GitCoin [56] used by the RadicalxChange community [2].



(d) The interface designed for gov4git [27].



(e) The interface used in the research by Cheng et al. [10].

Fig. 1. Recent implementations of interfaces applying the quadratic mechanism.

- Option list: A list of options contesting for votes.
- Vote Controls: Two buttons to increase and decrease votes associated with an option.
- Individual vote counts: Some representation of the number of votes associated with an option.
- Summary: A summary automatically calculates the cost of the total votes.

We constructed a text-based interface that included all five components but removed the use of emojis (i.e., thumbs up and thumbs down present in Figure 1a), progress bars, and other visualizations in the summary section (i.e., progress bars in Figure 1a and 1e or blocks presented in Figure 1c), and the visual cues for individual vote counts (i.e., the colored counts and icons present in Figure 1d and 1e). Prior literature suggests that the use of emojis might influence the interpretations of surveys [31] and decrease user satisfaction [71]. We also removed all visualization elements such as blocks, progress bars, and percentage indicators, as prior literature shows that not all data visualization elements reduce cognitive demand [32]. Last, we decided to present all the options on the same screen. Prior research emphasizes the importance of placing all the options on the same digital ballot screen to avoid losing votes (missing citations). This echoes the proverb "out of sight, out of mind," where individuals might be biased toward options that are shown to them and additional effort is required for individuals to retrieve specific information if options are hidden (citation needed).

These design decisions led to the interface shown in Figure 2. The interface contains the question prompt at the top of the screen. The options are presented as a list of options below. Survey respondents can update the votes, showing the cost and the number of votes, by selecting from a dropdown that provides all possible voting options and cost given the number of credits. A small summary box to the right of the interface shows the current total cost and the remaining credit for the respondent. To avoid ordering bias [22, 14], options are always randomly presented on the interface.

3.2 Interactive Interface (Interactive Condition)

The interactive interface, shown in Figure 3, builds additional interactive elements on top of the text interface to maintain consistency that allows comparison of the direct manipulation of the designed interactive elements. We designed two additional components: An additional organization step prior to voting and a drag-and-drop interface throughout the QS responding session informed through prior literature.

Organization Phase. The design of the organization phase in this interactive interface began from understanding Strack and Martin [64]'s research on how survey respondents process and respond to attitude surveys. They propose that individuals, after comprehending a question, would either recall a previous judgment or construct a new one. This organization phase is designed to guide respondents through the process of forming opinions on individual options incrementally and methodically.

The organizing interface, as shown on the left-hand side of Figure 3, begins by presenting all survey options one at a time right under the system prompt. Survey respondents select a response among three ordinal categories – lean positive, lean negative, or lean neutral. Once selected, the system will move that option to the respective category. Participants can skip the option if they do not want to indicate a preference. Options within the groups are draggable and rearrangeable to other groups should the participants want. In other words, we designed an interface to nudge survey respondents to review and group different options.

Three theoretical theories informed such design. First, to reduce cognitive load, showing participants one option at a time gated the amount of information presented to the survey respondent and thereby reduced the extraneous load [66]. The three possible options, positive, neutral, and negative, aim to scaffold participants in constructing their own choice architecture [43, 68], which strategically segments options into diverse and alternative choice presentations while avoiding the biases from defaults. The immediate feedback and the agency for participants to edit their 'categorization' stem from basic HCI principles [45]. This design underwent paper prototypes and various iterations, which all maintained the combination of these theoretical bases aimed at reducing cognitive load and scaffolding the decision-making process. We describe these iterations and the design process in Appendix A.

Interactive voting interface. Given that survey respondents conveyed a rough categorization of options, the voting interface presents the options as they were categorized on the interface, rather than in a randomly ordered manner like the text-based interface. As presented on the right-hand side of Figure 3, options are placed in the order of lean positive, lean neutral, lean negative, and skipped or undecided. Undecided options include options that remained in the organization queue. This could happen since survey respondents might have a pre-existing preference or do not want to organize their thoughts. The order within these categories is also preserved. Survey respondents can

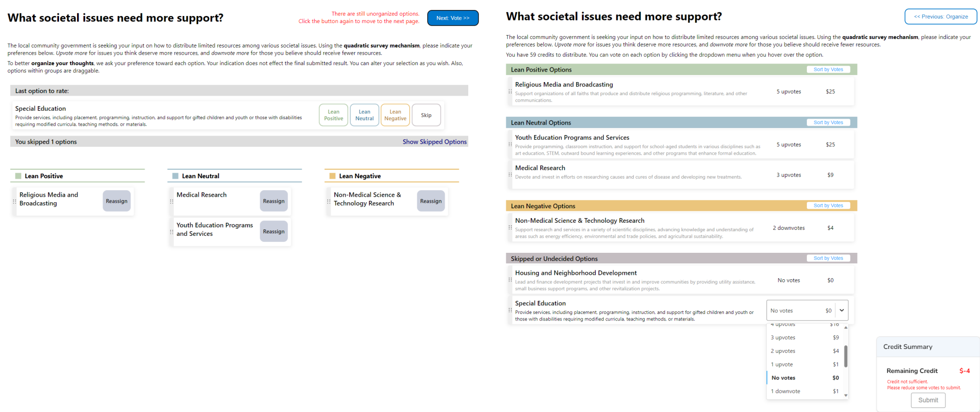


Fig. 3. The interactive interface

also choose to go back to the organization interface to focus on updating their choice architecture anytime throughout the survey.

The other difference for the interactive voting interface is that options are draggable for participants to reposition options. Each category also includes a sort-by-vote function to change the position of options within the same 'category.' Survey respondents are aware that these additional interface interaction designs will not influence the voting outcome.

These interactions aim to support the positional proximity in information organization. The goal is to partially automate (i.e., automatic positioning of groups) while providing a straightforward mechanism (i.e., drag-and-drop) such that 'similar' options and information are placed near each other, allowing participants to process decisions together, echoing research on the proximity compatibility principle [77], specifically spatial proximity [76], and mental compatibility.

While there are various interaction mechanisms to adjust the ordering of options, the notion of drag-and-drop has been explored widely in rank-based surveys. For instance, Krosnick et al. [36] demonstrated that replacing drag-and-drop with traditional number-filling rank-based questions improved participants' satisfaction with little trade-off in their time. A similar work embedded drag-and-drop as part of its ranking process [69]. Even though the drag-and-drop interface might lower the stability of the outcome, since we are not trying to capture the position of the options as our final solution, it is worth its trade-off for which survey respondents express a higher satisfactory affordability and ease of use [57].

Together, these design decisions lead to our belief that a two-step interactive interface with direct interface manipulation can reduce the cognitive load for survey respondents to form preference decisions when completing QS.

4 EXPERIMENT DESIGN

Based on the design decisions, we developed a QS interface using a React.js frontend and a Next.js backend powered on MongoDB. Both systems are open-sourced [1].

We recruited participants from a midwestern college town using online ads, digital bulletins, social media posts, physical flyers, and online newsletters. The study's researcher prioritized the non-student population to maximize participant diversity. This study is reviewed and approved by the college Institutional Review Board.

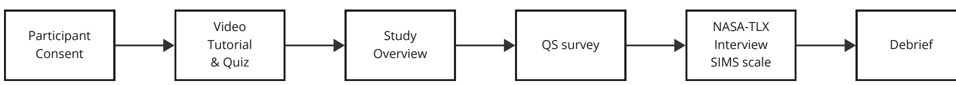


Fig. 4. Study protocol

Figure 4 shows a visual representation of the study protocol. Study participants are invited to the lab to participate in this study. The reason we made this experiment design decision is to minimize the influence of external factors that could affect the measurement of cognitive load. External factors, more prevalent in remote experiments or those conducted via platforms like MTurk, include potential multitasking or interruptions by others. An in-lab study also allows participants to operate across a consistent device that researchers have full control over. More specifically, the experiment involves participants operating on a 32-inch vertical monitor. This setup assured study participants, despite any condition in the study, can see all options on a QS, minimizing hidden information from an individual's decision-making process.

After consenting to the study, participants are invited to the study and they watch a pre-recorded video explaining the Quadratic mechanism and how QS operates. This video does not include any hints of either interface and how to operate the interface. Participants are then asked to complete a short quiz. The purpose of the quiz is to ensure that all participants fully understand how QS works. Participants are not screened out if they fail the quiz but are asked to rewatch the video or ask the researcher until they are able to select the correct answer. The device that the participant worked on is screen captured throughout the study.

The researcher then primes the participant that the purpose of this study is to assist local community organizers in understanding community members' preferences on a wide variety of societal issues so they can potentially distribute limited resources better. Participants would be randomly placed into one of the four groups:

- 6 options with a text-based interface
- 6 options with an interactive interface
- 24 options with a text-based interface
- 24 options with interactive interface

Participants will begin completing the survey independently, without the researcher's presence. Upon completion, they contact the researcher, who then requests they complete the NASA-TLX to assess cognitive load. This is followed by a short semi-structured interview to gain insights into the participants' experiences. This interview is audio recorded. Finally, participants complete the situational motivation scale (SIMS) to gauge motivation and a demographic survey. The session concludes with a debriefing and a $15 cash compensation for their participation. The debreifing

[1] link-to-github

explains to the participant regarding not disclosing the purpose of the survey is to measure cognitive load and interface design and allows for participants to ask any questions.

The study is designed as a between-subject study for two reasons. First, we aim to minimize the study fatigue that might occur given the complexity of responding to a QS. To complete a QS survey, participants can take up to 20 minutes. Thus, it is difficult to conduct back-to-back experiments that measure cognitive load. We choose not to ask participants to revisit the lab with several days in between, to reduce dropout rates and prevent demotivating participants from attending the in-person experiment, which might occur in a within-subject study design. Second, we aim to reduce the learning effect that is difficult to remove, especially concerning operating the interface and making decisions on the survey. Recall that preferences are constructed, we want to ensure that participants are not influenced by their previous preferences which can influence their perceived cognitive load.

In an ideal world, understanding participants' cognitive load across multiple options would require enumerating all possible numbers of options and eliciting the "breaking point" where the participant experiences cognitive overload. Unfortunately, this is not feasible. Iterating through all possible numbers of options is very costly, both in time and resources. Therefore, we refer to prior literature to inform our choice of 6 and 24 options, representing a short and long list of options. To decide the number for the short list, survey methods such as constant sum surveys and Analytic Hierarchy Process (AHP) recommend options fewer than ten and seven, respectively [42, 61, 60]. However, we are not aware of any specific works that justify these numbers. Saaty [60] associated this value with both the cognitive processing capacity of 7 ± 2 [41] and a theoretical proof using the consistency ratio of a pairwise comparison metric [59]. This informs our decision to contain a pair of dependent variables above and below seven options. We turn to experiments designed to study choice overload. A meta-analysis by Chernev et al. [11] surveyed 99 choice overload experiments ($N = 7202$) and summarized that 6 and 24 are the modal values for short and long lists when testing choice overload. These two values are likely rooted in the original choice overload experiment by Iyengar and Lepper [33]. The value six is often used in experiments to understand the effect of choice provision. The value 24 is the maximum number of ecologically valid jams produced by the jam company in the original study. We decided to follow suit with these two values, satisfying the previous decision to choose two values less than and greater than seven.

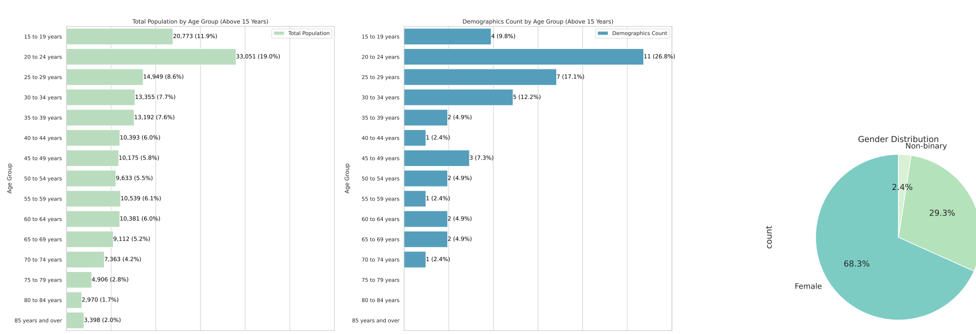
Next, we describe the context of the survey that participants completed. Participants were asked to complete a societal issue survey. We follow suit as described by Cheng et al. [10], believing that surveying societal issues is a good topic as it is relevant to every citizen and it is easy to convey that there are limited resources in the public sector to be prioritized across different sectors and areas. Participants across all four groups were presented with options randomly drawn from 26 societal issues. These issues were generated from the categories used by Charity Navigator [9], a non-profit organization that evaluates over 20 thousand charities in the United States. The full list of these societal issues is provided in Appendix B.

Last, we describe the two quantitative measurements taken during the study: cognitive load and motivation. At the time of this study, several methods existed to measure cognitive load, including performance measures, psychophysiological measures, subjective measures, and analytical measures [24]. Given the nature of QS, a task requiring a long period, adopting performance measures like secondary-task measures in our experiment proved challenging due to the difficulty of designing a secondary task. The secondary task must use the same cognitive resources as the primary tasks, and the cognitive resource for completing the survey would vary among participants. Similarly, psychophysiological measures such as pupil size [47] and ECG [28] can be highly sensitive to external environments and costly to obtain. Consequently, we relied primarily on subjective measures via self-report surveys and analytical measures like time and clicks collected via the interface. We adopted a paper-based weighted NASA Task Load Index (NASA TLX), a multidimensional scoring procedure using the weighted average of six subscale scores to represent overall workload. Weighted NASA-TLX uses a priori workload definitions from subjects to weight and average subscale ratings, requiring subjects to evaluate each weight's contribution to the workload of a specific task [29, 30, 7]. This approach reduces between-rater variability, indicating differences in workload definitions among raters within a task and variations in workload sources between tasks [7]. Despite criticisms regarding its validity and vulnerability, NASA-TLX is commonly used due to its low cost and ease of administration [24]. It has been tested on various experimental and lab tasks, and workload scores derived from these tests showed significantly less variability among evaluators than one-dimensional workload scores [58]. Thus, we chose NASA-TLX to measure cognitive load in our study.

In addition to NASA-TLX, we administered a situational motivation scale (SIMS) to measure participants' motivation (required citation). We posited that motivation would influence mental demand (required citation). SIMS, chosen for its widespread use, helps understand one's intrinsic motivation, extrinsic motivation, identified regulation, and external regulation, and was originally designed to measure self-determination. Both instruments were administered using pen-and-paper.

5 RESULTS

In this section, we present the results from our study, beginning with a description of the participants' demographics. This is followed by quantitative and qualitative results regarding cognitive load, and concluded with observations of participants' behaviors. The quantitative measures are directly copied from pen-and-paper surveys and captured through system logs during participant interactions. The interview results are transcribed and coded by the first author, with the codes then grouped into themes. The behavioral data is cleaned and processed, available as open data [2].



(a) Age distribution

(b) Gender distribution

Fig. 5. Two figures side by side

5.1 Demographics

We recruited a total of 41 participants, 10 for each experiment condition. One participant was removed due to data quality concerns. The age of the participants has a mean centered at 34.63 years old with the full distribution of the participant age distribution to the right of Figure 5a. The population of the county is presented to the left in the figure. We see that the recruited participants

[2]link-to-github

follow closely with the population only missing a few percentages in the 35-45 range making the recruited sample size slightly younger. Recruited participants were more female than male as shown in Figure 5b.

In terms of ethnicity, 51.2% of the respondents self-identify as White, followed by 26.8% as Asian, 4.9% as Hispanic, and 7.3% as African American. Additionally, 9.8% of participants identify as having mixed ethnicity.

5.2 Cognitive Load Results

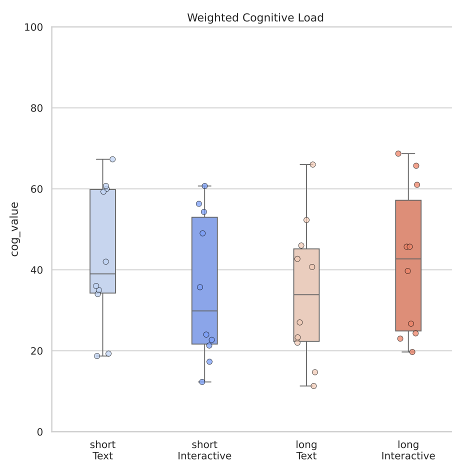


Fig. 6. NASA-TLX Results

We show the NASA-TLX weighted results in Figure 6. Qualitatively, there is a decrease in cognitive load when comparing short surveys between the text-based interface and interactive interfaces. Conversely, there is an increase in cognitive load when comparing the long survey between the text-based interface and interactive interfaces. However, we are not able to demonstrate statistical significance between the four groups using the Mann-Whitney U test. Reviewing the overall cognitive load, QS participants experienced a 'Somewhat High' and 'High' cognitive load regardless of length and interface [29]. (Need to get the exact percentage).

Next, we present the raw scores from the six measurements of NASA-TLX: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration in Figure 7. We also show the 95% confidence interval around the mean for each boxplot.

Mental Demand remains high across all four conditions with a wide variance of responses. The short interactive interface has the lowest median score across all conditions. We see no statistical significance between conditions. For Physical Demand, all conditions except Long Interactive showed minimal scores, with the Long Interactive condition having a noticeably higher median score. There is statistical significance between short and long interactive interfaces ($p < 0.01$) as well as text and interactive interfaces in long surveys ($p < 0.05$).

Temporal Demand showed more variation across groups. It is interesting to see that the median across the short text interface is among the highest compared to the rest of the groups, while the long-text interface showed the least Temporal Demand. Our statistical tests showed a significant

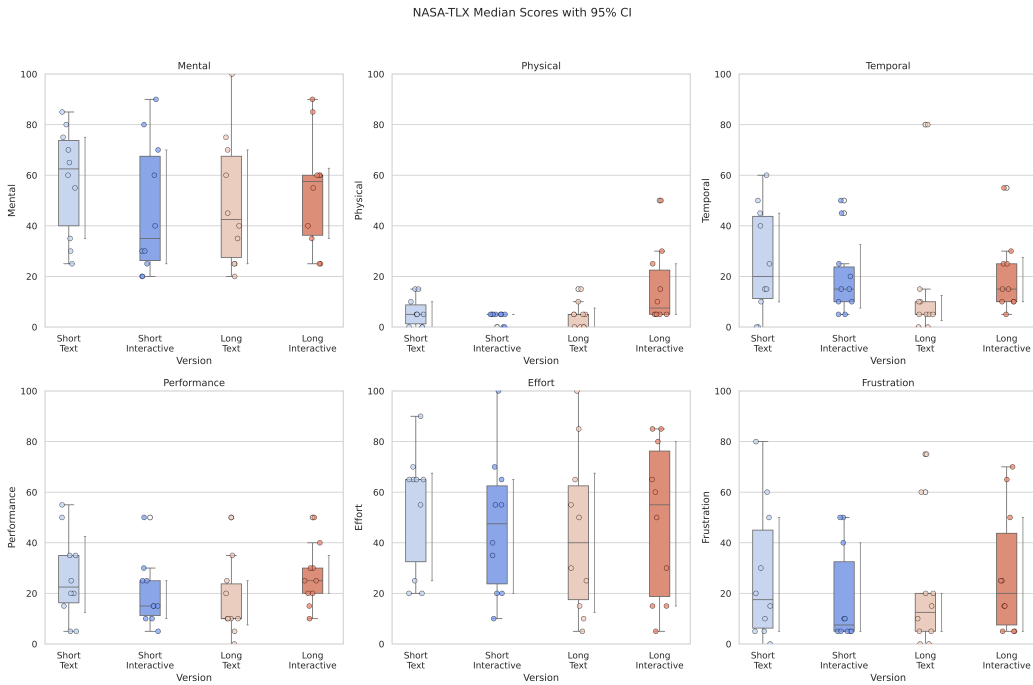


Fig. 7. NASA-TLX Results

difference between the long text interface and the long interactive interface ($p < 0.05$). Performance scores are relatively low across all conditions, indicating that participants experienced less cognitive load from performance. Effort scores echo the mental demand, showing a wide distribution with a high median across the four groups. Lastly, regarding Frustration, we observe qualitatively that the short text interface has a higher median value than the short interactive interface. The opposite is observed in longer surveys, where respondents in the long survey with a text interface experienced less frustration than the interactive version based on median values. We cannot establish statistical significance in the latter three aspects of NASA-TLX.

Based on these results, given the small sample size for each group of participants, we were not surprised that most results do not provide statistical significance in changes in cognitive load values. However, there are some trends that we capture via descriptive statistics. First, comparing the overall cognitive load and the breakdown of the sub-components between text interface and interactive interface across the short survey, we see a general trend in a reduction of cognitive load. Next, we are surprised by the upward trend between the text and interactive interface for the longer list. This is against our original hypothesis that under even complex situations, we should see a clearer portrait of how interactive interactions can reduce cognitive load. While it is possible that interactive interfaces can increase study participants' cognitive load, our qualitative results do not hint at this possibility. In addition, comparing the long and short survey in the text-based interface, it is counter intuitive to see a downward trend across cognitive load. Logically, choosing among more options would demonstrate a higher cognitive load.

These results lead us to further investigate the source of cognitive demands and participants' behaviors.

Mental Demand. This section dissects the sources where participants experienced mental demand.

Physical Demand. This section dissects the sources where participants experienced physical demand.

Temporal Demand. This section dissects the sources where participants experienced temporal demand.

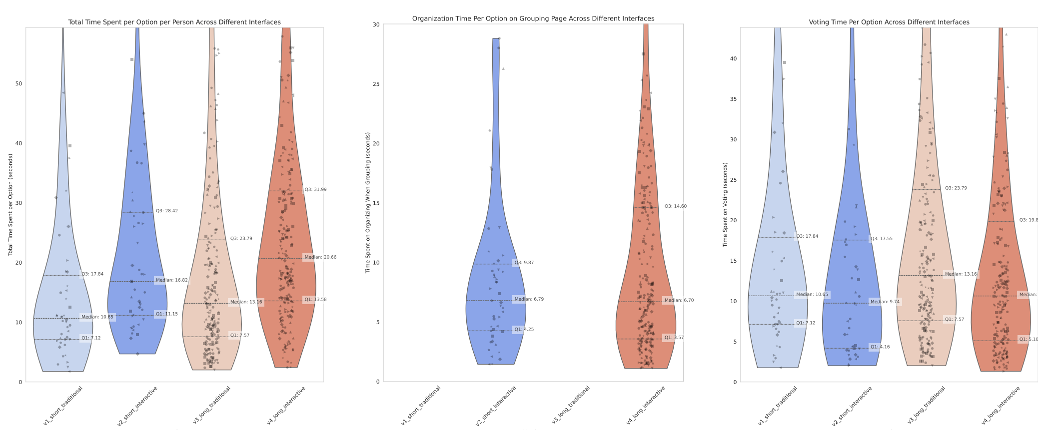
Performance. This section dissects the sources where participants experienced performance.

Effort. This section dissects the sources where participants experienced effort.

Frustration. This section dissects the sources where participants experienced frustration.

5.3 Interaction Behavior Analysis

To further investigate the cognitive load results, we focused on participants' behaviors during the survey. Specifically, we aim to understand the time participants spent on the options as well as when a participant makes changes to the survey. When a study participant clicks their mouse on the interface to complete some action, whether it is a drag-and-drop, updating votes, or placing options into a specific group, a timestamp and the payload of the update are stored in the log. The time difference between two actions is attributed to the time the participant took to decide and enact upon a behavior. While participants can be thinking about other things, this is the best proxy we have to study participant behaviors.



(a) Total Time per option (b) Organization Time per option (c) Voting Time per option

Fig. 8. Breakdown of time per option

5.4 Time Spent per Options

First, we define time spent per option. A participant can enact several actions related to the same option, for example, a participant might spend t_1 time to place the option into a 'lean positive' category; spend t_2 and t_3 time to drag and drop the options to reposition it on the interactive interface; spend t_4 and t_5 time to update the upvotes on that option. In this case, we would define voting time as $t_4 + t_5$ for that option, and organization time as $t_1 + t_2 + t_3$.

To reduce noise, we intentionally drop all the time participants spent on the first option in the organization phase or voting phase. The goal is to reduce the inclusion of time they spent on reading the prompt, forming their preference, or understanding the interface. We present the results in Figure 8 where each of the dots represents the time accumulated for an option that a participant interacted with. The violin plot shows the distribution of the dots and the three horizontal lines represent the median, 25th percentile, and 75th percentile of the time spent for that interface.

In Figure 8a, we observe that participants spent more time on the interactive interface than the text interface in both short and long surveys. A non-parametric statistical test supports such observation with $p < 0.01$ for short and $p < 0.0001$ for long surveys. This is not surprising because participants need to review the options and organize them in the interactive interface which takes more time. We break down the total time spent into organization time and voting time in Figure 8b and Figure 8c.

Once we separate the organization time (Figure 8b) and identify the voting time (Figure 8c), while there are no statistically significant differences between the text interface and the interactive interface in the short survey, we see a statistically significant reduction ($p < 0.01$) in voting time between the text interface and the interactive interface. In other words, our original hypothesis holds in which the two-step design process did facilitate participants in making their decisions.

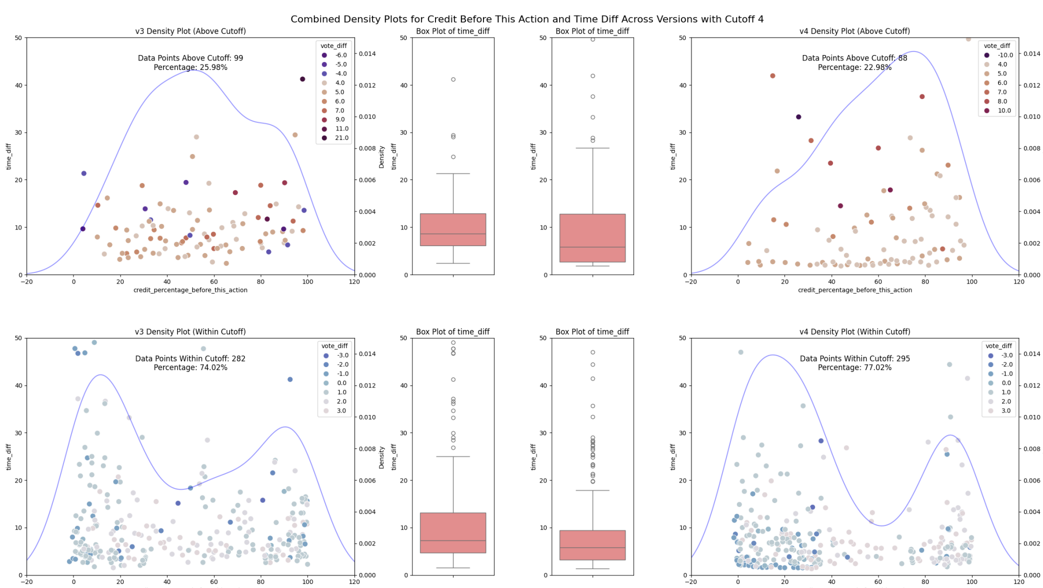


Fig. 9. Breakdown of voting actions (needs to update chart text)

5.5 Budget and Voting Behaviors

Next, we examine participants' voting behavior and how it changed throughout the progress. Given that we observe significant differences in voting time changes comparing text interface and interactive interface for the long option survey, we focus on deciphering the voting action changes between these two experiment conditions in this subsection.

Figure 9 plots the time of voting actions over the remainder of the participant's budget across the text and interactive interface. In other words, different from Quarfoot et al. [53] focusing on the number of accumulated votes over an individual's time, where they showed QV voters make more revisions than Likert Surveys, we focused on the budget scarcity which can influence QS respondents' behaviors.

We further separated the behaviors where participants made bigger changes or smaller changes to the option. In this figure, we define an adjustment of four or more votes as a large adjustment which we plotted in the first row of the Figure. Adjustments of three or fewer votes are considered small adjustments.

First, the plot demonstrated a clear cluster of voting actions in the bottom left corner of the interactive interface for small vote adjustments. In other words, participants made much smaller but more rapid adjustments when their budgets were running low. Second, larger adjustments are made when the participants have more options comparing the two plots on the first row. We interpret this behavior as participants in the interactive interface have constructed a clearer image of option preferences and, hence, have the ability to take larger strides in allotting their budget and deciding the number of votes at the beginning of the survey. Toward the end, participants using the interactive interface are then making fine-tuned adjustments to ensure that their preferences are reflected in their submissions.

5.6 Interface Comments

Finally, we present the qualitative responses related to the interface design and their experience working with QS across all experiment groups.

Organization is required and beneficial. Many participants (N=7) who responded to QS using the interactive interface expressed the helpfulness of the organization phase proactively when asked what they liked about the interface in general. In fact, half of the participants (N=5) in the long interactive interface group expressed such an opinion. Multiple participants (N=4, 3 from long interactive interface group) felt that the upfront introduction of all the topics allowed them to process and think about the full picture, thereby digesting all the information more comprehensively.

> I would say that (the interface) definitely (supported me), by being able to have a preliminary categorization of all the topics. First, it introduced me to all the topics, so that I can think about them like I can just kind of leave it there in my head space to think about and process [. . .] So being able to digest all the information prior to actually allocating the budget or completing the quadratic survey.
>
> – S009, long interactive interface.

Participants (N=4, 2 from long interactive interface group) mentioned that organization support them to allot the intensity of votes by helping them focus and prioritize options through ranking options. This excercise allows them to follow a clear decision making process that avoids confusion.

> If I had to choose a number like that in the beginning. That would have been really bad, but positive, neutral, negative. That was good enough.
>
> – S016, long interactive interface.

> I think . . . ranking at the beginning one's impression towards these issues helps to like determine how many votes should be put towards them.
>
> – S002, short interactive interface.

Last, one participants highlighted the one-at-a-time approach during the organization phase allowed thoughtful reflection to think about their attitude toward that option.

> Like, at the moment (during organization), when it gives you, like, rank it if it's positive or neutral or negative [. . .] it gives you time to just focus on that single thing and rank it based on how you feel at that moment.
>
> – S013, short interactive interface.

We see a call for organizational features proactively when asking participants using the text-based interface what features they wanted from the interface. Almost half of the participants (N=4) using the long text interface expressed some form or another that can help reduce the decision space when responding to the QS.

> If anything, I think I would like to be able to like, click and drag the categories themselves so I could maybe reorder them to like my priorities.
>
> – S025, long interactive interface.

> Because with this many (options), especially when I'm thinking ... Ok, where was (the option) ... Where was (the option) you know? Oh, that's right. Maybe I could give another up another upvote to the, you know whatever [...]
>
> – S028, long interactive interface.

Direct Manipulation Enhances Reflective Decision-Making. As the proximity of position are mostly determined by the categorization in the first phase of the interactive interface. Several participants mentioned how they used direct Manipulation in the software as a process for reflective thinking upon their decision making process. One participant mentioned:

> So I tried to make a ranking [...] and by creating this ranking, by dragging the related issues ...ĩ don't know ...t̃hat helped me organize my ideas.
>
> – S021, long interactive interface.

> I think the system was actually really helpful because I could just drag them. [...] Because when I was unsure, because if I couldn't drag them then I couldn't compare 2 options very well like side to side, because because this is pretty long list ... so if I couldn't drag it, then I would have a harder time organizing my thoughts, whereas with the dragging feature I can really compare them, I can drag this one up here, and then compare it to the top one versus like not being able to track it at all
>
> – S039, long interactive interface.

But more importantly, it acts as a process for reflective verification and iterative decision making. These can include post reflection after expressing the intesnsitve of preferences, or a preperation to decide on number of votes for the next option.

> So I would give the votes, and then I would drag and drop. [...] So I guess to see what my ranking look like. And see if I could give more money or not.
>
> – S021, long interactive interface.

> [...] this is something that's really important to me ... So I had the flexibility to move it to positive. So just having the kind of like shift in perception. [...] especially because when I was doing categorize categorization in the first step, [...] what I thought about it in the moment. [...] In the second step there was a shift in my perception of the issue just reflecting. So being able to change. That was really nice as well.
>
> – S009, long interactive interface.

Conversely, in the text interface, one participant proactively mentioned a request to add click and drag functionalities to the interface. The participants described such function to group by topic categorization and also priority placement through direct manipulation.

> If anything, I think I would like to be able to like, click and drag the categories themselves so I could maybe reorder them to like my priorities. And so I could maybe make that like a descending or ascending like list of like importance. [...] if I could pull that up to the top, say myself like click and drag it up there, I think then I would stack the things I think it would affect under it. So like, I would put then, like youth, pro-education programs and adult education and early childhood programs

> and kinda stack those altogether. [...] I would hope that money would trickle down and also increase all the rest of those things. So I would put less upvotes in there because I would hope to dribble out effect would kick in. [...] I would kind of make myself categories and subcategories out of this list. If I could organize it.
>
> – S025, long text interface.

**

6 DISCUSSION AND FUTURE WORKS

6.1 Construction of Preference

6.2 Design Implications

6.3 Limitations and Future Work

7 CONCLUSION

A EARLY PROTOTYPES OF THE STUDY

REFERENCES

[1] [n. d.] A New Way of Voting That Makes Zealotry Expensive - Bloomberg. https://www.bloomberg.com/news/articles/2019-05-01/a-new-way-of-voting-that-makes-zealotry-expensive. (). Retrieved Dec. 16, 2023 from.

[2] [n. d.] About RxC. https://www.radicalxchange.org/wiki/about/. (). Retrieved June 17, 2024 from.

[3] Duane F Alwin and Jon A Krosnick. 1985. The measurement of values in surveys: A comparison of ratings and rankings. *Public Opinion Quarterly*, 49, 4, 535–552.

[4] Madeline E. Bassetti, Gustavo Dias, Daniel L. Chen, Alan Mortoni, and Ritesh Das. 2023. Civicbase: An open-source platform for deploying Quadratic Voting for Survey Research. *AI Magazine*, 44, 3, 263–273. DOI: 10.1002/aaai.12103.

[5] Steven J. Brams and Peter C. Fishburn. 1978. Approval Voting. *American Political Science Review*, 72, 3, (Sept. 1978), 831–847. DOI: 10.2307/1955105.

[6] Vitalik Buterin, Zoë Hitzig, and E. Glen Weyl. 2019. A Flexible Design for Funding Public Goods. *Management Science*, 65, 11, (Nov. 2019), 5171–5187. DOI: 10.1287/mnsc.2019.3337.

[7] Brad Cain. 2007. A review of the mental workload literature. *DTIC Document*.

[8] Charlotte Cavaille and Daniel L Chen. [n. d.] Who Cares? Measuring Preference Intensity in a Polarized Environment.

[9] 2023. Charity Navigator: Animals. https://web.archive.org/web/20230323213631/https://www.charitynavigator.org/index.cfm?bay=s (Mar. 2023). Retrieved Dec. 16, 2023 from.

[10] Ti-Chung Cheng, Tiffany Li, Yi-Hung Chou, Karrie Karahalios, and Hari Sundaram. 2021. "I can show what I really like.": Eliciting Preferences via Quadratic Voting. *Proceedings of the ACM on Human-Computer Interaction*, 5, (Apr. 2021), 1–43. DOI: 10.1145/3449281.

[11] Alexander Chernev, Ulf Böckenholt, and Joseph Goodman. 2015. Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology*, 25, 2, (Apr. 2015), 333–358. DOI: 10.1016/j.jcps.2014.08.002.

[12] Dana Chisnell. 2016. Democracy Is a Design Problem. 11, 4.

[13] Frederick G. Conrad, Benjamin B. Bederson, Brian Lewis, Emilia Peytcheva, Michael W. Traugott, Michael J. Hanmer, Paul S. Herrnson, and Richard G. Niemi. 2009. Electronic voting eliminates hanging chads but introduces new usability challenges. *International Journal of Human-Computer Studies*, 67, 1, (Jan. 2009), 111–124. DOI: 10.1016/j.ijhcs.2008.09.010.

[14] M. P. Couper. 2001. Web survey design and administration. *Public Opinion Quarterly*, 65, 2, 230–253. DOI: 10.1086/322199.

[15] Kahneman Daniel. 2017. *Thinking, Fast and Slow*.

[16] Shaneé Dawkins, Tony Sullivan, Greg Rogers, E. Vincent Cross, Lauren Hamilton, and Juan E. Gilbert. 2009. Prime III: an innovative electronic voting interface. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (IUI '09). Association for Computing Machinery, New York, NY, USA, (Feb. 2009), 485–486. ISBN: 978-1-60558-168-2. DOI: 10.1145/1502650.1502727.

[17] 2015. Designing usable ballots | Center for civic design. https://civicdesign.org/fieldguides/designing-usable-ballots/. (June 2015). Retrieved June 17, 2024 from.

[18] Erik J Engstrom and Jason M Roberts. 2020. *The Politics of Ballot Design: How States Shape American Democracy*. Cambridge University Press.

[19] Sarah P. Everett, Kristen K. Greene, Michael D. Byrne, Dan S. Wallach, Kyle Derr, Daniel Sandler, and Ted Torous. 2008. Electronic voting machines versus traditional methods: improved preference, similar performance. In *Proceedings of*

the SIGCHI Conference on Human Factors in Computing Systems (CHI '08). Association for Computing Machinery, New York, NY, USA, (Apr. 2008), 883–892. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357195.

[20] Habiba Farzand, David Al Baiaty Suarez, Thomas Goodge, Shaun Alexander Macdonald, Karola Marky, Mohamed Khamis, and Paul Cairns. 2024. Beyond Aesthetics: Evaluating Response Widgets for Reliability & Construct Validity of Scale Questionnaires. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems* (CHI EA '24). Association for Computing Machinery, New York, NY, USA, (May 2024), 1–7. ISBN: 9798400703317. DOI: 10.1145/3613905.3650751.

[21] N. T. Feather. 1973. The measurement of values: Effects of different assessment procedures. *Australian Journal of Psychology*, 25, 3, (Dec. 1973), 221–231. DOI: 10.1080/00049537308255849.

[22] Robert Ferber. 1952. Order Bias in a Mail Survey. *Journal of Marketing*, 17, 2, 171–178. JSTOR: 1248043. DOI: 10.2307/1248043.

[23] Luis Mota Freitas and Wilfredo L. Maldonado. 2024. Quadratic funding with incomplete information. *Social Choice and Welfare*, (Feb. 2024). DOI: 10.1007/s00355-024-01512-7.

[24] Qin Gao, Yang Wang, Fei Song, Zhizhong Li, and Xiaolu Dong. 2013. Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, 56, 7, (July 2013), 1070–1085. DOI: 10.1080/00140139.2013.790483.

[25] Juan E. Gilbert, Jerone Dunbar, Alvitta Ottley, and John Mark Smotherman. 2013. Anomaly detection in electronic voting systems. *Information Design Journal (IDJ)*, 20, 3, (Sept. 2013), 194–206. DOI: 10.1075/idj.20.3.01gil.

[26] John T. Gourville and Dilip Soman. 2005. Overchoice and Assortment Type: When and Why Variety Backfires. *Marketing Science*, 24, 3, (Aug. 2005), 382–395. DOI: 10.1287/mksc.1040.0109.

[27] 2023. Gov4git: A Decentralized Platform for Community Governance. (Mar. 2023). Retrieved June 13, 2024 from.

[28] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. ACM, Copenhagen Denmark, (Sept. 2010), 301–310. ISBN: 978-1-60558-843-8. DOI: 10.1145/1864349.1864395.

[29] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In *Advances in Psychology*. Vol. 52. Elsevier, 139–183.

[30] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 9, (Oct. 2006), 904–908. DOI: 10.1177/154193120605000909.

[31] Susan C. Herring and Ashley R. Dainas. 2020. Gender and Age Influences on Interpretation of Emoji Functions. *ACM Transactions on Social Computing*, 3, 2, (June 2020), 1–26. DOI: 10.1145/3375629.

[32] Weidong Huang, Peter Eades, and Seok-Hee Hong. 2009. Measuring Effectiveness of Graph Visualizations: A Cognitive Load Perspective. *Information Visualization*, 8, 3, (Sept. 2009), 139–152. DOI: 10.1057/ivs.2009.10.

[33] Sheena S Iyengar and Mark R Lepper. 2000. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79, 6, 995.

[34] Julie A. Jacko, ed. 2012. *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*. (3. ed ed.). *Human Factors and Ergonomics*. CRC Press, Taylor & Francis, Boca Raton, Fla. ISBN: 978-1-4398-2944-8 978-1-4398-2943-1.

[35] N. D. Kieruj and G. Moors. 2010. Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of Public Opinion Research*, 22, 3, (Sept. 2010), 320–342. DOI: 10.1093/ijpor/edq001.

[36] Jon A Krosnick, Charles M Judd, and Bernd Wittenbrink. 2018. The measurement of attitudes. In *The Handbook of Attitudes*. Routledge, 45–105.

[37] Steven P Lalley and E Glen Weyl. 2018. Quadratic voting: How mechanism design can radicalize democracy. In *AEA Papers and Proceedings*. Vol. 108, 33–37.

[38] Seunghyun "Tina" Lee, Yilin Elaine Liu, Ljilja Ruzic, and Jon Sanford. 2016. Universal Design Ballot Interfaces on Voting Performance and Satisfaction of Voters with and without Vision Loss. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). Association for Computing Machinery, New York, NY, USA, (May 2016), 4861–4871. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858567.

[39] Sarah Lichtenstein and Paul Slovic, eds. 2006. *The Construction of Preference*. (1. publ ed.). Cambridge University Press, Cambridge. ISBN: 978-0-521-54220-3 978-0-521-83428-5.

[40] R. Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology*, 22 140, 55–55.

[41] George A. Miller. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 2, 81–97. DOI: 10.1037/h0043158.

[42] William F. Moroney and Joyce A. Cameron. 2019. *Questionnaire Design: How to Ask the Right Questions of the Right People at the Right Time to Get the Information You Need*. Human Factors and Ergonomics Society, (Feb. 2019). ISBN: 978-0-945289-55-5.

[43] Robert Münscher, Max Vetter, and Thomas Scheuerle. 2016. A Review and Taxonomy of Choice Architecture Techniques. *Journal of Behavioral Decision Making*, 29, 5, 511–524. DOI: 10.1002/bdm.1897.

[44] Ryan Naylor et al. 2017. First year student conceptions of success: What really matters? *Student Success*, 8, 2, 9–19.
[45] A Norman Donald. 2013. *The Design of Everyday Things*. MIT Press.
[46] Judith S. Olson and Wendy A. Kellogg, eds. 2014. *Ways of Knowing in HCI*. Springer, New York, NY. ISBN: 978-1-4939-0377-1 978-1-4939-0378-8. DOI: 10.1007/978-1-4939-0378-8.
[47] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*. ACM Press, Austin, Texas, 141. ISBN: 978-1-60558-994-7. DOI: 10.1145/1743666.1743701.
[48] John W. Payne, James R. Bettman, and Eric J. Johnson. 1988. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 3, (July 1988), 534–552. DOI: 10.1037/0278-7393.14.3.534.
[49] Martin Pielot and Mario Callegaro. 2024. Did You Misclick? Reversing 5-Point Satisfaction Scales Causes Unintended Responses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, (May 2024), 1–7. ISBN: 9798400703300. DOI: 10.1145/3613904.3642397.
[50] Eric A Posner and E Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.
[51] 2022. Quadratic Voting Frontend. Public Digital Innovation Space. (Jan. 2022). Retrieved Dec. 16, 2023 from.
[52] [n. d.] Quadratic Voting in Colorado. https://www.radicalxchange.org/wiki/colorado-qv/. (). Retrieved June 13, 2024 from.
[53] David Quarfoot, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein, and Ellen Konar. 2017. Quadratic voting in the wild: real people, real votes. *Public Choice*, 172, 1-2, 283–303.
[54] Whitney Quesenbery. 2020. Opinion | Good Design Is the Secret to Better Democracy. *The New York Times*, (Oct. 2020). Retrieved June 17, 2024 from.
[55] 2024. RadicalxChange/quadratic-voting. RadicalxChange. (May 2024). Retrieved June 17, 2024 from.
[56] [n. d.] Read the Whitepaper | Gitcoin. https://www.gitcoin.co/whitepaper/read. (). Retrieved June 17, 2024 from.
[57] Duncan Rintoul. [n. d.] Visual and animated response formats in web surveys: Do they produce better data, or is it all just fun and games?, 126.
[58] Susana Rubio, Eva Díaz, Jesús Martín, and José M. Puente. 2004. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology*, 53, 1, 61–86. DOI: 10.1111/j.1464-0597.2004.00161.x.
[59] Thomas L Saaty and Mujgan S Ozdemir. 2003. Why the magic number seven plus or minus two. *Mathematical and computer modelling*, 38, 3-4, 233–244.
[60] Thomas L. Saaty. 1987. Principles of the Analytic Hierarchy Process. In *Expert Judgment and Expert Systems*. Jeryl L. Mumpower, Ortwin Renn, Lawrence D. Phillips, and V. R. R. Uppuluri, editors. Springer, Berlin, Heidelberg, 27–73. ISBN: 978-3-642-86679-1. DOI: 10.1007/978-3-642-86679-1_3.
[61] Thomas L. Saaty and Kirti Peniwati. 2013. *Group Decision Making: Drawing Out and Reconciling Differences*. RWS Publications, (Nov. 2013). ISBN: 978-1-888603-22-4.
[62] Herbert A. Simon. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69, 1, 99–118. JSTOR: 1884852. DOI: 10.2307/1884852.
[63] Tobin South, Leon Erichsen, Shrey Jain, Petar Maymounkov, Scott Moore, and E. Glen Weyl. 2024. Plural Management. SSRN Scholarly Paper. Rochester, NY, (Jan. 2024). DOI: 10.2139/ssrn.4688040.
[64] Fritz Strack and Leonard L. Martin. 1987. Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys. In *Social Information Processing and Survey Methodology*. Recent Research in Psychology. Hans-J. Hippler, Norbert Schwarz, and Seymour Sudman, editors. Springer, New York, NY, 123–148. ISBN: 978-1-4612-4798-2. DOI: 10.1007/978-1-4612-4798-2_7.
[65] Kathryn Summers, Dana Chisnell, Drew Davies, Noel Alton, and Megan McKeever. 2014. Making voting accessible: designing digital ballot marking for people with low literacy and mild cognitive disabilities. In *2014 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE 14)*.
[66] John Sweller. 2011. Cognitive Load Theory. In *Psychology of Learning and Motivation*. Vol. 55. Elsevier, 37–76. ISBN: 978-0-12-387691-1. DOI: 10.1016/B978-0-12-387691-1.00002-8.
[67] Internet Team. [n. d.] Taiwan Digital Minister highlights country's use of technology to bolster democracy in FT interview. https://www.roc-taiwan.org/uk_en/post/6295.html. (). Retrieved June 13, 2024 from.
[68] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness. Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, CT, US, x, 293. ISBN: 978-0-300-12223-7.
[69] Jerry P Timbrook. 2013. *A Comparison of a Traditional Ranking Format to a Drag-and-Drop Format with Stacking*. PhD thesis. University of Dayton.

[70] Vera Toepoel and Frederik Funke. 2018. Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Mathematical Population Studies*, 25, 2, (Apr. 2018), 112–122. DOI: 10.1080/08898480.2018.1439245.

[71] Vera Toepoel, Brenda Vermeeren, and Baran Metin. 2019. Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on Substantive Response, Questionnaire Experience and Response Time. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 142, 1, (Apr. 2019), 57–74. DOI: 10.1177/0759106319834665.

[72] Amos Tversky and Daniel Kahneman. [n. d.] Judgments of and by Representativeness.

[73] Jonathan N. Wand, Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Michael C. Herron, and Henry E. Brady. 2001. The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida. *The American Political Science Review*, 95, 4, 793–810. Retrieved Dec. 16, 2023 from JSTOR: 3117714.

[74] Bert Weijters, Elke Cabooter, and Niels Schillewaert. 2010. The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 3, (Sept. 2010), 236–247. DOI: 10.1016/j.ijresmar.2010.02.004.

[75] Bert Weijters, Kobe Millet, and Elke Cabooter. 2021. Extremity in horizontal and vertical Likert scale format responses. Some evidence on how visual distance between response categories influences extreme responding. *International Journal of Research in Marketing*, 38, 1, (Mar. 2021), 85–103. DOI: 10.1016/j.ijresmar.2020.04.002.

[76] Christopher D Wickens and Anthony D Andre. 1990. Proximity compatibility and information display: Effects of color, space, and objectness on information integration. *Human factors*, 32, 1, 61–77.

[77] Christopher D Wickens and C Melody Carswell. 1995. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human factors*, 37, 3, 473–494.

[78] Ziang Xiao, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions. *ACM Transactions on Computer-Human Interaction*, 27, 3, (June 2020), 15:1–15:37. DOI: 10.1145/3381804.

[79] yehjxraymond. 2024. Yehjxraymond/qv-app. (Mar. 2024). Retrieved June 17, 2024 from.