Quadratic Survey Interface

ANONYMOUS AUTHOR(S)*

Here is the abstract that cites [1] and Somebody et al..

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), 30 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 [??]. While much research has explored the influence of the former two aspects, this research focuses on the design of a specific survey – Quadratic Surveys.

A **Quadratic Survey** is a surveying tool that employs the quadratic mechanism. In this paper, we use the term **Quadratic Mecahnism** to describe the mechanism where individuals express some intensity value bounded by a given budget using a quadratic formula. **Quadratic Voting** (QV), also known as plural voting, is the most well-known application of this mechanism. Quadratic Voting allows participants to allocate a finite amount of credits across a list of options, voting multiple times to demonstrate their strength of approval, provided the quadratic cost remains within their given budget [??]. In this paper, we use the term **Quadratic Survey** (QS) to focus on the surveys that use the quadratic mechanism to elicit individual preferences to gather public opinions. Recent work has demonstrated that QV can gauge public opinions [??] and can be transformed into surveys that outperform Likert scale surveys at eliciting individual preferences under resource-constrained scenarios [??].

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

Thus, the primary goal of this study is to present a novel interactive interface designed for quadratic surveys, which could presumably extend to other applications that utilize the quadratic mechanism. The Quadratic Mechanism is undeniably more complex than other voting and surveying

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

mechanisms like the Likert scale survey [??], where individuals select from a few responses, and Approval Voting [??], where participants mark as many options as they approve without constraints. Responding to a QS involves expressing numerical representations of a full set of constructed preferences. To mitigate such burden, the interface exploits human preference construction as Lichtenstein and Slovic [??] pointed out, when individuals do not have clear known preferences, they construct preferences in situ. This design involves scaffolding a two-step process involving an initial organization phase and a subsequent voting phase. YHC: Yi-Hung: Exploits? Is it good or bad if human construct preferences in-situ when they do not have clear known preferences. What do you mean by exploitation here?

Despite the advocacy of Quadratic Voting by Posner and Weyl [??], and its experimentation in various contexts such as the U.S. Colorado state government [??], the Democratic Caucus of the House of Representatives [??] in the U.S., government-sponsored hackathons [??], and the recent Gov4git [??], there is a notable gap that no peer-reviewed research has focused on the design perspective of such mechanisms. This increasing attention highlights the relevance and potential impact of QS.

The importance of design in surveying tools, the growing usage of applications on the quadratic mechanism, and the lack of research on the design regarding quadratic mechanisms that one could apply, motivated our main research question: *How can we design interactive interfaces to support participants in completing Quadratic Surveys?*

Quadratic Surveys and other quadratic mechanism powered applications allow individuals to allocate resources across multiple options, but the presence of many options can overwhelm participants, potentially compromising decision-making quality. Prior research in behavioral economics and marketing has pointed out the challenge of choice overload [??] and overchoice [??]. It can be difficult for decision-makers to reduce the number of options present in a survey. At the same time, reducing cognitive load and making survey less challenging is critical if the survey designer ought to reduce survey response errors and provide 'good enough' answers [??]. Effective design may mitigate these overload challenges, ensuring that the Quadratic Survey mechanism fulfills its potential to capture detailed and accurate preferences. Thus, more concretely, this study focused on answering the following three research questions:

- RQ1. How does the number of options on QS impact respondents' cognitive load?
- 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?

Before answering these research questions, we iteratively designed and built an interactive interface informed by prior literature in the questionnaire and survey response format. Then, we designed a two-by-two between-subject in-lab study where each group of participants experienced a QS with either a short or long list of options using a text-based or interactive-based interface. Participants' cognitive load was measured using NASA-TLX, a widely used assessment tool, followed by interviews. We recruited 41 participants from a Midwestern local community, asking participants their attitudes across a wide range of societal issues.

Contributions. In this study, proposed an interactive interface for QS and highlighted the importance of using a two-step "organize then vote" interactive interface for QS surveys with long lists of options. While we were not able to show statistically significant differences in overall cognitive load (NASA-TLX scores) between text-based and interactive interfaces, qualitative interviews revealed a shift in cognitive load sources from operational causes in the text-based interface to strategic and higher-level causes in the interactive interface. YHC: might sound weak for the contributions when you presented the negative results first? When there are more choices, the interactive interface allows for more frequent and incremental updates, supporting a more iterative and reflective decision-making process and indicating better engagement and understanding. This study also highlighted design recommendations for the use of QS and directions for the design of the quadratic mechanism. YHC: Yi-Hung: Please ignore if it is a common sense in CSCW community. I am confused when I see these terms here. what do you mean by operational causes? and what do you mean by higher-level causes?

In the remainder of this paper, we focus on the related works in section 2. Then, we detail the design decisions for the interactive QS interface, informed by the iterative design process and prior works, in section 3. Experiment design follows in section 4. Study findings and discussion are presented in section 5 and section 6.

2 Related works

As this research situates between three core elements – quadratic mechanisms, the importance of survey and voting interface design, and cognitive psychology – this section sequentially presents the critical grounds of each dimension.

2.1 Quadratic Survey and the Quadratic Mechanism

A Quadratic Survey (QS) is a surveying technique that presents a Quadratic Voting item for surveying an individual's attitude across a series of options. Both tools share 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 Somebody et al., 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. YHC: what is arrow's impossibility theorem, does everyone in CSCW know this? Citation maybe? 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 Somebody et al. and implemented as an open-source platform by Somebody et al.. 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 [??]. 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

Somebody et al.'s study on 4,500 participants' attitudes across ten public policies, highlighting differences between QV and Likert scale survey results. Somebody et al. applied quadratic surveys in Human-Computer Interaction (HCI) and subsequently showed QV's effectiveness in reflecting true preferences in monetary decision tasks. Somebody et al. used QV in educational research to gauge student opinions on factors affecting university success, and Somebody et al. 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 Somebody et al. applies the quadratic mechanism as part of the management framework to support networked authority, which was later applied to Gov4git [??]. Quadratic Funding focuses on the redistribution of funds following outcomes from consensus made using the quadratic mechanism [??]. Despite the breadth of applications, there is little attention investigating the user experience and interface design supporting individuals to express attitudes. This oversignt is critical as prior research in survey and voting interface design demonstrated substantial impacts on respondent behaviors and outcomes.

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 [??] highlighted the impact interface design can make on election outcomes. Researchers like Somebody et al., Somebody et al., and organizations like the Center for Civic Design, which publishes reports like "Designing Usable Ballots" [??], 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 [??]. 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 [??]. On the other hand, Somebody et al. 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 [??], Anywhere Ballot [??], and Prime III [??]. In addition, Somebody et al. investigated optimal touchpoints on voting interfaces, and Somebody et al. examined zoomable voting interfaces.

These findings underscore the profound impact 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, Somebody et al. 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 [??].

While the design of voting systems and question response format markedly influence voter behavior and decision accuracy, these interface elements also directly impact the cognitive load on users. An effective design would enhance usability and reduce cognitive challenges faced by survey respondents, especially in complex response mechanisms like QS.

2.3 Cognitive Challenges and Choice Overload

Despite the deep insight prior research learned about voting and surveying techniques and the robust mechanism demonstrated in theoretical applications of quadratic mechanisms, the inherent challenge that survey respondents are required to make many difficult decisions poses a unique cognitive challenge that no prior literature has tackled. Somebody et al. 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 a 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 [??]. 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 [??]. 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 [??].

Additionally, Somebody et al. 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 [??].

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 [??], and the 2019 Taiwan Presidential Hackathon featured 136 proposals [??]. These psychological and behavioral research highlighted the importance of understanding how individuals navigate and can potentially benefit from interfaces under long-list QS conditions.

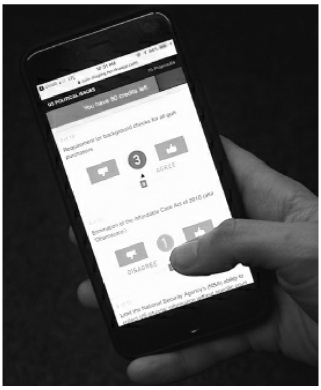
3 Interface Design

In this study, we developed an interactive interface for QS based on prior literature for the experiment condition. Since there were no studies and standard interfaces associated with quadratic survey-based tools, we designed and constructed two versions of the interface to study how interactive components influenced participants' cognitive load and behaviors. In the following subsections, we describe and justify these interface design decisions.

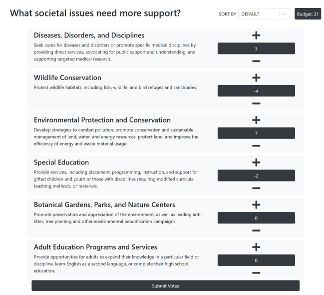
3.1 Text-based Interface

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

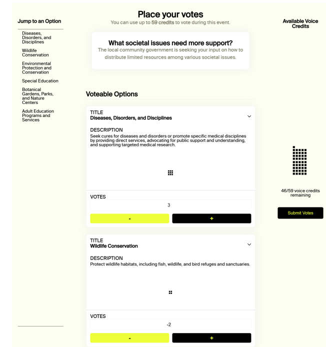
- Option list: A list of options contesting for votes.
- Vote Controls: Two buttons to increase and decrease votes associated with each option.
- Individual vote tally: A representation of votes associated with an option.



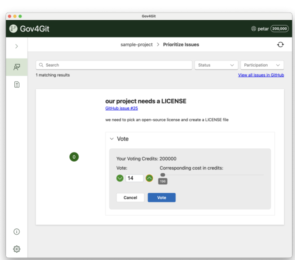
(a) Software designed by WeDesign used in [??]. Image taken from [??].



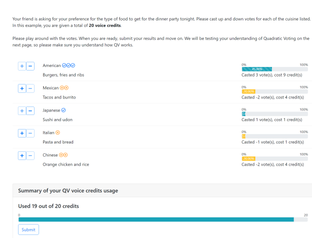
(b) An open-source QV interface [??] with a publicly available service.



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



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



(e) The interface used in the research by Somebody et al..

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

- Summary: A summary that automatically calculated the cost across options and the remaining budget.

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 suggested that the use of emojis might influence the interpretations of surveys [??] and decrease user satisfaction [??]. Prior literature also noted that not all data visualization elements reduce cognitive demand [??]. Even though effective visualization can aid decision making, it remains an open question that this study does not aim to address, thus we also removed all visualization elements such as blocks, progress bars, and percentage indicators. Lastly, different from all these interfaces, we decided to present all the options on the same screen. Prior research emphasized 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 contained the question prompt at the top of the screen. The options were presented in the list underneath the prompt. Survey respondents could update the votes by selecting from a dropdown that provided all possible voting options and cost given the number of credits. A small summary box to the right of the interface showed the current total cost and the remaining credits for the respondent. Options were always randomly presented on the interface to avoid ordering bias [??].

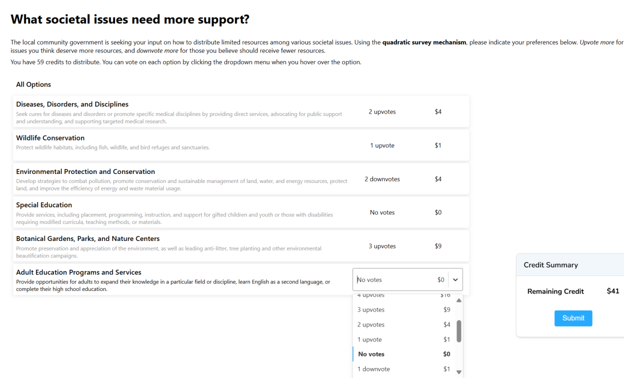


Fig. 2. The text-based interface

3.2 Interactive Interface

The design objective for the interactive interface was to facilitate preference construction and reduce cognitive load. The interactive interface, shown in Figure 3, built additional interactive elements on top of the text interface to maintain consistency that allowed 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.

A two phase approach. If preferences are constructed, by nature, they consist of a series of constructed decision-making processes [??]. Two major decision-making theories informed the design decision of a two-step interaction interface design: Somebody et al.'s Search for a Dominance Structure Theory (Dominance Theory) and Somebody et al.'s Differentiation and Consolidation Theory (Diff-Con Theory). The former suggested that decision-makers prioritize creating dominant choices to minimize cognitive effort by focusing on evidently superior options [??]. The latter described a two-phase process where decisions are formed by initially *differentiating* among alternatives and then *consolidating* these distinctions to form a stable preference [??]. Both theories guided the design decision in building the interactive experience to reduce initial decision dimensions and the mental procedures involved in emphasizing relatively important options and forming decisions.

Hence, the two-phase design – organize then vote – aimed to facilitate this cognitive journey explicitly. The first phase focused on differentiating and identifying dominant options, enabling survey respondents to preliminarily categorize and prioritize their choices. The second phase presented these categorized options in a comparable manner, with drag-and drop functionality, enhancing one's ability to consolidate preferences. This structured approach aimed to construct a clear decision-making procedure that reduced cognitive load and enhanced clarity and confidence in the decisions made.

Phase 1: Organization Phase. The goal of the organization phase was to support participants in identifying dominating options or partitioning options into differentiable groups. In this section, we first describe how the interaction worked, then we detail reasons for the different design decisions implemented.

The organizing interface, depicted on the left side of Figure 3, sequentially presented each survey option. Participants selected a response among three ordinal categories – lean positive, lean negative, or lean neutral. Once selected, the system moved that option to the respective category. Participants could skip the option if they did not want to indicate a preference. Options within the groups were draggable and rearrangeable to other groups should the participants wish.

Somebody et al.'s research showed that upon understanding a survey question, respondents either recalled a prior judgment or constructed a new one when completing an attitude survey. In addition, revealing one option at a time gated the amount of information presented to the survey respondent and thereby reduced the extraneous load [??]. This process allowed participants to form or express opinions on individual options incrementally.

The three possible options, positive, neutral, and negative, aimed to scaffold participants in constructing their own choice architecture [??], which strategically segmented options into diverse and alternative choice presentations while avoiding the biases from defaults. We believed that these three categories were sufficient for participants to segment the options. However, we chose not to limit the number of options one could place into a category to prevent restricting user agency, a core user interface design principle [??].

Immediate feedback displaying the placement of options and allowing participants to rearrange them via drag-and-drop adhered to key interface design principles [??]. At the same time, it allowed finer grain control for individuals to surface dominating options and create differentiating groups of options.

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.

Phase 2: Interactive Voting Phase. The objective of the voting phase was to facilitate the consolidation of differentiated options through interactive elements while reinforcing the differentiation across options constructed by participants from the previous phase. This facilitation was achieved by retaining the drag-and-drop functionality for direct manipulation of position and enabling sorting within each category.

Options were displayed as they were categorized within each category from the previous step and in the following section orders – lean positive, lean neutral, lean negative, and skipped or undecided as detailed on the right-hand side of Figure 3. The Skipped or Undecided category contained options left in the organization queue, possibly because survey respondents had a pre-existing preference or chose not to organize their thoughts further. The original order within these categories was preserved to maintain and reinforce the differentiated options. Respondents had the flexibility to return to the organization interface at any point during the survey to revise their choices.

In the interactive interface, options remained draggable, enabling participants to modify or reinforce their preference decisions as needed. Each category featured a sort-by-vote function that enabled reordering within the same category. Although these interactions did not influence the final voting outcome, they were designed to support consolidation and positional proximity in information organization. This design aimed to automate the grouping of similar options while providing an intuitive drag-and-drop mechanism, thereby facilitating decision-making by placing similar options near each other. This echoed the principles of the proximity compatibility principle, particularly emphasizing spatial proximity and mental compatibility [??]. The interface design

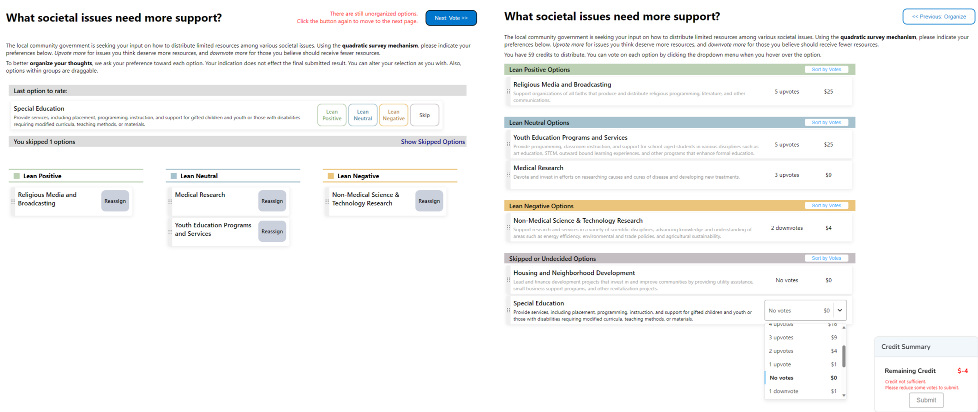


Fig. 3. The interactive interface

anticipated that participants would find it easier to consolidate their choices when similar options were positioned close together, thereby reducing cognitive load.

While multiple interaction mechanisms exist, drag-and-drop has been extensively explored in rank-based surveys. For instance, Somebody et al. demonstrated that replacing drag-and-drop with traditional number-filling rank-based questions improved participants' satisfaction with little trade-off in their time. Similarly, Somebody et al. found that integrating drag-and-drop into the ranking process, despite potentially reducing outcome stability, was justified by the increased satisfaction and ease of use reported by respondents. The trade-off was deemed worthwhile as QS did not use the final position of options as part of the outcome if it significantly enhanced user satisfaction and usability [??].

Together, these design decisions led to our belief that a two-step interactive interface with direct interface manipulation could 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 services were 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. When recruiting participants, we did not reveal that the goal of this study was to measure their cognitive load and study their behaviors, rather a study that elicited community members' attitudes on societal issues. The reason we withheld such information was to prevent response biases. This study was reviewed and approved by the college Institutional Review Board.

[1] link-to-github

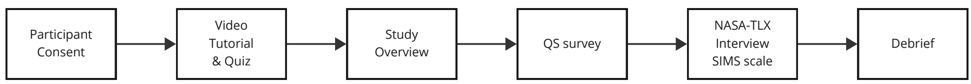


Fig. 4. Study protocol

Figure 4 shows a visual representation of the study protocol. Study participants were invited to the lab to participate in this study. The reason we made this experiment design decision was 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, included potential multitasking or interruptions by others. An in-lab study also allowed participants to operate across a consistent device that researchers had full control over. More specifically, the experiment involved participants operating on a 32-inch vertical monitor. This setup assured study participants, despite any condition in the study, could see all options on a QS, minimizing hidden information from an individual's decision-making process.

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

The researcher then primed the participant that the purpose of this study was to assist local community organizers in understanding community members' preferences on a wide variety of societal issues so they could potentially distribute limited resources better. Participants were 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 an interactive interface

Participants began completing the survey independently, without the researcher's presence. Upon completion, they contacted the researcher, who then requested they complete the NASA-TLX to assess cognitive load. This was followed by a short semi-structured interview to gain insights into the participants' experiences. This interview was audio recorded. The session concluded with a debriefing and a $15 cash compensation for their participation. The debriefing explained to the participant that not disclosing the purpose of the survey was to measure cognitive load and interface design and allowed for participants to ask any questions.

The study was designed as a between-subject study for two reasons. First, we aimed to minimize the study fatigue that might occur given the complexity of responding to a QS. To complete a QS survey, participants could take up to 20 minutes. Thus, it was difficult to conduct back-to-back experiments that measure cognitive load. We chose 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 aimed 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 wanted to ensure that participants were not influenced by their previous preferences which could influence their perceived cognitive load.

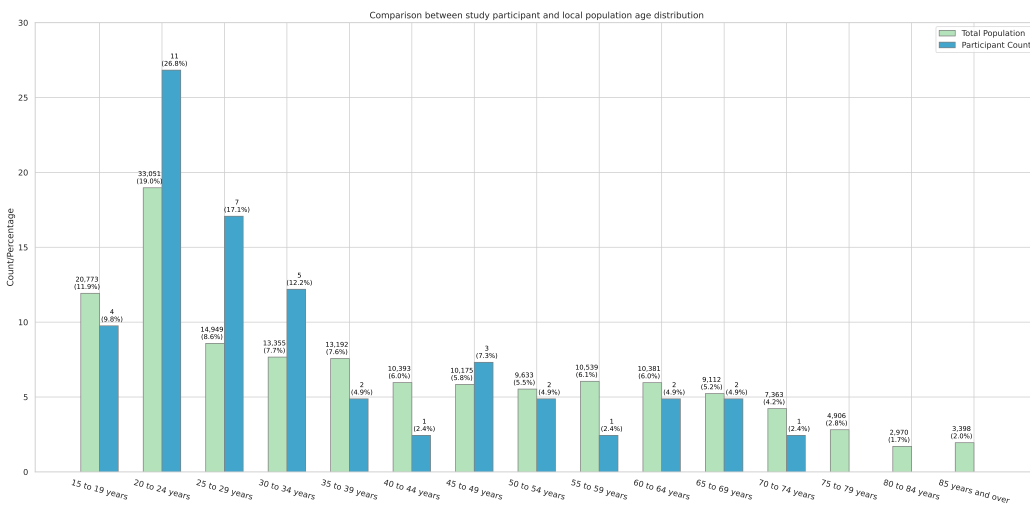
In an ideal world, understanding participants' cognitive load across multiple options would require enumerating all possible numbers of options and eliciting the "breaking point" where the participant experiences cognitive overload. Unfortunately, this was not feasible. Iterating through all possible numbers of options was very costly, both in time and resources. Therefore, we referred 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) recommended options fewer than ten and seven, respectively [??]. However, we were not aware of any specific works that justified these numbers. Somebody et al. associated this value with both the cognitive processing capacity of 7 ± 2 [??] and a theoretical proof using the consistency ratio of a pairwise comparison metric [??]. This informed our decision to contain a pair of dependent variables above and below seven options. We turned to experiments designed to study choice overload. A meta-analysis by Somebody et al. 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 were likely rooted in the original choice overload experiment by Somebody et al.. 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 followed suit as described by Somebody et al., 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 [??], 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.

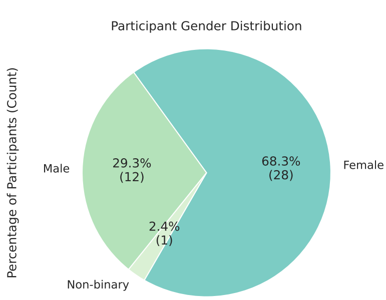
Last, we describe the quantitative measurements taken during the study: cognitive load. At the time of this study, several methods existed to measure cognitive load, including performance measures, psychophysiological measures, subjective measures, and analytical measures [??]. Given the nature of QS, 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 had to 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 [??] and ECG [??] could 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 used 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 [??]. This approach reduced between-rater variability, indicating differences in workload definitions among raters within a task and variations in workload sources between tasks [??]. Despite criticisms regarding its validity and vulnerability, NASA-TLX was commonly used due to its low cost and ease of administration [??]. It had 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 [??]. Thus, we chose NASA-TLX to measure cognitive load in our study.

5 Results

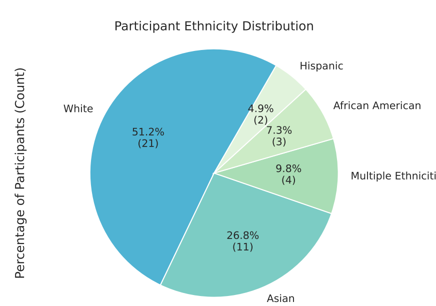
In this section, we present the results of our study, which are organized as follows: We begin with a description of participant demographics, followed by quantitative and qualitative assessments of cognitive load. We then provide a detailed analysis of participant behaviors. The section concludes with qualitative insights derived from participants' comments on their experiences and the interfaces used. Quantitative data were derived from pen-and-paper surveys and system logs captured during the study. Qualitative insights were generated from interviews, which were transcribed and thematically analyzed by the first author. All processed behavioral data are publicly available[2] to support transparency and facilitating further research.



(a) Age distribution



(b) Gender distribution



(c) Ethnicity distribution

Fig. 5. Demographic distributions: Age, Gender, and Ethnicity

[2]link-to-github

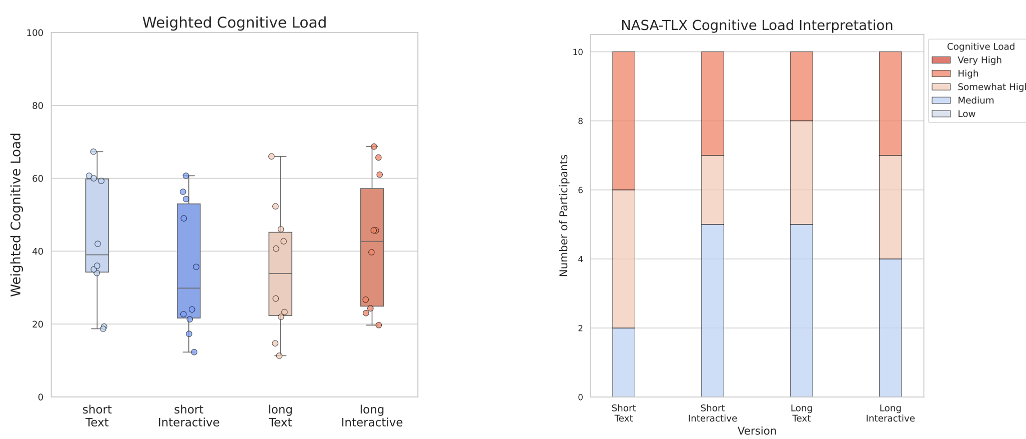
5.1 Demographics

We recruited a total of 41 participants, allocating ten to each experiment condition. Due to data quality concerns, we excluded one participant's data. The mean age of the participants was 34.63 years old, with a detailed age distribution presented alingside the county population distribution in Figure 5a. This comparison reveals that our sample closely matches the county's demographic profile, albeit with a slightly higher representation of younger adults, particularly in the 35-45 age range. As shown in Figure 5b, the majority of participants skewed toward female.

Regarding ethnicity, 51.2% of the participants identified as White, 26.8% as Asian, 7.3% as African American and 4.9% as Hispanic. Additionally, 9.8% of participants reported mixed ethnicity.

5.2 Cognitive Load Results

In this subsection, we present results in the order of the research questions. We first illustrate the NASA-TLX across four experiment conditions, providing interpretation of the results to answer RQ1 and RQ2a. Next, we report detailed findings across the six dimensions of NASA-TLX to answer RQ2b. Then, we present qualitative and quantitative findings on user behavior for RQ3. Finally, we present additional comments from participants regarding the interfaces.



(a) NASA-TLX Weight Score Distribution

(b) NASA-TLX Cognitive Interpretation

Fig. 6. NASA-TLX Results

We show the NASA-TLX weighted results in Figure 6a. A higher value refers to higher cognitive load. 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. We follow predefined mappings of NASA-TLX values to cognitive levels: low, medium, somewhat high, high, and very high, as listed by Somebody et al.. We show value interpretations in Figure 6b. The short text interface had the most participants ($N = 8$) rating their cognitive experience as somewhat high or above. The other three experiment groups showed similar cognitive load with about half of the participants experiencing medium cognitive laod and the other have somewhat high and high. No participants across all conditions expressed very high cognitive load.

These results partially answer our first two research questions. To our surprise, the longer survey did not introduce extraneous cognitive load despite the budget of the long QS increasing by 8 times and the options increasing fourfold. We deduct through a list of possible explainations. First, the interactive interface increases participants' cognitive load. However, we do not think this is the case. If it were, we would expect to see even more significant cognitive overload in the long interactive interface, resulting in lower cognitive load scores. Second, participants in the long text interface are cognitively overloaded, leading to satisficing behaviors due to the numerous decisions required to complete the task. We investigate if this is true in the following subsection. Third, we cannot rule out that the interface, contrary to our expectations, did not reduce cognitive load but rather shifts participants' cognitive load throughout the process of completing QS.

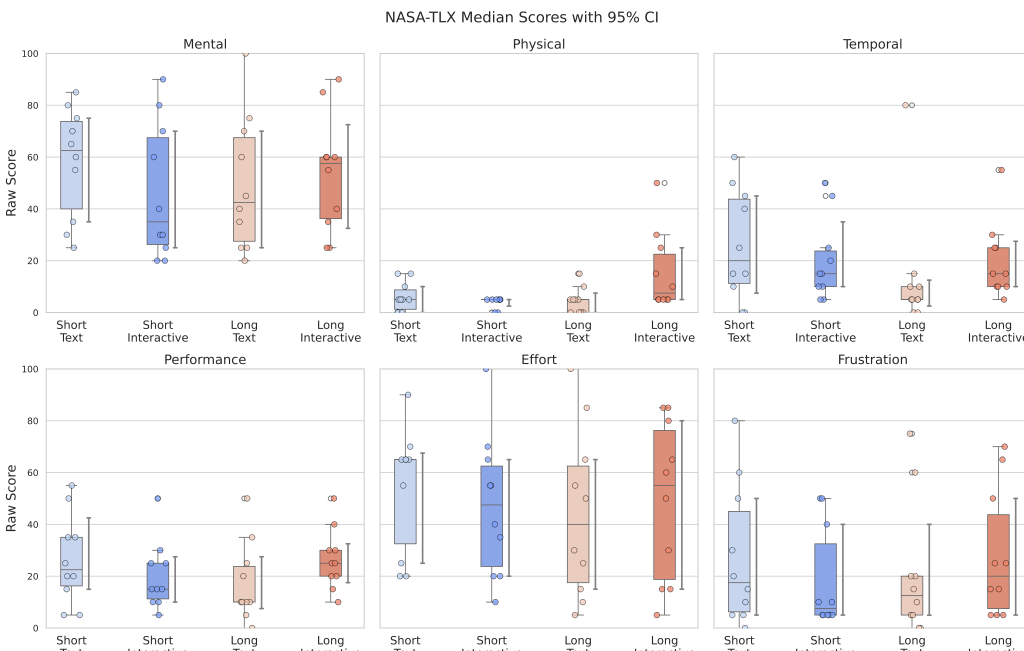


Fig. 7. NASA-TLX Results

5.3 Sources of Cognitive load

NASA-TLX consists of six weighted dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. To better understand the sources of cognitive demand and answer RQ2b, we highlight qualitative similarities and differences across four experiment conditions. This is followed by descriptive statistics and participant survey findings.

We present the raw NASA-TLX scale results in Figure 7. Next to each box plot, a line is present to denote the 95% confidence interval around the mean for each boxplot.

5.3.1 Mental Demand. Mental demand refers to the extent of mental and perceptual activity required. From the interview results, we identified two major sources contributing to mental demand: *Budget management* and *Preference construction.* Despite all experiment groups sharing these themes, we notice a distinct difference in the scope of *Preference construction*, especially when comparing long text and interactive interfaces.

Budget management. 17 participants expressed demand from trying to budget within limited credit ($N = 4$), track remaining credits ($N = 10$), and maximize the use of credits ($N = 8$). For example:

> *How many I got left that ... that I haven't voted on yet, and seeing if I and looking at the remaining credits, I'm trying to mentally divide that up before I start allocating upvotes and downvotes.*
>
> *– S006, long interactive interface*

> *And then I just wanted to make sure that I used all the credit that I had available to me, and also knowing that in order to like show your support for certain societal issues you had to like that was giving a tangential take away from other societal issues that you could support as well.*
>
> *– S032, short text interface.*

In the first quote, the participant struggles with not running out of credit while allocating credits to options they haven't yet attributed. The second quote highlights the challenge of maximizing spend while ensuring sufficient differentiation. All these factors relate to effective budget management.

Preference construction. Almost all participants ($N = 36$) experienced an increase in mental demand due to preference construction. This can be broken down into three sources: determining relative preference ($N = 16$), where participants focus on internal evaluation and comparison among different options; option prioritization ($N = 17$), where participants make trade-offs to identify high-priority options and translate internal preferences into a subset of options; and precise resource allocation ($N = 30$), where participants allocate specific values or adjustments to represent their preferences. For each of these sources, we show an example:

> *Figuring out my priorities, and how much I prioritize option 1 over option 2. What is the difference between those 2 on my priority list?*
>
> *– S002, short interactive interface, determining relative preference*

> *I knew which ones that I wanted to dedicate the most to, and I knew which one I wanted to dedicate the least to. But it was that middle area that was kind of a grey area.*
>
> *– S008, short interactive interface, option prioritization*

> *I'm not sure how to put into words ... like having to pick how many upvotes would go to each one*
>
> *– S023, long text interface, precice resource allocation*

While these sources are common across all experiment groups, when we focus on participants in the long QS condition, the sources of preference construction showed different scopes between participants using text and interactive interfaces. More specifically, participants ($N = 8$) in the long text interface tend to experience mental demand from preference construction by thinking about issues more narrowly and focusing on personal relevance. Conversely, participants ($N = 9$) in the long interactive interface experience higher mental demand from considering the broader societal impact and evaluating options more holistically. Only four participants in the long text interface expressed a holistic view, and three participants in the long interactive interface expressed a narrow and personal view.

> *Trying to figure out what upvotes I should give it you know ... compared to ... I even kind of went back compared to the other topics: <topic one> compared to <topic two>, and even with like <topic three>, I kind of went back and forth between those two. [...] So it was very mental tasking for me.*
>
> *– S015, long text interface*

> *[...] really going through the rest of the categories and deciding okay, which are the pressing issues of our time and which are the pressing issues for this particular society that that I live in. [...] You know these*

causes need a lot more funding, and and others can probably still have some sort of an impact, even with less resources.

– S019, long interactive interface

In the first quote, participants expressed mental demand narrowly focused on three options, trying to recall specific characteristics to differentiate among the options. In the second quote, participants consider how options play a role in society and the bigger picture, aiming to maximize impact. While these differences seem subtle, they indicate a shift in cognitive load. It is possible that exposing participants to all options during the organizing phase forced them to think through all options.

Circling back to mental demand related to budget management across these two experiment conditions, we find that long text interface participants focused on more operational behaviors such as:

So I wanted to be fair. [. . .] I actually took my calculator out and said [. . .] how much would it be if I equally distributed it and then how do I do that? Do I wanna do it all equally or not?

– S020, long text interface

compared to more procedures involving more strategic planning such as:

I wanted to make sure I wanted to give some credit to everything [. . .] I'm trying to make sure that I had without doing a lot of . . . I guess redos is trying to kind of get it right the first time on how I weight things.

– S032, long interactive interface

Strategic planning does not refer to gaming out others or 'winning' a game but rather to high-level thinking processes that consider strategies and plans to tackle a challenge, compared to operational tasks such as adjusting a specific vote value. While we did not notice significant differences in mental demand raw values (Figure 7, top left figure) across the four experiment groups, the different actions regarding budget management and preference construction show a shift in mental demand across experiment conditions.

5.3.2 Physical Demand. Physical demand refers to the physical effort required to complete a task, such as physical exertion or movement. Since this study involves participants sitting in front of a computer screen completing a survey, most participants reported minimal physical demand. We nonetheless report the sources of this minimal demand, which include reading text on the screen ($N = 4$), using the mouse ($N = 16$), and moving their head to navigate the vertical screen ($N = 5$). Participants emphasized that these demands were minimal, which is reflected in the low values reported in the NASA-TLX physical demand scores (Figure 7 middle top image.) Notably, 11 out of 20 participants that used the interactive interface mentioned physical demand from using the mouse, reflecting their increased interaction with the interface. This is further supported by the raw NASA-TLX physical demand scores, which show statistical significance between short and long interactive interfaces ($p < 0.01$) as well as between text and interactive interfaces in long surveys ($p < 0.05$) after running a Mann-Whitney U test.

5.3.3 Temporal Demand. Temporal demand refers to the time pressure felt by the participant while performing a task. A lower temporal demand suggests participants experience a slow and leisurely pace.

The themes we uncovered from the interviews consist of three main sources that lead to participants' increase in temporal demand. These include: *Budget*, *Decision Complexity*, and *Operational Efficiency.*

Budget. Budget is a lightly discussed theme that emerged across experiment conditions. Four participants mentioned budget increasing their temporal demand. Although budget can only decrease through spending, it is interesting that some participants expressed that the reduction in credit value created a sense of time pressure. Participants translated the increasing marginal cost of votes into higher temporal demand. As one participant said,

> *When the money was decreasing, as I was casting more upvotes or downvotes so as the money decreases I felt kind of rushed.*
>
> – *S034, long interactive interface*

Decision Complexity. Decision Complexity refers to when participants felt that there are many decisions to make. These causes are expressed in two forms—affirmative and negative. Affirmative perception refers to participants explicitly expressing that there are many decisions to make, while negative perception refers to participants describing concerns regarding the time and effort already invested in the survey.

> *So it didn't take too much time but obviously there was a lot of things to consider. So there was some temporal demand.*
>
> – *S022, short interactive interface*

> *[. . .] so at first it was like, 'Okay, this is fine.' But then on the end, I was like, maybe I should just hurry up and make a decision. So it's like at first it would been here, but then I kinda moved up near the end when I was hanging a waffling between my upvotes.* – *S024, short text interface*

The former quote pointed out participants making many decisions , while the latter highlighted the increase in temporal demand due to an expected devoted time. What we found important was that each experiment group had participants expressing both perspectives on decision complexity as a source of temporal demand. However, half of the participants ($N = 5$) in the short text interface and half of the participants ($N = 5$) in long interactive interface expressed concerns due to decision complexity. The long interactive interface involved all five participants registering an affirmative perspective. This is not surprising because participants in the long interactive interface had the most actions needed for organizing and voting. On the other hand, it is interesting to observe that four of these five participants in the short text interface expressed a negative perspective. This indicates that participants in this group are highly sensitive to their sunk cost effect.

Operational Efficiency. Unlike decision complexity, which refers to the abundance of decisions to be made, operational efficiency refers to specific and concrete operations or goals. For example, completing the survey, executing an operation, or accomplishing a specific task like updating vote values.

> *I wanna get through things in an efficient manner which doesn't necessarily mean I rush it. But it does mean that I do things expeditiously. Especially. I'd like to think I'm somewhat computer-savvy. And so to be able to move through this quickly and efficiently. I do take pride in, but it's all personal stuff. It's not nothing outwardly influencing me.*
>
> – *S032, short text interface*

> *I want the credit done but I don't want to be overthinking.*
>
> – *S013, short text interface*

The former quote refers to the participant aims to operate swiftly on the interface, not specifically related to decision making. Similarly, the latter focuses on using the credit to complete a specific goal. When asked about temporal demand, 11 participants (five from interactive and six from text interface) out of 20 who responded to the short survey expressed operational efficiency resulting in temporal demand, compared to just five (three from text and two from interactive interface) out of 20 in the long interface group.

Taking *Decision Complexity* and *Operational Efficiency* altogether, we interpret that the participants in the short survey misperceived the task as simple, seeing just six options on the screen, and thus anticipated the task to be simple and easily completed. We observe similar patterns from the NASA-TLX temporal demand raw values (Figure 7). The short text interface shows a relatively higher demand across the four groups, reflecting the demand from both decision complexity and operational efficiency. This is followed by the short interactive interface, affected by operational efficiency, and the long interactive interface, affected by decision complexity. The long text interface showed the least amount of temporal demand. Our statistical tests showed a significant difference between the long text interface and the long interactive interface ($p < 0.05$) after a Mann-Whitney U test.

It is also worth noting that three participants from the 20 who responded to the long survey mentioned that the vertical screen's ability to see all options facilitated direct comparisons and transparency about the entirety of the task, which reduced the temporal demand.

> *(Seeing) all at once I can see how many there are, so it's kind of like I can kind of tell when I will be done.*
>
> *– S041, long text interface*

5.3.4 Performance. Performance refers to how the person perceived if they successfully completed the task. A lower value refers to a good performance and vice versa. We find less differences between experiment groups qualitatively and quantitatively. However, there are notable takeaways that we can derive from the data.

First, we identified two sources of performance demand: *Operational Actions* and *Social Responsibility* from the interviews.

Operational Actions. Similar to previous demands, operational actions refer to specific and executable procedures participants can perform in the survey. These sources are shared across experiment groups. Six participants reported feeling pressured to spend all their credits or ensure they stayed within budget. Five participants were concerned that their choices did not accurately reflect their true preferences. Additionally, six participants mentioned experiencing performance demand due to the limited time, energy, and resources available, which ties into other cognitive demands. Here we show two examples:

> *I don't think I did it perfectly, because I didn't have 0 remaining credits.*
>
> *– S024, short text interface, budget management*

> *I'm concerned that it's not as reflective of my views as I wanted to be like, or I was concerned about it. [...] I was concerned that maybe it didn't.*
>
> *– S041, long text interface, preference reflectiveness*

Social Responsibility. Social Responsibility is a noteworthy source of perforamce demand, categorized into accounting *decision-maker responsibility* (N=8) and from considering *uncertainity of the outcome* (N=3). The former refers to individuals feeling guilty because they weren't able to avoid because of specific tradeoffs or that they want to be fair. For example,

> *I don't want people to think that I just like don't care about <ethnicity> people at all. I also don't think like government funding should go towards like religious organizations. You know what I mean. So I don't want somebody to think that like, I just don't care about <ethnicity> people.*
>
> *– S041, long text interface, decision-maker responsibility*

In this quote, the participant placed themselves inside the shoes of a member of the government, rather than a citizen expressing their own attitudes. This shift in roles introduced the perforamce demand, however, it demonstrated that QS shared decision maker's dilemma to individual survey respondents. This characteristic extends to the latter which further to the participants trying to forsee an outcome:

> *If I was actually running a government funding [...] I don't know how this (the survey results) might actually affect people. Some of these things might be unpopular or bad, or have outcomes that I didn't forsee.*
>
> *– S027, short interactive interface, uncertainity of the outcome*

Similar to the previous source, social responsibility is also shared across experiment groups. Considering the raw NASA-TLX scores(Figure 7), participants expressed similar levels of perforamce score. This aligns with the interview results where most participants felt positive about their final submission. This result is expected because the task is designed to reflect their preferences, not to measure performance. We further analyzed the types of satisfactory statements regarding performance.

We identified three types of satisfactory statements regarding self reported performance:

- *Did their best* refers to statements where a participant stated they exhausted their maximum effort to complete the task.
- *Feel good* efers to statements where a participant who expressed positive emotions or satisfaction about their performance or the outcome.
- *Good enough* refers to statements where a participant acknowledged that their performance or the outcome was acceptable or satisfactory, but not necessarily perfect or the best possible.

We found approximately the same number of participants in each of the four experiment groups expressed *Good enough*. Meanwhile, participants using the interactive interface across short and long groups had almost double the number of participants ($N = 11$) who expressed *Feel good* compared to the text interface ($N = 6$). On the other hand, the text interface had slightly more participants ($N = 5$) who expressed *Did their best* compared to the interactive interface ($N = 3$).

This result highlights a few important takeaways. First, participants from all experiment groups expressed satisficing behaviors (*Good enough*) with no particular group reporting a higher frequency of this behavior. Second, participants using the text interface are experiencing challenges that make them feel they have to do their best to complete the task. Last, participants using the interactive interface are generally positive about their experience and the outcome.

5.3.5 Effort. Effort refers to the amount of work required to achieve a level of performance. It includes the intensity of both mental and physical resources expended during the task.

Similar to our analysis for mental demand, we code the source of effort into to major categories: *Operational Tasks* and *Strategic Planning*.

Operational Tasks. Similiar to performance, we focus on operational tasks that contributed to effort with a narrow focus including: navigating the interface, managing the budget at an operational scale (i.e., making sure not to run out of budget, making specific updates between two options), or translating an opinion to a quantifiable adjustment on the survey. These narrower low level operations involves taking effort to making updates or actions related to the interface itself. We show two examples associated with different aspects of operational tasks that influence precieved effort:

> *And then I wanted to bump up (an option) maybe to 4 or <option> to 5 and realize I couldn't. My point (number of votes) had to like back down a little bit . . . So that would be effort came in of how do I want to really rearrange this to make it (the budget spending) maximize?*
>
> *– S029, short text interface*

> *So it was like it was very . . . I have to put a lot of effort in terms of you know . . . think about each dimension that if I give one credit to <option name> whether it will affect my credits on <another option name>.*
>
> *– S005, long text interface*

Notably, 14 of the 20 participants using the text interface expressed overwhelminly mentioned sources related to such sources, compared to less than half of the participants ($N = 7$) from the interactive interface, with the lowest mention by the long interactive interface group ($N = 2$). We review the other category before making interpretations.

Strategic Planning. Opposite to operational tasks, strategic planning follow definitions established for mental demand which involces higher level strategies to complete the survey. We further derive two distinct types of planning: *personal* and *global*. *Personal strategic planning* refers to taking effort to translate preferences onto the survey without considering governing values or broader beliefs. For example, this participant expressed effort from retrieving past experiences to inform their decisions:

> *[...] having that prior experience and being able to quickly link it to a tangible thing that I've experienced in my personal life.*
>
> *– S032, short text interface*

> *And really the bulk of the effort was how to rank order these (options) and allocate the resources behind the upvotes so that I can accurately depict what I want ... say, a committee to focus on and allocate actual fungible resources, too.*
>
> *– S019, long interactive interface*

While the difference in the number of citations to personal strategic planning are less pronounced across groups, the interactive interface (N=13) still scores slightly higher counts compared to the text interface (N=9). *Global strategic planning*, on the other hand, involves participants formulating strategies to align with broader, communal values. This includes ensuring fairness, considering the impact of different options on the entire community, and evaluating the complex relationships between various options. For example:

> *I think, imagining the trying to imagine every outcome trying to to imagine what what else would be encompassed, encompassed by each category.*
>
> *– S027, short interactive interface*

> *Hey, even though I don't really like this idea. But what if they're important? They sort of kind of deserve some attention ... that's why I think I have the effort here.*
>
> *– S037, long interactive interface*

Both examples shows considerations beyond personal experiences, considering outcomes or social values. We notice that nearly twice as many participants (N=7) in the interactive interface expressed effort from global strategic efforts compared to the text interface (N=4). Altogether, we observe more participants using the interactive interface (N=17) reported sources of strategic effort compared to those using text-based interfaces (N=11).

Qualitative analysis in this subsubsection added clear evidence that the source of cognitive demand for effort differs between text and interactive interfaces, similiar to mental demand. Participants using the interactive interface focus less on operational tasks and more on strategic planning, specifcially global strategic planning, where they think about options holistically and beyond the option itself. This is in contrast to participants using the text interface, who focus more on operational tasks and a narrower strategic planning scope. The raw NASA-TLX effort scores (Figure 7) can then be explained that even though reported values are similar across the four experiment groups, the sources of effort differ between text and interactive interfaces.

5.3.6 Frustration. Fustration is the last dimension of NASA-TLX. It refers to the extent to which the participant is annoyed, irritated, or discouraged during the task.

Following the previous analysis, we categorize the sources of frustration into three major themes: *Operational Actions* and *Strategic Planning*

Operational Actions. Similar to the previous definitions, 15 participants highlighted this source for frustration. Six participants expressed frustration regarding credit management (i.e., overspending budget); four participants mentioned had trouble deciding the final value for the options; three participants are frustrated because they need to make multiple decisions; five participants were frustrated with the quadratic mechanism; four participants are frustrated trying to understand the content of the option or how the option connects to them. For example,

> *I was slightly frustrated when doing the task, probably because there was a budget that we kind of had to stick with it.*
>
> *– S001, long text interface, quadratic mechanism*

> *i think just frustration [...] because when i was making like the decisions on how many upvotes I could put in each section, I was running out of credits.*
>
> *– S013, short interactive interface, budget management*

These demonstrated participants frustration because they are hindered by not being able to complete specific operational actions or constraints presented by QS. What is noteable is that all experiment groups had almost half of the participants express operational frustration compared to only two participants from the long text interface group. It is not clear why they did not encounter similiar frustration.

Strategic Planning. For frustration, we further derived strategic planning into two types: *lower-level* and *higher-level*. For the former, Four participants expressed conflict between their personal preferences and what they believe would be other people's preferences. Eight participants experienced conflict between making tradeoffs among a few options. For example:

> *Because I know that's important to other people. But it just doesn't to me.*
>
> *– S010, short interactive interface*

> *I would have loved to have given more to other groups ... and I felt stressed like [...] well ... it's a group that you know is still ... you know ... important [...]* *– S020, long text interface*

These quotes showed participants trying to ahere to lower-level strategies such as considering personal prefernces or making trade-offs within a smaller scope. Compare to *higher-level strategic planning*, where six participants expressed conflicts that touch on the broader society and their core values of looking at the broader scope. Eight participants felt frustrated because they were forced to make trade-offs among *all* options instead of a few. For example,

> *I had to consider how I feel towards that ... how religious media broadcasting is being used in like today's society ... today's political environment. So yeah ... you really have to consider what is important to you. – S020, long text interface, value conflicts*

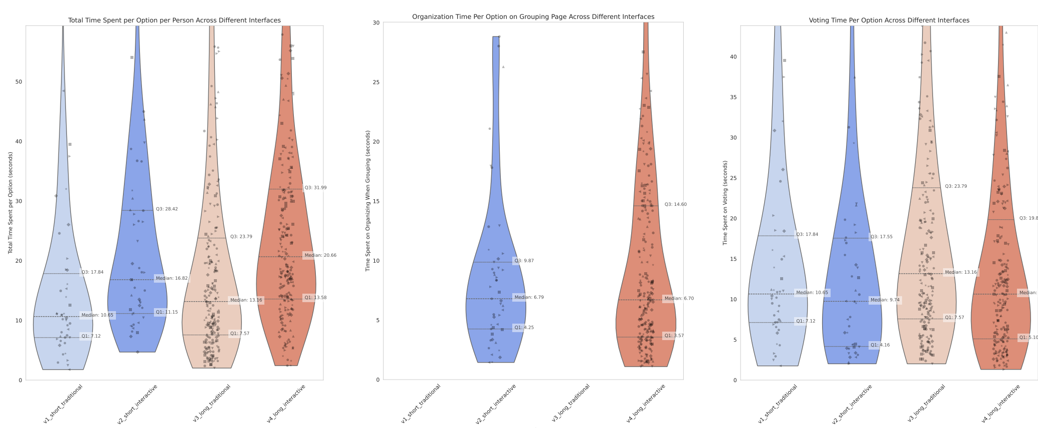
> *I think the frustration is ... I wish that we could help all of these causes, but you know it's just like anything else. You can't do everything and when it's not ... I feel like it's hard to quantify how much some of these things should be supported versus others. So when you're talking about upvotes and things that's challenging to me, it's frustrating.* *– S026, long interactive interface, considering all options*

Fustration that stemmed from strategy planning are spread across all experimental conditions. Reflecting on the raw NASA-TLX scores (Figure 7), We only see a slight difference in less frustration from the long text interface participants compared with the rest of the participants, likey due to the less frustration from operational tasks. Thus, we interpret that frustration comes more from individual's ability to discern and make decisions and not necessarily tied to specific methods in the construction of preference.

5.3.7 Summary. To recap, the analysis identified the different sources of cognitive load experienced by participants. More specifically, it highlighted differences across experimental conditions. Interactive interfaces, especially long ones, drive participants to adopt a holistic view and encourage higher-level deliberation, indicated by increased mental demand and effort. Conversely, participants perceived more operational demand when completing specific tasks using the text interface. Mental demand, effort, and temporal demand highlighted the urgency participants felt to complete tasks swiftly. This distinction doesn't mean one group of participants excludes the other group's demands, but it highlights that the main source of demand shifts with different interfaces.

5.4 Interaction Behavior Analysis

To answer RQ3 and collect evidence of shifts in participants' cognitive sources, we analyze their behaviors during the survey. We aim to understand the time participants spend on options and when they make changes. When a participant clicks their mouse on the interface to complete an action, such as 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. In this subsection, we analyze these log data. We acknowledge that the time difference between two actions indicates the time the participant took to decide and act. Although participants might be thinking about other things, this is our best proxy to study their behaviors.



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

Fig. 8. Breakdown of time per option

5.5 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.

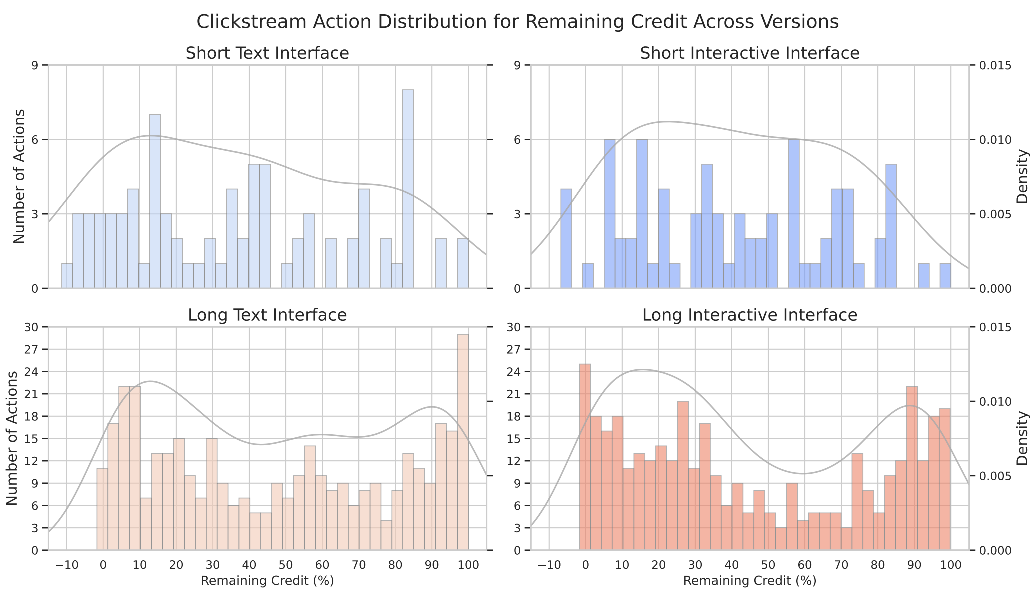


Fig. 9. voting actions across all options (needs to update chart text, remove normalization, and change the dot colors.)

5.6 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 across all four groups. In other words, different from Somebody et al. 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.

In this plot, we see two distinct patterns between the short survey and the long survey in terms of participant behaviors. Only in the long surveys did participants exhibit more actions when the budget was abundant and when it began to run out, with the long interactive interface being more significant.

Thus, we further separated the behaviors where participants made bigger changes or smaller changes to the option, specifically for the long version. In Figure 10 , we define an adjustment of four

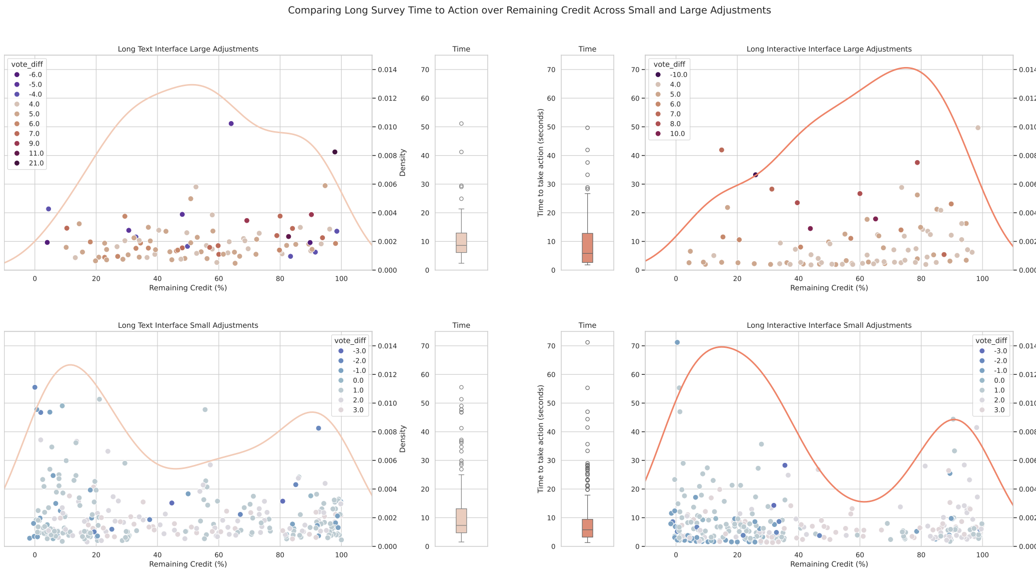


Fig. 10. Breakdown of voting actions (needs to update chart text)

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, we are able to surface the bimodal action distribution in both plots, with a even stronger signal for long interactive interface participants. Second, 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.

Iterative Support from Interactive Interface. Among all the interviews, when discussing about their experience of the interface, five particpants pointed out the importance of flexibility on the interface and how they took an incremental and iterative approach to navigating their attitude expression. All these participants are using the interactive interface. While this does not mean the study participant using the text interface did not use an iterative approach, but this highlighted the interactive interface encouraged the participants to make iterative and incremental updates. As one participant pointed out:

> *I like the fact that it remembers everything that you know. If if you make a mistake, that you don't lose all the work that you've already done. so I think that's very important is that it's an iterative process.*
> *– S019, long interactive interface.*

5.7 Interface Comments

Given our findings from cognitive load and behavior change, we futher analysis the participant's feedback on the interfaces and challenges they faced. In this subsection, we detail our findings.

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.

Automatic calculation as a critical component. More than one-third of participants (N=14) from all four experiment condition highlighted the importance of automated calculation from two perspectives: *deriving cost* and *Keeping track of spent.* Two participant highlighted the importance of automated calculation regarding the cost for each vote.

I really like having the costs of all the votes displayed. When you select the dropdown menu and ranked in order.

– S002, short interactive interface.

12 participants highlighted the summarization box and the automated summation of the current credit spent allowed them to focus on managing their next voting decision and express their preferences.

I thought I have [. . .] (to) do all the numbers or calculation myself as a part of checking my ability of doing mathematics. But I guess you have taken care of that really well, so I could really really see that how much credit has left, and [. . .] how well I should allocate [. . .] I said that credit summary to be very specific. The credit summary section was really wonderful in doing all the calculation on that end.

– S005, long text interface.

I like that I don't have to make the calculation of the dollars that it does it automatically. So if I had to do it myself it would be more tedious. And so I think that that effort and frustration and mental demand would be much higher. So I appreciate that that calculation occurs automatically and very easily.

– S017, short interactive interface.

6 Discussion and Future Works

6.1 Bounded Rationality and interface design

One core repeated theme that emerged throughout participants' responses during the interview relates to Bounded Rationality. In earilier sections, we highlighted the challenge of multiple options presented on the quadratic survey. Now, we also consider the budget avaliable to survey respondents, which credits become a broad space of possible voting options, adds additional layers and numbers of decision to make. This additional set of decisions are highlighed by several participants when they expressed apprication of the dop down menu showing all possible options with their costs precalculated.

However, the dropdown does not mitigate the bigger challenge of bounded rationality. Bounded rationality [??] highlights individuals' cognitive limitations to process and utilize information and therefore formulate and solve complex problems. This sea of decision making requires participants to recall and scramble many information at once which is extremely difficult.

So I did say, Okay, you know, you thought of enough things, you know, and so it wasn't the most effort I could put in because again, that would have been diminishing returns. I tried to think of enough things that I could make, make a meaningful decision and then move on.

– S036, short text interface.

The byproduct of bounded rationality often translate to individuals satisficing behaviors [??], creating Heuristics [??], overreliance on defaults [??], and problem decompositionn [??].

Satisficing is the most common behavior observed among the participants, which refers to survey respondents making decisions that are not optimal but rather complete a 'good enough' decision. The same participant `036` when asked about demand from performance then continues to describe:

I think that that's just not a realistic expectation (to be perfect), but I felt satisfied. [...] I felt like that (the response) was satisfied, but not perfect. Cause perfect is not a reality

– S036, short text interface.

Problem decomposition and dimension reduction is the other common behavior that we observed. Severl participant would create a duo-dimension grouping, dispite the group they are in. Participants would have categories that cluster similiar topics (i.e., all the topics related to health vs. humanitarian), and categories that depict the positivity of their preference (i.e., positive v.s. negative). The goal of highlighting bounded rationality is not to critice or exploit the possible biases that this mechanism might introduce, but highlight the importance of designing interface interventions to prevent survey respondents from enacting decisions that differs from their true preferences.

For example, the design of showing one option at a time in the interactive interface lowers the possibility for participants to be influenced by the default positions of options. One participant from the short text interface said,

Honestly, if medical research [...] I think if it was the first option, the first thing I saw, I probably would have given it more [...] because medical research [...] to me this seems like the most important, but I think if if it was the first one I saw, I think it would automatically gave it a lot more.

– S003, short text interface.

Another example comes from another participant from the long text interface.

I think the categories were kind of in the same location. The environment stuff is at the bottom. Education policy is like in the top half. So I think I just looked and determined (my votes) that way.

– S035, long text interface.

Recall that the options presented on the survey are randomly generated; even though there are some options related to the environment and education at the relevant location, participants were inferring the options to these topics. Active management of the options forced participants to think about their rough preference for each option at minimal cognitive requirements and the repositioning of options allowed participants to focus on subsets of the options during their decison making process. These are reflected in the positve responses from the interface comments on organization and direct manipulation.

6.2 Quadratic mechanism is challenging

On the other hand, the interface did not include elements that help participants kick off their voting process. One of the most difficult challenges for participants is for them to decide 'how many votes' to begin with. This challenge does not refer to the relative vote, but the starting vote. Some participants would begin by first equally distributing their credits to all options and then make adjustments (find quote), some participants established 1, 2, and 3 votes as three 'tiers' of votes as starting points, and a small handful of participants, out of our surprise, used the number of votes in the tutorial (which showed an example with 4 upvotes as the highest value), as their anchor.

This seemingly arbitrary voting decisions echos prior literature's discussion on whether an absolute value exists for an individual. Coherent arbitrariness [??], similar to the anchoring effect in marketing, refers to participants' willingness to pay can be influenced by an arbitrary value. However, the ordinal utility remains intact among the set of preferences.

Participants are also required to navigate between three elements: budget, credit, votes, and thinking about how the results would impact the 'shared resource.' This is not straightforward.

> *[. . .] get rid of the Upvote column or just get rid of the word upvote and just really focus on the money column. Listen. You're an organization or your participant. You have X amount of dollars you need to. You can only distribute X amount of dollars to these these causes. So you have to figure out which ones get the most, which ones don't get as much. [. . .]*
> *Interviewer: So when you're operating this interface. Do you feel that the more votes you're giving to a cause you're actually spending more on it?*
> *Yeah.*

– S003, short text interface.

Recall that this survey aims to assist community organizers in distributing resources to a societal cause. This participant decided to 'skip' over the quadratic formulation and the concept that their votes are governed by the quadratic formulation, drawing a direct translation between votes and the resources to which community organizers ought to contribute. While this does not invalid the power of quadratic mechanism, it builds frustration and friction for participants to construct a clean picture of how to make voting decisions.

Budget related sources draw across mental demand, temporal demand, preference demand, and fustration. These span from making sure to keep within budget to recovering from overbudgeting. While prior literature

Adds the final layer to operating the quadratic survey. While prior scarcity literature [??] believes that values and careful decisions are derived from limited resources, prospect theory [??] also highlight a higher negative value of *precieved* utility for individuals when cuts ought to be made.

These three major challenges do not threaten the integrity of Quadratic Survey and relavent tools using this mechanism, but as we demonstrated in the results section, across all experiment conditions, the NASA-TLX scales show medium to high cognitive load even for the short interactive interface. In other words, we believe that the improvement of the Quadratic Survey's ability to elicit more accurate preferences, yet, it comes at the cost of higher cognitive load.

6.3 Construction of Preference on QS

We believe that even if the interface did not significantly reduce the cognitive load from participants, the interface *shifted* the cognitive focus onto contributing upon more in-depth preference construction and more fine-tuning.

We show that participants constructed their preferences in situ. While some participants came in with some existing preference (i.e., environmental issues are important), participants need to reconsider aspects of the options they had not expected, or they need to map options on the survey to their beliefs.

> *[...] the other part of the mental demanding was probably trying to associate with (what) I'm concerned in soci(ety) [...] is that question able to deal with my social concerns like, for example, climate change [...] How does that fit in?*
>
> *– S006, long interactive interface*

> *I mean, it's not necessarily a challenge, but it's interesting to see: 'Oh, there are other aspects that I never care about.' And actually ... some people care <an option>. Sure. Why? Why (should) I spend money on that? That's the first thought that comes to mind.*
>
> *– S037, long interactive interface*

Both quotes highlighted the cause of construction of preference, but also highlighted the individuals reflecting on their personal preferences. As stated in prior works [??], QS, by listing a list of options bounded by a common credit, forced participants to consider within options. This is supported by the qualitative analysis as one of the main causes of mental demand – *preference construction.*

Next, the long text and interactive interface participant behavior analysis surfaced participants, despite sharing a similar number of actions, small adjustments on the votes are clustered toward budget depiction with lesser time spent. These fine grain adjustments represented participants are not making ad-hoc decisions as they complete QS, rather they are deciding how they ought to better utilize the remainder of the budget. This is not as obvious in the short survey likely because of the limited options (hence budget), there are less decision space and adjustments that individual participants can make. We were still able to identify the bi-modal interaction pattern but it is less clear if there are differences in the clusters.

We believe that the bimodal behavior observed in the voting actions across groups is the realization of the Differentiation and Consolidation Theory presented by Somebody et al.. The theory segmented the decision-making process into two steps: differentiation and consolidation. The former supports individuals on focusing on the differences and eliminating less favorable alternatives. The latter is a process where individuals strengthen their commitment to the chosen option, even mentally. The bimodality reflected participants differentiating some options through change of votes in the beginning of the survey and then consolidating with smaller adjustments toward the end. The two-stage decision making is stronger with the interactive interface because part of the differentiation is completed in the organization phase. As one participant mentioned:

> *I only start from the positive one [...] I finish everything ... and then I move to the second part (the netural box). [...] I want to focus on these and make sure that resources are at least they get the attention they want. And if there's surplus and they can move to the second part*
>
> *– S037, long interactive interface*

6.4 QS Usage and Design Recommendation and Future Works

This study proposed an interface that supports thought organization and preference construction. While this interface was not able to show a decrease in cognitive load significantly, this study identified additional challenges and insights into how survey respondents complete QS. We also identify open directions to support individual decision making for collective outcomes.

6.4.1 QS for critical evaluations. This study highlighted the complex cognitive challenges and in-depth consideration when ranking and rating across options using QS, even in a short survey. Similar to survey respondents needing to make trade-offs across options, researchers and agencies that wanted the additional insights and alignment to respondent preferences need to make trade-offs in assuring survey respondents have the cognitive capacity to complete such surveys rigorously. We recommend that QS should be designed for specific use cases that require critical evaluations, i.e., investment decisions or settings where participants have enough time to think and process the survey, i.e., revealing the options ahead of time for preference construction.

6.4.2 Using organization processes. This study demonstrated differences among the source of various demands shifted from operational causes to strategic and higher-level causes. This shift in preference constructions highlighted how an additional organizational phase with direct manipulation capability allowed survey respondents to access higher-level critical thinking. We believe that such behavior shift should not only relate to QS, but extends to other ranking-based surveying tools such as rank-choice voting and constant sum surveys. Further work should examine if implementing such functionality altered survey respondent's mental model.

6.5 Support for absolute credit decision

Deciding the *absolute* amount of credits in QS is very demanding. Designing interfaces and interactions to support the cold start challenge as well as helping deciding the absolute vote value yet considering limiting direct influences is an open question.

7 Conclusion

In this study, we proposed a novel interface for quadratic surveys and studied the cognitive load of survey respondents when using this interface or not across a short and long survey. Our results highlighted that while the introduction or organization phase before conducting the survey did not reduce cognitive load significantly, participants' behaviors had shifted with more engaging and higher-level considerations. We believe that the interactive interface is critical for survey respondents.

A Early prototypes of the study