Quadratic Survey Interface

ANONYMOUS AUTHOR(S)*

Here is the abstract that cites [52] and Posner and Weyl [52].

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), 39 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Effectively capturing individuals' responses, attitudes, and preferences is the cornerstone of forming consensus within computer-supported collaborative work (CSCW)and is crucial for studying human subjects in the CSCW community. Recently, Quadratic Voting [52] for collective decision-making has emerged in both public [60, 54, 74] and private sectors [27]. We introduce the term **Quadratic Surveys** (QS) in this paper to describe concurrent research that explores similar techniques to better survey individual preferences [55, 9]. The design of any attitude-capturing tool significantly affects how individual express their attitudes [17, 82, 36, 78, 19, 85, 51]. However, to date, no peer-reviewed research has examined the design perspective of similar preference-eliciting mechanisms. *How can we design interactive interfaces to support participants in completing Quadratic Surveys?*

In this study, we introduce a novel interactive QS interface (Figure **??**) that nudges QS survey respondents to take a two-step 'organize-then-vote' approach to preference construction and elicitation. QS employs a more complex mechanism than other voting and survey methods, requiring survey respondents to express numerical representations of a full set of constructed preferences. To mitigate this burden, we developed the interactive interface through several iterations, grounding our design decisions in theories of human preference construction. As Lichtenstein and Slovic [41] point out, individuals often construct preferences in situ when their preferences are not clearly defined, necessitating trade-offs, or quantifying opinions. This interface aims to scaffold the 'differentiation and construction' decision-making process Svenson [72], supporting the preference construction journey through a two-step approach: first organizing, then voting.

The QS and QV-like applications' capability to allow for better preference expression [9] across multiple options at once directly contributes to challenges such as choice overload [34] and over-choice [26]. The increase in choices can raise cognitive load, leading to more frequent survey response errors and respondents opting for 'good enough' rather than 'optimal' answers [40, 4]. In some cases, it is impractical for decision-makers to reduce the number of options in a survey if their goal is to elicit fine-grained individual preferences simultaneously. Consequently, this research

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

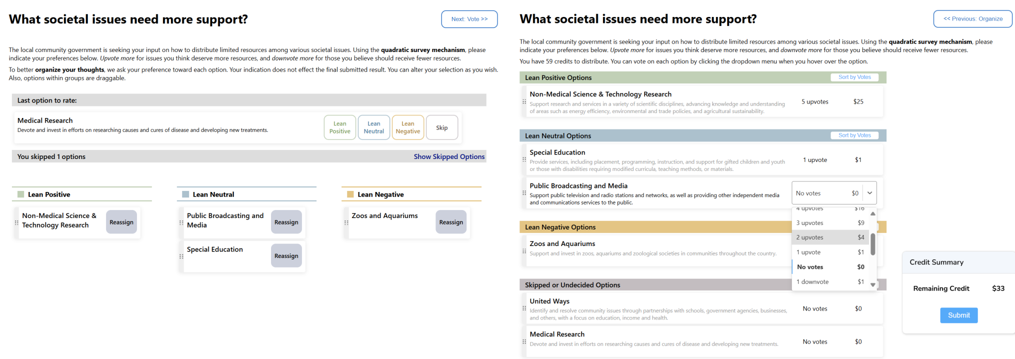


Fig. 1. A novel two-phase interactive interface for Quadratic Surveys detailed in Sec. 3.2

aims to examine the influence of interactive interfaces on QS at different option lengths. Thus, we seek to answer the following research questions:

- RQ1. How does the number of options in QS impact respondents' cognitive load?
- RQ2a. How does the two-phase interactive interface influence QS respondents' cognitive load compared to a text interface?
- RQ2b. Across the two interfaces, what are the sources of cognitive load?
- RQ3. What are differences in QS respondents' behaviors when coping with long lists of options across the two-phase interactive interface and the text interface?

We designed a two-by-two between-subject in-lab study to answer these research questions. Each group of participants experienced a QS with either a short or long list of options using a text or interactive interface we designed. Participants' cognitive load was measured using NASA-TLX, followed by interviews that focused on the different elements of cognitive load. Clicksteam data was collected from participants as they completed the QS. We recruited 41 participants from a Midwestern local community, asking participants their attitudes across a wide range of societal issues. Study results analyzed and interviews are transcribe, coded, and thematically analyzed by the author.

Contributions. This study introduces the first interactive interface specifically designed for Quadratic Surveys and Quadratic Voting-like applications, addressing a significant gap in existing research tools. Additionally, we conduct the first in-depth qualitative analysis which identifyed key factors contributing to cognitive load among survey respondents, leading to practical design recommendations. Although our findings did not show statistically significant differences in cognitive load as measured by NASA-TLX, our two-step interactive interface proved effective in facilitating more holistic and reflective decision-making, as evidenced by qualitative interview results. This scaffolding design promotes incremental updates and fosters deeper engagement, ultimately enhancing understanding and decision quality.

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

This research sits in the intersection of three core areas: quadratic surveys, survey and voting interface design, and choice overload and its cognitive challenges. In this section, we review related works in each of these areas.

2.1 Quadratic Survey and the Quadratic Mechanism

We introduce the term Quadratic Survey (QS) to describe the use of Quadratic Voting (QV) techniques in surveys to collect individual attitudes. As Quarfoot et al. [55] demonstrated, QV effectively gauges public opinions. Unlike traditional surveys that elicit either rankings *or* ratings, QS allows detailed presentations of *both* by casting multiple votes for or against options, incurring a quadratic cost. Cheng et al. [9] showed that this mechanism aligns individual preferences more accurately with their behaviors than Likert Scale surveys, especially in resource-constrained scenarios.

The concept of QS is rooted in the **Quadratic Mechanism**, a theoretical framework designed to encourage truthful revelation of individual preferences through a quadratic cost function [28]. This framework gained popularity through **Quadratic Voting (QV)**, also known as plural voting, which uses a quadratic cost function in a voting framework to facilitate collective decision-making [38]. While QV mitigates the tyranny of the majority in traditional voting systems, QS adapts these strengths to encourage truthful preference expression in surveys. QV is not subject to Arrow's impossibility theorem, which states that no voting system can perfectly aggregate individual preferences without trade-offs [45], because it does not require aggregating rankings.

To illustrate how QS works, we formally define the mechanism as follows: Each survey respondent is allocated a fixed budget, denoted as B, to distribute among various options. Participants can cast n votes for or against each option k. The cost c_k for each option k is derived as:

$$c_k = n_k^2 \quad \text{where} \quad n_k \in \mathbb{Z}$$

The total cost of all votes must not exceed the participant's budget:

$$\sum_k c_k \leq B$$

Survey results are determined by summing the total votes for each option:

$$\text{Total Votes for Option } k = \sum_{i=1}^{S} n_{i,k}$$

where S is the total number of participants, and $n_{i,k}$ is the number of votes cast by participant i for option k. Each additional vote for each option increases the marginal cost linearly, encouraging participants to vote proportionally to their level of concern for an issue [52].

In recent years, empirical studies on QV have expanded to different domains [47, 7]. Applications based on the quadratic mechanism have also grown, such as Quadratic Funding, which redistributes funds based on outcomes from consensus made using the quadratic mechanism [5, 22]. Recent work by South et al. [69] applies the quadratic mechanism to networked authority management, later used in Gov4git [27]. However, despite the growth in depth and breath of applications using the quadratic mechanism, little attention has been given to the user experience and interface design that support individuals in expressing their preference intensity.

2.2 Survey, Questionnaire, and Voting Design

The relative lack of research in quadratic mechanism and QS interface design is concerning, as prior research in survey and questionnaire interfaces demonstrated substantial impact on the response and individual's experience on even seemingly minor design decisions.

Research in the marketing and research community studying survey and questionnaire design, usability, and interactions focuses on understanding the infuence of styles and question presentation, or 'Response Format,' of a survey or questionnaire. Weijters et al. [83] 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 [77] These research highlighted outcomes are influenced by the different designs.

Given that voting, similiar to surveys and questionnaires are designed to elicit individual choices, we turn to voting interface literature. Voting interfaces can have an even more substantial influence on behaviors and outcomes.The notorious butterfly ballot [81] is one example of this – Wand et al. [81] argue that an atypical ballot design may have caused enough accidental votes to swing the 2000 U.S. Presidential Election. Researchers like Engstrom and Roberts [17], Chisnell [11], and organizations like the Center for Civic Design, which publishes reports like "Designing Usable Ballots" [16], stress the importance of interface design and how it can influence democratic processes. 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 [17]. 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 [56]. On the other hand, Everett et al. [18] 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 [39], Anywhere Ballot [71], and Prime III [15]. In addition, Gilbert et al. [25] investigated optimal touchpoints on voting interfaces, and Conrad et al. [12] examined zoomable voting interfaces.

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. Lichtenstein and Slovic [41] 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.

Daniel [14] 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 [67, 50, 80]. This behavior refers to when an individual settles on a 'good enough' solution rather than 'optimal' response. 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 [34].

Additionally, Alwin and Krosnick [2] 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 [20].

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 [**NewWayVoting**], and the 2019 Taiwan Presidential Hackathon featured 136 proposals [53]. 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

Since there are no prior studies and standard interfaces for tools using the quadratic mechanism, we designed, iterated, and developed an interactive interface for QS based on the prior literature and multiple iterations. In this section, we first describe the iterative prototyping process, detail the final design of the interactive interface, and finally, present the text-based interface designed for comparison in this study.

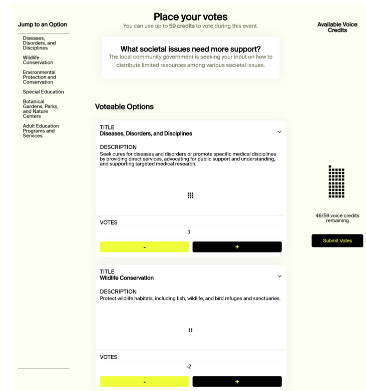
3.1 Initial Prototypes

The initial prototyping involves collecting the interface designs for existing quadratic mechanism-based software. Iterative pretests informed each subsequent design. We present these iterations, which aimed to enhance user experience in the preference construction process in the subsubsections. Alternative design considerations are presented in Appendix A

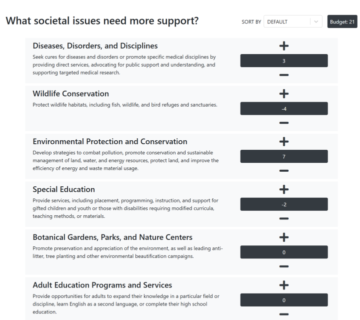
3.1.1 Paper prototype: visualizing trade-offs. The original paper prototype aimed to help visualize survey respondents' tradeoffs among options. Early on, we didn't know which components made QS more difficult than other survey techniques. We began by surveying existing interfaces (Figure 2 other than Figure 2d which did not exist near the writing of this paper). All four interfaces consists of these common components:

- Option list: A list of options contesting for votes.
- Vote Controls: Buttons to increase and decrease votes associated with each option.
- Individual vote tally: A representation of votes associated with an option.
- Summary: A summary that automatically calculated the cost across options and the remaining budget.

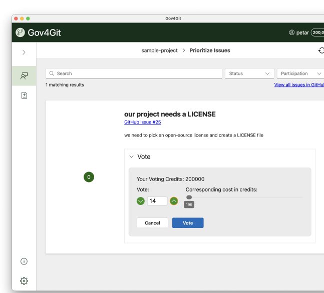
(a) Software by WeDesign, used in the first empirical QV research [55]. Little information is available about the software, except for an image from [52]. In the image, each prompt has thumbs up and down icons to update the vote in the center. The remaining budget appears as a progress bar at the top.



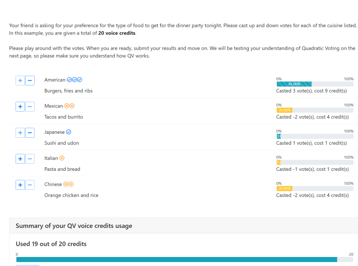
(b) An open-sourced QV interface [57] forked from GitCoin [58], used by the RadicalxChange community [1]. This interface presents total credits as small blocks. Votes are updated using plus and minus buttons, with numerical counts shown under each option and surface area as costs.



(c) An open-source QV interface [86] offers a publicly available service. Options show only the current number of votes, with credits displayed in the top right corner. This interface does not show the costs of votes but supports sorting options.



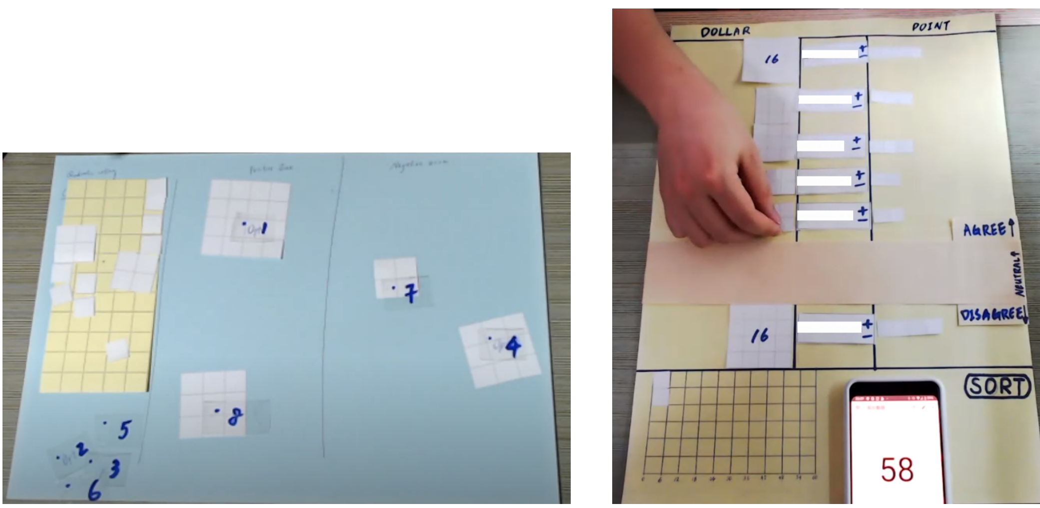
(d) The interface designed for gov4git [27] updates votes using arrows under each option, with the associated cost shown as a percentage bar to the right. A search bar exists for searching specific pull requests or issues.



(e) The interface used in the research by Cheng et al. [9] employs the most visual components. Icons depict the current number of votes, with progress bars signifying the current spending.

Fig. 2. Recent interface implementations for applications using quadratic mechanism.

To brainstorm ways to help survey respondents manage tradeoffs across options, we decomposed these options and explore several innovative layouts. Initially, we thought trade-offs is the core cause of cognitive load. In this paper, we show two versions of the paper prototypes in Figure 3. In both figures, costs are represented by blocks, similiar to Figure 2b. We imagine the survey respondents to drag and position options in the space provided unstructurally (Figure 3a) or structurally (Fig 3b). Similar to the seminal debate on direct manipulation vs. interface agents [66], the research team was unsure how much control survey respondents should have over the positioning of the options to aid the decision-making process that considers trade-offs. Different from prior interfaces, we used placements of the interface to denote positive or negative number of votes. After several pretests, we learned that the main process participants aim to do throughout the survey is establishing *relative* preferences across the options, rather then thinking so much about tradeoffs.



(a) In this paper prototype, issues are denoted by different numbers that appear on mouseover. Pretest respondents can move options anywhere in the two sections of the interface, one denoting positive and one negative. The blocks represent the cost for each option, with no indication of the number of current votes. The credits are shown in the yellow box on the left.

(b) This paper prototype separates the positive and negative areas with a 'band' at the center. Undecided options are placed inside this band. The cost and the votes on both sides of the interface are denoted by small blocks. The budget is shown in the yellow box below the interface with a numerical counter.

Fig. 3. Initial paper prototypes designed for QS interface

3.1.2 Prototype 1: Ranking-Vote. In our first prototyped tool, we aim to help survey respondents rank options to establish relative preferences before voting. As shown in Figure 4, our prototype allows respondents to move options before finalizing their votes. During our pretest, we found that respondents rarely moved the options and some questioned the need for a full ranking since it didn't affect the QS submission. Many didn't realize the options were draggable until we pointed it out. The main insight from this prototype is that creating a full rank is *not* essential for establishing *relative* preferences, leading us to consider selecting a subset of options instead of requiring a full rank among all options.

3.1.3 Prototype 2: Select-then-Vote. Based on feedback from Prototype 1, we shifted from *allowing* option ranking to a two-phase process where individuals *intentionally* select options to express

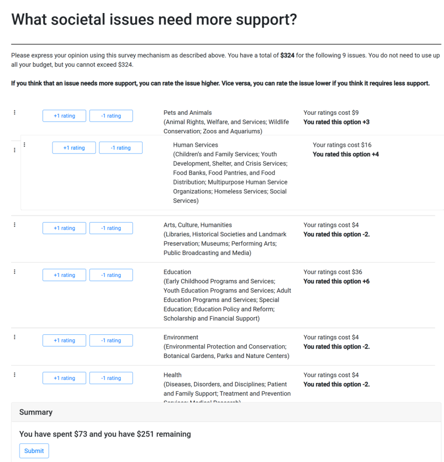
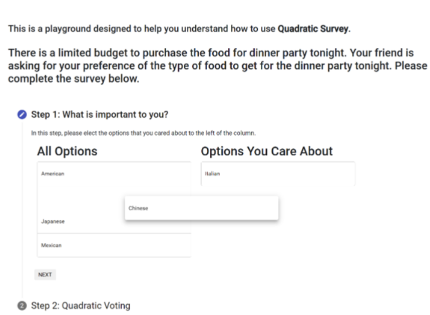


Fig. 4. A Ranking-Vote Prototype: Each option is draggable to position in a specific location amongst the full list of options. Votes can be updated using the bottons to the right of the interface with vote count and costs to the right of the interface. A summary box is placed sticky to the bottom of the screen.

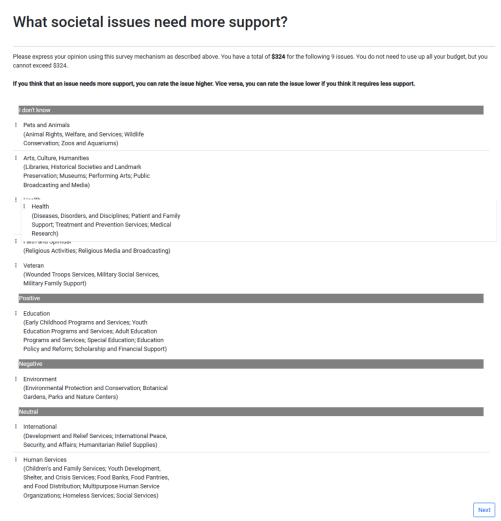


(a) Options are drag and dropped to the 'Option You Care About' box.

(b) The previous step collapses showing all voting options.

Fig. 5. A Select-then-Vote Prototype: This prototype introduces a two-step voting process. As shown in Fig. 5a, the first step involves selecting options for further consideration. Important options are placed at the top of the list for voting shown in Fig. 5b, but options can be placed anywhere on the list if desired. The rest of the controls remain the same as the previous prototype.

opinions before voting. As shown in Figure 5, survey respondents select options they cared about (Figure 5a) and the interface positioned these options at the top of the list for voting (Figure 5b). During this process, we noticed several issues. First, many survey respondents placed most options into the 'option you care about' category, defeating the design's purpose. Second, there were no indicators distinguishing between selected and remaining options. Respondents did not notice their selections were kept at the top in the voting stage and were unsure why Step 1 was necessary if all options were shown again. This informed two takeaways: selecting options to vote on is too coarse to construct relative preferences, and there needs to be a clearer distinction and connection between the two phases.



(a) The Organization Interface: Options are shown initially in the first bin labeled as 'I don't know.' Survey respondents can then drag and drop these options into the latter bins: Positive, Neutral, or Negative. On this interface, only the details of each option are shown.

(b) The Voting Interface: Voting controls appear on the left side of each option, showing the current votes and associated costs on the right. A budget summary is stuck at the bottom of the screen.

Fig. 6. Organize-then-Vote Prototype: Participants first organize their thoughts into categories, then vote on the options.

3.1.4 Prototype 3: Organize-then-Vote. Figure 6 shows the last prototype where we built on the previous takeaway by providing finer-grain groupings and creating a clear connection between option organization and voting position. Specifically, we provided three categories: Positive, Negative, and Neutral. Initially, respondents see all options under the section labeled 'I don't know,' which includes only the option descriptions. We ask respondents to move these options into the categories below. Voting controls and information appear on each option once respondents move to the subsequent page, forming a clear connection between option groups, position, and voting controls.

Feedback indicated that survey respondents are comfortable with this two-phase organize-then-vote design. Several user experience issues emerged, but they were addressable without significantly modifying this interaction structure. These issues include: First, dragging and dropping all options into different categories is cumbersome and can mislead respondents into thinking this is a ranking process, which is not the goal. Second, the position of unorganized options at the top of the voting list is counterintuitive. Third, the voting controls are disconnected from the option summaries, dividing attention between the left and right sides of the screen.

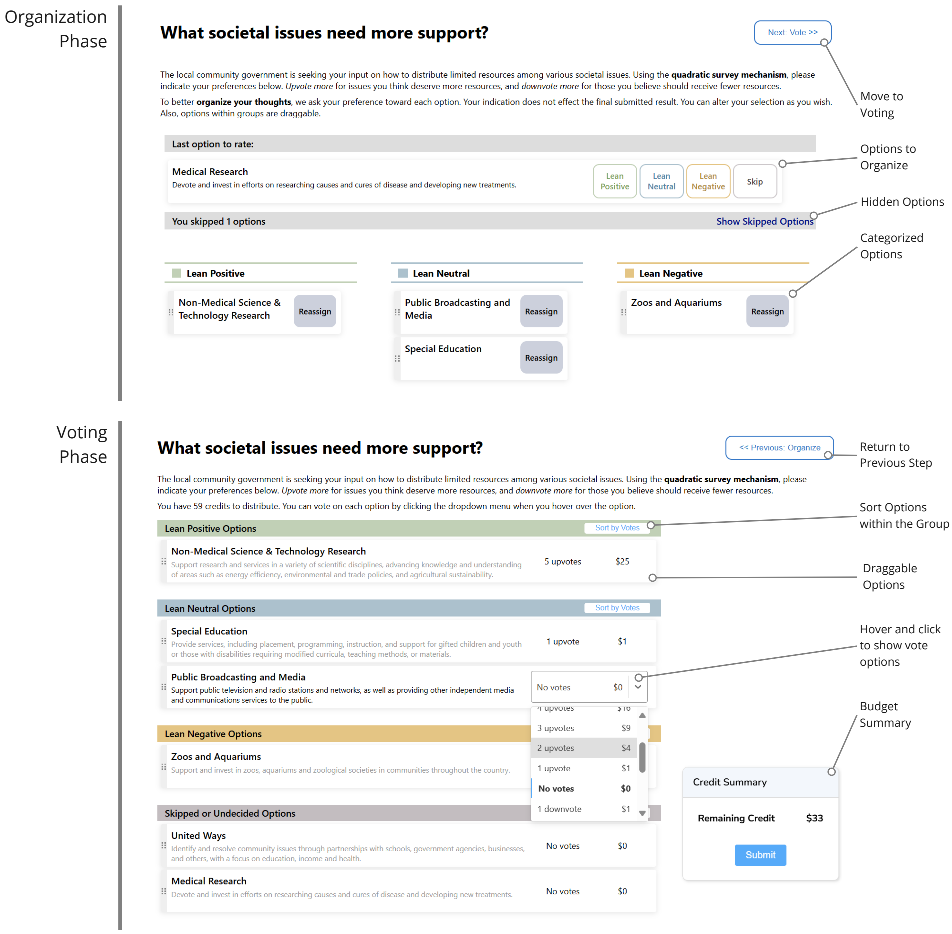


Fig. 7. The Interactive Interface: The interface consists of two phases. Survey respondents can navigate between phases using the top right button. In the organization phase, respondents will be presented with one option at a time where they can choose among four choices: Lean Positive, Lean Neutral, Lean Negative, or Skip. Skipped options are hidden and can be evaluated later. Options that are chosen will be listed below. Items can be dragged and dropped across categories or returned to the stack. In the voting phase, options are listed in the order of the four categories. When hovering over each option, respondents can select a vote for that option using the dropdown. Each dropdown contains the cost associated with the vote. A sort button allows ascending sorting within each category. A summary box tracks the remaining credit balance.

3.2 Finalizing the Interactive Interface

In the previous subsection, we highlighted critical prototype iterations that informed the final two-phase interactive process that defines the user journey. Now we present the final interactive interface and describe how it operates. In this subsection, we provide supporting evidence from prior literature that we previously omitted. These pieces of literature were omitted for clarity and focus in the previous subsection but will be reintroduced here. We also discuss the aesthetic design choices that emerged throughout the iterations.

3.2.1 Justifying a two-phase approach. Recall the ultimate objective of the interactive interface is to facilitate preference construction and reduce cognitive load. The interactive interface, shown in Figure 7 consists of two steps: An organization phase and a voting phase. Throughout both phases, survey respondents can drag-and-drop options across the presented option list.

A two phase approach. If preferences are constructed, by nature, they consist of a series of constructed decision-making processes [41]. Two major decision-making theories informed the design decision of a two-step interaction interface design: Montgomery [43]'s Search for a Dominance Structure Theory (Dominance Theory) and Svenson [72]'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 [43]. 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 [72]. Both theories echoed 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 top half of Figure 7, 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.

Strack and Martin [70]'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 [73]. This process allowed participants to form or express opinions on individual options incrementally. This design also mitigated the original concern from prototype 3 where participants accidentaly treat the orgnaizing task as a ranking task.

The three possible options, positive, neutral, and negative, aimed to scaffold participants in constructing their own choice architecture [46, 75], 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 [48].

Immediate feedback displaying the placement of options and allowing participants to rearrange them via drag-and-drop adhered to key interface design principles [48]. At the same time, it allowed finer grain control for individuals to surface dominating options and create differentiating groups of options.

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 bottom half of Figure 7. 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. This new ordering sequence mitgated the concern from prototype 3 where options without an category are left at the top of the voting interface. 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 [84]. The interface design 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, Krosnick et al. [37] demonstrated that replacing drag-and-drop with traditional number-filling rank-based questions improved participants' satisfaction with little trade-off in their time. Similarly, Timbrook [76] 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 [59].

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.

3.2.2 Aesthetic Design Decisions. There are three aesthetic design decisions that made it to the final interface. First, we decided to remove all visual elements. Prior literature suggested that the use of emojis might influence the interpretations of surveys [32] and decrease user satisfaction [78]. Prior literature also noted that not all data visualization elements reduce cognitive demand [33]. 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.

Second, the final interface has all options presented on the screen at the same time, intentionally. Unlike all the prototypes and existing interfaces, prior literature emphasized the importance of placing all the options on the same digital ballot screen to avoid losing votes. 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.

Last, we decided to use a dropdown positioned to the right of the option such that control of votes and the budget summary are placed near one another. The layout of the votes and cost was inspired by online shopping cart checkout interfaces where quantities are supplied next to the itemized costs followed by the total checkout amount. We chose a dropdown after iterating with two alternative input methods (Figure 8): the original click-based buttons and a wheel-based implementation. The former design requires survey respondents to click multiple times to reach their desired vote values. Thus, we wanted to look for a solution to aid respondents in reaching their intended value faster. A wheel-based approach allows intuitive control of the votes by using the wheel on the mouse and clicking to fine-tune the values. However, in our early pilot studies, not all participants were familiar with wheel control, thus we opted for a dropdown menu for vote selection.

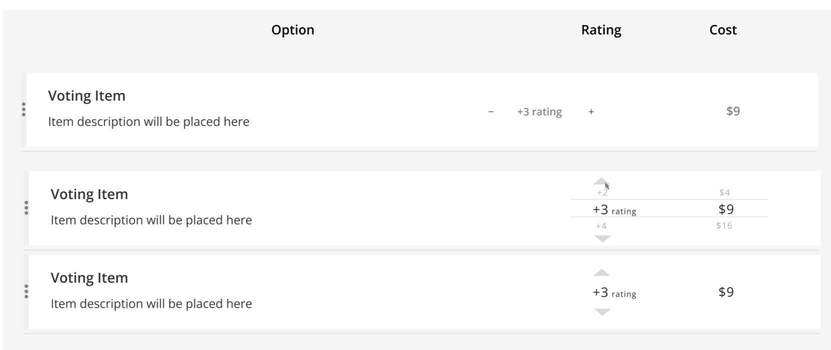


Fig. 8. The click-based and wheel-based designs for vote control included: The former mirrors traditional vote control used in other quadratic voting interfaces, where each click represents an increase in votes. The latter allows control through both clicks and mouse wheel rotation.

3.3 Text-based Interface

To study how the interactive components influenced participants' cognitive load and behavior, we removed the two-phase component and the drag-and-drop features for the text-based interface. The text-based interface shares the other functionalities of the interactive interface, as shown in Figure 9. The interface contained the question prompt at the top of the screen. The options were presented in a list underneath the prompt. Survey respondents could update the votes by selecting from a dropdown that provided all possible voting options and costs given the number of credits available. A small summary box to the right of the interface showed the current total cost and the remaining credits for the respondent. Options were randomly presented on the interface to avoid ordering bias [21, 13].

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

[1]link-to-github

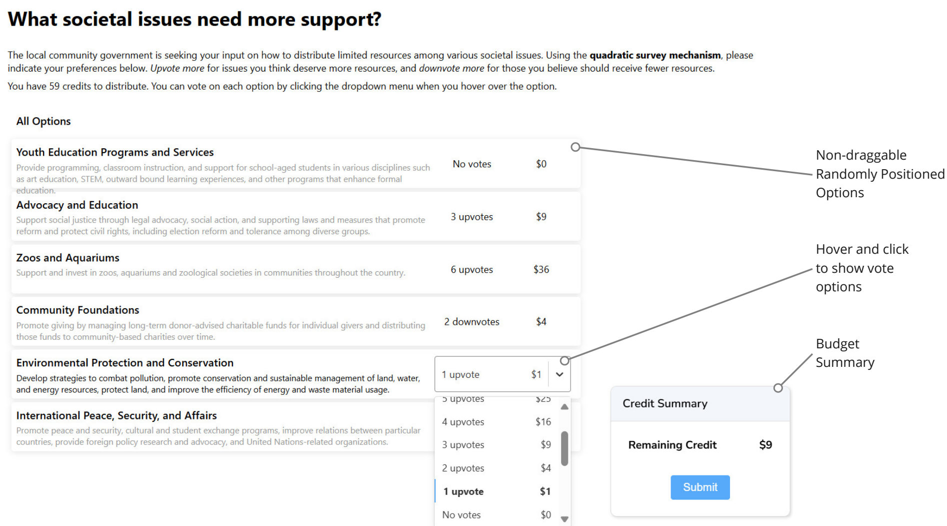


Fig. 9. The text-based interface: This interface is based on the interactive version but does not include the two-phase interactive support and lacks the drag-and-drop functionality. Options are randomly positioned.

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.

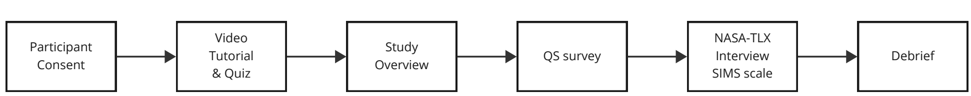


Fig. 10. Study protocol

Figure 10 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 [44, 64, 63]. However, we were not aware of any specific works that justified these numbers. Saaty [63] associated this value with both the cognitive processing capacity of 7 ± 2 [42] and a theoretical proof using the consistency ratio of a pairwise comparison metric [62]. 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 Chernev et al. [10] 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 Iyengar and Lepper [34]. 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 Cheng et al. [9], 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 [8], 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 [23]. 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 [49] and ECG [29] 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 [30, 31, 6]. This approach reduced between-rater variability, indicating differences in workload definitions among raters within a task and variations in workload sources between tasks [6]. Despite criticisms regarding its validity and vulnerability, NASA-TLX was commonly used due to its low cost and ease of administration [23]. 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 [61]. Thus, we chose NASA-TLX to measure cognitive load in our study.

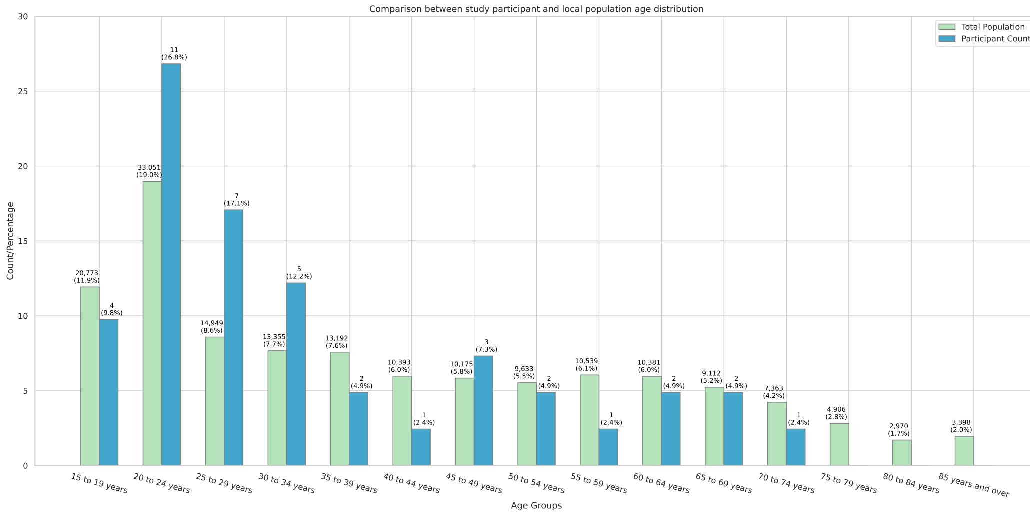
5 Results

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 conducted after participants completed the survey task. Interview were transcribed and thematically analyzed by the first author. All processed behavioral data are publicly available[2] to support transparency and facilitate further research.

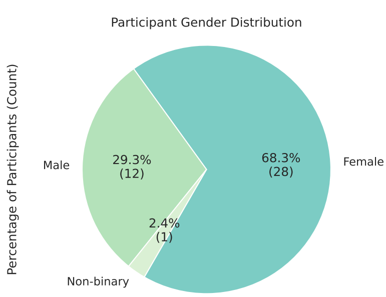
5.1 Demographics

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

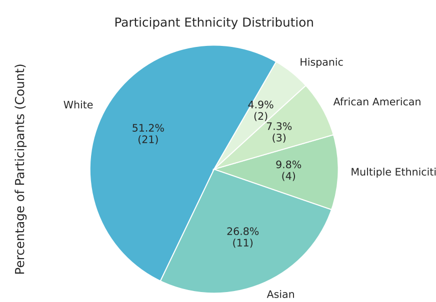
[2]link-to-github



(a) Age distribution



(b) Gender distribution



(c) Ethnicity distribution

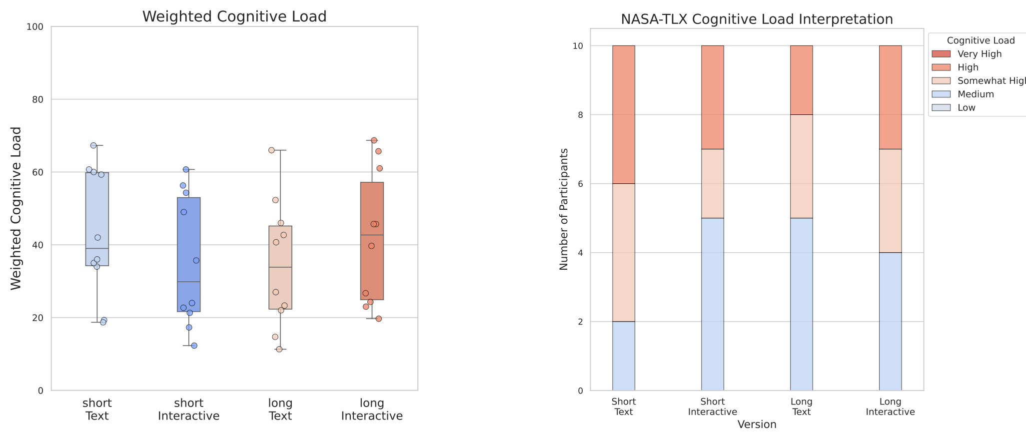
Fig. 11. Demographic distributions: Age, Gender, and Ethnicity

years old, with a detailed age distribution presented alongside the county population distribution in Figure 11a. 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 11b, 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 discuss the distribution of the aggregate NASA-TLX scores across the four experiment conditions, providing interpretation of the results to answer RQ1 and RQ2a. Next, report the NASA-TLX scores disaggregated into their composite six dimensions 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. 12. NASA-TLX Results

We show the NASA-TLX weighted results in Figure 12a. A higher value refers to higher cognitive load. Qualitatively, the interactive interfaced decreased the cognitive load for the short survey, but seems to have increased the cognitive load for the long survey. A Mann-Whitney U-Test indicates that these differences are not statistically significant. We follow predefined mappings of NASA-TLX values to cognitive levels: low, medium, somewhat high, high, and very high, as listed by Hart and Staveland [30]. We show value interpretations in Figure 12b. 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 load and the others experiencing somewhat high or high loads. No participants in any conditions expressed experiencing very high cognitive loads.

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.

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.

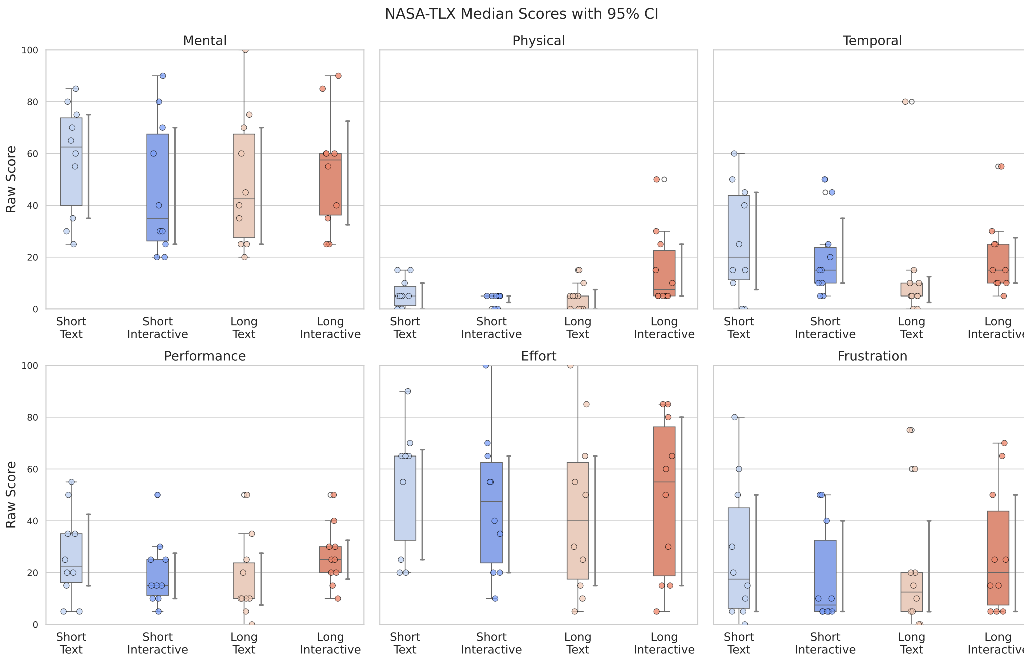


Fig. 13. NASA-TLX Results

We present the raw NASA-TLX scale results in Figure 13. 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 13, 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 13 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 13). 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 13), 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 13) 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*

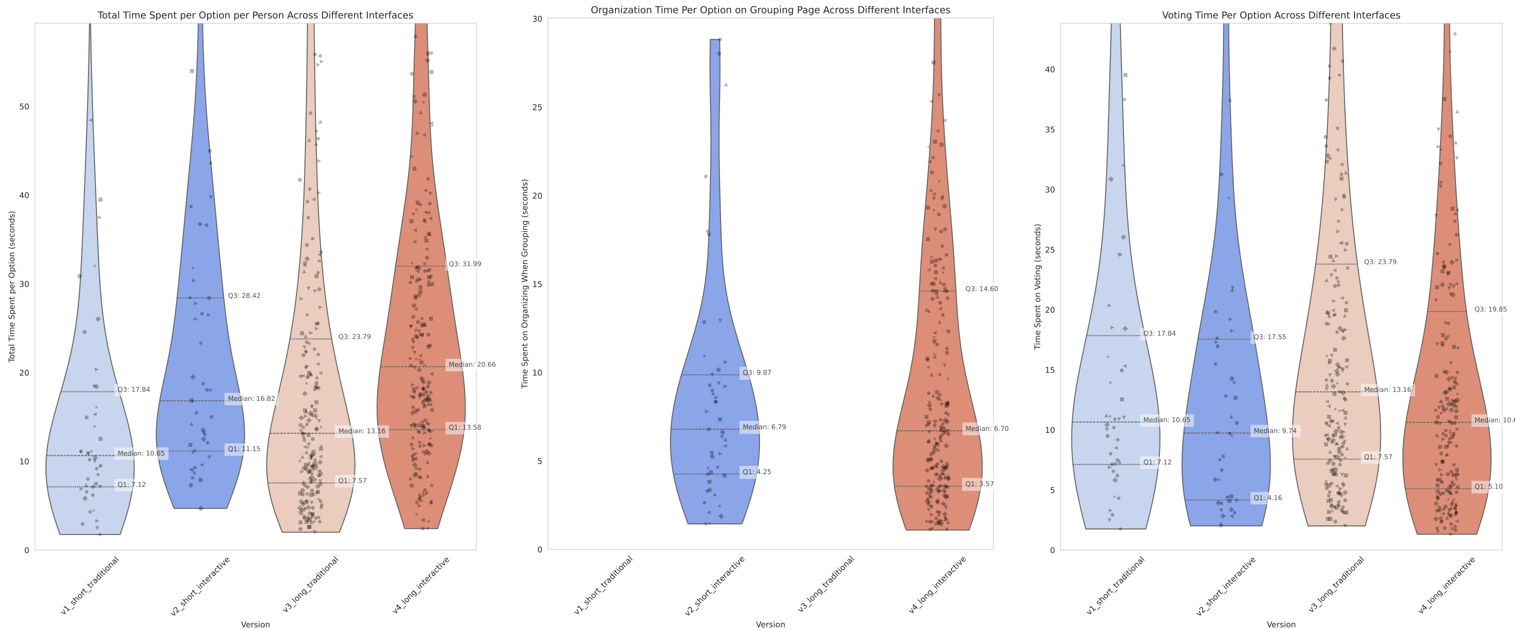
> *I think the frustration is ... I wish that we could help all of these causes, but you know it's just like anything else. You can't do everything and when it's not ... I feel like it's hard to quantify how much some of these things should be supported versus others. So when you're talking about upvotes 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 13), 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. 14. 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 14 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 14a, 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 14b and Figure 14c.

Once we separate the organization time (Figure 14b) and identify the voting time (Figure 14c), 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.

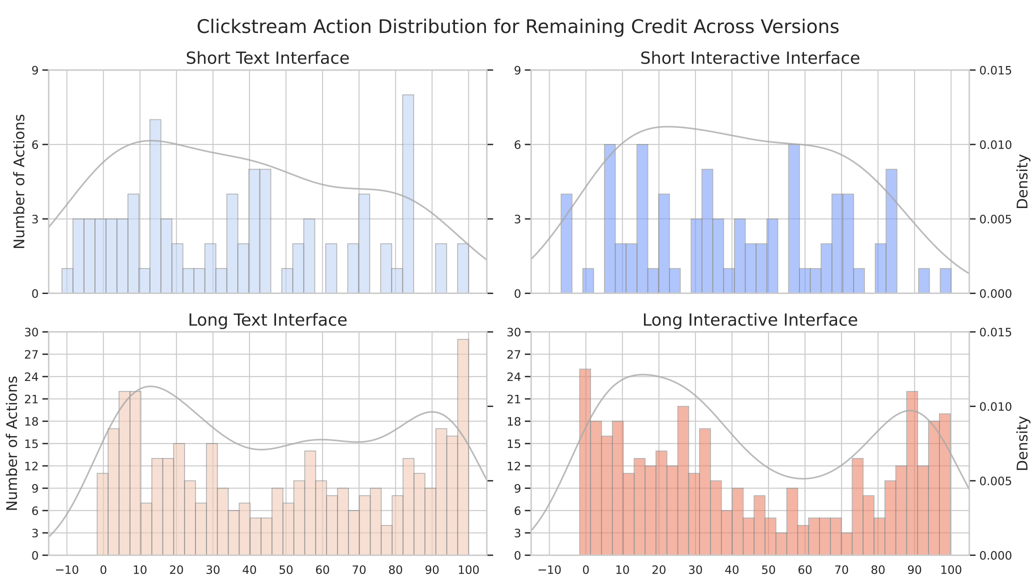


Fig. 15. voting actions across all options (needs to update chart text, remove normalization, and change the dot colors.)

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 15 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 Quarfoot et al. [55] 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.

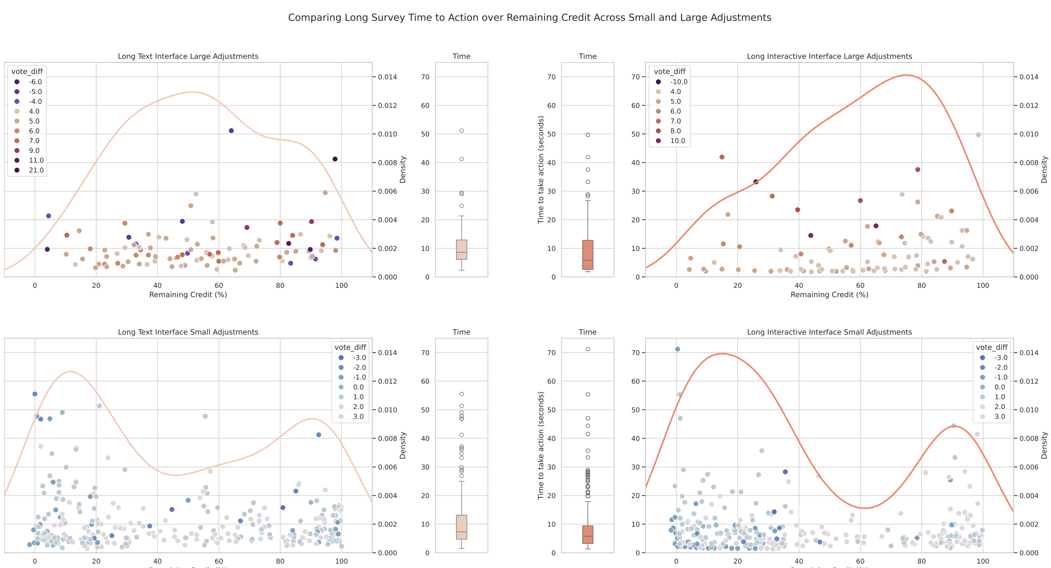


Fig. 16. Breakdown of voting actions (needs to update chart text)

Thus, we further separated the behaviors where participants made bigger changes or smaller changes to the option, specifically for the long version. In Figure 16 , we define an adjustment of four or more votes as a large adjustment which we plotted in the first row of the Figure. Adjustments of three or fewer votes are considered small adjustments.

First, 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.*

6 Discussion and Future Works

The goal of this study is to propose an interactive interface for QS and understand how the number of survey options, along with the interface, influences individuals' cognitive load and behaviors. Our results, presented in section **??**, revealed that longer surveys did not increase cognitive load among participants. We further analyzed the sources of cognitive load and identified different causes across several dimensions. Participants using the interactive interface demonstrated more strategic planning and holistic thinking compared to those using the text-based interface, who focused more on operational tasks. Additionally, the behavioral analysis presented in section **??** showed that participants using the long interactive interface made more frequent, small, iterative updates, indicating a shift in their cognitive load focus.

In the discussion section, we address three key topics: first, we explore where participants experienced bounded rationality and how the current interface design supported decision-making. Next, we examine how participants constructed their preferences and how direct manipulation within the interface supported this process. Finally, we discuss how the interface design mitigated some challenges of quadratic mechanisms and identify remaining issues. Following these discussions, we provide recommendations for using this tool and propose design improvements for future development.

6.1 Bounded rationality and interface design

One core repeated theme that emerged throughout participants' responses during the interview relates to bounded rationality. Bounded rationality defined by Simon [67] refers to the idea that individuals' cognitive limitations limited one's ability to use and process all available information, leading to a suboptimal resolution when decision making. When participants respond to a QS, they are faced with multiple options presented on the quadratic survey as well as the abundance of budget. Since the remaining budget translates to possible votes one can select to apply to an option, this adds additional numbers of decisions to make.

> *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.*

This quote exemplifies participants expressing their bounded rationality, trying to limit their effort to derive to their decision. Even though the drop-down menu showing all possible pre-calculated vote-credit values was a relief for a few participants so they do not need to search for the bounds, this sea of decision-making requires participants to recall and scramble information at once, which is extremely difficult. The byproduct of bounded rationality often translates to individuals satisficing behaviors [24], creating heuristics [79], over reliance on defaults [75], and problem decomposition [68].

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 are other common behaviors we observed. Several participants would create duo-dimension groupings, despite the experiment group they are in. Participants would have categories that cluster similar topics (i.e., all topics related to health vs. humanitarian) and categories that depict the positivity of their preference (i.e., positive vs. negative). Highlighting bounded rationality aims not to critique or exploit possible biases but to emphasize the importance of designing interface interventions to prevent survey respondents from making decisions that differ from their true preferences. For example, the design of showing one option at a time in the interactive interface lowers the possibility of participants being 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.*

The influence of bounded rationality highlights how critical and beneficial organization on the interface is. 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 the 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 the long interactive interface group) mentioned that organization support helped them to allot the intensity of votes by helping them focus and prioritize options through ranking. This exercise 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 also see a call for organizational features from participants using the text-based interface. Almost half of the participants (N=4) using the long text interface expressed a desire for features 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.*

Active management of the options forced participants to think about their rough preference for each option at minimal cognitive requirements. The repositioning of options allowed participants to focus on subsets of the options during their decision-making process.

6.2 Construction of Preference on QS

Since the interface supported some participants in managing their limited cognitive ability to make decisions, as shown in the previous subsection, we argue that the interactive interface *shifted* the cognitive focus onto contributing to more in-depth preference construction and fine-tuning, even if it did not significantly reduce the cognitive load. Here we provide more evidence.

Literature from Lichtenstein and Slovic [41] identifies three types of difficult decision-making scenarios: when one's preferences are not clearly defined, necessitating trade-offs, or quantifying opinions. We first show that participants constructed their preferences in situ. While some participants had existing preferences (e.g., environmental issues are important), they needed to reconsider aspects of the options or map them 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, one with a prior preference and the other without, illustrate how participants identified where the options on QS fall on their preference spectrum. Additionally, QS requires participants to utilize a common budget across options, necessitating trade-offs, as noted in prior works [9, 47]. Finally, selecting the final value involves quantifying opinions. Participants are thus engaged in a difficult *preference construction* decision-making process.

Behavior analysis in section **??** of participants using the long text and interactive interfaces revealed that they made small adjustments on the votes, clustered toward budget depiction with lesser time spent. These fine-grain adjustments indicated that participants are making less ad-hoc decisions; rather, they are deciding how to better utilize the remainder of the budget when the budget runs low. We identified a bi-modal interaction pattern.

We argue that the bimodal behavior observed in the voting actions across groups aligns with the Differentiation and Consolidation Theory presented by Svenson [72] as we previously invisioned. Recall that the theory states that decision making contains a differentiation stage involving identifying differences and eliminating less favorable alternatives, while the consolidation stage strengthens commitment to the chosen option. In the interactive interface, we see a stronger bimodal indicating the two stages, compared to the text interface. 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*

These findings show the role of the two-phase interactive design in scaffolding the preference construction process. Additionally, several participants mentioned how the direct manipulation functionality, allowing individuals to drag and drop options for repositioning, supports their reflective thinking during preference construction. One participant noted:

> *So I tried to make a ranking [. . .] and by creating this ranking, by dragging the related issues . . . I don't know . . . that 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 this is a 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 not being able to track it at all. – S039, long interactive interface.*

More importantly, it acts as a process for reflective verification and iterative decision-making. These can include post-reflection after expressing the intensity of preferences or preparation to decide on the 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 looks 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 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 participant 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.*

Throughout the preference construction journey, we confirm that the two-stage interactive interface and the direct manipulation through drag-and-drop facilitated participants in constructing and reflecting on their preferences, adhering to preference construction theory.

6.3 Quadratic mechanism is challenging

From the qualitatie interview, we found that many participants found automatic calculation critical. More than one-third of participants (N=14) from all four experiment conditions highlighted the importance of automated calculation from two perspectives: *deriving cost* and *keeping track of spent*. Two participants 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.*

Twelve participants highlighted the summarization box and the automated summation of the current credit spent, allowing them to focus on managing their next voting decision and expressing 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.*

However, the interface did not include elements that helped participants kick off their voting process. One of the most difficult challenges for participants when completing to QS is deciding 'how many votes' to begin with. This challenge does not refer to the relative vote, but the starting vote. Some participants began by equally distributing their credits to all options and then made adjustments, some established 1, 2, and 3 votes as three 'tiers' of votes as starting points, and a small handful of participants, surprisingly, 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 decision echoes prior literature's discussion on whether an absolute value exists for an individual. Coherent arbitrariness [3], similar to the anchoring effect in marketing, refers to participants' willingness to pay, which can be influenced by an arbitrary value. However, the ordinal utility remains intact among the set of preferences. We also made similiar observations in the prototyping stage of the interactive interface.

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 invalidate the power of the quadratic mechanism, it builds frustration and friction for participants to construct a clear picture of how to make voting decisions.

Budget-related sources draw across mental demand, temporal demand, preference demand, and frustration. These span from making sure to keep within budget to recovering from overbudgeting. While prior scarcity literature [65] believes that values and careful decisions are derived from limited resources, prospect theory [35] also highlights a higher negative value of *perceived* utility for individuals when cuts ought to be made.

These three major challenges do not threaten the integrity of the Quadratic Survey and relevant 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 comes at the cost of higher cognitive load.

6.4 QS Usage and Design Recommendations and Future Work

With a deeper understanding of how survey respondents interact with QS and the sources of cognitive load, we recognize that while this interface may not significantly reduce cognitive load, it represents a crucial step toward constructing better interfaces to support individuals responding to QS. In this subsection, we outline usage and design recommendations applicable to all applications using the quadratic mechanism and highlight directions for future work.

6.4.1 Usage Recommendation: QS for Critical Evaluations. Our study highlighted the complex cognitive challenges and in-depth consideration required when ranking and rating options using QS, even in a short survey. Similar to survey respondents needing to make trade-offs across options, researchers and agencies seeking additional insights and alignment with respondent preferences must ensure that survey respondents have the cognitive capacity to complete such surveys rigorously. We recommend designing QS for specific use cases requiring critical evaluations, such as investment decisions or settings where participants have ample time to think and process the survey. For instance, revealing the options ahead of time can aid in preference construction.

6.4.2 Design Recommendations.

Use Organization Phases for Quadratic Mechanism Applications. Our study demonstrated that preference construction can shift from operational to strategic and higher-level causes. An additional organizational phase with direct manipulation capability allows survey respondents to engage in higher-level critical thinking. We believe this approach should extend beyond QS to other ranking-based surveying tools, such as rank-choice voting and constant sum surveys. Further research should examine how implementing such functionality alters survey respondents' mental models.

Facilitate Differentiation through Categorization, Not Ranking. Participants in our study were less inclined to provide a full rank unless necessary. The final 'rank' of option preferences often emerged as a byproduct of their vote allocation, constructed in situ. Therefore, for survey designs to be effective in constructing preferences, it is more important to facilitate differentiation than to focus on direct manipulation solely for fine-tuning. Emphasizing categorization can better support participants in articulating their preferences.

6.4.3 Future Work: Support for Absolute Credit Decision. Deciding the absolute amount of credits in QS is highly demanding. Designing interfaces and interactions that address the cold start challenge and help participants decide the absolute vote value, while considering ways to limit direct influences, remains an open question. Future research should explore innovative solutions to support participants in making these complex decisions effectively.

By implementing these recommendations and pursuing future research directions, we can improve the usability and effectiveness of QS and other quadratic mechanism-powered applications, ultimately aiding respondents in making more informed and accurate decisions.

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 Alternative prototypes for QS

here are some alternatives

References

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

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

[3] Dan Ariely, George Loewenstein, and Drazen Prelec. 2003. "Coherent Arbitrariness": Stable Demand Curves Without Stable Preferences*. *The Quarterly Journal of Economics*, 118, 1, (Feb. 2003), 73–106. DOI: 10.1162/00335530360535153.

[4] Herbert Bless, Gerd Bohner, Traudel Hild, and Norbert Schwarz. 1992. Asking Difficult Questions: Task Complexity Increases the Impact of Response Alternatives. *European Journal of Social Psychology*, 22, 3, (May 1992), 309–312. DOI: 10.1002/ejsp.2420220309.

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

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

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

[8] 2023. Charity Navigator. https://www.charitynavigator.org/index.cfm?bay=search.categories. (May 2023). Retrieved Dec. 16, 2023 from.

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

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

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

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

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

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

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

[16] 2015. Designing usable ballots | Center for civic design. https://civicdesign.org/fieldguides/designing-usable-ballots/. (June 2015). Retrieved June 17, 2024 from.

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

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

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

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

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

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

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

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

[24] Gerd Gigerenzer and Daniel G. Goldstein. 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 4, 650–669. DOI: 10.1037/0033-295X.103.4.650.

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

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

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

[28] Theodore Groves and John Ledyard. 1977. Optimal Allocation of Public Goods: A Solution to the "Free Rider" Problem. *Econometrica*, 45, 4, 783–809. JSTOR: 1912672. DOI: 10.2307/1912672.

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

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

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

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

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

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

[35] Daniel Kahneman and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47, 2, 263–291. JSTOR: 1914185. DOI: 10.2307/1914185.

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

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

[38] Steven P Lalley, E Glen Weyl, et al. 2016. Quadratic voting. *Available at SSRN*.

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

[40] Timo Lenzner, Lars Kaczmirek, and Alwine Lenzner. 2010. Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, 24, 7, 1003–1020. DOI: 10.1002/acp.1602.

[41] Sarah Lichtenstein and Paul Slovic, eds. 2006. *The Construction of Preference*. (1. publ ed.). Cambridge University Press, Cambridge.

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

[43] Henry Montgomery. 1983. Decision Rules and the Search for a Dominance Structure: Towards a Process Model of Decision Making. In *Advances in Psychology*. Vol. 14. Elsevier, 343–369. DOI: 10.1016/S0166-4115(08)62243-8.

[44] William F. Moroney and Joyce A. Cameron. 2019. *Questionnaire Design: How to Ask the Right Questions of the Right People at the Right Time to Get the Information You Need.* Human Factors and Ergonomics Society, (Feb. 2019).
[45] Michael Morreau. 2014. Arrow's theorem.
[46] Robert Münscher, Max Vetter, and Thomas Scheuerle. 2016. A Review and Taxonomy of Choice Architecture Techniques. *Journal of Behavioral Decision Making*, 29, 5, 511–524. DOI: 10.1002/bdm.1897.
[47] Ryan Naylor et al. 2017. First year student conceptions of success: What really matters? *Student Success*, 8, 2, 9–19.
[48] A Norman Donald. 2013. *The Design of Everyday Things.* MIT Press.
[49] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10.* ACM Press, Austin, Texas, 141. DOI: 10.1145/1743666.1743701.
[50] John W. Payne, James R. Bettman, and Eric J. Johnson. 1988. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 3, (July 1988), 534–552. DOI: 10.1037/0278-7393.14.3.534.
[51] Martin Pielot and Mario Callegaro. 2024. Did You Misclick? Reversing 5-Point Satisfaction Scales Causes Unintended Responses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, Honolulu HI USA, (May 2024), 1–7. DOI: 10.1145/3613904.3642397.
[52] Eric A Posner and E Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society.* Princeton University Press.
[53] 2022. Quadratic Voting Frontend. Public Digital Innovation Space. (Jan. 2022). Retrieved Dec. 16, 2023 from.
[54] [n. d.] Quadratic Voting in Colorado. https://www.radicalxchange.org/wiki/colorado-qv/. (). Retrieved June 13, 2024 from.
[55] David Quarfoot, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein, and Ellen Konar. 2017. Quadratic voting in the wild: real people, real votes. *Public Choice*, 172, 1-2, 283–303.
[56] Whitney Quesenbery. 2020. Opinion | Good Design Is the Secret to Better Democracy. *The New York Times*, (Oct. 2020). Retrieved June 17, 2024 from.
[57] 2024. RadicalxChange/quadratic-voting. RadicalxChange. (May 2024). Retrieved June 17, 2024 from.
[58] [n. d.] Read the Whitepaper | Gitcoin. https://www.gitcoin.co/whitepaper/read. (). Retrieved June 17, 2024 from.
[59] Duncan Rintoul. [n. d.] Visual and animated response formats in web surveys: Do they produce better data, or is it all just fun and games?, 126.
[60] Adam Rogers. 2019. Colorado Tried a New Way to Vote: Make People Pay—Quadratically | WIRED. *Wired*, (Apr. 2019). Retrieved June 22, 2024 from.
[61] Susana Rubio, Eva Díaz, Jesús Martín, and José M. Puente. 2004. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Applied Psychology*, 53, 1, 61–86. DOI: 10.1111/j.1464-0597.2004.00161.x.
[62] Thomas L Saaty and Mujgan S Ozdemir. 2003. Why the magic number seven plus or minus two. *Mathematical and computer modelling*, 38, 3-4, 233–244.
[63] Thomas L. Saaty. 1987. Principles of the Analytic Hierarchy Process. In *Expert Judgment and Expert Systems.* Jeryl L. Mumpower, Ortwin Renn, Lawrence D. Phillips, and V. R. R. Uppuluri, editors. Springer, Berlin, Heidelberg, 27–73. DOI: 10.1007/978-3-642-86679-1_3.
[64] Thomas L. Saaty and Kirti Peniwati. 2013. *Group Decision Making: Drawing Out and Reconciling Differences.* RWS Publications, (Nov. 2013).
[65] Anuj K. Shah, Eldar Shafir, and Sendhil Mullainathan. 2015. Scarcity frames value. *Psychological Science*, 26, 4, 402–412.
[66] Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *Interactions*, 4, 6, (Nov. 1997), 42–61. DOI: 10.1145/267505.267514.
[67] Herbert A. Simon. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69, 1, 99–118. JSTOR: 1884852. DOI: 10.2307/1884852.
[68] Herbert A. Simon. 1996. *The Sciences of the Artificial.* (3rd ed ed.). MIT Press, Cambridge, Mass.
[69] Tobin South, Leon Erichsen, Shrey Jain, Petar Maymounkov, Scott Moore, and E. Glen Weyl. 2024. Plural Management. SSRN Scholarly Paper. Rochester, NY, (Jan. 2024). DOI: 10.2139/ssrn.4688040.
[70] Fritz Strack and Leonard L. Martin. 1987. Thinking, Judging, and Communicating: A Process Account of Context Effects in Attitude Surveys. In *Social Information Processing and Survey Methodology.* Recent Research in Psychology. Hans-J. Hippler, Norbert Schwarz, and Seymour Sudman, editors. Springer, New York, NY, 123–148. DOI: 10.1007/978-1-4612-4798-2_7.
[71] Kathryn Summers, Dana Chisnell, Drew Davies, Noel Alton, and Megan McKeever. 2014. Making voting accessible: designing digital ballot marking for people with low literacy and mild cognitive disabilities. In *2014 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE 14).*

[72] Ola Svenson. 1992. Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychologica*, 80, 1-3, (Aug. 1992), 143–168. DOI: 10.1016/0001-6918(92)90044-E.

[73] John Sweller. 2011. Cognitive Load Theory. In *Psychology of Learning and Motivation*. Vol. 55. Elsevier, 37–76. DOI: 10.1016/B978-0-12-387691-1.00002-8.

[74] Internet Team. [n. d.] Taiwan Digital Minister highlights country's use of technology to bolster democracy in FT interview. https://www.roc-taiwan.org/uk_en/post/6295.html. (). Retrieved June 13, 2024 from.

[75] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness. Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press, New Haven, CT, US, x, 293.

[76] Jerry P Timbrook. 2013. *A Comparison of a Traditional Ranking Format to a Drag-and-Drop Format with Stacking*. PhD thesis. University of Dayton.

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

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

[79] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 4157, 1124–1131. Retrieved June 21, 2024 from JSTOR: 1738360.

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

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

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

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

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

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

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