Organize, Then Vote: Exploring Cognitive Load in Quadratic Survey Interfaces

ANONYMOUS AUTHOR(S)*

Quadratic Surveys (QS) elicit more accurate individual preferences than traditional surveys, such as Likert-scale surveys. However, the cognitive load associated with QS has hindered its adoption in digital surveys for collective decision-making. We introduce a two-phase "organize-then-vote" QS interface based on decision-making and preference construction theories designed to lessen the cognitive load. Since interface design significantly impacts survey results and accuracy, our design scaffolds survey takers' decision-making while managing the cognitive load imposed by QS. In a 2x2 between-subject in-lab study on public resource allotment, we compared our interface with a traditional text interface across QS with 6 (short) and 24 (long) options. Our interface reduced satisficing behaviors arising from cognitive overload in long QS conditions. Participants using our interface in the long QSs shifted their cognitive effort from mechanical operations to constructing more comprehensive preferences. This research clarifies how human-centered design improves preference elicitation tools for collective decision-making.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**; **Collaborative and social computing design and evaluation methods**; User studies; **HCI design and evaluation methods**; **Interactive systems and tools**; **Empirical studies in interaction design**.

Additional Key Words and Phrases: Quadratic Survey; Survey Response Format; User Interface; Preference Construction; Cognitive Load

ACM Reference Format:
Anonymous Author(s). 2024. Organize, Then Vote: Exploring Cognitive Load in Quadratic Survey Interfaces. *Proc. ACM Hum.-Comput. Interact.* 1, 1 (September 2024), 42 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Designing intuitive survey interfaces is crucial for accurately capturing respondents' preferences, which directly impact the quality and reliability of the data collected. Recent Human-Computer Interaction (HCI) studies highlight how certain survey response formats can increase errors [1, 2] and influence survey effectiveness [3]. In this paper, our goal is to introduce an effective interface for **Quadratic Surveys (QS)**, a survey tool designed to elicit preferences more accurately than traditional methods [4]. Despite the promise of QS, there has been no research on designing interfaces to support its unique quadratic mechanisms [5], where participants must rank and rate items — a task that poses significant cognitive challenges. To popularize QS and ensure high-quality data, this paper addresses the question: *How can we design interfaces to support participants in completing Quadratic Surveys (QS) more effectively?*

We envision an effective interface that navigates participants through the complex mechanism and preference construction process, tailored to the unique challenges of QS. QS improves accuracy in individual preference elicitation compared to traditional methods like Likert scales by requiring participants to make trade-offs using a fixed budget of credits, where purchasing k votes for an option in QS costs k^2 credits [6, 4]. This quadratic cost structure forces respondents to carefully evaluate their preferences, balancing the strength of their support or opposition against the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

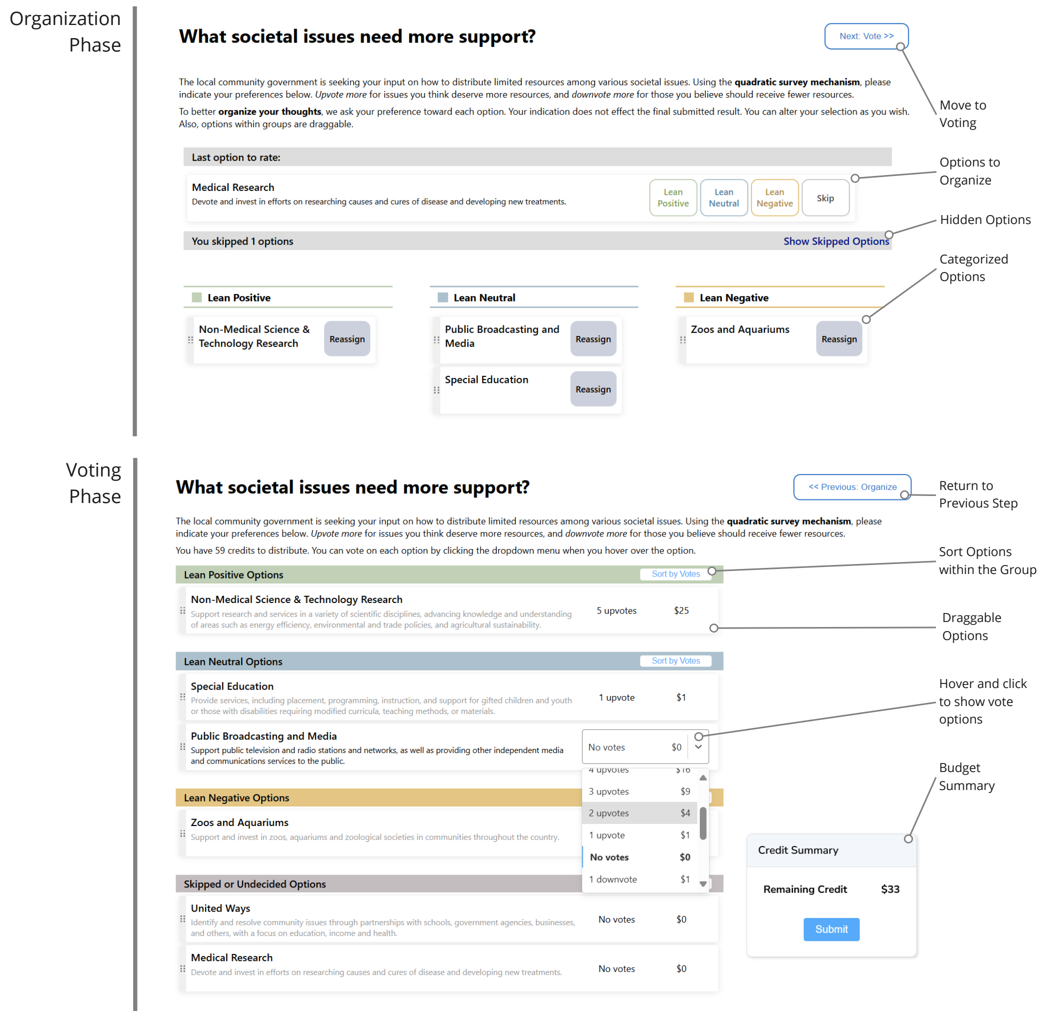


Fig. 1. The Two-Phase Interface: The interface consists of two phases. Survey respondents can navigate between phases using the top right button. In the organization phase, the interface presented one option at a time to the respondents, and they chose four choices: “Lean Positive”, “Lean Neutral”, “Lean Negative”, or “Skip”. Skipped options are hidden and can be evaluated later. The chosen options 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.

limited budget. While this cost structure forces them to make thoughtful trade-offs, this complexity also increases cognitive load, making it mentally taxing to weigh costs, evaluate options, and construct rankings [7]. Moreover, QS, often referred to as Quadratic Voting (QV) in voting scenarios, can involve hundreds of options [8, 9], increasing the risk of cognitive overload and satisficing [10, 11, 12]. We propose that an effective interface tailored to support users through these complex trade-offs can reduce cognitive load and facilitate better preference construction.

To date, existing quadratic mechanism-powered applications simply present options, allow vote adjustments and automatically calculate votes, costs, and budget usage. These designs focused heavily on the mechanics operating the tool, rather than supporting possible challenges these application users faced. Survey interface literature, while addressing decision-making and usability, most focus on traditional surveys that do not share the unique option-to-option trade-offs that QS introduces [13, 14, 15, 16, 17, 1]. More specifically, managing cognitive load [18, 19, 16, 20, 21] and scaffolding challenging tasks [22, 23, 24, 25] like preference construction, are opportunities where user interfaces have shown their promises. As a result, it remains unclear how different interface designs might support QS in reducing cognitive load and aiding preference construction.

After multiple design iterations, we propose a novel interactive two-phase "organize-then-vote" QS interface (referred to two-phase interface for short, Figure 1) that integrates three key elements. First, the interface scaffolds the preference construction process by having participants initially categorize the survey options into "Lean Positive," "Lean Neutral," or "Lean Negative." This serves as a cognitive warm-up, easing participants into the more complex QS voting task. Second, the interface arranges the options according to these categorizations, providing a structured visual layout. Third, participants can refine the positions of these options using drag-and-drop functionality, giving them greater control and agency in the preference-construction process. These design features are aligned with preference construction theory and build upon prior research in interface design to reduce cognitive load and enhance user engagement.

To explore how these interface elements mitigate the cognitive load and support preference construction in Quadratic Surveys, we pose the following research questions:

- RQ1. How does the number of options in Quadratic Surveys impact respondents' cognitive load?
- RQ2a. How does the two-phase interface impact respondents' cognitive load compared to a single-phase text interface?
- RQ2b. What are the similarities and differences in sources of cognitive load across the two interfaces?
- RQ3. What are the differences in Quadratic Survey respondents' behaviors when coping with long lists of options across the two-phase interface and the single-phase text interface?

We invited 41 participants to a lab study comparing our two-phase interface with a baseline to understand how different interface designs and option lengths (6 options or 24 options) impact cognitive load. Qualitative findings, measured using the NASA Task Load Index (NASA-TLX) and semi-structured interviews, revealed that participants using the two-phase interface experienced cognitive demand more from strategic, holistic thinking compared to personal relevance and operational tasks, particularly in longer surveys. Quantitative results showed that, although participants spent more time per option, they made faster decisions during the voting phase, suggesting a more efficient distribution of cognitive effort. We concluded that the two-phase interface mitigated cognitive overload in long QS surveys and shifted mental load toward more strategic thinking, reducing reliance on mental shortcuts like satisficing [10].

Contributions. We contribute to the HCI community by proposing the first interface specifically designed for QS and QV-like applications, aimed at reducing cognitive challenges and scaffolding preference construction through a two-phase interface with direct manipulation. Before our work, no research had explored QS interfaces, particularly for long surveys prone to cognitive overload. Few studies in HCI address interfaces for surveys and questionnaires. Our study demonstrated how user interfaces can facilitate preference construction in situ and reduce satisficing behaviors by promoting incremental updates and deeper engagement with survey items through interface elements. Additionally, our interview is also the first in-depth qualitative analysis of user experiences with Quadratic Mechanism applications, identifying key factors contributing to cognitive load. The impact of our contribution extends beyond QS, offering design implications for other preference-elicitation tools in complex scenarios. By making QS easier to use and more accurate, our design also encourages wider adoption among researchers and practitioners. Finally, our work lays the groundwork for future studies on qualitative insights and future interface designs of quadratic mechanisms, supporting decision-makers in eliciting accurate respondents' preferences.

2 Related Work

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

2.1 Quadratic Survey and the Quadratic Mechanism

We introduce the term **Quadratic Survey (QS)** to describe surveys that utilize the quadratic mechanism to collect individual attitudes. The **quadratic mechanism** is a theoretical framework designed to encourage the truthful revelation of individual preferences through a quadratic cost function [5]. 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 [26].

To illustrate how QS works, we formally define the mechanism: each survey respondent is allocated a fixed budget, denoted by B, to distribute among various options. Participants can cast n votes for or against 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:

$$\sum_k c_k \leq B$$

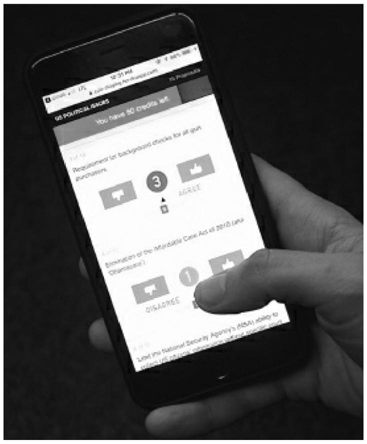
Survey results are determined by summing the total votes for each option:

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

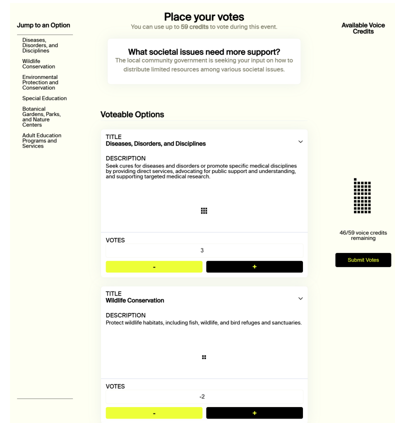
where S represents the total number of participants, and $n_{i,k}$ is the number of votes cast by participant i for option k. Each additional vote for each option increases the marginal cost linearly, encouraging participants to vote proportionally to their level of concern for an issue [27].

QS adapts these strengths of the quadratic mechanism in *voting* to encourage truthful expression of preferences in *surveys*. Unlike traditional surveys that elicit either rankings *or* ratings, QS allows for *both*, enabling participants to cast multiple votes for or against options, incurring a quadratic cost. Cheng et al. [4] showed that this mechanism aligns individual preferences with behaviors more accurately than Likert Scale surveys, particularly in resource-constrained scenarios like prioritizing user feedback on user experiences.

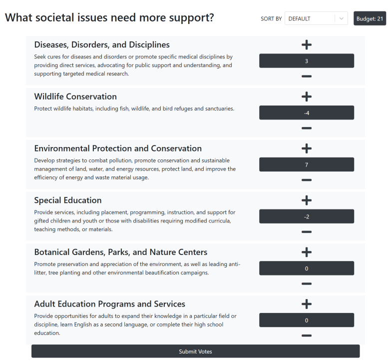
In recent years, empirical studies on QV have expanded into various domains [28, 29]. Applications based on the quadratic mechanism have also grown, including Quadratic Funding, which redistributes funds based on outcomes from consensus made using the quadratic mechanism [30, 31]. Recent work by South et al. [32] applies the quadratic mechanism to networked authority management, later used in Gov4git [33]. Despite the increasing breadth and depth of applications utilizing the quadratic mechanism, little attention has been paid to user experience and interface design,



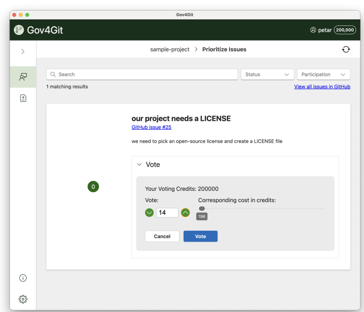
(a) Software by WeDesign, used in the first empirical QV research [6]. Little information is available about the software, except for an image from Posner and Weyl [27]. In the image, each prompt has thumbs up and down icons to update the vote in the center. The remaining budget appears as a progress bar at the top.



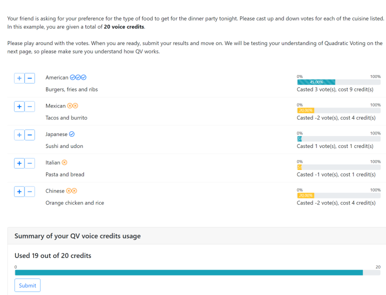
(b) An open-sourced QV interface [34] forked from GitCoin [35], used by the RadicalxChange community [36]. This interface presents total credits as small blocks. Votes are updated using plus and minus buttons, with numerical counts shown under each option and surface area as costs.



(c) An open-source QV interface [37] 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 [33] 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. [4] employs the most visual components. Icons depict the current number of votes, with progress bars signifying the current spending.

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

which support individuals in expressing their preference intensity. Our work aims to address this by designing interfaces supporting quadratic mechanisms.

2.2 Design Implications from Surveys and Voting Interfaces

We began by examining existing QV applications (Fig. 2), which share the same mechanism as QS, and identifying shared interface components. All QV interfaces generally include:

- Option list: A list of options for voting.
- Vote controls: Buttons to increase or decrease votes for each option.
- Individual vote tally: A display of the votes cast per option.
- Summary: An auto-generated summary of costs and the remaining budget.

These components present options, calculate costs, and allow vote adjustments. However, these interfaces focus purely on mechanics without little understanding of voter's usability needs or offering cognitive support to help them complete the task effectively. In addition, the HCI community conducted few research [38, 39] on survey and questionnaire interfaces components, with more work focusing more on alternative input modalities like bots, voice, and virtual reality [40, 41, 2, 42]. Thus, we turn to marketing and research literature for insights into how digital survey interface elements can impact user behavior and usability.

Research in the marketing and research communities focusing on survey and questionnaire design, usability, and interactions examines the influence of presentation styles and 'response format.' Weijters et al. [43] 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 analog scales that operate on a point-and-click principle perform better [44]. These studies show how even small design changes can have a large impact on usability, highlighting the importance of designing interfaces that prioritize human-centered interaction rather than focusing solely on functionality.

Voting interfaces are a specialized type of survey interface that not only elicit individual choices but often have consequential impacts. For example, the butterfly ballot, an atypical design, may have influenced the outcome of the 2000 U.S. Presidential Election [45]. Research has shown that ballot interfaces can significantly influence democratic processes [13, 46, 47]. Several studies also highlighted how voting interface designs shift voter decisions [13], reduce usability errors [48, 49], or improve interaction [50, 51, 52, 53, 54]. We provide more details to these voting interfaces in the Appendix A.

From the QV implementations, response format literature, and voting interfaces, we identified how interfaces significantly influence respondent behavior, decision accuracy, and cognitive load. While these systems are functional, they lack the human-centered design needed to reduce cognitive load and make them truly usable, rather than simply operable. These burdens are especially problematic for complex systems like QS, where high cognitive demands may deter researchers and users alike. Developing effective, human-centered interfaces for QS could enhance usability, reduce cognitive overload, and increase adoption in both research and practical applications.

2.3 Cognitive Challenges and Choice Overload

The challenge of respondents making difficult decisions using quadratic mechanisms remains unexplored in the literature. Lichtenstein and Slovic [7] identified three key elements that make decisions difficult. These elements include making decisions in unfamiliar contexts, being forced to make trade-offs due to conflicting choices, and quantifying the value of one's opinions. QS fits all three elements: participants may encounter unfamiliar options set by the decision-maker, are constrained by budgets that require trade-offs, and cast final votes as numerical values. Thus, we believe QS introduces high cognitive load.

According to cognitive load theory, cognitive load refers to the demands placed on a user's working memory during the interaction process, which significantly influences the usability of the system [55, 56]. Cognitive overload can adversely affect performance [57], leading individuals to rely on heuristics rather than deliberate, logical decision-making [58]. When presented with excessive information, such as too many options, individuals 'satisfice', settling for a 'good enough' solution rather than an optimal one [10, 11, 12]. Subsequently, too many options can overwhelm individuals, resulting in decision paralysis, demotivation, and dissatisfaction [59].

Additionally, Alwin and Krosnick [60] highlighted that the use of ranking techniques in surveys can be time-consuming and potentially more costly to administer. These challenges are compounded when ranking numerous items, requiring substantial cognitive sophistication and concentration from survey respondents [61].

Notable applications of Quadratic Voting include the 2019 Colorado House, which considered 107 bills [62], and the 2019 Taiwan Presidential Hackathon, which featured 136 proposals [63]; both used a single QV question with hundreds of options. Psychological and behavioral research highlights the importance of understanding how individuals navigate and benefit from new interfaces under long-list QS conditions. These empirical applications of QV suggest QS's potential to elicit individual preferences, emphasizing the need to study cognitive load and interface design.

3 Quadratic Survey Interface Design

In this section, we present the QS interface. Using components from existing QV interfaces (Fig 2) and insights from prior literature, we iterated through paper prototypes and three design pre-tests, detailed in Appendix B. In our initial paper prototyping iterations, participants struggled to *rank* relative preferences among options and *rate* the degree of trade-offs between them. In this study, we focus on addressing the former challenge, which pertains to preference construction.

3.1 'Organize-then-Vote': The Two-Phase Interface

3.1.1 Justifying a two-phase approach. The main objective of the two-phase interface is to facilitate preference construction and reduce cognitive load. As shown in Figure 1, the interface consists of two steps: an organization phase and a voting phase. In both phases, survey respondents can drag and drop options across the presented list.

A two-phase approach. Preferences are shaped through a series of decision-making processes [7]. Two major decision-making theories informed this two-step interaction interface design: Montgomery [64]'s Search for a Dominance Structure Theory (Dominance Theory) and Svenson [65]'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 [64]. 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 [65]. Both theories supported the design decision to reduce the dimensions during the initial decision process and help emphasize relatively important options to form 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 clearly superior options or partitioning choices into distinguishable groups. In this section, we first describe how the interaction works, then we detail the reasons for the implemented design decisions.

The organizing interface, depicted on the top half of Figure 1, sequentially presents each survey option. Participants select a response among three ordinal categories – "Lean Positive", "Lean Negative", or "Lean Neutral". Once selected, the system moves that option to the respective category. Participants can skip the option if they do not want to indicate a preference. Options within the groups are draggable and rearrangeable to other groups should the participants wish.

To support preference formation, respondents are shown one option at a time, allowing them to either recall a prior judgment or construct a new one based on the presented choices [66]. Limiting the information presented this way also helps reduce cognitive load by preventing overload from too many options [67]. This incremental process ensures that participants form opinions on individual options, addressing an early prototype issue where the organizing task was mistakenly treated as a ranking task.

The three possible options — Lean Positive, Lean Neutral, and Lean Negative — aim to scaffold participants in constructing their own choice architecture [68, 69], which strategically segments options into diverse and alternative choice presentations while avoiding biases from defaults. We believed that these three categories were sufficient for participants to segment the options. We do not limit the number of options one can place in each category to prioritize user agency, allowing participants full control over how they organize their preferences [70]. Immediate feedback displays the placement of options and allows participants to rearrange them via drag-and-drop, adhering to key interface design principles [70]. At the same time, it allows 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 is to facilitate the consolidation of differentiated options through interactive elements while reinforcing the differentiation across options constructed by participants in the previous phase. This facilitation is achieved by retaining the drag-and-drop functionality for direct manipulation of position and enabling sorting within each category.

Options are displayed as they are categorized within each category from the previous step and in the following section — Lean Positive, Lean Neutral, Lean Negative, and Skipped or Undecided — as detailed on the bottom half of Figure 1. The Skipped or Undecided category contains options left in the organization queue, possibly because survey respondents have a pre-existing preference or chose not to organize their thoughts further. The original order within these categories is preserved to maintain and reinforce the differentiated options. This ordering sequence mitigated early prototype concerns where uncategorized options were left at the top of the voting interface confusing survey respondents. Respondents have the flexibility to return to the organization interface at any point during the survey to revise their choices.

In the voting interface, options are draggable, allowing participants to modify or reinforce their preference decisions as needed. Each category features a sort-by-vote function for reordering within the group, which, although it doesn't affect the final outcome, supports information organization and consolidation. Both features aim to group similar options automatically and emphasize proximity, reducing cognitive load by following the proximity compatibility principle to enhance decision-making [71].

While multiple interaction mechanisms exist, drag-and-drop has been extensively explored in rank-based surveys. For instance, Krosnick et al. [72] 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 [73] 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 [74]. Together, these design decisions led to our belief that a two-phase interface with direct interface manipulation could reduce the cognitive load for survey respondents to form preference decisions when completing QS.

In addition, we made three aesthetic design decisions. First, we removed visual elements like icons, emojis, progress bars, and vote visualizations, as prior research indicated that emojis could influence survey interpretations and reduce user satisfaction [75, 16]. While effective visualizations can aid decision-making, this study does not aim to address that question. 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 each survey option for ease of access to the budget summary when determining the votes. 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. After testing two alternatives (Figure 3) input methods—click-based buttons, which required multiple clicks, and a wheel-based design, which offered intuitive control but was unfamiliar to some participants—we opted for a more accessible dropdown menu for vote selection.

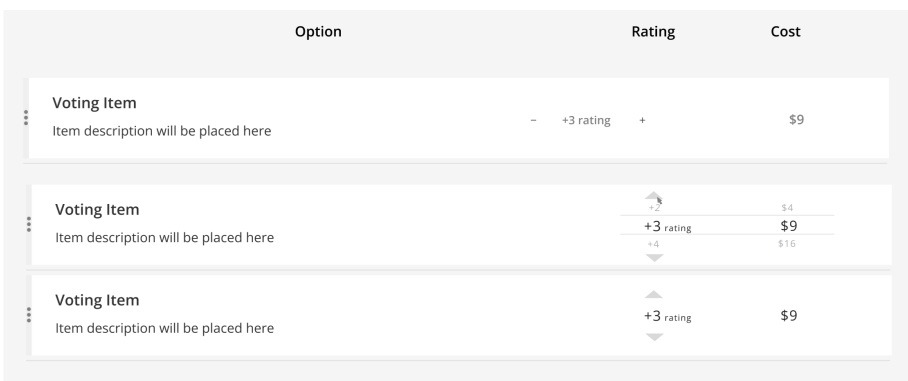


Fig. 3. Alternative vote control. The click-based design (upper) mirrors traditional vote control used in other QV interfaces, where each click controls one vote. The wheel-based design (the latter two) allows control through both clicks and mouse wheel rotation.

3.2 Baseline Interface: Single-Phase Text Interface

We designed the single-phase text interface (referred to as text interface for short) as our control condition to compare how the interactive components influenced participants' cognitive load and behavior. The text-based interface includes all the functionalities of the two-phase interface except for the two-phase interactive design and drag-and-drop

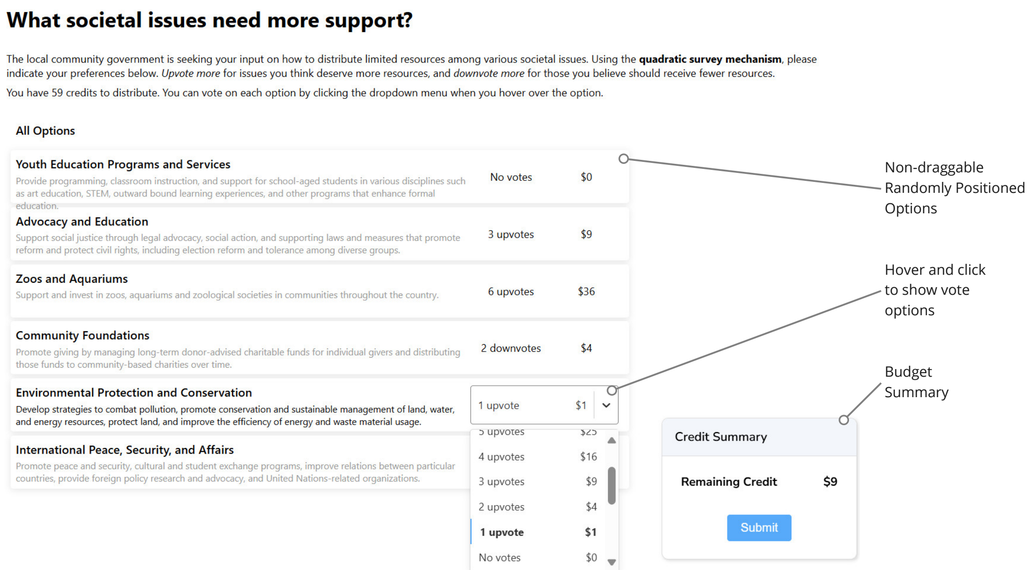


Fig. 4. The text-based interface: This interface is based on the interactive version but does not include the two-phase interactive support and lacks the drag-and-drop functionality. Options are randomly positioned.

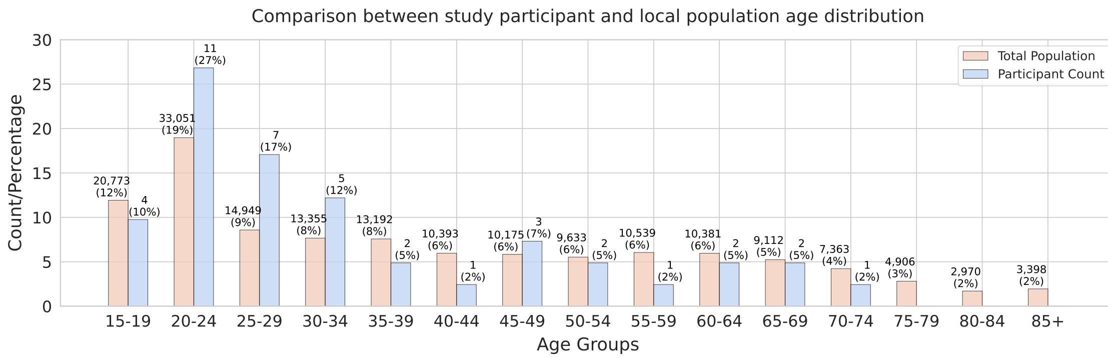
feature (Figure 4.) The interface displays the question prompt at the top of the screen. The options appear in a list underneath the prompt. Survey can update their votes by selecting from a dropdown that lists all possible voting options and costs, based on the number of available credits. A small summary box on the right of the interface displays the current total cost and the remaining credits for the respondent. The interface randomly presents options to prevent ordering bias [76, 77].

Both experimental interfaces were developed with a ReactJS frontend and a NextJS backend powered by MongoDB. We open-source both interfaces.[1]

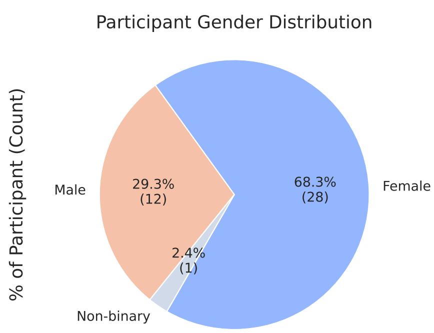
4 Experiment Design

We recruited 41 participants, with one excluded[2] due to data quality concerns, from a United States college town using online ads, digital bulletins, social media posts, online newsletters, and physical flyers in public spaces beyond campus. To ensure participant diversity, we prioritized non-students by selectively accepting them as we monitored demographics. Study participants' mean age was 34.63 years old, with an age distribution similar to the county's demographic profile (Figure 5a) albeit a slightly higher representation of younger adults. Gender and race demographics are aggregated in Figure 5b and 5c. The study was framed as focusing on societal attitudes to avoid response bias. The university's Institutional Review Board reviewed and approved this study.

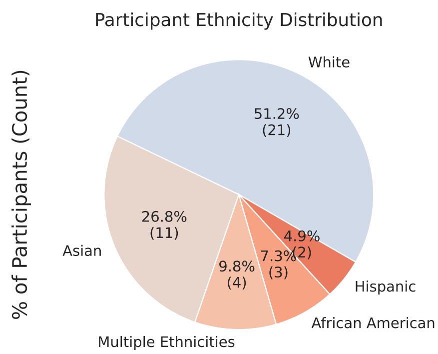
[1]link-to-github
[2]The participant reported not completing the survey seriously because they believed the experiment was fake.



(a) Age distribution of the study participants were similar to the locale's demographic profile.



(b) Gender distribution of our participants skewed towards female participants.



(c) Ethnicity distribution remains diverse with fewer Hispanic and African American participants.

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

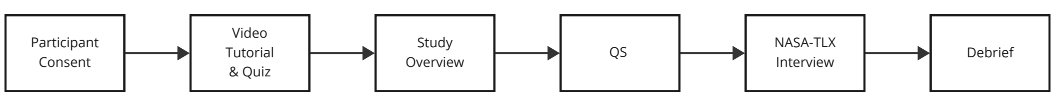


Fig. 6. Study protocol: Participants are asked to learn about the mechanism of QS after consenting to the study. The researcher explained the study overview and asked participants to complete the QS. A NASA-TLX survey followed by interviews to understand participants' cognitive load. We debriefed participants after the study.

Figure 6 visually represents the study protocol. Participants completed the study in the lab to control for external influences. Participants used a 32-inch vertical monitor displaying all options. After consenting, participants watched a video explaining the quadratic mechanism without hints of interface operation followed by a quiz to ensure understanding. Participants rewatched the video or consulted the researcher until they could select the correct answers. The participant's screen was captured throughout the study. The researcher primed the participant that the study aimed to help local community organizers understand preferences on societal issues to better allocate resources. Participants were randomly assigned to one of four groups:

- Short Text (ST): A text interface with 6 options. ($N = 10$)
- Short Two-Phase (SP): A two-phase interface 6 options. ($N = 10$)
- Long Text (LT): A text-based interface 24 options. ($N = 10$)
- Long Two-Phase (LP): A two-phase interface with 24 options. ($N = 10$)

Participants completed the survey independently, without the researcher's presence. They then contacted the researcher for the NASA-TLX survey, followed by a short audio-recorded semi-structured interview. The session concluded with a debriefing and a $15 cash compensation, during which participants were informed of the study goal on cognitive load and interface design.

We made several experimental design choices. First, we selected a between-subject design to minimize study fatigue, considering the complexity of QS, and avoiding the learning effect that could influence how participants evaluated the options. Second, we chose the context of public resource allotment, where participants expressed their preferences regarding their preference across 6 or 24 societal issue options, following the methodology of Cheng et al. [4]. These issues are relevant to all citizens and effectively demonstrate the need to prioritize limited public resources. We curated 26 societal issues used by Charity Navigator [78] which evaluates over 20,000 charities in the United States. The interface randomly presents options from this list to participants. Appendix C contains the full list.

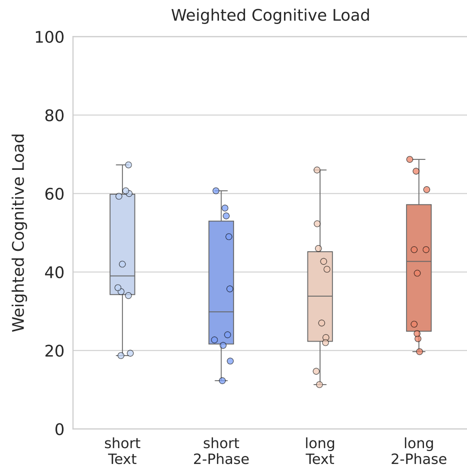
We decided to test 6 and 24 options, representing short and long lists, as identifying the 'breaking point' for cognitive overload would require impractical time and resource commitments. Constant sum surveys and the Analytic Hierarchy Process (AHP) recommend fewer than ten and seven options, respectively [79, 80, 81]. Miller [82]'s classic work on cognitive processing capacity and Saaty and Ozdemir [83]'s theoretical proof supported the use of 7 ± 2 items. A meta-analysis by Chernev et al. [84] identified 6 and 24 as common values for short and long lists in choice overload studies, rooted in the original experiment by Iyengar and Lepper [59].

Finally, we deployed self-report subjective surveys and analytical measures (i.e., time and clickstream data). We adopted the paper-based weighted NASA Task Load Index (NASA TLX), a widely used multidimensional tool that averages six subscale scores to represent overall workload after completing a task [85, 86, 87]. NASA-TLX is favored for its low cost and ease of administration [88], with less variability compared to one-dimensional workload scores [89], making it suitable for our study. Given the extended nature of QS, we did not choose to measured cognitive load using performance measures, psychophysiological measures, subjective measures, and analytical measures [88]

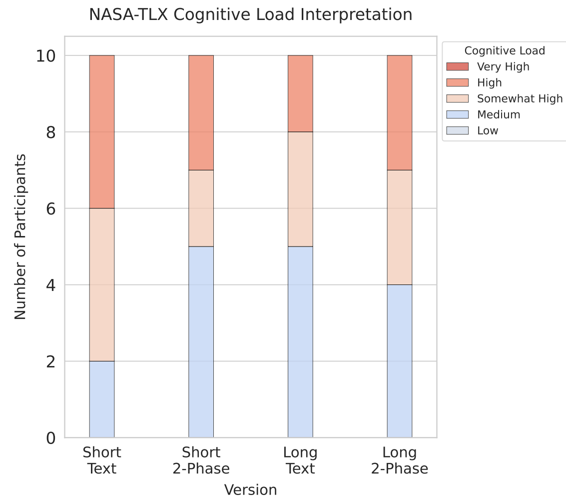
5 Cognitive Load and Sources across Experiment Conditions

This section presents cognitive load across experiment groups and the sources contributing to each cognitive load dimension. Due to the smaller sample size, we focus on descriptive statistics and qualitative assessments. We present quantitative data from survey tasks and qualitative insights from post-survey interviews.

The first author conducted an inductive thematic analysis [90] after transcribing the interviews. Snippets were coded based on research questions and topics of interest, and similar codes were merged to form themes. When differences emerged across experiment conditions and hypotheses were formed, the first author applied a deductive coding process to the relevant text snippets. We begin by presenting an overview of the cognitive load findings.



(a) NASA-TLX Weight Score: The Long Two-Phase Interface exhibits the highest weighted cognitive load with a median of 42.70, a mean of 42.02. This is higher than the long text interface, which has a median cognitive load of 33.85 and a mean of 34.60. However, the short text interface demonstrates a higher cognitive load with a median of 39.00, a mean of 43.23, compared to the short two-phase interface, which has a median of 29.85, a mean of 35.36. The standard deviation is similar across groups at around 18.



(b) NASA-TLX Cognitive Interpretation: More participants in the short text interface, totaling 8, reported a somewhat high or above cognitive load, which is significantly higher compared to the 5 participants who reported similarly for the short two-phase interface. However, the long two-phase interface saw slightly more participants, 6 in total, reporting somewhat high or above cognitive load compared to the long text interface.

Fig. 7. This figure shows the box plot results for weighted NASA-TLX scores across experiment groups and participant counts based on individual score interpretations. In 7a, we observe a downward trend in cognitive load for the short QS, while the long QS shows an upward trend. Interestingly, there is a counterintuitive downward trend between short and long text interfaces. In 7b, these trends are clearer when NASA-TLX scores are grouped into five tiers.

5.1 Overall Cognitive Load

To answer our research question on how the number of options in QS (**RQ1**) and the interface (**RQ2a**) impacted cognitive load, we derive the weighted NASA-TLX scores across the four experiment conditions. We show these results in Figure 7. Weighted NASA-TLX uses a continuous 0-100 score, with higher values indicating greater cognitive load. We use predefined mappings of NASA-TLX scores to cognitive levels: low, medium, somewhat high, high, and very high, as listed by Hart and Staveland [85]. We show value interpretations in Figure 7b.

Surprisingly, the long text interface had lower mean ($\mu = 34.60$) and median ($\tilde{x} = 33.85$) cognitive load scores than both the short text ($\mu = 43.23$, $\tilde{x} = 39.00$) and long two-phase interfaces ($\mu = 42.02$, $\tilde{x} = 42.70$). Additionally, the two-phase interface decreased cognitive load for the short survey but increased it for the long survey. Notably, the short text interface had the most participants ($N = 8$) reporting somewhat high or higher cognitive loads. In contrast, the other conditions had a more balanced distribution, with about half reporting medium to high loads.

We acknowledge that these results may not fully reflect actual cognitive load due to potential noise from factors such as small sample size, task nature, or participants' interpretation of the scale. To explore the similarities and sources of cognitive load across interfaces (**RQ2b**), we turn to qualitative insights from post-task interviews based on NASA-TLX cognitive load dimensions. Among the six NASA-TLX dimensions, we focus on mental demand, temporal demand, and frustration while summarizing key findings regarding physical demand, performance, and effort for brevity. Appendix D provides further details on all six dimensions.

5.2 Sources of Mental Demand

We began by examining mental demand, which refers to the amount of mental and perceptual activity required to complete a task. Interview results indicated that the primary drivers of participants' mental demand were *Budget management* and *Preference construction.*

5.2.1 Mental Demand Source #1: Budget management. 14 participants expressed demand from budgeting within limited credit (S032 *[...] for certain societal issues you had to ... take away from other societal issues that you could support.* , $N = 5$), tracking remaining credits (S006 *[...] looking at the remaining credits, I'm trying to mentally divide that up before I start allocating* , $N = 10$), and maximizing credit use (S032 *[...] I used all the credit that I had available* , $N = 8$).

We categorized budget management-induced mental demand as either operational (single interface-level action, e.g., using the last remaining credit) or strategic (higher-level goal, e.g., evenly distributing credits across options). Participants who completed longer surveys more frequently reported operational causes, suggesting that an increase in survey options induced short-term thinking.

5.2.2 Mental Demand Source #2: Preference construction. All but one participant reported increased mental demand due to preference construction. We further break it down into three sources: comparative preference evaluation (i.e., evaluating relative importance between options; S002 *Figuring out ... how much I prioritize option 1 over option 2* , $N = 16$), resource-constraint prioritization (i.e., trading off between options due to resource constraints, S005 *[...] very hard to take decisions ... because I felt that multiple options deserve equal amount of credit ... but you have given very limited amount of credit.* , $N = 17$), and deciding the exact number of votes (S023 *[...] having to pick how many upvotes would go to each one* , $N = 30$).

Almost all participants mentioned preference construction as a source of mental demand, supporting our claim that preference construction when completing QS is a difficult and mentally demanding task. Notably, more participants using the text interface reported mental demand from deciding the exact number of votes compared to the two-phase interface (18 vs. 12). We conjecture that the first pass on the survey items and organizing them helped participants construct preliminary preferences and reduced their mental demands when they started allocating votes.

In addition, when categorizing preference construction-induced mental demand, participants ($N = 8$) in the long text interface tend to consider a smaller scope that focuses on personal relevance. Conversely, participants ($N = 9$) in the long two-phase interface considered the broader societal impact and evaluated options more comprehensively. Compare the following two quotes, where one focused on adjusting credits between two options and the other reflecting across broader societal values:

> *Trying to figure out what upvotes I should give [...] went back and forth between those two. [...] it was very mentally tasking for me.*
> S015 (LT)

> *[...] really having to think, especially with so many different societal issues. How do I personally prioritize them? And to what extent do I prioritize them?* S009 (LI)

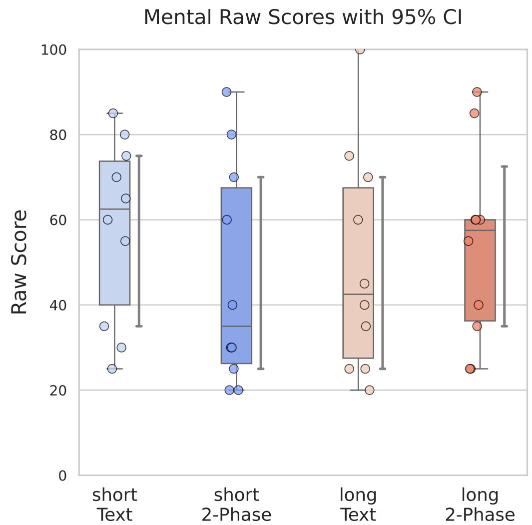


Fig. 8. Mental Demand Raw Score: Across all four experiment groups, participants' reported mental demand is spread across a wide range with many participants experiencing high mental demand.

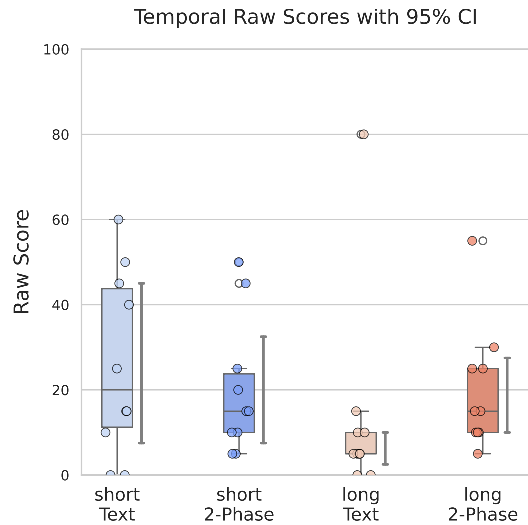


Fig. 9. Temporal Demand Raw Score: The short text interface results in the highest temporal demand, while the long text interface is the lowest. Two-phase interfaces show moderate temporal demand, suggesting that interactive elements allowed participants to pace themselves better.

Inspecting both causes, we did not notice significant differences in mental demand raw values (Figure 8) across the four experiment groups, especially between the interfaces across long QS. However, the experiment groups' mental demands came from different sources.

5.3 Sources of Temporal Demand

Temporal demand measures the time pressure participants feel during a task. Lower demand indicates participants taking a more leisurely pace. The main sources of increased temporal demand relate to time pressure on *Decision-Making* (`S024` *maybe I should just hurry up and make a decision.* , $N = 15$) and *Operational Tasks* (`S032` *to be able to move through this quickly and efficiently* , $N = 16$) (Table 5). Additionally, some participants mention *Budget management* as a source of temporal demand (`S034` *as the money decreases I felt kind of rushed* , $N = 4$).

5.3.1 Two-phase Interface Reduced Temporal Demand on Short QS. The raw NASA-TLX values in Fig 9 show that participants in the short text interface reported the highest temporal demand. Five participants expressed concerns about time spent on decision-making, feeling themselves invested more time and effort than expected, prompting them to rush. However, the two-phase interface reduced this, with only one participant in the short survey group reporting similar concerns.

5.3.2 Long QS on Text Interface Showed the Lowest Temporal Demand. Surprisingly, participants in the long text interface exhibited the lowest temporal demand (Fig. 9) despite making more decisions and operations compared to the short text group. Two possible explanations might explain this counter-intuitive result. First, more participants in the short survey group ($N = 7$) expressed a desire to complete the task efficiently, compared to just one participant ($N = 1$) in the long survey group, saying things like:

I wanna get through things in an efficient manner [. . .] to move through this quickly and efficiently. S032 (ST)

Second, satisficing behaviors may explain the lower temporal demand in the long text condition. Participants may have experienced cognitive overload from the long list of options and, as a result, spent less time than expected on decisions. For example:

I didn't really do the math, so I was like $2 is not that much left so I tried my best to use up most of it. S035 (LT)

We will discuss this possibility further in Section 7.1.

5.3.3 *Two-phase Interface Increased Temporal Demand in Long QS.*

Despite the unexpectedly low temporal demand in long QS with a text interface, two-phase interface participants have indifferent temporal demand across survey lengths. All five participants who mentioned a feeling of time pressure on decision-making in the long two-phase group described the pressure affirmatively. This means their pressure stemmed from having too many remaining decisions to make (S022 *So it didn't take too much time. but obviously there were a lot of things to consider, so there was some temporal demand.*), not from the time they have already spent (i.e., framed negatively) as that in the short text group.

5.4 Source of Frustration

Frustration refers to the extent to which the participant is annoyed, irritated, or discouraged during the task. We identified either *Operational Actions* (e.g., credit management ($N = 6$) and managing quadratic vote costs ($N = 5$)), or *Societal Concerns* (e.g., regretful trade-offs ($N = 8$) or pessimism about other's vote ($N = 6$)) as sources of frustration.

In general, the frustration derived from societal concerns did not seem strongly affected by any of the experimental conditions. We saw some discrepancies with respect to operational action-driven frustration. The long text interface condition had the fewest participants expressing operational frustration, with half expressing no frustration, mirroring the trends in the actual scores (Figure 10). Similar to the finding that the long text group has the lowest temporal demand, this is counter-intuitive as more options and dense text are known to lead to more frustration in interface design [91]. Participants engaging in satisficing behaviors in the long text interface may explain this phenomenon – prior literature [92, 93] indicates that satisficers tend to be less frustrated and happier than maximizers.

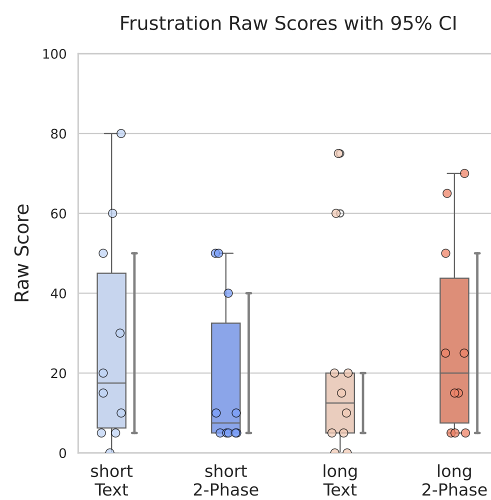


Fig. 10. Frustration Raw Score: Participants other than the long text interface highlighted several operational tasks that led to frustration. All groups share causes from strategic planning.

5.5 Physical Demand, Effort and Performance

Physical demand refers to the physical effort required to complete a task, such as physical exertion or movement. The two-phase interface experienced higher physical demand from increased mouse usage.

Effort refers to how hard participants felt they worked to achieve the level of performance they did. Qualitative analysis showed participants using the two-phase interface, regardless of length, considered options more comprehensively and felt less effort in completing operational tasks. Almost all participants ($N = 9$) from the long two-phase interface spent effort planning a strategy to complete tasks with many ($N = 7$) considered options comprehensively and beyond the immediate task (i.e., considering the broader community impact of their choices).

Performance refers to a person's perception of their success in completing a task. An interesting element that contributed to their cognitive load comes from concerning social responsibility. They wonder how their final vote counts would be perceived by others (S041 *I don't want people to think that I just like don't care about <ethnicity> people at all*) or influence real-world decision-making (S027 *Some of these things might . . . have outcomes that I didn't foresee*). In addition, when analyzing how participants describe their performance, we categorize them into indications of satisficing behaviors("good enough"), exhausting their effort (i.e., "done their best,"), or feeling positive (i.e., "feeling good.") We observed twice as many participants using the two-phase interface to report the positive feeling about their final submission ($N = 11$ v.s. $N = 6$).

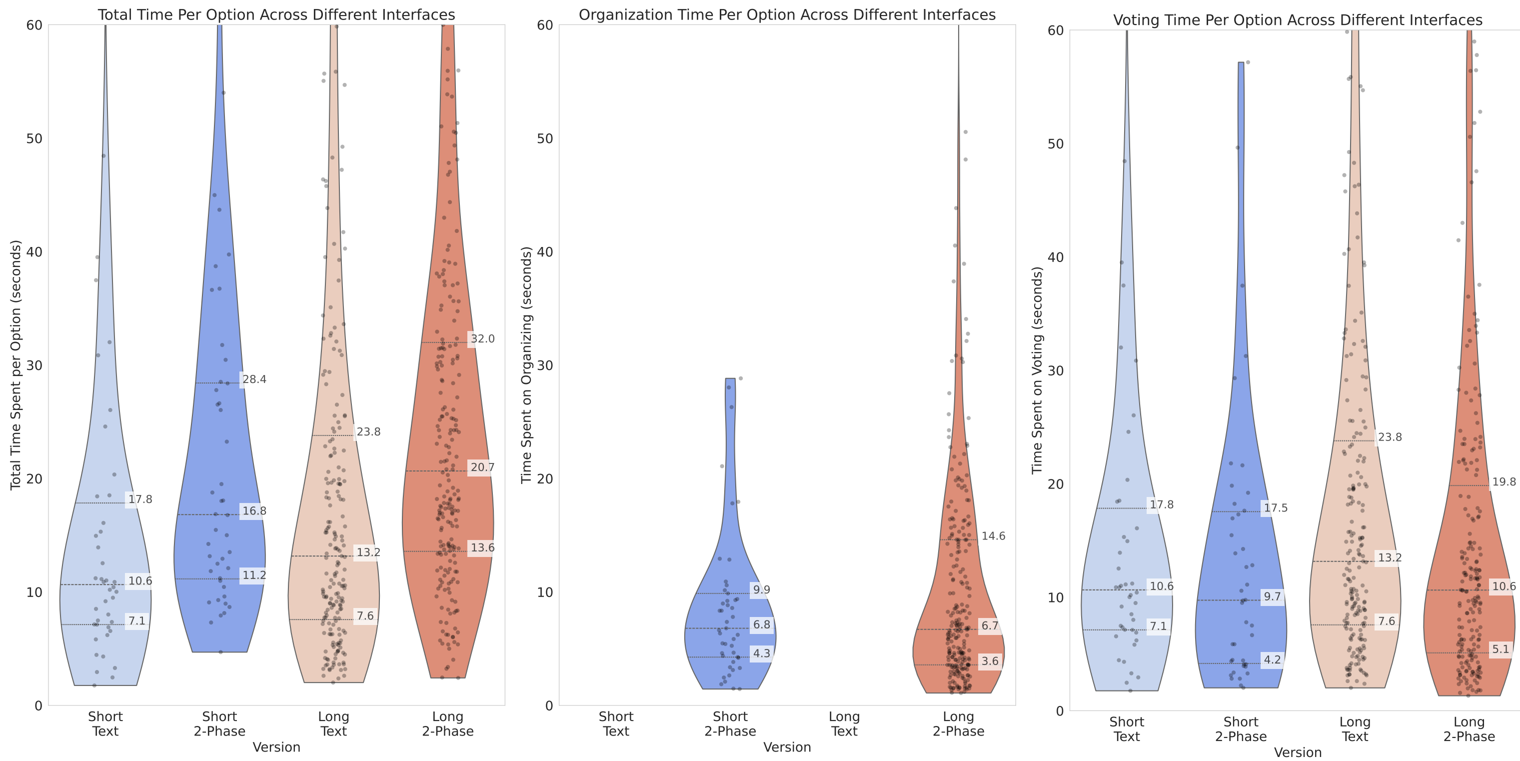
5.6 Summary across all cognitive load dimensions

Overall, participants using the two-phase interface tended to think more comprehensively and critically, while those using the text interface focused more on operational tasks. This is reflected by *mental demand* findings where participants using the text interface reported greater difficulty in determining the precise number of votes. In contrast, participants using the long two-phase interface were more likely to consider broader societal impacts and evaluate options holistically. *Effort* echoes similar findings. In terms of *physical demand*, participants using the two-phase interface, particularly in longer surveys, experienced higher demand due to increased mouse usage, which was expected. Additionally, participants using the short text interface wanted to complete the task quickly and reported the highest *temporal demand*, while those using the long text interface reported the lowest. Moreover, participants using the long text interface exhibited the least amount of *frustration* from operational causes compared to other experimental conditions. Thus, we suspect that participants who completed the long QS on a text interface engaged in satisficing behaviors, based on the counter-intuitive results showing they had the lowest temporal demand and frustration levels. Finally, in relation to *performance*, participants using the two-phase interface more frequently reported positive feelings about their final submissions, suggesting greater confidence in their decision-making process. We will interpret these results in the discussion section. To better understand participants' behavior, we analyze click-stream data across experimental conditions in the next section.

6 Behavior Results

To understand how interfaces and options influenced survey response behaviors (**RQ3**), we investigate time-to-action and remaining credit differences across experiment conditions. Time-to-action is a widely used metric in decision sciences, where longer decision time often indicates more complex cognitive processing [94]. Additionally, resource allocation strongly influences decision-making. Cheng et al. [4] showed that the number of given credits influences the validity of QV. Decision science studies like Shah et al. [95] and [96] showed how scarcity influences decisions, increases risk aversion, and adds cognitive load. These measures serve as proxies for participant behavior, and all analyzed data is publicly available[3] for transparency and to facilitate further research.

[3]link-to-github



(a) Total Time per option: We identified that the two-phase interface skewed slightly higher than the text interface, as expected. This discrepancy can be attributed to the extra organization step required in the two-phase interface, leading to a slightly longer overall completion time per option.

(b) Organization Time per option: Only the two-phase interface includes an organization phase, hence the other experimental conditions do not exhibit any accumulated organization time. This figure reinforces that organization time was isolated to the two-phase design.

(c) Voting Time per option: We observe statistically significant faster voting times for the long QS in the two-phase interface. This suggests that the two-step design improves voting efficiency, especially for longer surveys.

Fig. 11. Time per option across all experiment conditions. The violin plots visualize time spent per option across the experimental conditions, with dots representing total time per option and horizontal lines indicating the median and interquartile ranges. While total time was slightly higher for the two-phase interface, the distinct phases (organization and voting) allowed for a more structured approach, particularly benefiting longer surveys.

6.1 Time Spent per Option

Our first analysis focuses on understanding how much time participants spent per option across different stages and experiment conditions. Based on the QS system log, we extracted the following detail: *the option* involved in the interaction, *the type of interaction* (such as updating a certain number of votes), and *the time* between this interaction and the previous one. Each dot on Figure 11 is a specific type of time a participant spent for one option.

Total time. Total time is the aggregate of all time spent on each option across both phases. Each dot on Figure 11a visualizes the total time. Participants spent slightly more time per option on the two-phase interface than the text interface. A non-parametric Mann-Whitney U test showed a small effect size (long QS: $p < 0.0000001$, Rank-biserial: -0.304, Cohen's d: 0.030; short QS: $p = 0.01$, Rank-biserial: -0.37, Cohen's d: 0.082). This is expected as the two-phase interface has an additional step of organizing the options.

We further break down the total time spent into organization time and voting time. To minimize noise, we intentionally dropped all the time participants spent on the first option in the organization phase or voting phase. The goal is to exclude time spent on reading the prompt, forming their preference, or understanding the interface.

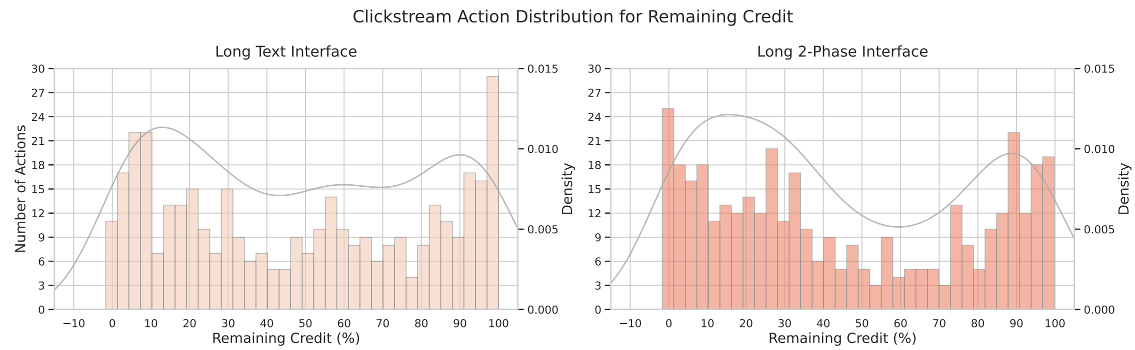
Organization time. Organization time covers both placing options into categories and the drag-and-drop time during the organization phase. Illustrated in Figure 11b, we observed minimal difference in organization time per option between short and long surveys, as the interface shows options one at a time for categorization. It also suggests that even for longer surveys, the organization functionalities did not significantly impact the time participants needed for an option.

Voting time. Voting time strictly refers to the time participants took to update vote values for each option. As shown in Figure 11c, participants spent significantly less time voting on the two-phase interface than on the text interface with a small effect size in the long QS ($U = 24053$, $p < 0.005$, Rank-biserial: 0.167, Cohen's d: 0.017), but not in the short survey ($p > 0.4$, Power=0.051). This supports our hypothesis that the two-step design in the two-phase interface facilitates more efficient decision-making, especially in longer surveys.

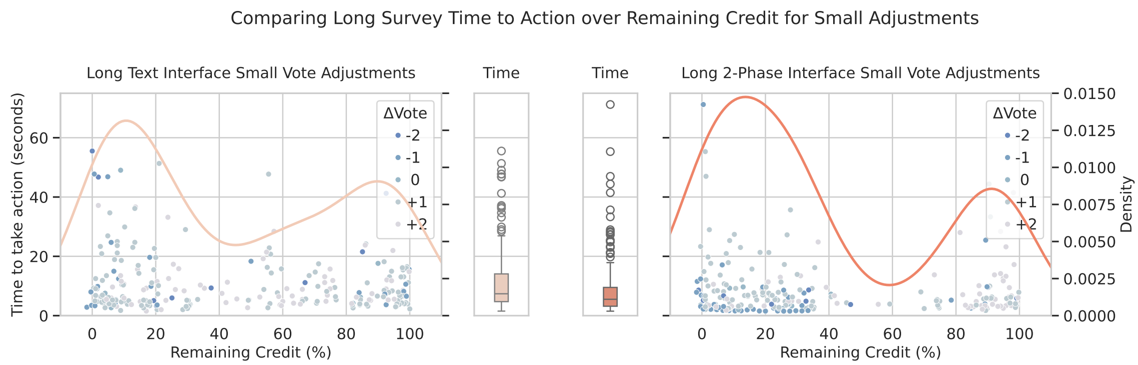
6.2 Budget and Voting Behaviors

We further breakdown and highlight key differences of how long QS participant's voting behaviors when credits changed their voting behaviors with detailed analysis in Appendix F. Figure 12a first shows the number of vote adjustments at a given remaining budget across the two interfaces. We then plotted the vote adjustments of two or fewer votes, which is 10% of the possible values one can choose among the maximum of 21 votes in Figure 12b. A kernel density estimate (KDE) plot is provided to visualize the trends and compare how it changed when we plotted only small vote adjustments.

In long surveys, participants exhibited more actions both when the budget was abundant and when it began to run out. This pattern was more pronounced with the long two-phase interface. In fact, the bimodal distribution is more pronounced in the two-phase interface. This suggests that participants make small adjustments both at the beginning and toward the end of the QS. However, the two-phase interface shows more frequent and faster edits towards the end. Visually, dots are more clustered in the long two-phase interface for small vote adjustments compared to the long text interface. The Mann-Whitney U Test on the time spent on small vote adjustments showed significant differences ($U = 13037$, $p < 0.001$), with a small effect size (Rank-biserial: 0.227, Cohen's d: 0.195) and a power of 0.381. This indicates that participants had a clearer idea of how to distribute their credits across the options.



(a) This plot counts the number of voting actions when there are x percentages of credits remaining. A KDE plot is provided to help better understand the action distribution.



(b) This plot further separates participants' interaction behavior based on the number of votes participants adjusted. We observed a bimodal interaction pattern across long QS when small vote adjustments are made.

Fig. 12. Comparison of voting actions and participant behavior in different survey interfaces. Subplot (a) shows the overall distribution of actions based on the remaining credit, while subplot (b) further differentiates the interaction based on the size of vote adjustments.

Five participants highlighted how the interface supported their incremental iterative approach during the interview whom all used the two-phase interface. As one participant pointed out:

> *I like the fact that it remembers everything that you know. [. . .] that's very important is that it's an iterative process.* 🗨 `S019 (LI)`

In summary, participants spent more time on the two-phase interface compared to the text interface in both short and long surveys. Across the two-phase interfaces, organization time remained consistent. While voting time did not differ between interfaces for the short survey, participants voted more quickly on the two-phase interface in the long survey, confirming the hypothesis that the two-step design enhances decision-making efficiency. Voting behaviors indicated more frequent actions when the budget was abundant and nearly exhausted, particularly in the long two-phase interface. Additionally, the analysis revealed more frequent and faster small vote adjustments towards the end of the QS in the two-phase interface, demonstrating an iterative and incremental approach.

7 Discussion and Future Works

In the discussion section, we interpret results related to cognitive load and survey respondent behaviors. Specifically, we aim to understand why results in Section 5 revealed that long text interface participants did not experience higher cognitive load, contrary to expectations referencing prior literature and insights from Section 6.

We focus on three key topics: what two-phase interface elements influenced behavior, how these elements supported preference construction, and the remaining design challenges. Additionally, we provide recommendations for using QS and suggest design improvements. We concluded that the two-phase interface prevented satisficing behaviors and promoted strategic, holistic thinking, unlike the text-based interface, which leaned toward operational tasks. Behavioral analysis showed that long two-phase interface participants made frequent, small updates, shifting their cognitive focus.

7.1 Interpretation of results: Two-phase interface limits satisficing during cognitive overload

Participants using the long two-phase interface reported slightly higher cognitive load compared to those using the long text interface. In comparison, the short two-phase interface resulted in a lower cognitive load. This suggests that the two-phase interface alone does not inherently increase cognitive load.

One explanation is that the two-phase interface reduces cognitive load in the short QS but not in the longer one. In longer surveys, the interactive nature of the two-phase interface may require participants to perform more operations without significantly altering their decision-making processes. However, our findings suggest that participants using the two-phase interface in longer surveys responded more efficiently. Across both short and long surveys, participants using the two-phase interface demonstrated iterative and efficient fine-tuning preferences, indicating deeper survey engagement. Therefore, we reject the claim that the two-phase interface increases cognitive load in the long survey.

An alternative explanation is that while the two-phase interface limits satisficing during cognitive overload, the complexity of the long QS may have counteracted this effect, preventing a full reduction in cognitive load. This suggests that long QS participants resorted to satisficing behaviors to manage their cognitive overload. Qualitative results support this explanation: participants using the two-phase interface engaged in broader, more strategic considerations, while those using the text interface focused on operational tasks. For instance, participants using the long two-phase interface experienced less cognitive burden from precise voting and managed their time better. Overall, the two-phase interface participants demonstrated deeper and more critical thinking, as they engaged in more strategic decision-making.

In contrast, participants using the long text interface satisficed due to cognitive overload from having to decide and allocate more credits for more options. However, fewer participants in this group reported high cognitive load compared to those in the short-text interface. This group also experienced the least temporal demand and frustration with operational tasks, despite spending a similar amount of time per option as participants using the text interface. These counterintuitive findings pointed to the remaining plausible explanation that long-text interface participants used satisficing to manage the overload.

In summary, the interactive components of the two-phase interface likely prevented mental shortcuts [58, 10, 11, 12], resulting in a cognitive *shift* towards deeper reflection and decision-making. Thus, while QS with many options might lead to satisficing, the two-phase interface redirects participants' attention, facilitating the decision-making process that promotes comprehensive preference construction. In the following section, we examine these specific elements that guided participants to achieve this.

7.2 Bounded rationality and interface design

Bounded rationality emerged as a core theme in participants' responses, highlighting how cognitive limitations lead to sub-optimal decision-making due to their inability to process all available information [10]. This often led to *satisficing* behaviors, where participants settle for *good enough* but not *optimal* decisions [97].

For instance, S036 described making meaningful choices while acknowledging the limits of their effort:

> *[...] 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 [...] and then move on. [...] I felt like that (the response) was satisfied, but not perfect. Cause perfect is not a reality.* S036 (ST)

This illustrates typical satisficing decision-making, where participants settle for suboptimal choices. In long QS, the significant increase in decision points—due to the numerous vote and credit options—was so overwhelming to some that even features like pre-calculated vote-credit values in drop-down menus provided relief, as they helped avoid the cognitively demanding task of searching for bounds.

Participants often relied on *heuristics* [98] and *defaults* [69], behaviors less common among two-phase interface participants. Presenting one option at a time reduced reliance on defaults and encouraged deeper reflection, as illustrated by S013, who highlighted how the organization phase supported their option preferences construction:

> *[...] it (organization phase) gives you time to just focus on that single thing and rank it based on how you feel at that moment.* S013 (SI)

Conversely, text interface participants, like S003, described how default placements influenced their decisions:

> *Honestly, if medical research [...] was the first one I saw, I think it would automatically give it a lot more.* S003 (ST)

The three key elements of the organization phase—presenting options one at a time, grouping them into categories, and enabling drag-and-drop—worked together to structure participant preferences. These elements aligned with cognitive strategies like *problem decomposition* [99] and *dimension reduction* which reduces cognitive overload. Bounded rationality illuminates the importance of decision-making support interfaces rather than being a critique of human behaviors. One participant explained how the organization phase broke down complex decisions into manageable steps:

> *[...] being able to have a preliminary categorization of all the topics. First, it introduced me to all the topics, [...] to think about and process [...] being able to digest all the information prior to actually allocating the budget or completing the quadratic survey.* S009 (LI)

Participants using the two-phase interface, especially in the long version, organized options along dimensions such as topics (e.g., health vs. humanitarian) and preferences (positive vs. negative) before voting. Others expressed that the upfront introduction of all options and the ability to rank and group them helped manage their cognitive load effectively. In contrast, almost half of the participants using the long text interface, like S028, expressed a desire for features that can help reduce the decision space when responding to the QS, further supporting the importance of these organizational design elements:

> *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 (LT)

This quote reflected participants' need to manually track and revisit options, which occupies cognitive load, without a more structured interface.

In summary, individual's bounded rationality encouraged participants to exhibit *satisficing* behaviors, *heuristics*, and *defaults* when responding to QS. Showing all options upfront, one at a time, and repositioning options based on participants' rough preferences prevented participants from using defaults and heuristics. The two-phase organization actively scaffolds participants' decision-making process, supporting efficiency *problem decomposition* and *dimension reduction*. Together, these elements in the two-phase interface design prevented *satisficing* behaviors and supported participants in making more informed decisions through a more strategic planning and holistic thinking process.

7.3 Construction of Preference on Quadratic Survey

Completing QS is a series of difficult decision tasks Lichtenstein and Slovic [7]. Svenson [65]'s differentiation and consolidation theory helps explain how participants process these decisions. The decision process begins with differentiation, where participants identify differences and eliminate less favorable options, followed by consolidation, which strengthens their commitment to selected choices. This theory aligns with how the two-phase interface helps participants decompose options into categories, effectively reducing decision complexity.

Participants started by constructing preferences in situ, especially regarding options they hadn't previously considered:

> *[...] '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?*
> S037 (LI)

Those using the text interface, lacking the interactive tools, found it challenging to facilitate differentiation as S025 noted:

> *I would like to be able to like, click and drag the categories themselves so I could maybe reorder them to like my priorities. [...] make myself categories and subcategories out of this list ... If I could organize it.* S025 (LT)

In contrast, the two-phase interface allowed participants to express at least one dimension of differentiation more easily. The drag-and-drop feature helped blend this differentiation into the consolidation phase. Not only does participants drag-and-drop options post voting to reflect and assure a correct vote allocation, it also enables participants, like S039, to make fine-grain comparisons between options:

> *I think the system was actually really helpful because I could just drag them. [...] I can really compare them, I can drag this one up here, and then compare it to the top one [...]* S039 (SI)

The bi-modal behavior observed in the long interactive interface participants provided in the results aligns with the differentiation and consolidation framework. Participants in the two-phase interface began differentiating options earlier allowing them to later adjust fine-grain votes. The faster and smaller vote updates indicated participants consolidating. The less prominent bi-modal behavior from the long text interface participants implied that the interface guided this decision framework as participant 037 explained:

> *I only start from the positive one [...] I finish everything ... and then I move to the second part (the neutral 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 (LI)

These evidences explain how the organization phase and the drag-and-drop features supported differentiation and consolidation, scaffolding a decision-making framework enabling deeper engagement.

In summary, participants construct their preferences as they complete QS. We observed behaviors and qualitative insights that align with the differentiation and consolidation theory in decision-making. Our interface scaffolded many of the differentiation stages through pre-voting organization and some consolidation phases through drag-and-drop, explaining how the two-phase approach supports preference construction to yield effective QS responses.

7.4 Opportunities for better budget management

Budget management is a recurring theme in participant interviews. While they appreciated the automatic calculation feature in modern QV interfaces, we identified three challenges for future QS interfaces: *cognitive load*, *the cold-start problem*, and *navigating between budget, votes, and outcome*.

7.4.1 Automatic calculation is critical. Over one-third of participants ($N = 14$) from all four experiment conditions emphasized the importance of automated calculation for deriving costs and tracking expenditures. For example:

> *I thought I have [...] (to) do all the numbers or calculations myself as a part of checking my ability of doing mathematics. [...] I said that credit summary to be very specific. The credit summary section was really wonderful in doing all the calculations on that end.*
> `S005 (LT, keeping track of spent)`

The quotes marked the importance that QS must be facilitated by computer-supported interfaces.

7.4.2 The coldstart problem. We notice from the study that one of the biggest challenges for participants is deciding 'how many votes' to start with. This challenge pertains to the initial vote, not the relative vote. Some participants began by equally distributing their credits to all options and then made adjustments. Others established 1, 2, and 3 votes as starting points. A small handful surprisingly used the tutorial's example of 4 upvotes as their anchor.

This arbitrary voting decision echoes discussions in prior literature about the existence of an absolute value for individuals. Coherent arbitrariness [100], similar to the anchoring effect in marketing, refers to participants' willingness to allocate votes, which can be influenced by an arbitrary value. However, the ordinal utility remains intact among the set of preferences.

7.4.3 Navigating Between Budget, Votes, and Actual Impact. The third challenge is participants' confusion between budget, votes, and outcomes, despite understanding their definitions. One participant stated:

> *[...] 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 causes. So you have to figure out which ones get the most, which ones don't get as much. [...]*
> *Interviewer: [...] Do you feel that the more votes you're giving to a cause you're actually spending more on it?*
> *Yeah.* `S003 (ST)`

Participants like `S003` bypassed the quadratic formulation, directly translating votes to resource allocation. While this does not invalidate the power of the quadratic mechanism, it causes frustration and friction for participants to construct a clear picture of how to make voting decisions. Future interfaces should better communicate these relationships to facilitate respondents' trade-offs.

In summary, while the interface supports budget management through automated cost calculation, participants still face cognitive load from managing the budget. The cold-start problem and the confusion between budget, votes, and actual impact are open questions for future research. These challenges highlight the need for better budget management support to complete the QS interface.

7.5 Quadratic Survey Usage, 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 current 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.

7.5.1 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. QS should be designed for critical evaluations, such as investment decisions, or situations where participants have ample time to think and process the survey. For instance, revealing the options ahead of time can aid in preference construction.

7.5.2 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.

7.5.3 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.

8 Limitations

Evaluating the QS interface is challenging due to its novelty. During the study, we identified several limitations that require further research.

Understanding results influence on decision-makers. Further research is required to understand how the QS interface impacts decision-makers and broader societal resource distributions. Since QS is still in its early stages, we prioritize its widespread adoption and usage before attempting a comprehensive assessment of its influence on decision-making. Future studies will examine how decision-makers interpret and use QS data, as well as its broader implications for societal decisions.

Individual differences in cognitive capacity. Variations in individual cognitive capacity influenced participants' cognitive scores. For example, participants with more experience in decision-making might be able to manage multiple options more effectively. A within-subject study could clarify cognitive load shifts, but deconstructing established preferences and altering options further complicates this. Thus, we opted for this in-depth, between-subject study, although the small sample size may introduce noise that distorts the actual cognitive load. Future research should quantify the impact of different QS interfaces. In addition, participants completed this study in a controlled lab environment with options displayed on a large screen. Future work should also explore how individuals respond to QS on smaller devices in a less controlled environment.

Limited experience with QS. Participants had no prior experience with the QS interface. Following a tutorial and quiz, participants proceeded to complete tasks using the QS interface. While participants understood the QS mechanics, familiarity with the interface still influences strategies and cognitive load. As quadratic mechanisms become more prevalent, future research can compare novices and experts.

Duration between clicks to represent decision-making. Click duration may include time spent considering other options, so it must be treated as an approximate measure of decision-making time. For instance, deciding between two options may take longer for the first option and less time for the second. Despite its limitations, this approach provides valuable insights into decision-making within our experimental constraints.

9 Conclusion

In this study, we designed and evaluated a novel two-phase "Organize-then-Vote" interface aimed at guiding Quadratic Survey (QS) respondents in effectively constructing their preferences. Through an in-lab study employing NASA-TLX and interviews, we explored how this two-phase interface influenced individuals' cognitive load and survey response behaviors when engaging with societal issues of varying lengths. The interface's organization and voting phases, designed to reduce cognitive overload by structuring the decision-making process, allowed respondents to differentiate between options before voting. Results revealed that the two-phase design decreased reliance on satisficing behaviors and encouraged more iterative and reflective preference construction, even though it did not clearly reduce overall cognitive load. Nonetheless, this design shift promoted deeper engagement and strategic thinking compared to the text-based interface, especially in longer surveys, by distributing cognitive effort more effectively. Quantitative results confirmed that participants, particularly those responding to the longer survey, exhibited more frequent fine-tuning of their votes, reflecting the iterative nature fostered by the interface. By integrating the organization and drag-and-drop functions, the interface facilitated both preference differentiation and consolidation, making it easier for respondents to refine their decisions. This two-phase interface design supports the development of future software tools that facilitate preference construction and promote the broader adoption of Quadratic Surveys. Future research should explore how to better support individuals in deciding the allocation of budget and design interfaces for smaller devices.

References

[1] 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.

[2] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, (May 2019), 1–12. DOI: 10.1145/3290605.3300316.

[3] Muhsin Ugur, Dvijesh Shastri, Panagiotis Tsiamyrtzis, Malcolm Dcosta, Allison Kalpakci, Carla Sharp, and Ioannis Pavlidis. 2015. Evaluating smartphone-based user interface designs for a 2d psychological questionnaire. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 275–282.

[4] 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.

[5] 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.

[6] 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.

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

[8] Adam Rogers. 2019. Colorado Tried a New Way to Vote: Make People Pay—Quadratically | WIRED. *Wired*, (Apr. 2019). Retrieved June 22, 2024 from.

[9] 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.

[10] 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.

[11] 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.

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

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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] Christian Jilek Paula Gauselmann Yannick Runge and Tobias Tempel. 2023. A relief from mental overload in a digitalized world: How context-sensitive user interfaces can enhance cognitive performance. *International Journal of Human–Computer Interaction*, 39, 1, 140–150. eprint: https://doi.org/10.1080/10447318.2022.2041882. DOI: 10.1080/10447318.2022.2041882.

[19] Sharon Oviatt. 2006. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM International Conference on Multimedia*, 871–880.

[20] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2021. To reuse or not to reuse? A framework and system for evaluating summarized knowledge. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW1, (Apr. 2021). DOI: 10.1145/3449240.

[21] Helena M Reis et al. 2012. Towards reducing cognitive load and enhancing usability through a reduced graphical user interface for a dynamic geometry system: An experimental study. In *2012 IEEE International Symposium on Multimedia*. IEEE, 445–450.

[22] Benjamin Lafreniere, Andrea Bunt, and Michael Terry. 2014. Task-centric interfaces for feature-rich software. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (OzCHI '14). Association for Computing Machinery, New York, NY, USA, 49–58. DOI: 10.1145/2686612.2686620.

[23] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW1, (Apr. 2021). DOI: 10.1145/3449161.

[24] Emin İbili. 2019. Effect of augmented reality environments on cognitive load: pedagogical effect, instructional design, motivation and interaction interfaces. *International Journal of Progressive Education*, 15, 5, 42–57.

[25] Amy X. Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proc. ACM Hum.-Comput. Interact.*, 2, CSCW, (Nov. 2018). DOI: 10.1145/3274465.

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

[27] Eric A Posner and E Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.

[28] Ryan Naylor et al. 2017. First year student conceptions of success: What really matters? *Student Success*, 8, 2, 9–19.

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

[30] 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.

[31] 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.

[32] 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.

[33] 2023. Gov4git: A Decentralized Platform for Community Governance. (Mar. 2023). Retrieved June 13, 2024 from.

[34] 2024. RadicalxChange/quadratic-voting. RadicalxChange. (May 2024). Retrieved June 17, 2024 from.

[35] [n. d.] Read the Whitepaper | Gitcoin. https://www.gitcoin.co/whitepaper/read. (). Retrieved June 17, 2024 from.

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

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

[38] Syavash Nobarany, Louise Oram, Vasanth Kumar Rajendran, Chi-Hsiang Chen, Joanna McGrenere, and Tamara Munzner. 2012. The design space of opinion measurement interfaces: exploring recall support for rating and ranking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2035–2044.

[39] Paul Van Schaik and Jonathan Ling. 2007. Design parameters of rating scales for web sites. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 14, 1, 4–es.

[40] Jing Wei, Weiwei Jiang, Chaofan Wang, Difeng Yu, Jorge Goncalves, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding how to administer voice surveys through smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 6, CSCW2, (Nov. 2022). DOI: 10.1145/3555606.

[41] Aman Khullar et al. 2021. Costs and benefits of conducting voice-based surveys versus keypress-based surveys on interactive voice response systems. In *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies* (Compass '21). Association for Computing Machinery, New York, NY, USA, 288–298. DOI: 10.1145/3460112.3471963.

[42] Martin Feick, Niko Kleer, Anthony Tang, and Antonio Krüger. 2020. The virtual reality questionnaire toolkit. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 68–69.

[43] 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.

[44] 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.

[45] 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.

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

[47] 2015. Designing usable ballots | Center for civic design. https://civicdesign.org/fieldguides/designing-usable-ballots/. (June 2015). Retrieved June 17, 2024 from.

[48] Whitney Quesenbery. 2020. Opinion | Good Design Is the Secret to Better Democracy. *The New York Times*, (Oct. 2020). Retrieved June 17, 2024 from.

[49] 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.

[50] 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.

[51] 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)*.

[52] 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.

[53] 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.

[54] 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.

[55] Graham Cooper. 1998. Research into cognitive load theory and instructional design at UNSW. (1998).

[56] Stoo Sepp, Steven J. Howard, Sharon Tindall-Ford, Shirley Agostinho, and Fred Paas. 2019. Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory. *Educational Psychology Review*, 31, 2, (June 2019), 293–317. DOI: 10.1007/s10648-019-09461-9.

[57] Antonio Drommi, Gregory W Ulferts, and Dan Shoemaker. 2001. Interface design: A focus on cognitive science. In *The Proceedings of ISECON 2001*. Vol. 18.

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

[59] 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.

[60] 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.

[61] 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.

[62] Peter Coy. 2019. A New Way of Voting That Makes Zealotry Expensive - Bloomberg. *Bloomberg*, (May 2019). Retrieved Dec. 16, 2023 from.

[63] 2022. Quadratic Voting Frontend. Public Digital Innovation Space. (Jan. 2022). Retrieved Dec. 16, 2023 from.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] A Norman Donald. 2013. *The Design of Everyday Things*. MIT Press.

[71] 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.

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

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

[74] 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.

[75] 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.

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

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

[78] 2023. Charity Navigator. https://www.charitynavigator.org/index.cfm?bay=search.categories. (May 2023). Retrieved Dec. 16, 2023 from.

[79] 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).

[80] Thomas L. Saaty and Kirti Peniwati. 2013. *Group Decision Making: Drawing Out and Reconciling Differences*. RWS Publications, (Nov. 2013).

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

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

[88] 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.

[89] 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.

[90] Judith S. Olson and Wendy A. Kellogg, eds. 2014. *Ways of Knowing in HCI*. Springer, New York, NY. doi: 10.1007/978-1-4939-0378-8.

[91] Jakob Nielsen. 1997. How users read on the web.

[92] Evan Polman. 2010. Why are maximizers less happy than satisficers? Because they maximize positive and negative outcomes. *Journal of Behavioral Decision Making*, 23, 2, 179–190. DOI: 10.1002/bdm.647.
[93] Barry Schwartz, Andrew Ward, John Monterosso, Sonja Lyubomirsky, Katherine White, and Darrin R. Lehman. 2002. Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology*, 83, 5, 1178–1197. DOI: 10.1037/0022-3514.83.5.1178.
[94] John W. Payne, James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9781139173933.
[95] Anuj K. Shah, Eldar Shafir, and Sendhil Mullainathan. 2015. Scarcity frames value. *Psychological Science*, 26, 4, 402–412.
[96] Ernst-Jan de Bruijn and Gerrit Antonides. 2022. Poverty and economic decision making: a review of scarcity theory. *Theory and Decision*, 92, 1, (Feb. 2022), 5–37. DOI: 10.1007/s11238-021-09802-7.
[97] 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.
[98] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 4157, 1124–1131. Retrieved June 21, 2024 from JSTOR: 1738360.
[99] Herbert A. Simon. 1996. *The Sciences of the Artificial*. (3rd ed ed.). MIT Press, Cambridge, Mass.
[100] 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.

A Voting Interface Breakdown

Compared to digital survey interfaces, there exists a rich literature on voting interfaces, which we argue is a special type of survey interface. We categorize these related works into three main categories detailed below:

Designs that shifted voter decisions: For example, states without straight-party ticket voting (where voters can select all candidates from one party through a single choice) exhibited higher rates of split-ticket voting [13]. Another example from the Australian ballot showing incumbency advantages is where candidates are listed by the office they are running for, with no party labels or boxes.

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 [48]. On the other hand, Everett et al. [49] showed the use of incorporating physical voting behaviors, like lever voting, into graphical user interfaces.

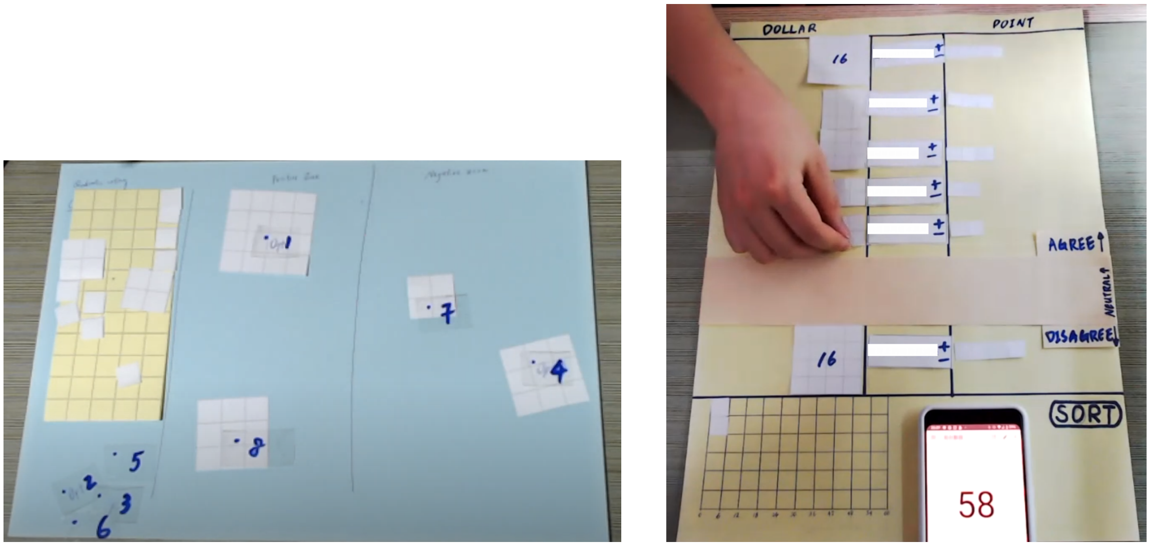
Designs that incorporated technologies: Other projects like the Caltech-MIT Voting Technology Project have sparked research to address accessibility challenges, resulting in innovations like EZ Ballot [50], Anywhere Ballot [51], and Prime III [52]. In addition, Gilbert et al. [53] investigated optimal touchpoints on voting interfaces, and Conrad et al. [54] examined zoomable voting interfaces.

B Interface design process

In this section, we outline the design process leading to our final interface. As mentioned in the paper, our design iteration began from existing QV applications in the wild.

B.1 Prototype 1: Ranking-Vote

Considering that relative preference is often through ranking items, we tested whether ranking options before voting would help establish an individual's relative preference in our prototype 1. This prototype allowed respondents to reposition options before voting. Pretests revealed that respondents rarely moved the options and questioned the necessity of full ranking, as it did not influence their QS submission. Additionally, many were unaware that options



(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. 13. Initial paper prototypes designed for QS interface

were draggable until shown. This insight indicates that full ranking is unnecessary for establishing relative preferences. Therefore, we decided to ask respondents to select a subset of options instead of requiring a full rank among all options.

B.2 Prototype 2: Select-then-Vote

Based on feedback from Prototype 1, instead of *allowing* individuals to rank options, Prototype 2 implemented a two-phase process that *intentionally* asks respondents to select options to express opinions before voting. As shown in Figure 15, survey respondents selected their preferred options (Figure 15a), and the interface positioned these options at the top of the list for voting (Figure 15b). We identified several issues during the prototype 2 pretest: many respondents marked most options as 'options they care about,' which undermined the design's purpose. Additionally, the lack of clear distinction between selected and unselected options confused respondents about the necessity of Step 1. Thus, we need a clearer distinction and connection between the two phases to effectively construct relative preferences.

B.3 Prototype 3: Organize-then-Vote

Figure 16 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: Lean Positive, Lean Negative, and Lean 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

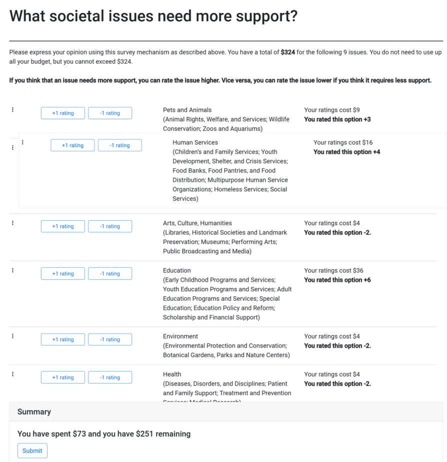
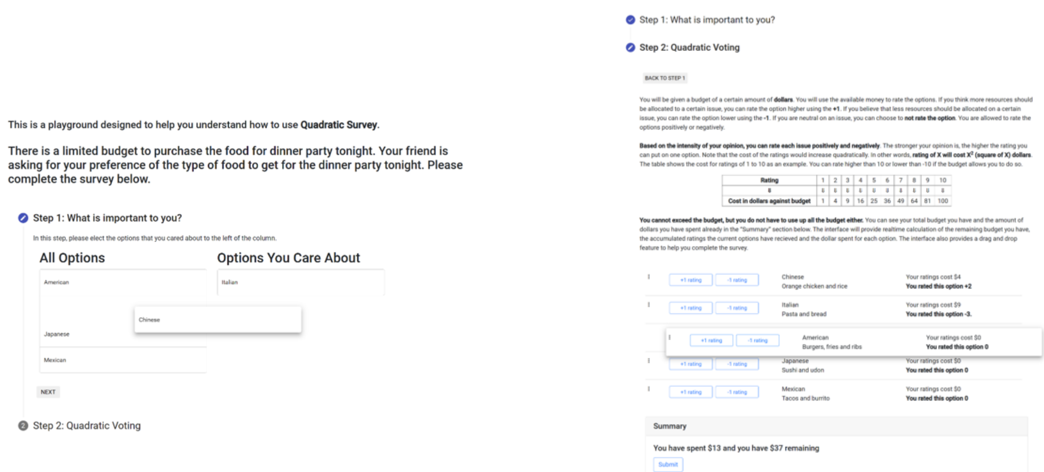


Fig. 14. A Ranking-Vote Prototype: The goal of this prototype is to test whether ranking options prior to voting help establish an individual's relative preferences and reduce effort when voting. Each option is draggable to position in a specific location amongst the full list of options. Votes can be updated using the buttons 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.



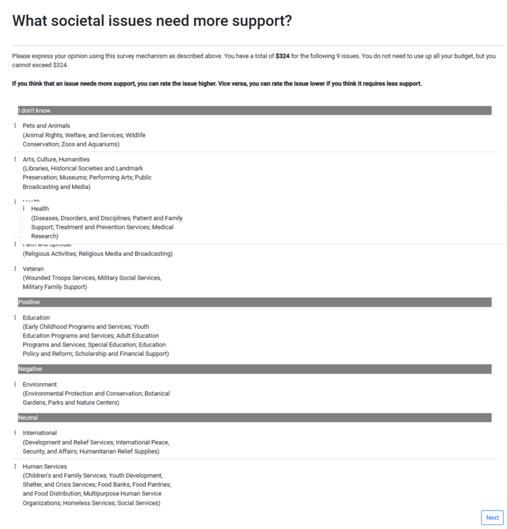
(a) Options are dragged and dropped to the 'Option You Care About' box.

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

Fig. 15. A Select-then-Vote Prototype: The goal of this prototype is to nudge participants to focus on a subset of options to vote, rather than ranking all of them. This prototype introduces a two-step voting process. As shown in Fig. 15a, the first step involves selecting options for further consideration. Important options are placed at the top of the list for voting shown in Fig. 15b, but options can be placed anywhere on the list if desired. The rest of the controls remain the same as the previous prototype.

below. Voting controls and information appear on each option once respondents move to the subsequent page, forming a clear connection between option groups, positions, and voting controls.

Feedback indicated that survey respondents are comfortable with the two-phase organize-then-vote design, demonstrating it as a central strategy for our interface development. However, several areas for enhancement were identified:



(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: Lean Positive, Lean Neutral, or Lean Negative. Only the details of each option are shown on this interface.

(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. 16. Organize-then-Vote Prototype: The goal of this prototype is to encourage participants to derive finer grain categories among options before voting. Survey respondents first organize their thoughts into categories and then vote on the options.

First, the dragging and dropping mechanism in the organization phase is cumbersome and may inadvertently suggest a ranking process, contrary to our intentions. Second, placing 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. These insights guided refinements in the final two-phase interface, adhering to the two-phase organize-then-vote design framework.

C List of Options

We provide the full list of options presented on the survey.

- **Animal Rights, Welfare, and Services:** Protect animals from cruelty, exploitation and other abuses, provide veterinary services and train guide dogs.
- **Wildlife Conservation:** Protect wildlife habitats, including fish, wildlife, and bird refuges and sanctuaries.
- **Zoos and Aquariums:** Support and invest in zoos, aquariums and zoological societies in communities throughout the country.
- **Libraries, Historical Societies and Landmark Preservation:** Support and invest public and specialized libraries, historical societies, historical preservation programs, and historical estates.
- **Museums:** Support and invest in maintaining collections and provide training to practitioners in traditional arts, science, technology, and natural history.

- **Performing Arts:** Support symphonies, orchestras, and other musical groups; ballets and operas; theater groups; arts festivals; and performance halls and cultural centers.
- **Public Broadcasting and Media:** Support public television and radio stations and networks, as well as providing other independent media and communications services to the public.
- **Community Foundations:** Promote giving by managing long-term donor-advised charitable funds for individual givers and distributing those funds to community-based charities over time.
- **Housing and Neighborhood Development:** Lead and finance development projects that invest in and improve communities by providing utility assistance, small business support programs, and other revitalization projects.
- **Jewish Federations:** Focus on a specific geographic region and primarily support Jewish-oriented programs, organizations and activities through grantmaking efforts
- **United Ways:** Identify and resolve community issues through partnerships with schools, government agencies, businesses, and others, with a focus on education, income and health.
- **Adult Education Programs and Services:** Provide opportunities for adults to expand their knowledge in a particular field or discipline, learn English as a second language, or complete their high school education.
- **Early Childhood Programs and Services:** Provide foundation-level learning and literacy for children prior to entering the formal school setting.
- **Education Policy and Reform:** Promote and provide research, policy, and reform of the management of educational institutions, educational systems, and education policy.
- **Scholarship and Financial Support:** Support and enable students to obtain the financial assistance they require to meet their educational and living expenses while in school.
- **Special Education:** Provide services, including placement, programming, instruction, and support for gifted children and youth or those with disabilities requiring modified curricula, teaching methods, or materials.
- **Youth Education Programs and Services:** Provide programming, classroom instruction, and support for school-aged students in various disciplines such as art education, STEM, outward bound learning experiences, and other programs that enhance formal education.
- **Botanical Gardens, Parks, and Nature Centers:** Promote preservation and appreciation of the environment, as well as leading anti-litter, tree planting and other environmental beautification campaigns.
- **Environmental Protection and Conservation:** Develop strategies to combat pollution, promote conservation and sustainable management of land, water, and energy resources, protect land, and improve the efficiency of energy and waste material usage.
- **Diseases, Disorders, and Disciplines:** Seek cures for diseases and disorders or promote specific medical disciplines by providing direct services, advocating for public support and understanding, and supporting targeted medical research.
- **Medical Research:** Devote and invest in efforts on researching causes and cures of disease and developing new treatments.
- **Patient and Family Support:** Support programs and services for family members and patients that are diagnosed with a serious illness, including wish granting programs, camping programs, housing or travel assistance.
- **Treatment and Prevention Services:** Provide direct medical services and educate the public on ways to prevent diseases and reduce health risks.

- **Advocacy and Education:** Support social justice through legal advocacy, social action, and supporting laws and measures that promote reform and protect civil rights, including election reform and tolerance among diverse groups.
- **Development and Relief Services:** Provide medical care and other human services as well as economic, educational, and agricultural development services to people around the world.
- **Humanitarian Relief Supplies:** Specialize in collecting donated medical, food, agriculture, and other supplies and distributing them overseas to those in need.
- **International Peace, Security, and Affairs:** Promote peace and security, cultural and student exchange programs, improve relations between particular countries, provide foreign policy research and advocacy, and United Nations-related organizations.
- **Religious Activities:** Support and promote various faiths.
- **Religious Media and Broadcasting:** Support organizations of all faiths that produce and distribute religious programming, literature, and other communications.
- **Non-Medical Science & Technology Research:** Support research and services in a variety of scientific disciplines, advancing knowledge and understanding of areas such as energy efficiency, environmental and trade policies, and agricultural sustainability.
- **Social and Public Policy Research:** Support economic and social issues impacting our country today, educate the public, and influence policy regarding healthcare, employment rights, taxation, and other civic ventures.

D Detailed Qualitative Cognitive Load Breakdown

In addition to the discussion on cognitive load sources presented in the main text, we provide additional details on the six cognitive dimensions. Among all dimensions, we also provide the codes representing different types of demand in a table form. The shaded cells represent the percentage of participants citing each source of mental demand, allowing for comparison within columns. The abbreviations in the columns: ST (Short Text Interface), SP (Short Two-phase Interface), LT (Long Text Interface), and LP (Long Two-phase Interface). Short and Long refer to the sum across both interfaces; Text and Inter refer to the sum across both survey lengths. We include Sparklines for comparisons across these experiment groups.

E Sources of Mental Demand

Table E lists all the mental demand codes.

E.1 Sources of Physical Demand

Physical demand refers to the physical effort required to complete a task, such as physical exertion or movement. Most participants reported minimal physical demand ($N = 32$), reflected in the low NASA-TLX physical demand scores (Figure 17). Notably, 11 out of 20 participants who used the two-phase 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 (Figure 17), which show a significant visual difference between short and long two-phase interfaces as well as between text and two-phase interfaces in long surveys. Table 2 presents all the relevant codes across experiment groups.

Table 1. This table lists all the causes participants mentioned as contributing to their Mental Demand.

[Mental Demand]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
Budget Management	14	3	3	5	3	6	8	8	6
Budget within limited credit	5	2	2	1	0	4	1	3	2
Track remaining credits	10	2	2	3	3	4	6	5	5
Maximize credit usage	8	2	3	2	1	5	3	4	4
Operational	12	3	2	4	3	5	7	7	5
Strategic	7	2	4	1	0	6	1	3	4
Preference Construction	39	10	9	10	10	19	20	20	19
Determining relative preference	16	4	4	5	3	8	8	9	7
Option prioritization	17	6	4	3	4	10	7	9	8
Precise resource allocation	30	9	6	9	6	15	15	18	12
Narrow - Consider a few options/personal causes	23	6	6	8	3	12	11	14	9
Broad - Considering all options or higher order values	23	5	5	4	9	10	13	9	14
Demand from Experiment Setup	24	6	6	6	6	12	12	12	12
Many options on the survey	6	0	0	3	3	0	6	3	3
QS Mechanism	4	2	0	2	0	2	2	4	0
Recalling experience or understanding options	20	5	6	4	5	11	9	9	11
Justification or Reflection on response	8	2	2	1	3	4	4	3	5
External Factors	12	3	1	4	4	4	8	7	5
Demand due to Interface	8	2	2	0	4	4	4	2	6
Increase	4	1	1	0	2	2	2	1	3
Decrease	4	1	1	0	2	2	2	1	3

Table 2. Physical Demand Causes: Most participants expressed little or no physical demand. Results reflected that participants in the long two-phase interface required more actions, hence the higher mention of mouse usage as a source.

[Physical]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
Reading	4	0	2	1	1	2	2	1	3
Mouse	16	3	5	2	6	8	8	5	11
Vertical Screen	4	1	0	1	2	1	3	2	2
None/Little	32	8	9	8	7	17	15	16	16

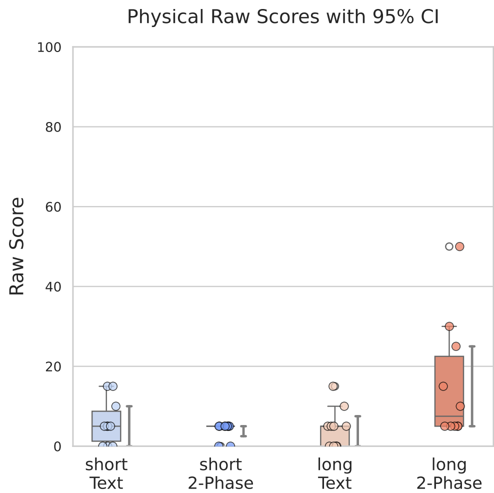


Fig. 17. Physical Demand Raw Score: Participants other than the long two-phase interface reported minimal physical demand. The long two-phase interface had the highest physical demand, likely due to increased mouse clicks and extended time spent looking at the vertical screen.

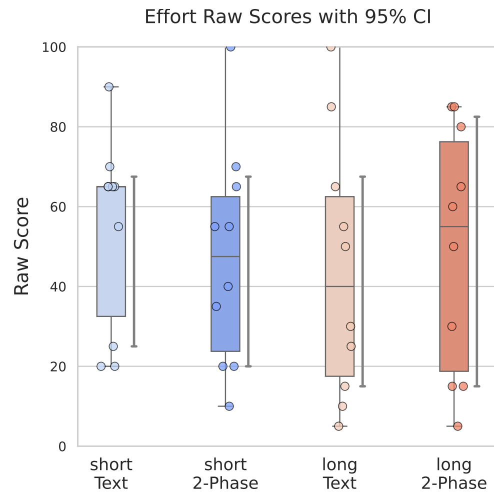


Fig. 18. Effort Raw Score: Effort scores shows indifference across groups.

Table 3. Effort Sources: Participants using the text interface focused more on operational tasks, while those using the two-phase interface focused more on strategic planning.

[Effort]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
Operational	21	6	5	8	2	11	10	14	7
Strategic	28	6	8	5	9	14	14	11	17
Personal	22	4	7	5	6	11	11	9	13
Global	11	2	3	2	4	5	6	4	7
None/Little/a bit	9	2	1	3	3	3	6	5	4

E.2 Source of Effort

Effort refers to how hard participants felt they worked to achieve the level of performance they did. Since effort includes both mental and physical resource intensity, refer to Section 5.2 and Appendix E.1 for definitions. Raw NASA-TLX effort scores (Figure 18) showed a similar spread across experiment groups, the qualitative analysis showed more distinction that participants using the two-phase interface considered options more comprehensively and felt less effort on completing operational tasks, similar to what we found on mental demands (Section 5.2). Table 3 contains codes.

E.2.1 Effort Source #1: Operational Tasks. 14 of the 20 participants using the text interface mentioned Operational Tasks as effort sources, compared to 7 using the two-phase interface, with the lowest mention by the long two-phase interface group ($N = 2$). Quotes below illustrated participants putting in effort to manipulate the interface.

I wanted to bump up (an option) maybe to 4 or <option> to 5 and realize I couldn't. [. . .] 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*

E.2.2 Effort Source #2: Strategic Planning. Different from Operational Tasks, 11 participants in the text interface compared to 17 participants described strategic planning as sources of effort, with almost all participants ($N = 9$) from the long two-phase interface. We further categorize strategic planning into *narrow* and *broad* scopes as we did for mental demand section 5.2. Participants using the two-phase interface ($N = 7$) had nearly mentioned double ($N = 4$) times regarding global strategies. For example:

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 two-phase interface*

Table 4. Performance Causes: Most causes are shared across experiment conditions. We provided qualitative interpretations of their own performance assessments.

[Performance]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
Operational Action	13	2	3	3	5	5	8	5	8
Budget Control	6	1	1	2	2	2	4	3	3
Preference Reflection	6	1	1	2	2	2	4	3	3
Limited Resources	5	1	2	1	1	3	2	2	3
Social Responsibility	8	2	2	2	2	4	4	4	4
Decision maker	7	1	2	2	2	3	4	3	4
Outcome Uncertainty	7	1	2	2	2	3	4	3	4
Performance Assessment									
Did their best	8	2	1	3	2	3	5	5	3
Feel Good	17	3	5	3	6	8	9	6	11
Good Enough	10	2	2	3	3	4	6	5	5

E.3 Source from Performance

Performance refers to a person's perception of their success in completing a task. Lower values mean good perceived performance; higher values mean poor perceived performance. We found minimal qualitative differences between experiment groups regarding factors influencing perceived performance. Two influencing factors emerged: *Operational Actions* and *Social Responsibility*. Despite most participants reporting positively on their performance, nuances exist in how different groups interpret their performance.

E.3.1 Operational Actions. Operational actions, like the theme presented in temporal demand, refer to specific, executable procedures participants perform in the survey. This could involve: pressure to spend all credits or stay within budget ($N = 6$), fears that final vote choices did not reflect true preferences($N = 5$), or concerns that they had finished the task inefficiently ($N = 6$).

E.3.2 Social Responsibility. Social responsibility-based concerns around performance came up when participants reflected on how their final vote counts would be perceived by others (S041 🗨 *I don't want people to think that I just like don't care about <ethnicity> people at all*) or influence real-world decision-making (S027 🗨 *Some of these things might . . . have outcomes that I didn't foresee*).

All groups cited social responsibility as source to evaluate effort. Raw NASA-TLX scores (Figure 19) show participants had indistinguishable performance scores. This aligns with the interview results where most participants felt positive about their final submission.

Fig. 19. Performance Demand Raw Score: Participants showed indifferent performance raw scores across experiment conditions, all trending toward satisfactory.

To dig deeper, we also analyzed participants' language when they described their performance. Expressions like "good enough" may be indicative of satisficing behaviors – our results suggest participants are satisfied at similar rates regardless of the interface. 1/4 of the participants in the text interface expressed "done their best," referring to exhausting their effort. Participants who used a two-phase interface were generally more positive about their final outcome – they were twice as likely to report "feeling good" about their final results ($N = 11$ v.s. $N = 6$).

E.4 Temporal Demand

Table E.4 lists all the mental demand codes.

Table 5. Temporal Demand Sources: Decision-making and Operational Tasks are the main causes. Participants framed their decision-making sources differently.

[Temporal]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
Budget Management	4	0	1	1	2	1	3	1	3
Decision Making	15	5	2	3	5	7	8	8	7
Affirmative	9	0	2	2	5	2	7	2	7
Negative	8	5	1	2	0	6	2	7	1
Operational	16	5	6	3	2	11	5	8	8
Task completion	8	2	2	3	1	4	4	5	3
Being efficient	8	3	4	0	1	7	1	3	5

E.5 Frustration

Table E.5 lists all the mental demand codes.

Table 6. Frustration Sources: Frustration comes from different levels of strategic operations or operational tasks.

[Fustration]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
Strategic	17	4	4	5	4	8	9	9	8
Higher-level	11	3	2	3	3	5	6	6	5
x Conflict between personal preference and broader society and common values	6	1	1	2	2	2	4	3	3
x Trade-offs among all options	8	3	1	2	2	4	4	5	3
Lower-Level	10	3	3	2	2	6	4	5	5
x Conflict between personal preference and	4	1	2	0	1	3	1	1	3
x Trade-offs among a few options	8	2	2	2	2	4	4	4	4
Operational	15	4	5	2	4	9	6	6	9
Credit management	6	2	3	1	0	5	1	3	3
Adhering to the Quadratic Mechanism	5	2	1	1	1	3	2	3	2
Deciding number of votes for an option	4	2	0	0	2	2	2	2	2
Making multiple decisions	3	2	0	0	1	2	1	2	1
Understanding Option	4	0	3	0	1	3	1	0	4
None/Little	16	4	5	5	2	9	7	9	7

F Additional voting behavior data

The reason why we decided to focus on the percentage of remaining credits comes from prior literature 'scarcity frames value' [95], a driver that makes researchers believe makes quadratic voting more accurate [4]. We did not follow Quarfoot et al. [6] in counting accumulated votes over time due to varying total times across individuals.

In this section, we complement the main text, showing a few additional figures. Figure 20 follows the main text, showing all the voting actions over the remaining credit for the four experiment conditions. Here we see two distinct patterns between the short survey and the long survey in terms of participant behaviors. In long surveys, participants exhibited more actions both when the budget was abundant and when it began to run out. This pattern was more pronounced with the long two-phase interface. This difference is why we further focused on the long QS group.

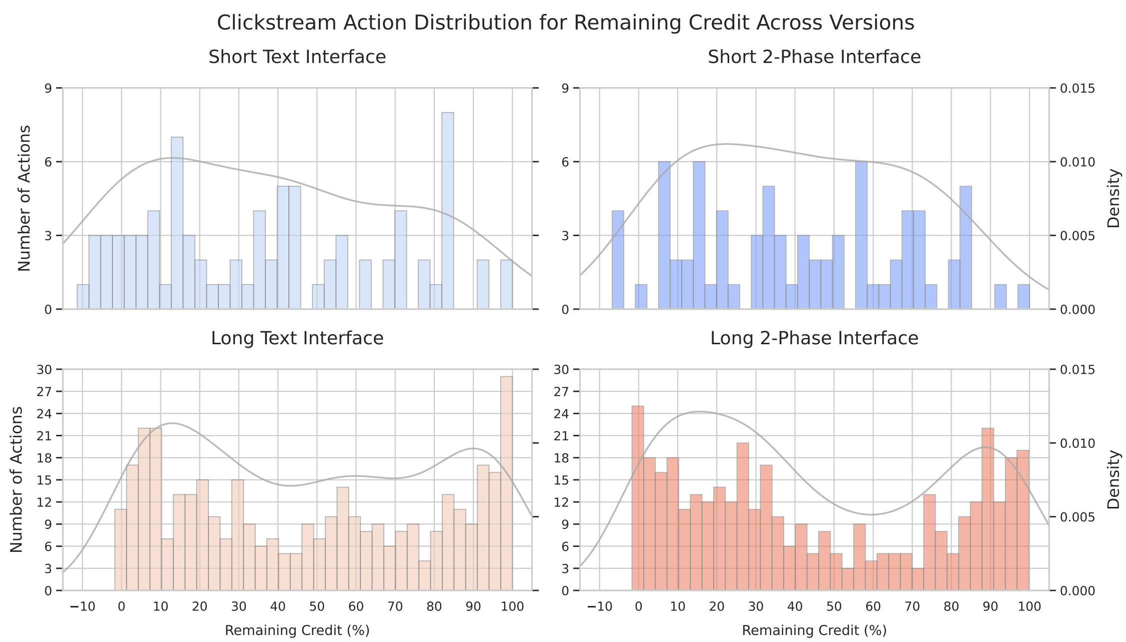


Fig. 20. This plot counts the number of voting actions when there are x percentages of credits remaining. A KDE plot is provided to help better understand the action distribution.

In addition to the small adjustment plots presented in Figure 12b, Figure 21 presents the comparison between when participants make small or large vote adjustments at different budget levels. Revisiting the KDE curve in the second row in Figure 20 and the curve of the second row in Figure 21 show a stronger bimodal distribution for small vote adjustments across interfaces. In fact, the bimodal distribution is more pronounced in the two-phase interface. This suggests that participants make small adjustments both at the beginning and toward the end of the QS. However, the two-phase interface shows more frequent and faster edits towards the end. In comparison, participants also made more large vote adjustments early on that spread more equally compared to the text interface. This indicates that participants had a clearer idea of how to distribute their credits across the options.

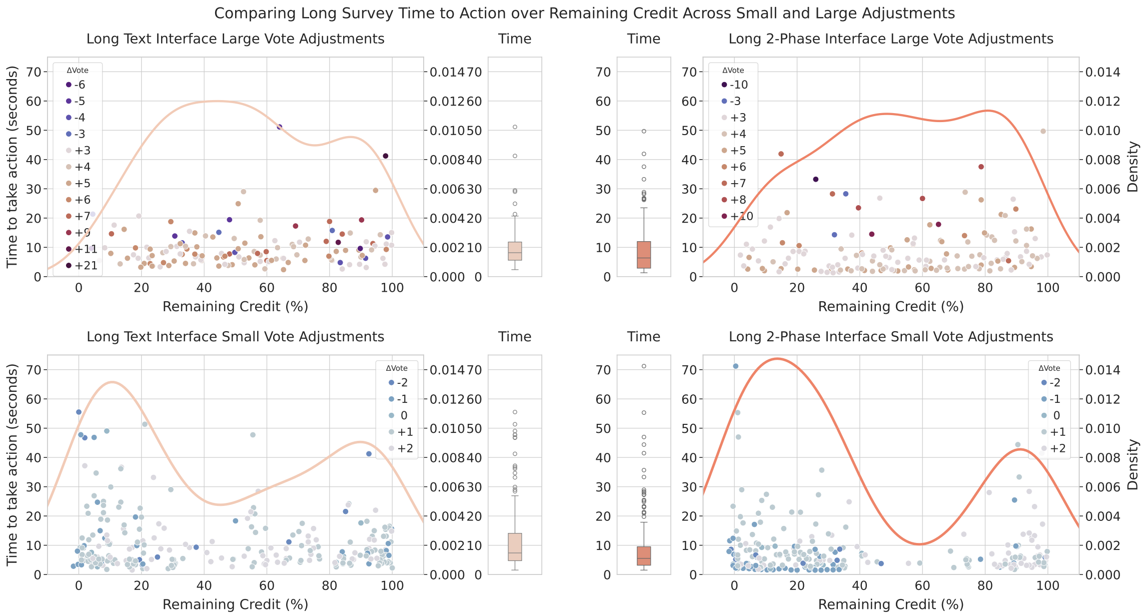


Fig. 21. This plot further separates participants' interaction behavior based on the number of votes participants adjusted. We observed a bimodal interaction pattern across long QS when small vote adjustments are made.