Quadratic Survey Interface

ANONYMOUS AUTHOR(S)*

This study introduces a new two-phase interface for Quadratic Surveys (QS) designed to scaffold users' decision-making process in order to better manage the cognitive load associated with QS's complexity. QS's complexity has greatly hindered its widespread adoption in surveying the public's opinions on societal issues for collective decision-making, despite its ability to elicit more accurate individual preferences compared to traditional surveys like the Likert-scale surveys as prior CSCW research has shown. Prior work shows that survey interface design significantly influences results and accuracy. To realize QS's full potential, we iteratively designed a two-phase "organize-then-vote" two-phase interface for QS based on decision-making and preference construction theories. Through a 2x2 between-subject in-lab study in a public resource allotment decision context, we compared this novel interface with a traditional text interface across two QSs with different lengths, one short (6 options) and one long (24 options). We found that our two-phase interface reduced participants' satisficing behaviors caused by cognitive overload in long QS conditions. In addition, we noted a shift in participants' cognitive effort from operating QS to constructing more comprehensive preferences when participants used our two-phase interface for long QS. This research contributes to CSCW by demonstrating how human-centered digital interface design can enhance the effectiveness of preference elicitation tools in collective decision-making on societal issues.

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 Voting; Survey Response Format; Collective decision-making; Preference Construction

ACM Reference Format:
Anonymous Author(s). 2024. Quadratic Survey Interface. *Proc. ACM Hum.-Comput. Interact.* 1, 1 (July 2024), 44 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Surveys are a ubiquitous tool for collective decision-making problems. Fundamentally, surveys allow decision-makers to aggregate the opinions of crowds. States utilize referendums to form policy decisions; organizations such as Pew Research Center in the United States deploy surveys to identify public perspectives on societal challenges; and city council meetings provide public forums where community members can voice their concerns. However, to ensure that the signals decision-makers receive are high quality, it is imperative that survey tools accurately capture the attitudes of survey-takers. In the domain of CSCW and HCI, researchers have demonstrated how interactive design can affect the quality of participant responses. For example, the use of voice assistants to elicit user feedback [1] and recent work showed that survey response format can increase errors [2].

Quadratic Surveys (QS) have arisen as a promising new survey tool for eliciting survey-taker preferences out of a list of items. In this paper, we define **Quadratic Surveys (QS)** as a surveying tool that employs a modification of the quadratic mechanism [3]. In QS, participants are given a

ACM 2573-0142/2024/7-ART
https://doi.org/10.1145/nnnnnnn.nnnnnnn

fixed budget of credits to spend on votes indicating support or opposition for each item. Purchasing k votes for an option in QS expense k^2 credits. Because participants are able to allocate multiple votes for or against a particular item, QS fundamentally asks participants to both rank (determine a relative preference) and rate (determine strength of preference) survey items. A recent study in CSCW research demonstrated the benefits of using QS over traditional Likert-scale surveys for social resource allotment and user experience elicitation – it more accurately reflects individual preferences because the fixed budget forces survey-takers to make trade-offs between different survey items [4]. Quadratic Surveys, referred to as Quadratic Voting when used for voting, have been deployed in the wild, both in the public [5, 6] and private sectors [7]. From research to empirical use, one of QS's goals is to better survey public opinions on societal issues. Despite the promise of QS, however, its complexity has limited its practical use. QS takes significantly longer for survey-takers to complete than a corresponding Likert-scale survey, likely due to the difficulties of reasoning around the quadratic vote cost and tradeoff-thinking induced by QS.

Prior work points to interface design as a promising tool for managing QS complexity [8, 9, 10, 11, 12, 13, 2]. The goal of Quadratic Surveys is to produce higher quality responses – the interface should facilitate, not obstruct, this outcome. Further, as Lichtenstein and Slovic [14] note, survey-takers may not always have deeply held or clear preferences prior elicitation – instead, they may construct preferences in situ. As such, the QS interface can play a crucial role in helping participants structure their thinking. Therefore, we ask: *How can we design interfaces to support participants in completing Quadratic Surveys?*

In this work, we explore ranking, selection, and organization approaches to improving QS over three design iterations. These design iterations resulted in a novel interactive QS interface (Figure 6) that has QS respondents to take a two-step 'organize-then-vote' approach to expressing their preferences. Participants are first shown survey options one-at-a-time, and are asked to categorize them into a three-tiered preference category ("Positive", "Neutral", or "Negative"). These categories are then used to order items in the QS voting page, and participants are able to drag-and-drop items to further organize them within these bins. Effectively, our interface forces raters to complete a lower-effort Likert rating task before attempting the full QS, acting as both a cognitive warm-up, and a means for identifying a convenient physical ordering of QS items.

We evaluate this new interface through a 2x2 between-subject in-lab study involving 41 participants where participants completed a societal issue survey, following the methodology described by Cheng et al. [4]. Because prior work has indicated that the number of survey options can exacerbate QS cognitive load [15, 16], participants completed a QS with either a short or long list of options using either a baseline traditional QS interface, or our novel two-phase interface. During the lab study, we measured participants' cognitive load using the NASA Task Load Index (NASA-TLX) and conducted interviews to identify how interface components affected different aspects of cognitive load. We also collected clickstream data to get more fine-grained insights into how participants approach QS.

Through this experiment, we answered the following research questions:

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

Contributions. We contributed to CSCW by proposing the first interface specifically designed for QS and QV-like applications to help promote their adoption in surveying public opinions for collective decision-making on societal issues. No prior research has investigated interfaces for QS, especially long ones that lead to cognitive overload. Our two-stage organize-then-vote interface facilitates critical decision-making and limits satisficing behaviors. This design promotes incremental updates and deeper engagement, enhancing understanding and decision quality. Second, we conducted the first in-depth qualitative analysis identifying key factors contributing to cognitive load among respondents to surveying tools that use the Quadratic Mechanism. Our qualitative findings identified design challenges for QS, driving further research directions.

2 Related Work

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

2.1 Quadratic Survey and the Quadratic Mechanism

We introduce the term **Quadratic Survey (QS)** to describe surveys that use the Quadratic Mechanism to collect individual attitudes. The **Quadratic Mechanism** is a theoretical framework designed to encourage truthful revelation of individual preferences through a quadratic cost function [3]. 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 [17]. Quarfoot et al. [18] demonstrated that QV effectively gauges public opinions and mitigates the tyranny of the majority in traditional voting systems. In addition, QV is not subject to Arrow's impossibility theorem, which states that no voting system can perfectly aggregate individual preferences without trade-offs [19], because it does not require aggregating rankings.

Quadratic Survey (QS) adapts these strengths of the Quadratic Mechanism in *voting* to encourage truthful preference expression in *surveys*. Unlike traditional surveys that elicit either rankings *or* ratings, QS allows detailed presentations of *both* by casting multiple votes for or against options, incurring a quadratic cost. Cheng et al. [4] showed that this mechanism aligns individual preferences more accurately with their behaviors than Likert Scale surveys, especially in resource-constrained scenarios.

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

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

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

$$\sum_k c_k \leq B$$

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

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

where S is the total number of participants, and $n_{i,k}$ is the number of votes cast by participant i for option k. Each additional vote for each option increases the marginal cost linearly, encouraging participants to vote proportionally to their level of concern for an issue [20].

In recent years, empirical studies on QV have expanded to different domains [21, 22]. Applications based on the quadratic mechanism have also grown, such as Quadratic Funding, which redistributes funds based on outcomes from consensus made using the quadratic mechanism [23, 24]. Recent work by South et al. [25] applies the quadratic mechanism to networked authority management, later used in Gov4git [7]. However, despite the growth in depth and breadth of applications using the quadratic mechanism, little attention has been given to the user experience and interface design that support individuals in expressing their preference intensity.

2.2 Survey, Questionnaire, and Voting Design

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

Research in the marketing and research community studying survey and questionnaire design, usability, and interactions focuses on understanding the influence of styles and question presentation, or 'Response Format,' of a survey or questionnaire. Weijters et al. [26] demonstrated that horizontal distances between options are more influential than vertical distances, with the latter recommended for reduced bias. Slider bars, which operate on a drag-and-drop principle, show lower mean scores and higher nonresponse rates compared to buttons, indicating they are more prone to bias and difficult to use. In contrast, visual analogue scales that operate on a point-and-click principle perform better [27] These research highlighted outcomes are influenced by the different designs.

Given that voting, similar to surveys and questionnaires are designed to elicit individual choices, we turn to voting interface literature. Voting interfaces can have an even more substantial influence on behaviors and outcomes. The notorious butterfly ballot [28] is one example of this – Wand et al. [28] argue that an atypical ballot design may have caused enough accidental votes to swing the 2000 U.S. Presidential Election. Researchers like Engstrom and Roberts [8], Chisnell [29], and organizations like the Center for Civic Design, which publishes reports like "Designing Usable Ballots" [30], stress the importance of interface design and how it can influence democratic processes. We group this literature into three main categories: designs that shifted voter decisions, designs that influenced human errors, and designs that incorporated technologies to improve usability.

Designs that shifted voter decisions: For example, states without the option for straight-party ticket voting (the option to circle an option that votes for all the candidates in the same party) exhibited higher rates of split-ticket voting [8]. Another example from the Australian ballot with an office block and no party box (having a box that clearly segments the position that the candidates are competing for) has been shown to enhance incumbency advantages.

Designs that influenced errors: Butterfly ballots increased voter errors because voters could not correctly identify the punch hole on the ballot. Splitting contestants across columns increases the chance for voters to overvote [31]. On the other hand, Everett et al. [32] showed the use of incorporating physical voting behaviors, like lever voting, into GUI interfaces.

Designs that incorporated technologies: Other projects like the Caltech-MIT Voting Technology Project have sparked research to address accessibility challenges, resulting in innovations like EZ Ballot [33], Anywhere Ballot [34], and Prime III [35]. In addition, Gilbert et al. [36] investigated optimal touchpoints on voting interfaces, and Conrad et al. [37] examined zoomable voting interfaces.

While the design of voting systems and question response format markedly influence voter behavior and decision accuracy, these interface elements also directly impact the cognitive load on users. Research like Galesic [38] showed that burden on survey respondents increases dropouts. An effective design would enhance usability and reduce cognitive challenges faced by survey respondents, especially in complex response mechanisms like QS.

2.3 Cognitive Challenges and Choice Overload

Despite the deep insight prior research learned about voting and surveying techniques and the robust mechanism demonstrated in theoretical applications of quadratic mechanisms, the inherent challenge that survey respondents are required to make many difficult decisions poses a unique cognitive challenge that no prior literature has tackled. Lichtenstein and Slovic [14] laid out the three key elements that make decisions difficult. They include people making decisions within an unfamiliar context, people forced to make tradeoffs due to conflicts among choices, and people quantifying values for their opinions. QS fits into the description of all three elements, as participants can face options placed by the decision maker which they have never seen before. Participants are bounded by budgets that force them to make tradeoffs, and the final votes are presented in values. Hence, we believe that QS introduces a high cognitive load.

Daniel [39] demonstrated that cognitive overload can adversely affect performance, for instance, causing individuals to rely more on heuristics rather than engaging in deliberate and logical decision-making. In addition, some researchers believe that preferences are constructed in situ just as memories are. Thus, when too much information is presented to an individual, they can 'satisfice' their decisions [40, 41, 42]. This behavior refers to when an individual settles on a 'good enough' solution rather than 'optimal' response. This overload can happen because of the presence of too many options. Subsequently, too many options can lead to individuals feeling overloaded, leading to decision paralysis, demotivation, and dissatisfaction [43].

Additionally, Alwin and Krosnick [44] highlighted that the use of ranking techniques in surveys can be time-consuming and potentially more costly to administer. These challenges are compounded when there are numerous items to rank, requiring substantial cognitive sophistication and concentration from survey respondents [45].

In several notable applications of Quadratic Voting, a single QV question can have hundreds of options. In addition to the 2019 Colorado House of Representatives considered 107 bills [46] inroduced in the introduction, the 2019 Taiwan Presidential Hackathon featured 136 proposals [47]. Psychological and behavioral research highlights the importance of understanding how individuals navigate and benefit from new interfaces under long-list QS conditions.

These empirical uses of QV hint at how QS can elicit individual preferences. This underscores the need to study the cognitive load of QS respondents and how interface design can alleviate these challenges.

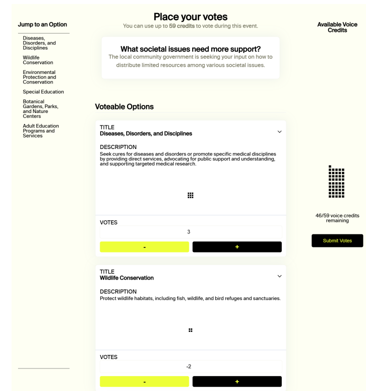
3 Interface Design

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

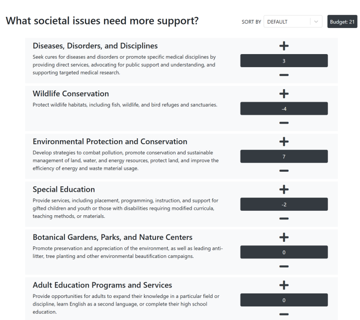
3.1 QV applications interfaces in the wild and early paper prototypes

Since QS builds upon the QV mechanism, we being our design iteration based on existing QV applications in the wild. For coherence and brevity, we chose not to describe the paper prototype iteration in depth given it's exploratory nature.

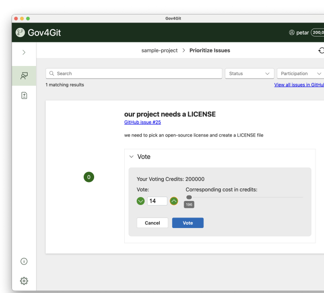
(a) Software by WeDesign, used in the first empirical QV research [18]. Little information is available about the software, except for an image from [20]. In the image, each prompt has thumbs up and down icons to update the vote in the center. The remaining budget appears as a progress bar at the top.



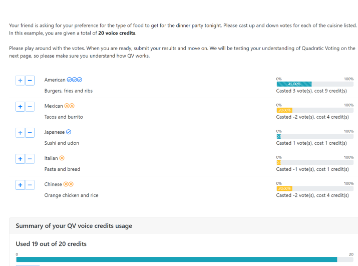
(b) An open-sourced QV interface [48] forked from GitCoin [49], used by the RadicalxChange community [50]. This interface presents total credits as small blocks. Votes are updated using plus and minus buttons, with numerical counts shown under each option and surface area as costs.



(c) An open-source QV interface [51] 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 [7] 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.



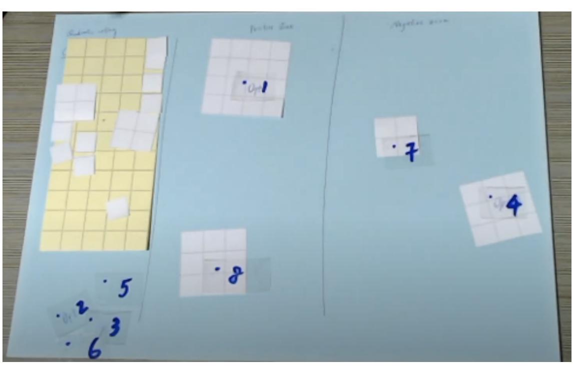
(e) The interface used in the research by Cheng et al. [4] employs the most visual components. Icons depict the current number of votes, with progress bars signifying the current spending.

Fig. 1. Recent interface implementations for applications using quadratic mechanism.

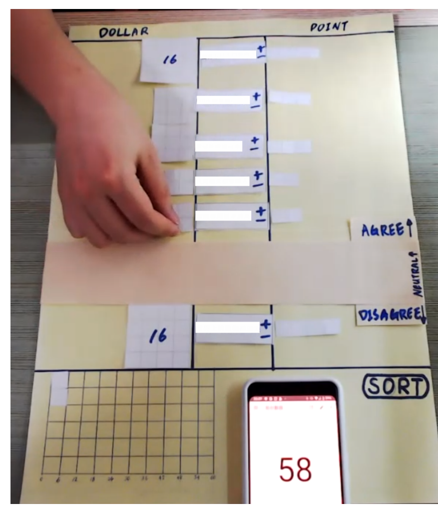
Existing interfaces consist standardized components shown in Figure 1 [1]. All interfaces shared these key components:

- Option list: A list of options contesting for votes.
- Vote controls: Buttons to increase and decrease votes associated with each option.
- Individual vote tally: A representation of votes associated with an option.
- Summary: An auto-generated summary of costs and remaining budget.

Most components were focused on presenting information and obtaining votes, only the summary addressed budget constraints. During our initial paper prototyping iterations (Figure 2), we tested different presentations of these components with added features to assist information organization. Pretest respondents felt QS was challenging due to identifying relative preferences among options and deciding the degree of trade-offs between options. In this study, we focus on the first challenge to inform our interface design iterations.



(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. 2. Initial paper prototypes designed for QS interface

3.1.1 Prototype 1: Ranking-Vote. Considering that relative preference is often through ranking items, we tested whether ranking options before voting would help establish 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 were draggable until shown. This insight indicates that full ranking is unnecessary for establishing relative preferences. Therefore, we decided to ask respondents select a subset of options instead of requiring a full rank among all options.

[1]Figure 1d did not exist until the writing of this paper.

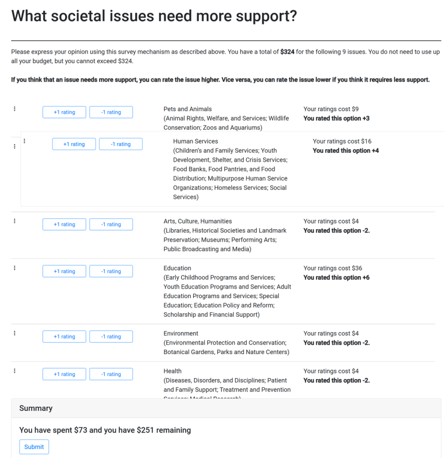
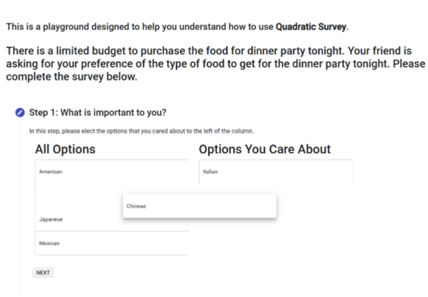


Fig. 3. A Ranking-Vote Prototype: The goal of this prototype is to test whether ranking options prior to voting help establish individual's relative preferences, and reduce effort when voting. Each option is draggable to position in a specific location amongst the full list of options. Votes can be updated 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.

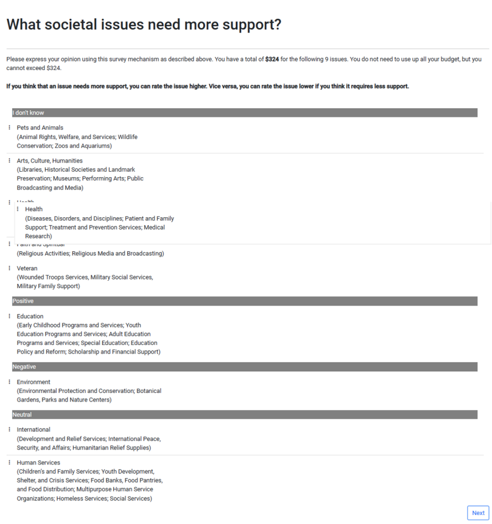


(a) Options are dragged and dropped to the 'Option You Care About' box.

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

Fig. 4. 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. 4a, the first step involves selecting options for further consideration. Important options are placed at the top of the list for voting shown in Fig. 4b, but options can be placed anywhere on the list if desired. The rest of the controls remain the same as the previous prototype.

3.1.2 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 4, survey respondents selected their preferred options (Figure 4a), and the interface positioned these options at the top of the list for voting (Figure 4b). We identified several issues during 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.



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

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

Fig. 5. Organize-then-Vote Prototype: The goal of this prototype is to encourage participants at deriving finer grain categories among options before voting. Survey respondents first organize their thoughts into categories, then vote on the options.

3.1.3 Prototype 3: Organize-then-Vote. Figure 5 shows the last prototype where we built on the previous takeaway by providing finer-grain groupings and creating a clear connection between option organization and voting position. Specifically, we provided three categories: Positive, Negative, and Neutral. Initially, respondents see all options under the section labeled 'I don't know,' which includes only the option descriptions. We ask respondents to move these options into the categories below. Voting controls and information appear on each option once respondents move to the subsequent page, forming a clear connection between option groups, 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: 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.

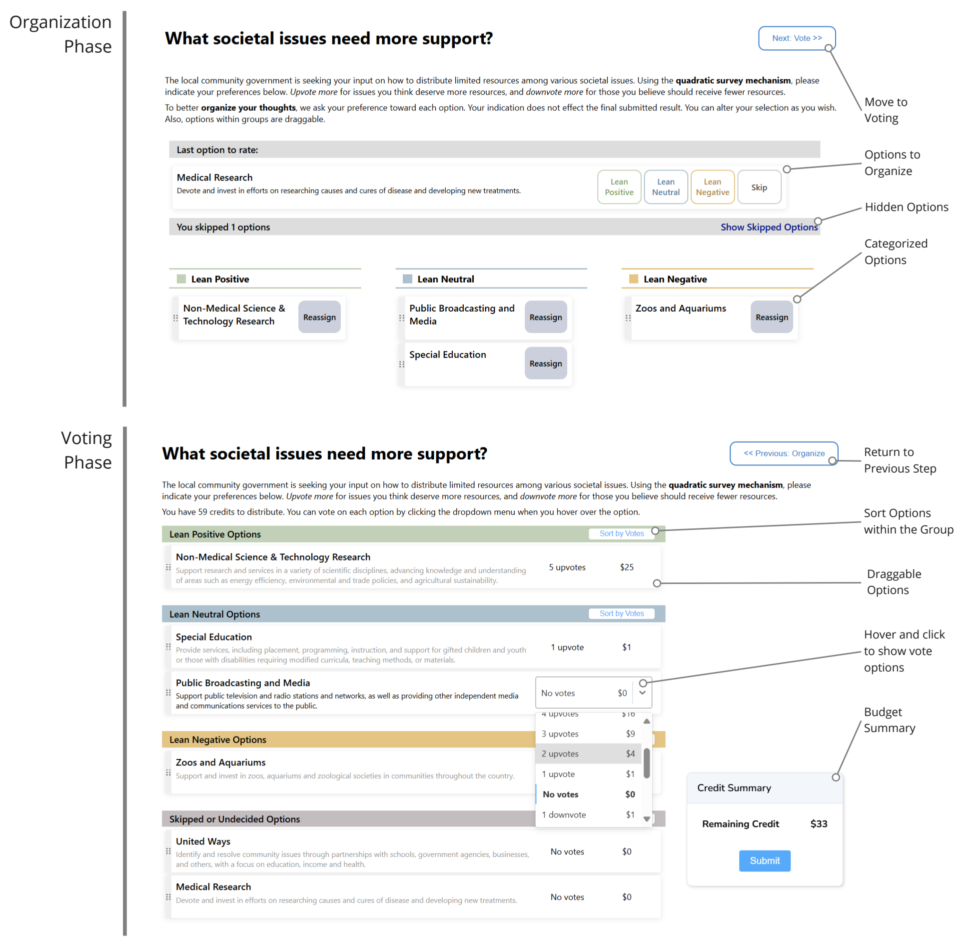


Fig. 6. The Two-Phase Interface: The interface consists of two phases. Survey respondents can navigate between phases using the top right button. In the organization phase, respondents will be presented with one option at a time where they can choose among four choices: Lean Positive, Lean Neutral, Lean Negative, or Skip. Skipped options are hidden and can be evaluated later. Options that are chosen will be listed below. Items can be dragged and dropped across categories or returned to the stack. In the voting phase, options are listed in the order of the four categories. When hovering over each option, respondents can select a vote for that option using the dropdown. Each dropdown contains the cost associated with the vote. A sort button allows ascending sorting within each category. A summary box tracks the remaining credit balance.

3.2 Finalizing the Two-Phase Interface

In the previous subsection, we highlighted critical prototype iterations that informed the final two-phase interactive process that defines the user journey. We now present the final two-phase interface, its operations, and the supporting literature for comprehensive understanding. Then, We also discuss the aesthetic design choices that emerged throughout the iterations.

3.2.1 Justifying a two-phase approach. Recall the ultimate objective of the two-phase interface is to facilitate preference construction and reduce cognitive load. The two-phase interface, shown in Figure 6 consists of two steps: An organization phase and a voting phase. Throughout both phases, survey respondents can drag-and-drop options across the presented option list.

A two-phase approach. If preferences are constructed, by nature, they consist of a series of constructed decision-making processes [14]. Two major decision-making theories informed the design decision of a two-step interaction interface design: Montgomery [52]'s Search for a Dominance Structure Theory (Dominance Theory) and Svenson [53]'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 [52]. 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 [53]. Both theories echoed the design decision in building the interactive experience to reduce initial decision dimensions and the mental procedures involved in emphasizing relatively important options and forming decisions.

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

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

The organizing interface, depicted on the top half of Figure 6, sequentially presented each survey option. Participants selected a response among three ordinal categories – lean positive, lean negative, or lean neutral. Once selected, the system moved that option to the respective category. Participants could skip the option if they did not want to indicate a preference. Options within the groups were draggable and rearrangeable to other groups should the participants wish.

Strack and Martin [54]'s research showed that upon understanding a survey question, respondents either recalled a prior judgment or constructed a new one when completing an attitude survey. In addition, revealing one option at a time gated the amount of information presented to the survey respondent and thereby reduced the extraneous load [55]. This process allowed participants to form or express opinions on individual options incrementally. This design also mitigated the original concern from prototype 3 where participants accidentally treated the organizing task as a ranking task.

The three possible options, positive, neutral, and negative, aimed to scaffold participants in constructing their own choice architecture [56, 57], which strategically segmented options into diverse and alternative choice presentations while avoiding the biases from defaults. We believed that these three categories were sufficient for participants to segment the options. However, we chose not to limit the number of options one could place into a category to prevent restricting user agency, a core user interface design principle [58].

Immediate feedback displaying the placement of options and allowing participants to rearrange them via drag-and-drop adhered to key interface design principles [58]. At the same time, it allowed finer grain control for individuals to surface dominating options and create differentiating groups of options.

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

Options were displayed as they were categorized within each category from the previous step and in the following section orders – lean positive, lean neutral, lean negative, and skipped or undecided as detailed on the bottom half of Figure 6. The Skipped or Undecided category contained options left in the organization queue, possibly because survey respondents had a pre-existing preference or chose not to organize their thoughts further. The original order within these categories was preserved to maintain and reinforce the differentiated options. This new ordering sequence mitigated the concern from prototype 3 where options without a category are left at the top of the voting interface. Respondents had the flexibility to return to the organization interface at any point during the survey to revise their choices.

In the voting interface, options remained draggable, enabling participants to modify or reinforce their preference decisions as needed. Each category featured a sort-by-vote function that enabled reordering within the same category. Although these interactions did not influence the final voting outcome, they were designed to support consolidation and positional proximity in information organization. This design aimed to automate the grouping of similar options while providing an intuitive drag-and-drop mechanism, thereby facilitating decision-making by placing similar options near each other. This echoed the principles of the proximity compatibility principle, particularly emphasizing spatial proximity and mental compatibility [59]. The interface design anticipated that participants would find it easier to consolidate their choices when similar options were positioned close together, thereby reducing cognitive load.

While multiple interaction mechanisms exist, drag-and-drop has been extensively explored in rank-based surveys. For instance, Krosnick et al. [60] 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 [61] 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 [62].

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.

3.2.2 Aesthetic Design Decisions. There are three aesthetic design decisions that made it to the final interface. First, we decided to remove all visual elements. Prior literature suggested that the use of emojis might influence the interpretations of surveys [63] and decrease user satisfaction [11]. Prior literature also noted that not all data visualization elements reduce cognitive demand [64]. Even though effective visualization can aid decision-making, it remains an open question that this study does not aim to address, thus we also removed all visualization elements such as blocks, progress bars, and percentage indicators.

Second, the final interface has all options presented on the screen at the same time, intentionally. Unlike all the prototypes and existing interfaces, prior literature emphasized the importance of placing all the options on the same digital ballot screen to avoid losing votes. This echoes the proverb "out of sight, out of mind," where individuals might be biased toward options that are shown to them, and additional effort is required for individuals to retrieve specific information if options are hidden.

Last, we decided to use a dropdown positioned to the right of the option such that control of votes and the budget summary are placed near one another. The layout of the votes and cost was inspired by online shopping cart checkout interfaces where quantities are supplied next to the itemized costs followed by the total checkout amount. We chose a dropdown after iterating with two alternative input methods (Figure 7): the original click-based buttons and a wheel-based implementation. The former design requires survey respondents to click multiple times to reach their desired vote values. Thus, we wanted to look for a solution to aid respondents in reaching their intended value faster. A wheel-based approach allows intuitive control of the votes by using the wheel on the mouse and clicking to fine-tune the values. However, in our early pilot studies, not all participants were familiar with wheel control, thus we opted for a dropdown menu for vote selection.

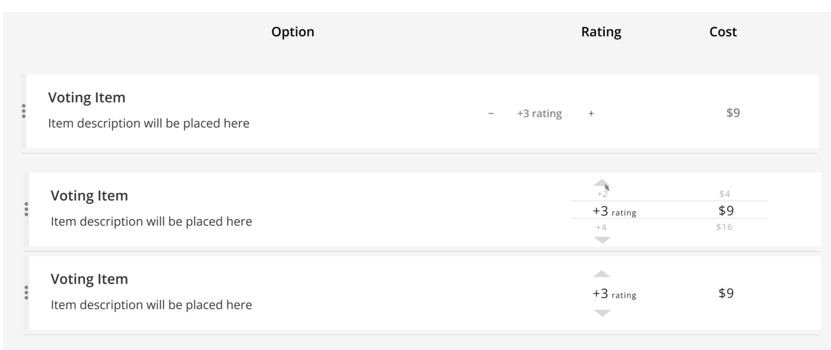


Fig. 7. The click-based and wheel-based designs for vote control included: The former mirrors traditional vote control used in other quadratic voting interfaces, where each click represents an increase in votes. The latter allows control through both clicks and mouse wheel rotation.

3.3 Text-based Interface

To study how the interactive components influenced participants' cognitive load and behavior, we removed the two-phase interactive design and the drag-and-drop features for the text-based interface. The text-based interface shares the other functionalities of the two-phase interface, as shown in Figure 8. The interface contained the question prompt at the top of the screen. The options were presented in a list underneath the prompt. Survey respondents could update the votes by selecting from a dropdown that provided all possible voting options and costs given the number of credits available. A small summary box to the right of the interface showed the current total cost and the remaining credits for the respondent. The interface randomly presented options to avoid ordering bias [65, 66].

Both experiment interfaces are developed with a React.js frontend and a Next.js backend powered on MongoDB. Both services were open-sourced [2].

[2]link-to-github

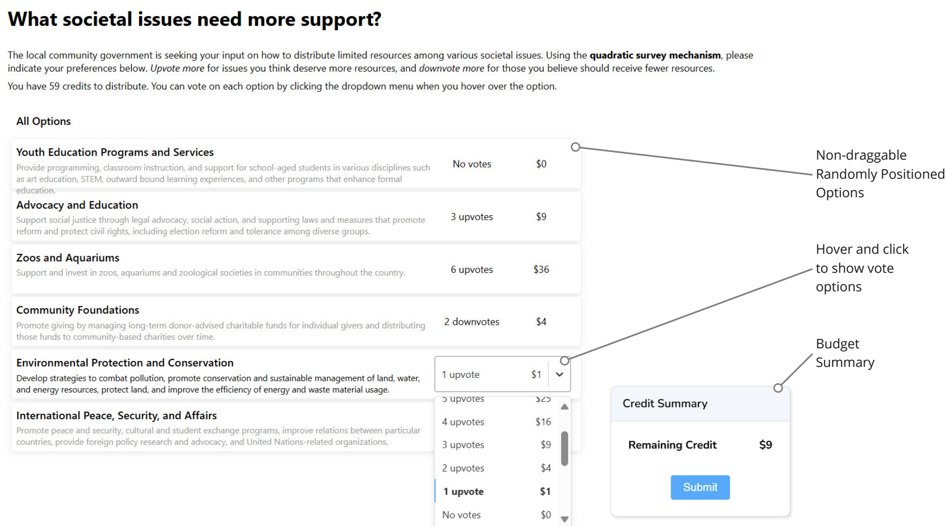


Fig. 8. The text-based interface: This interface is based on the interactive version but does not include the two-phase interactive support and lacks the drag-and-drop functionality. Options are randomly positioned.

4 Experiment Design

In this section, we detail the overall experiment design and explain the reasoning behind specific experiment decisions.

4.1 Experiment Protocol

We recruited participants from a United States college town using various methods: online ads, digital bulletins, social media posts, physical flyers, and online newsletters. To ensure diversity, we prioritized non-students by placing physical flyers in public spaces like restaurants, cafes, and libraries beyond campus. As we monitored respondent demographics, we began selectively accepting non-student participants. If a respondent self-identified as a student, we thanked them and informed them of our current priority for non-students, though some self-identified students were still accepted. To prevent response bias, we framed the study as focusing on attitudes toward societal issues rather than measuring cognitive load and behaviors. The college Institutional Review Board reviewed and approved this study.

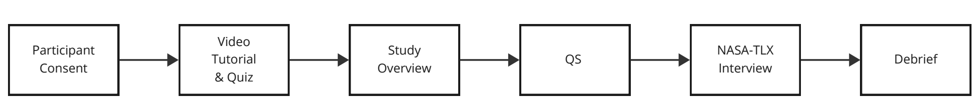


Fig. 9. Study protocol: Participants are asked to learn about the mechanism of QS after consenting to the study. The researcher explains the study overview and asked participants to complete QS. A NASA-TLX survey followed by interviews to understand participant's cognitive load. A debrief happens before ending the study.

Figure 9 visually represents the study protocol. Participants were invited to the lab to minimize external influences on cognitive load measurement and ensure consistency between sessions. The experiment involved participants using a 32-inch vertical monitor, ensuring all options on a QS were visible to minimize hidden information during decision-making. After consenting to the study, participants watched a pre-recorded video explaining the Quadratic mechanism and how QS operates without hints of interface operation. Participants then completed a short quiz to ensure their understanding of the QS mechanism. Those who failed to answer all questions correctly were asked to rewatch the video or consult 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:

- 6 options with a text-based interface
- 6 options with an two-phase interface
- 24 options with a text-based interface
- 24 options with an two-phase interface

Participants completed the survey independently, without the researcher's presence. Upon completion, they contacted the researcher. The researcher distributed the NASA-TLX cognitive load measurement and a short semi-structured interview about their experiences. This interview was audio recorded. The session concluded with a debriefing and a $15 cash compensation. During the debriefing, participants were informed that the survey's true purpose was to measure cognitive load and interface design, and they could ask any questions.

4.2 Experiment Design Choices

This subsection details the justifications behind each experiment design decision. We made four major decisions:

4.2.1 A between subject study. We chose a between-subject study design for two main reasons. First, to minimize study fatigue given the complexity of responding to a QS survey, which could take up to 20 minutes. Conducting back-to-back experiments to measure cognitive load would be impractical, and asking participants to revisit the lab after several days would likely increase dropout rates and demotivate participants from attending in-person sessions. Second, we aimed to reduce the learning effect, which is challenging to eliminate, especially concerning interface operation and decision-making in the survey. Since preferences are constructed, we wanted to ensure that participants were not influenced by their previous preferences, which could affect their perceived cognitive load.

4.2.2 Deciding number of survey options. Ideally, understanding participants' cognitive load across multiple options would require enumerating all possible numbers of options to identify the "breaking point" where cognitive overload occurs. However, this approach is not feasible due to time and resource constraints. Therefore, we referred to prior literature to inform our choice of 6 and 24 options, representing short and long lists. Constant sum surveys and the Analytic Hierarchy Process (AHP) recommend fewer than ten and seven options, respectively [67, 68, 69]. Some refer to the value 7 based on human's cognitive processing capacity of 7 ± 2 [70] and a theoretical proof using the consistency ratio of a pairwise comparison metric [71]. A meta-analysis by Chernev et al. [72] surveyed 99 choice overload experiments and summarized that 6 and 24 are the modal values for short and long lists when testing choice overload. These two values were likely rooted in the original choice overload experiment[3] by Iyengar and Lepper [43]. Thus, we adopted these values to align with established research.

4.2.3 Context of the Study. Participants completed a societal issue survey, following the methodology described by Cheng et al. [4]. Surveying societal issues is relevant to every citizen, and easily conveys the concept of limited public sector resources that need prioritization across different areas. We curated a list of societal issues directly from the categories used by Charity Navigator [73], a non-profit organization that evaluates over 20 thousand charities in the United States. Participants across all four groups were presented with options randomly drawn from 26 societal issues. The full list of these societal issues is provided in Appendix A.

4.2.4 Using NASA-TLX to measure cognitive load. Several methods exist to measure cognitive load, including performance measures, psychophysiological measures, subjective measures, and analytical measures [74]. Given the nature of QS, a task requiring a long period, performance measures like secondary-task measures were challenging to implement due to the difficulty in designing a suitable secondary task. Psychophysiological measures such as pupil size [75] and ECG [76] are highly sensitive to external environments and costly to obtain. Consequently, we relied on subjective measures via self-report surveys and analytical measures like time and clicks collected via the interface. We adopted the paper-based weighted NASA Task Load Index (NASA TLX), a multidimensional scoring procedure using the weighted average of six subscale scores to represent the overall workload. This process requires subjects to evaluate each weight's contribution to the workload of a specific task [77, 78, 79] which reduced between-rater variability [79]. Despite criticisms regarding its validity and vulnerability, NASA-TLX is commonly used due to its low cost and ease of administration [74]. It has been tested on various experimental and lab tasks, showing significantly less variability among evaluators than one-dimensional workload scores [80]. Thus, we chose NASA-TLX to measure cognitive load in our study.

5 Cognitive Load and Sources across Experiment Conditions

In this section, we present the cognitive load across experiment groups and the sources contributing to each cognitive load dimension. Given the limited number of participants, we focus on descriptive statistics and qualitative assessments of cognitive load. Quantitative data includes metrics from the survey tasks, while qualitative insights come from post-survey interviews transcribed and analyzed by the first author.

To analyze the qualitative data, the first author conducted an inductive thematic analysis process [81]. They coded snippets from each transcript based on specific research questions and topics of interest for the qualitative analysis. Similar codes were merged within each research question or topic to form relevant themes. When differences were hypothesized, they applied a deductive coding process to text snippets related to a specific research question or topic of interest.

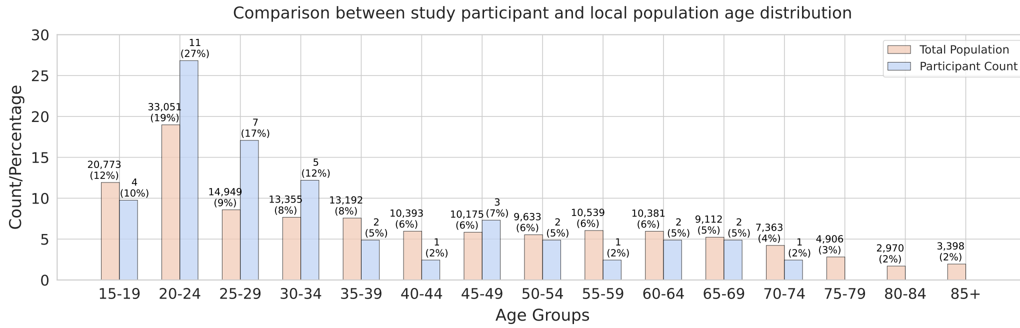
The results for this section are organized as follows: We start with participant demographics and then provide an overview of our cognitive load findings. We then examine the six dimensions used in the NASA-TLX survey: mental demand, physical demand, temporal demand, performance, effort, and frustration.

5.1 Demographics

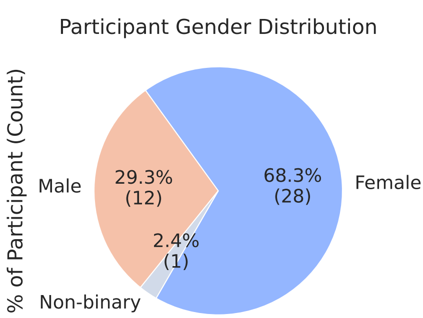
We recruited a total of 41 participants, allocating ten to each experiment condition. Due to data quality concerns, we excluded one participant's data[4]. The mean age of the participants was 34.63

[3]We believe that the original value decision was due to the limitations of the jam flavors.

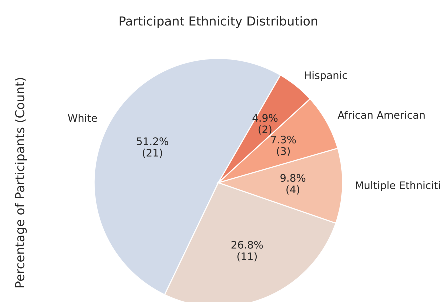
[4]The participant reported that they did not complete the survey seriously since they think the experiment is fake.



(a) Age distribution of the participants skewed slightly younger despite following similar trends as the county population.



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



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

Fig. 10. Demographic distributions: Age, Gender, and Ethnicity

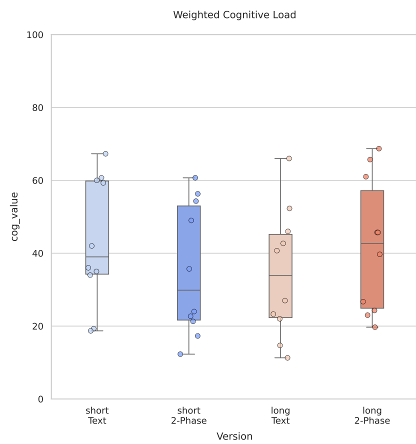
years old, with a detailed age distribution presented alongside the county population distribution in Figure 10a. This comparison reveals that our sample closely matches the county's demographic profile, albeit with a slightly higher representation of younger adults, particularly in the 35-45 age range. As shown in Figure 10b, the majority of participants skewed toward females.

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

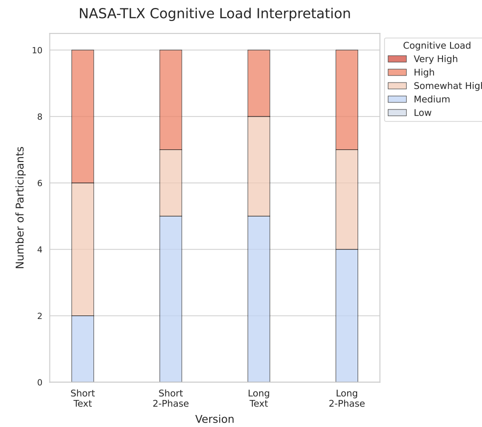
5.2 Overall Cognitive Load

To answer **RQ1** and **RQ2a**, we derive the weighted NASA-TLX scores across the four experiment conditions. We show these results in Figure 11. Weighted NASA-TLX uses a continuous 0-100 score, with higher values indicating greater cognitive load. We use predefined mappings of NASA-TLX values to cognitive levels: low, medium, somewhat high, high, and very high, as listed by Hart and Staveland [77]. We show value interpretations in Figure 11b.

Surprisingly, cognitive load scores were lower for the long interface than the short interface. Further, the two-phase interface decreased cognitive load compared to the text interface for the



(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, 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. Standard deviation are 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. 11. This figure shows the box plot results for weighted NASA-TLX scores across experiment groups and participant counts based on individual score interpretations. In 11a, 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 11b, these trends are clearer when NASA-TLX scores are grouped into five tiers.

short survey but increased it for the long survey. This is evident from the median cognitive load decrease from 39.00 to 29.85, with more participants reporting lower cognitive load using the two-phase interface. The short text interface had the most participants ($N = 8$) rating their cognitive load as somewhat high or above. The other three conditions had similar distributions, with about half experiencing medium and half somewhat high or high loads.

We acknowledge the possibility that the elicited values are pure noise and do not reflect the actual cognitive load. This could be due to the small sample size, the nature of the task, or the participants' understanding of the cognitive load scale. However, to dig deeper into this phenomenon, we turn to qualitative insights from the post task interviews.

5.3 Sources of Mental Demand

Key Differences: First, slightly more participants using the text interface reported mental demand from precisely determining the number of votes for options compared to the two-phase interface. Second, when it comes to long QS, participants using the long two-phase interface considered broader societal impacts and evaluated options holistically, while those in the long text interface focused on personal relevance and individual issues.

Mental demand refers to the degree of mental and perceptual activity required to complete a task. Interview results showed primary drivers of participants' mental demand were *Budget management* and *Preference construction* [5].

5.3.1 Mental Demand Source #1: Budget management. 14 participants expressed demand from budgeting within limited credit (S032 *[...] in order to like show your support 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 just wanted to make sure that I used all the credit that I had available to me* , $N = 8$).

We further categorized budget management-induced mental demand as either operational (corresponding to a single interface-level action, e.g., using the last remaining credit) or strategic (achieving a higher-level goal, e.g., evenly distributing credits across options). We noticed that long survey participants tended to report operational causes of mental demand rather than strategic ones, which may indicate that the larger number of survey options induced more short-term thinking.

5.3.2 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 *[...] was 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 precise resource allocation (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 the theory that preference construction is a difficult and mentally demanding task. Notably, more participants using the text interface reported mental demand from precise resource allocation compared to the two-phase interface (18 vs. 12). We conjecture that the first pass on the survey items in the organization phase helped participants reduce their mental demands in this area once they got around to vote allocation.

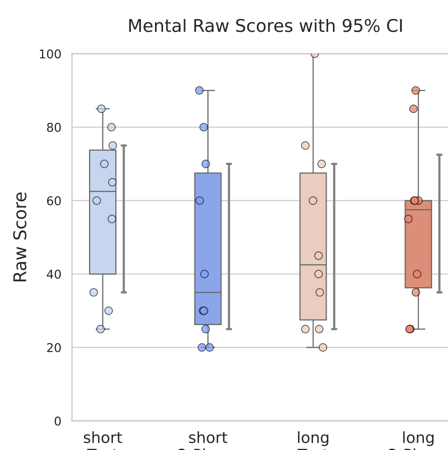


Fig. 12. Mental Demand Raw Score: Across all four experiment groups, participants' reported mental demand is spread across a wide range with many participants experiencing high mental demand.

[5] The full table is in Appendix B.1.

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 [...] I kind of went back and forth between those two. [...] So 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)`

Inspecting both causes, while we did not notice significant differences in mental demand raw values (Figure 12) across the four experiment groups, especially between the interfaces across long QS, they exhibited different sources contributing to their mental demand.

5.4 Sources of Physical Demand

> **Key Differences:** Two-phase interface experienced higher physical demand from increased mouse usage.

Physical demand refers to the physical effort required to complete a task, such as physical exertion or movement. Since this study involves participants sitting in front of a computer screen completing a survey, most participants reported minimal physical demand($N = 32$). We nonetheless report the sources of this minimal demand, which include reading text on the screen ($N = 4$), using the mouse ($N = 16$), and moving their head to navigate the vertical screen ($N = 5$) [6]. Participants emphasized that these demands were minimal, which is reflected in the low values reported in the NASA-TLX physical demand scores (Figure 13) 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 13), 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 1. 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

[6]The full table showing the distribution is in Appendix B.2

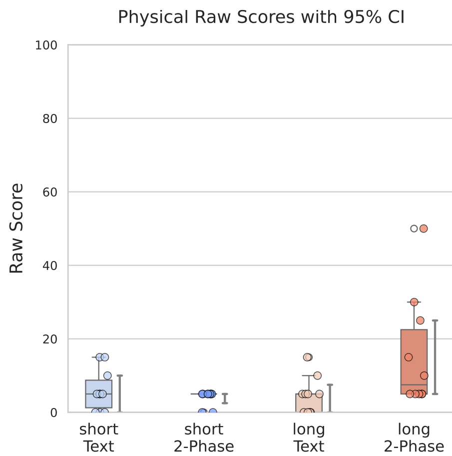


Fig. 13. Physical Demand Raw Score: Participants other than the long two-phase interface reported minimal physical demand. The long two-phase interface had the highest physical demand, likely due to increased mouse clicks and extended time spent looking at the vertical screen.

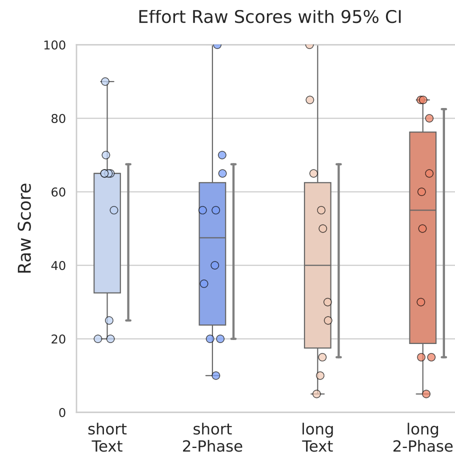


Fig. 14. Effort Raw Score: Effort scores shows indifference across groups.

5.5 Source of Effort

key Differences: First, participants in the text interface associated effort with operational tasks more often than participants from the two-phase interface. Conversely, participants in the two-phase interface cited more sources from strategic planning than those in the text interface. We observed that participants experienced effort when considering a comprehensive view while using the two-phase interface.

Effort refers to how hard participants felt they worked to achieve the level of performance they did. Because effort includes the intensity of both mental and physical resources expended during the task, we refer readers to section 5.3 and section 5.4 to understand the definitions of these sources.

While the raw NASA-TLX effort scores (Figure 14) 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 5.3).

5.5.1 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 ($N = 2$).

5.5.2 Effort Source #2: Strategic Planning. Different from Operational Tasks, 11 participants in the text interface compared to 17 participants describing 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 5.3. Participants using the two-phase interface ($N = 7$) had nearly mentioned double ($N = 4$) times regarding global strategies.

5.6 Sources of Temporal Demand

> **Key Differences:** First, participants using the short text interface wanted to complete the task quickly and reported the highest temporal demand. The two-phase interface lowered the temporal demand for short QS. Second, participants using the long text interface showed the lowest temporal demand score. The two-phase interface increased the temporal demand for long QS.

Temporal demand measures the time pressure participants feel during a task. Lower demand indicates participants felt comfortable taking a more leisurely pace. We categorize the main sources of increased temporal demand as the 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). *Budget* also came up occasionally (S034 *as the money decreases I felt kind of rushed* , $N = 4$).

5.6.1 Two-phase Interface Reduced Temporal Demand on Short QS. The raw NASA-TLX values in Fig 15 show that participants in the Short Text Interface condition reported the highest temporal demand among all. They tended to frame their concerns over time spent on making decisions negatively. Five participants thought the time and effort they had invested were more than expected and felt a need to rush (S024 *maybe I should just hurry up and make a decision.*). The two-phase interface reduced this feeling of needing to rush – only one participant who used our two-phase interface in the short survey expressed similar temporal concerns.

5.6.2 Long QS on Text Interface Showed the Lowest Temporal Demand. Based on the raw NASA-TLX values in Fig 15, participants in the long text interface exhibited the lowest temporal demand, explaining why this group had the lowest overall cognitive load as discussed in Section 5.2. This is counter-intuitive since participants in this condition made more decisions and operations compared to the short text group. There are two potential explanations. First, we noticed that more participants who experienced a short survey expressed a desire to complete the task efficiently ($N = 7$) than those in the long survey groups ($N = 1$). They often expressed things like:

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

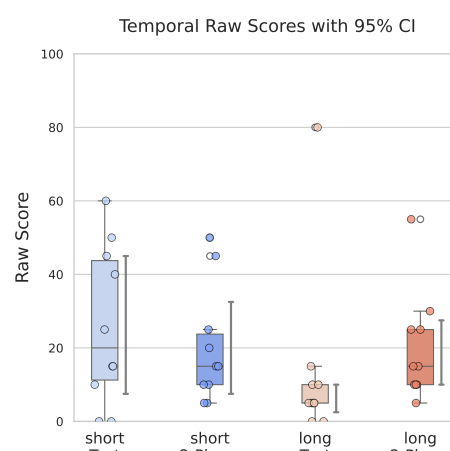


Fig. 15. Temporal Demand Raw Score: The short text interface results in the highest temporal demand, while the long text interface is the lowest. Two-phase interfaces, show moderate temporal demand, suggesting that interactive elements allowed participants to pace themselves better.

A second possible explanation for the unexpectedly low temporal demand in the Long Text condition was the participants' satisficing behaviors.

> *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)

Due to cognitive overload from the long list of options, participants may have spent less time than they expected on the decisions. We will discuss this possibility in-depth in Section 7.1.

5.6.3 Two-phase Interface Increased Temporal Demand in Long QS. Despite the unexpectedly low temporal demand in long QS with a text interface, completing the long QS with our two-phase interface raised the temporal demand to the level of short QS groups. 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 was 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.

5.7 Source of Frustration

> **Key differences:** We observed evidence that participants in the long text interface showed the least amount of frustration from operational causes compared to other experiment conditions.

Frustration refers to the extent to which the participant is annoyed, irritated, or discouraged during the task. Sources of frustration were generally related to either *Operational Actions* like credit management ($N = 6$) and dealing with the quadratic vote costs ($N = 5$), or *Societal Concerns*, like wishing they did not have to make tradeoffs ($N = 8$) or feeling pessimistic about how others will vote ($N = 6$). We provide the full table in Appendix B.5

In general, the frustration derived from societal concerns did not seem strongly affected by any of the experimental conditions. With respect to operational action-driven frustration, however, we saw some discrepancies. 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 16). 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. Participants engaging in satisficing behaviors in the long text interface may explain this phenomenon – prior literature [82, 83] indicates that satisficers tend to be less frustrated and happier than maximizers.

5.8 Source of Performance

> **Key Differences:** 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

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 arose from the interviews: *Operational Actions* and *Social Responsibility* [7]. Despite most participants reporting positively on their performance, nuances exist in how different groups interpret their performance.

5.8.1 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

[7] The full performance table is at Appendix B.3

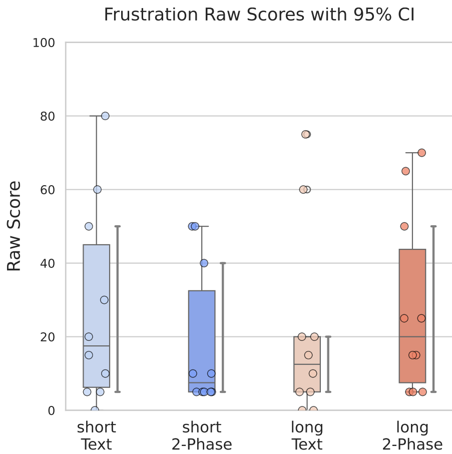


Fig. 16. Fustration Raw Score: Participants other than the long text interface highlighted seversl operational tasks that led to fustration. All groups share causes from strategic planning.

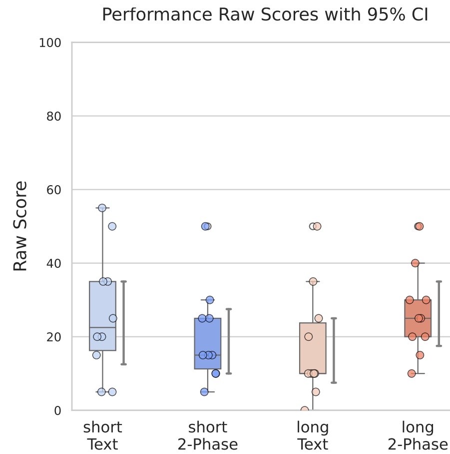


Fig. 17. Performance Demand Raw Score: Participants showed indifferent performance raw scores across experiment conditions, all trending toward satisfactory.

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 ($N = 6$).

5.8.2 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 17) show participants had indistinguishable performance scores. This aligns with the interview results where most participants felt positive about their final submission.

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 ($N = 11$ vs $N = 6$).

5.9 Summary across all cognitive load dimensions

In this subsection, we gathered differences across dimensions and synthesized them into a list.

- **Mental Demand:** Participants using the text interface reported higher mental demand from determining the number of votes for options. Those using the long two-phase interface considered broader societal impacts and evaluated options holistically.
- **Physical Demand:** Physical demand was higher for participants using the two-phase interface due to increased mouse usage.

- **Effort:** Effort sources varied, with text interface participants focusing more on operational tasks, and two-phase interface participants engaging more in strategic planning, reflecting deeper, more comprehensive consideration of options.
- **Temporal Demand:** Temporal demand was highest in the short text interface, where participants aimed to complete tasks quickly, while the long text interface showed the lowest temporal demand.
- **Frustration:** Frustration levels from operational causes were lowest in the long text interface.
- **Performance:** Participants using the two-phase interface felt more positive about their performance, being twice as likely to report "feeling good" about their results compared to the text interface users.

Overall, participants using the two-phase interface tend to think more comprehensively and critically, while those using the text interface focus more narrowly on operational tasks. In addition, we suspect that participants who completed the long QS on a text interface engaged in satisficing behaviors based on the counter-intuitive results that they had the lowest temporal demand and frustration level. We will interpret these results in the discussion section. To better understand participants' behavior, we analyze click-stream data across experiment conditions in the next section.

6 Behavior Results

To answer RQ3, we investigate time-to-action and remaining credit differences across experiment conditions. Time-to-action is a widely used metric in decision sciences to understand individual behaviors. For example, Payne et al. [84] theorized that longer decision time represents more complex and deeper cognitive processing. 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. [85] and [86] showed how scarcity influences decisions, increases risk aversion, and adds cognitive load. In this section, we use these two measures as proxies to understand participant behaviors. We publicly shared all analyzed participant interaction data[8] to support transparency and facilitate further research.

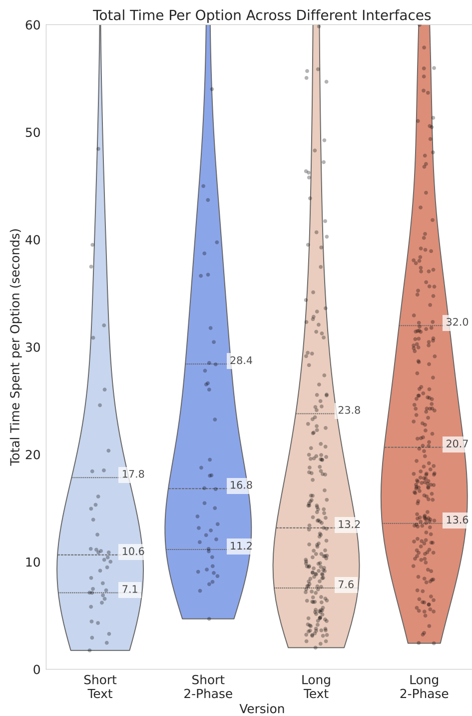
6.1 Time Spent per Options

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 can derive the following detailed logs of participant actions: *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.

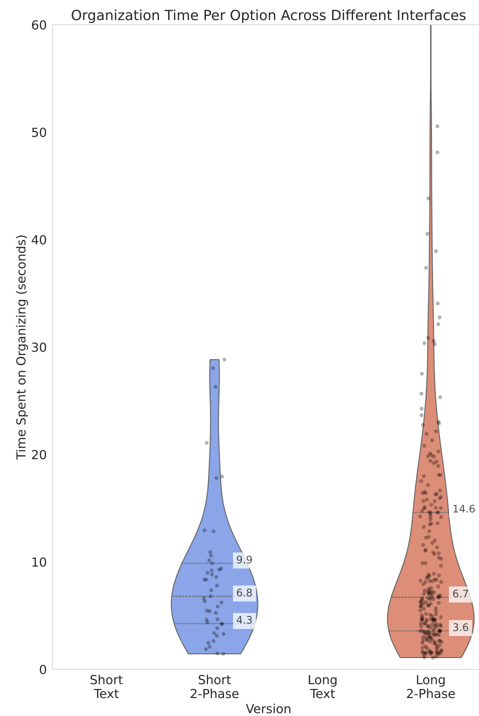
We aggregate all the time spent on each option as the total time spent for that option. Organization time includes the time participants spent placing options into preference categories and the drag-and-drop time associated with each option during the organization phase. Voting time strictly refers to the time participants took to update vote values for each option. To minimize noise, we intentionally drop 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.

Figure 18 each dot represents one option for one participant. Figure 18a shows total time, figure 18b shows organization time, and figure 18c shows voting time. The violin plot shows the distribution of the dots and the three horizontal lines represent the median, 25th percentile, and

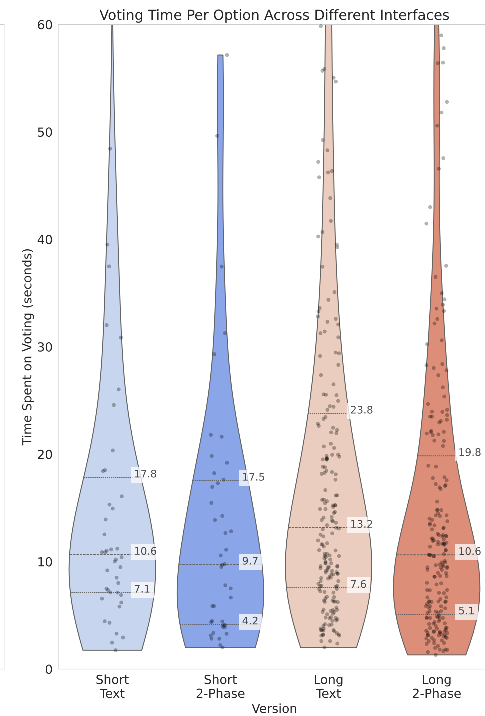
[8]link-to-github



(a) Total Time per option: We identified that two-phase interface skewed slightly higher than the text interface. This is reasonable since participants in these two conditions requires additional time to complete the organizing steps.



(b) Organization Time per option: Since only two-phase interfaces has an organization phase, the other two experiment groups does not have any accumulated time.



(c) Voting Time per option: We see a slight downward trend between the short QS, and a statistical faster voting time for the long two-phase interface.

Fig. 18. Time per option across all experiment conditions. In each of these violin plots, a dot represents the total time a participant spent on that option. Fig 18a one dot represents the total time a person spent for one options. Figure 18b and Figure 18c are decomposition of the total time.

75th percentile of the time spent for that interface. We limited the y-axis to 1 minute to improve visualization clarity.

Participants spent slightly more time per option on the two-phase interface than the text interface in both short and long surveys. 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 break down the total time spent into organization time and voting time in Figure 18b and Figure 18c.

We observed minimal difference in organization time (Figure 18b) between short and long surveys. The interface was designed with this in mind, given that options are shown one at a time, and participants can drag and drop them into the preference categories when needed. In terms of the voting time (Figure 18c), 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

To further analyze participant behaviors, we break down the aggregated time from the previous analysis and examine fine-grain interactions. Specifically, we examine if there are differences among behavior across interfaces. As we outlined, credit scarcity might influence decision making. Figure 19 plots the time of voting actions over the remainder of the participant's budget across the text and two-phase interface across all four groups. Each bar shows the number of actions accumulated across participants at specific percentages of remaining credits. A KDE plot is provided to better visualize the trends. We chose not to follow Quarfoot et al. [18] focusing on the number of accumulated votes over an individual's time given that each individual's total time spent differ across experiment conditions.

Comparing experiment groups, we see fewer differences in the short QS but different interaction distributions between the two interfaces in the long QS. Given the significant differences in voting time between the text and two-phase interface for the long QS, we focus on deciphering the voting action changes between these two conditions in this subsection.

In Figure 19, 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. We further separated the behaviors where participants made large or small changes to the options, specifically for the long version. In Figure 20, we define an adjustment of four or more votes as large, which we plotted in the first row of the figure. Adjustments of two or fewer votes are considered small, which is 10% of the possible values one can choose among the maximum of 21 votes.

Instead of showing the number of actions, we plotted all actions against the time it took to make them. Revisiting the KDE curve in the second row in Figure 19 and the curve of the second row in Figure 20 which represents the small vote adjustment across both interfaces, we see a stronger bimodal action distribution. 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 towards 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

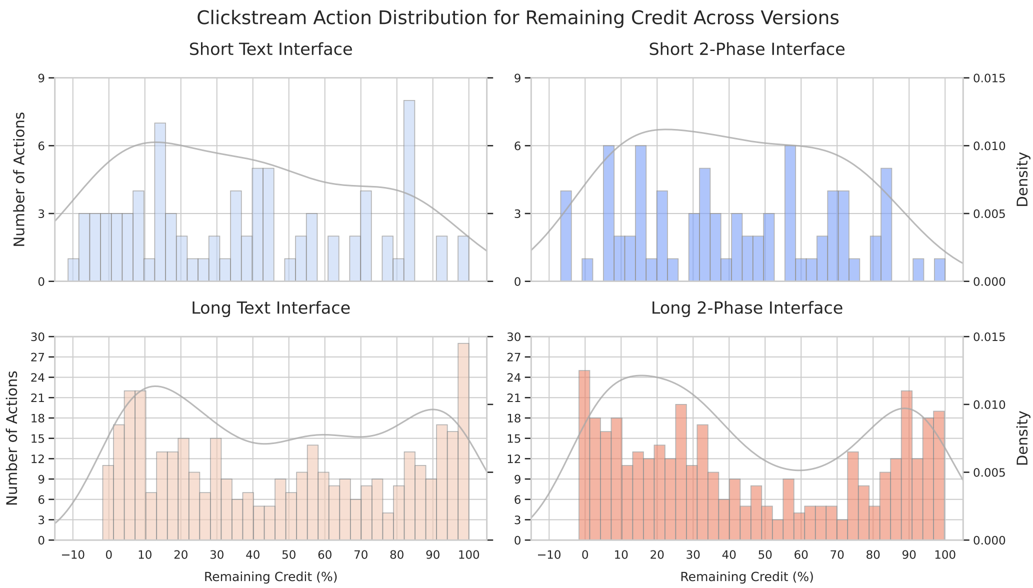


Fig. 19. This plot counts the number of voting actions when there are x percentage of credits remaining. A KDE plot is provided to help better understand the action distribution.

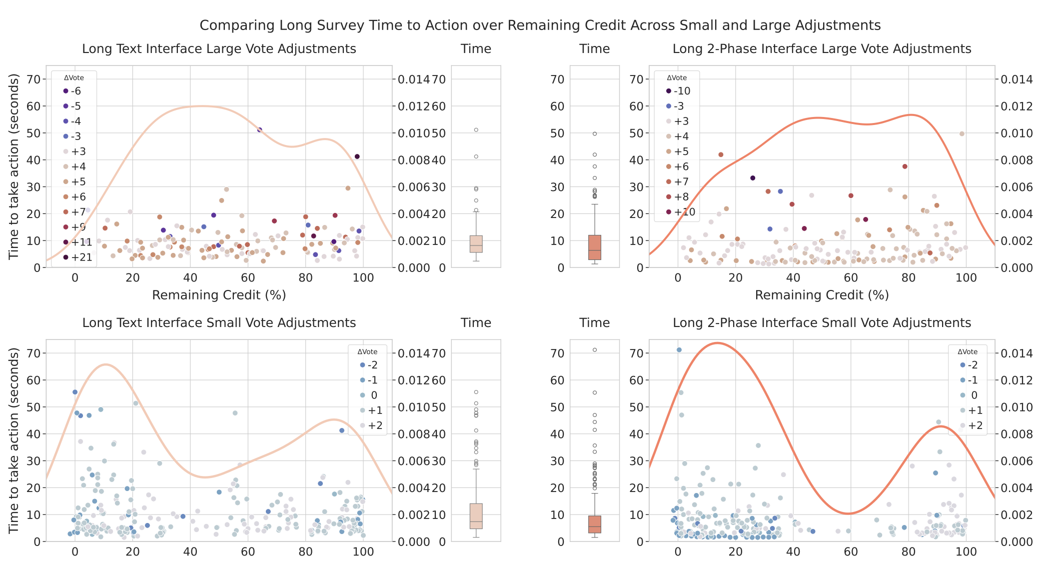


Fig. 20. This plot further separates participants interaction behavior based on the number of votes participants adjusted. We observed a bimodal interaction pattern across long QS when small vote adjustments are made.

effect size (Rank-biserial: 0.227, Cohen's d: 0.195) and a power of 0.381. Based on the KDE plots in the first row of Figure 20, 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.

In interviews, five participants highlighted the importance of the interface's flexibility and their use of an incremental, iterative approach. All these participants used the two-phase interface. While this doesn't mean participants using the text interface didn't take an iterative approach, it highlights that the two-phase interface encouraged iterative and incremental updates. As one participant pointed out:

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

To summarize, 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

This study proposes an two-phase interface for QS to better understand how the number of survey options and interface influences individuals' cognitive load and behaviors in the context of social resource allotment. Our results, presented in section 5 and 6, revealed surprising results that longer surveys did not increase cognitive load.

In the discussion section, we first interpreted the result across cognitive load and survey respondent behaviors. Then we explore three key topics: first, we explore specific two-phase interface elements that influenced participant behaviors and their relation to bounded rationality. Next, we examine how participants built their preferences and how direct manipulation within the interface supported them. Finally, we discuss how the interface design mitigated challenges of quadratic mechanisms and identifies remaining issues. We provide recommendations for using this tool and propose design improvements for future development.

We drew the conclusion that the two-phase interface prevented statisficing behaviors when participants experienced cognitive overload. In addition, participants using the two-phase interface demonstrated more strategic planning and holistic thinking compared to those using the text-based interface, who focused more on operational tasks. Additionally, the behavioral analysis showed that participants using the long two-phase interface made more frequent, small, iterative updates, indicating a shift in their cognitive load focus.

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

We start by referencing Figure a to examine participants' cognitive scores. Participants using the long two-phase interface reported slightly higher cognitive scores compared to the long text interface, while the short two-phase interface scores were lower. If the long two-phase interface increased cognitive load, we would expect a similar increase with the short two-phase interface compared to the short text interface. However, we observed lower cognitive load in the short two-phase interface, suggesting alternative explanations.

One explanation is that the two-phase interface reduces cognitive load in the short QS but not in the long QS. The increase in cognitive load might result from the interface's interactivity, as participants complete more operations without altering their decision-making processes. Our experiment results contradict this. Behavioral data shows a slight decrease in voting time for 2-phase interface participants responding to the long survey. Participants in the two-phase interface exhibited distinct behaviors, such as clear bimodal action distribution across remaining credits and faster iterative adjustments when credits were low. We interpret this as participants fine-tuning their preferences to align with their choices. Qualitative interviews on mental demand (Sec. 5.3) and effort (Sec. 5.5) revealed that participants in the long two-phase interface faced higher-level strategic challenges, unlike the text interface's lower-level operational tasks. Therefore, we reject the claim that the two-phase interface increases cognitive load in the long survey.

Next, we consider that the two-phase interface limits participants' satisficing when faced with cognitive overload. The two-phase interface lessened the cognitive load in the short QS, but in the long QS, the complexity of tasks diverted participants' focus, so we observed no reduction in cognitive load. Participants in the long QS group satisficed due to cognitive overload, which indicated lower cognitive load. Our results support this: Participants using the text interface handled more operational tasks and voting decisions, while those using the two-phase interface considered more comprehensive aspects. Evidence shows that in the long two-phase interface, participants engaged in deeper cognitive processing and were less burdened by precise voting. Participants using the two-phase interface invested more effort on deeper and more critical thinking. We interpret the temporal demand data to show that the two-phase interface helped participants pace themselves better and we see no sign of negative descriptions from long two-phase interface participants regarding temporal demand.

In addition, we also show that participants in the long text interface are satisficing because of cognitive overload from having too many options. Participants in the long text interface dealt with more text to read, more credits to allocate, and more options to consider. Recall results from Section 5 and Section 6, and consider common sense; it's counterintuitive that this group had fewer participants experiencing high cognitive load compared to the short text interface. This group also experienced the least temporal demand (Sec. 5.6) while showing no difference in time spent per option compared to the text interface (Figure 18c). Participants in the long text interface also expressed the least frustration with operational tasks (Sec 5.7). These counterintuitive findings pointed to the only plausible explanation that participants are exhibiting satisficing behaviors.

In short, as prior literature pointed out, these interactive components may have prevented participants from taking these mental shortcuts, which would typically reduce measured cognitive load [39, 40, 41, 42]. This prevention could result in a higher cognitive load in the long two-phase interface compared to the long text interface, reflected by our observation that participants' cognitive load shifted from some dimensions to others, maintaining a overall higher cognitive load.

Thus, we claim that QS with many options leads to satisficing, but a two-phase interface can limit it, redirecting participants' cognitive attention to facilitate decision-making that constructs a more comprehensive preference across options. In the following section, we examine the specific elements that guided participant's completing QS.

7.2 Bounded rationality and interface design

One core repeated theme that emerged throughout participants' responses during the interview relates to bounded rationality. Bounded rationality defined by Simon [40] refers to the idea that individuals' cognitive limitations limited one's ability to use and process all available information, leading to a sub-optimal resolution when decision making. When participants respond to a QS, they are faced with multiple options presented on the quadratic survey as well as the abundance of budget. Since the remaining budget translates to possible votes one can select to apply to an option, this adds additional numbers of decisions to make. Even though the drop-down menu showing all possible pre-calculated vote-credit values was a relief for a few participants so they do not need to search for the bounds, this sea of decision-making requires participants to recall and scramble information at once, which is extremely difficult. The byproduct of bounded rationality often translates to individuals satisficing behaviors [87].

> *So I did say, Okay, you know, you thought of enough things, you know, and so it wasn't the most effort I could put in because again, that would have been diminishing returns. I tried to think of enough things that I could make, make a meaningful decision and then move on. [...] I think that that's just not a realistic expectation (to be perfect), but I felt satisfied. [...] I felt like that (the response) was satisfied, but not perfect. Cause perfect is not a reality.* S036 (ST)

Participant expressed their bounded rationality and their decision to satisfice, making *good enough* and not *optimal* decisions.

Cognitive overload also encouraged individuals to create heuristics [88] and over reliance on defaults [57]. Thus, the design of showing one option at a time in the two-phase interface lowers the possibility of participants being influenced by the default positions of options or apply heruisitcs to a narrowing set of options and allowed thoughtful reflection to think about their attitude toward that option.

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

It is important to note that bounded rationality aims not to critique or exploit possible biases but to emphasize the importance of designing interface interventions to prevent survey respondents from making decisions that differ from their true preferences. For example, problem decomposition [89] and dimension reduction is a strategic approach when individuals face cognitive overload. Several participants would create a two axis groupings, despite the experiment group they are in. Participants would have categories that cluster similar topics (i.e., all topics related to health vs. humanitarian) and categories that depict the positivity of their preference (i.e., positive vs. negative). The different between experiment condition only lies in whether these groupings are possible to represent on the interface.

> *Honestly, if medical research [...] I think if it was the first option, the first thing I saw, I probably would have given it more [...] because medical research [...] to me this seems like the most important, but I think if... if it was the first one I saw, I think it would automatically gave it a lot more.* S003 (ST)

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

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

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

> *If I had to choose a number like that in the beginning. That would have been really bad, but positive, neutral, negative. That was good enough.* S026 (LI)

> *I think … ranking at the beginning one's impression towards these issues helps to like determine how many votes should be put towards them.* S002 (SI)

Conversly, participants using the text-based interface requested for organizational features. Almost half of the participants (N=4) using the long text interface expressed a desire for features that can help reduce the decision space when responding to the QS.

> *If anything, I think I would like to be able to like, click and drag the categories themselves so I could maybe reorder them to like my priorities.* S025 (LI)

> *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 (LI)

In summary, individual's bounded rationality encouraged participants to exhibit satisficing behaviors, heuristics, and defaults when responding to QS. By showing all options upfront, one-at-a-time, and repositioning options based on participant's rough preferences prevented participants from using defaults and heuristics. The two-phase organization actively scaffolds participant's decision making process supporting efficienceproblem 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 QS

When completing QS, deciding the number of votes for one QS option fits squarely into Lichtenstein and Slovic [14]'s characteristic of difficult decisions. Svenson [53]'s differentiation and consolidation theory explains how participants behave and quote. Recall that the theory states that decision making contains a differentiation stage involving identifying differences and eliminating less favorable alternatives, while the consolidation stage strengthens commitment to the chosen option. Hence, participants decompose options into categories, effectively reducing the decision dimension to mitigate difficulties.

Participants started by constructing preferences in situ, especially regarding options they had not thought about:

> *I mean, it's not necessarily a challenge, but it's interesting to see: 'Oh, there are other aspects that I never care about.' And actually … some people care <an option>. Sure. Why? Why (should) I spend money on that?* S037 (LI)

During this process, participants try to identify differences between options and differentiate. Hence, we observe the two axis groupings. The two-phase interface effectively facilitated the expression of at least one dimension. Participants without such support expressed difficulty:

> *If anything, I think I would like to be able to like, click and drag the categories themselves so I could maybe reorder them to like my priorities. And so I could maybe make that like a descending or ascending like list of like importance. […] if I could pull that up to the top, say myself like click and drag it up there, I think then I would stack the things I think it would affect under it. So like, I would put then, like youth, pro-education programs and adult education and early childhood programs and kinda stack those altogether. […] I would kind of make myself categories and subcategories out of this list … If I could organize it.* S025 (LT)

The ability to drag-and-drop options in the two-phase interface blends this differentiation process into the consolidation phase. For example, placing options next to one another facilitated fine-grain differentiation:

> *I think the system was actually really helpful because I could just drag them. [...] if I couldn't drag it, then I would have a harder time organizing my thoughts, whereas with the dragging feature I can really compare them, I can drag this one up here, and then compare it to the top one versus not being able to track it at all.* `S039 (SI)`

We also observe participants differentiating and consolidating through their behavior. Recall the bi-modal behavior in Figure 20, since participants in the two-phase interface had completed initial differentiation early in the voting process, they assign votes to highlight these differences. Conversely, participants in the text interface made larger updates of the votes slightly later. As participants began consolidating their preferences, participants in the long two-phase interface maintain their mental capacity to make fine-grain adjustments given faster iterations and small vote updates. Hence, we see a stronger bimodal pattern indicating the two stages in the long two-phase interface, compared to the long text interface. One participant explicitly expressed their strategy:

> *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)`

As participants begin consolidating their preferences using votes, drag-and-drop are occasionally used to reflect on their expressed preferences. In other words, reflecting and reassuring their decisions. When asking `S021` why they would drag-and-drop an option after voting, the participant responded:

> *So I guess to see what my ranking look like ... and see if I could give more money or not.* `S021 (LI)`

One participant confirmed this mapping by describing their approach to QS on the long two-phase interface, highlighting the differentiation, the consolidation, and reflection:

> *[...] this (option) is something that's really important to me ... So I had the flexibility to move it to positive. [...] especially because when I was doing categorization in the first step, [...] what I thought about it in the moment. [...] In the second step there was a shift in my perception of the issue just reflecting ... So being able to change ... That was really nice as well.* `S009 (LI)`

In summary, participants construct their preferences as they complete QS. We observed behaviors and qualitative insights that align to the differentiation and consolidation process in decision making. Our interface scaffolded many of the differentiation stages through pre-voting organization and some consolidation phases through drag-and-drop. Thus, this explains 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 really like having the costs of all the votes displayed. When you select the dropdown menu and ranked in order.* `S002 (SI, deriving cost)`

> *I thought I have [...] (to) do all the numbers or calculation 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 calculation on that end.* `S005 (LT, keeping track of spent)`

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

7.4.2 Cognitive load from budget management. Section 5 reveals that participants experienced cognitive load due to budget management. Behavioral economists like Shah et al. [85] believe that values and careful decision-making are derived from limited resources, which introduce higher cognitive load. Prospect theory [90] highlights the higher negative value of *perceived* utility when cuts must be made. This evidence underscores the heightened cognitive load participants experience when managing their budget. Our interface lacks tools to mitigate this, making it a crucial area for future work.

7.4.3 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 [91], 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.

This challenge remains an open question, and future research can explore possibilities to scale ordinal utility to map it to individual preferences without biasing participants.

7.4.4 Navigating Between Budget, Votes, and Actual Impact. The third challenge refers to participants' confusion between budget, votes, and actual outcomes. While participants are clear with the definitions of these elements, it is not clear for them to make effective decisions. 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 these causes. So you have to figure out which ones get the most, which ones don't get as much. [...]*
> *Interviewer: So when you're operating this interface. Do you feel that the more votes you're giving to a cause you're actually spending more on it?*
> *Yeah.* S003 (ST)

Recall that this survey aims to assist community organizers in distributing resources to a societal cause. This participant decided to 'skip' over the quadratic formulation and the concept that their votes are governed by the quadratic formulation, drawing a direct translation between votes and the resources to which community organizers ought to contribute. While this does not invalidate the power of the quadratic mechanism, it causes frustration and friction for participants to construct a clear picture of how to make voting decisions. Thus, future interface development should consider how to communicate the relationships to highlight trade-offs survey respondents are making.

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 QS Usage and Design Recommendations and Future Work

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

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. We recommend designing QS for specific use cases requiring critical evaluations, such as investment decisions or settings where participants have ample time to think and process the survey. For instance, revealing the options ahead of time can aid in preference construction.

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 need further research.

Understanding results influence on decision-makers. We need further research to understand how the QS interface impacts decision-makers and broader societal resource distributions. Since QS is new, we prioritize widespread adoption and usage before fully assessing its influence on decision-making. Future studies will examine how decision-makers interpret and use QS data, and its broader implications for societal decisions.

Individual differences in cognitive capacity. Variations in individual cognitive capacity influenced the cognitive scores from participants. Ideally, a within-subject study would clarify how cognitive load shifts across QS interfaces. However, deconstructing established preferences is near impossible. Changing options across domains shifts cognitive load further, and QS completion is lengthy. Thus, we designed this in-depth, between-subject study even though it may reflect noise rather than actual cognitive load due to the small sample size. Future research will quantify the impact between different QS interfaces.

Limited experience with QS. Participants lacked prior QS experience. After a tutorial and quiz, participants completed QS. Although participants understood QS mechanics, familiarity influences strategies and cognitive load. As quadratic mechanisms become more prevalent, especially in software engineering, future research will compare novices and experts.

Using duration between clicks to represent decision-making. Recognizing that click duration may include time spent considering other options, we must consider it as an approximate measure of decision-making time. For instance, deciding on the number of votes for two options may cause longer time for the first option and shorter for the second. Thus, we consider click duration an approximate decision-making measure, acknowledging it may include deliberation time. Despite its imperfections, this approach offers valuable insights into decision-making within our experimental constraints.

9 Conclusion

Surveys enable decision-makers to aggregate crowd opinions. In this study, we use QS to elicit individual responses in the context of social resource allotment. After multiple design iterations, we propose an two-phase interface for QS. We then examined its influence on individuals' cognitive load and behaviors when faced with societal issues of varying lengths. In a 2x2 between-subject study, we had participants experience either a long or short QS using a text-based or two-phase interface. NASA-TLX questionnaires and interviews revealed that participants using the two-phase interface for a long QS demonstrated a more comprehensive and critical evaluation of societal issues, despite not experiencing a lower cognitive load. Participants using the long text interface experienced cognitive overload, which led to satisficing behaviors or mental shortcuts. Analyzing click-stream data, we identified that participants made fine-grain iterations using the long two-phase interface when credits were low. We demonstrate that a two-phase, organize-then-vote interface can scaffold the complex decision-making process, helping individuals express their opinions for collective societal decisions. Through the iterative design process and detailed interviews, we identified future directions and design recommendations for collective decision-making applications using the quadratic mechanism.

References

[1] Ziang Xiao, Sarah Mennicken, Bernd Huber, Adam Shonkoff, and Jennifer Thom. 2021. Let Me Ask You This: How Can a Voice Assistant Elicit Explicit User Feedback? *Proceedings of the ACM on Human-Computer Interaction*, 5, CSCW2, (Oct. 2021), 1–24. DOI: 10.1145/3479532.

[2] 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.

[3] 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.

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

[6] 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.

[7] 2023. Gov4git: A Decentralized Platform for Community Governance. (Mar. 2023). Retrieved June 13, 2024 from.

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

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

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

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

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

[18] 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.

[19] Michael Morreau. 2014. Arrow's theorem.

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

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

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

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

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

[30] 2015. Designing usable ballots | Center for civic design. https://civicdesign.org/fieldguides/designing-usable-ballots/. (June 2015). Retrieved June 17, 2024 from.

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

[32] 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.
[33] 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.
[34] 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)*.
[35] 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.
[36] 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.
[37] 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.
[38] Mirta Galesic. 2006. Dropouts on the Web: Effects of Interest and Burden Experienced During an Online Survey. *Journal of Official Statistics*, 22, 2, (June 2006), 313. Retrieved July 3, 2024 from.
[39] Kahneman Daniel. 2017. *Thinking, Fast and Slow*.
[40] 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.
[41] 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.
[42] Amos Tversky and Daniel Kahneman. [n. d.] Judgments of and by Representativeness.
[43] 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.
[44] 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.
[45] 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.
[46] Peter Coy. 2019. A New Way of Voting That Makes Zealotry Expensive - Bloomberg. *Bloomberg*, (May 2019). Retrieved Dec. 16, 2023 from.
[47] 2022. Quadratic Voting Frontend. Public Digital Innovation Space. (Jan. 2022). Retrieved Dec. 16, 2023 from.
[48] 2024. RadicalxChange/quadratic-voting. RadicalxChange. (May 2024). Retrieved June 17, 2024 from.
[49] [n. d.] Read the Whitepaper | Gitcoin. https://www.gitcoin.co/whitepaper/read. (). Retrieved June 17, 2024 from.
[50] [n. d.] About RxC. https://www.radicalxchange.org/wiki/about/. (). Retrieved June 17, 2024 from.
[51] yehjxraymond. 2024. Yehjxraymond/qv-app. (Mar. 2024). Retrieved June 17, 2024 from.
[52] 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.
[53] 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.
[54] 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.
[55] 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.
[56] 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.
[57] 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.

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

[59] 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.

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

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

[62] 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.

[63] 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.

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

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

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

[67] 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).

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

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 2023. Charity Navigator. https://www.charitynavigator.org/index.cfm?bay=search.categories. (May 2023). Retrieved Dec. 16, 2023 from.

[74] 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.

[75] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*. ACM Press, Austin, Texas, 141. DOI: 10.1145/1743666.1743701.

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

[77] 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.

[78] 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.

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

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] John W. Payne, James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge. DOI: 10.1017/CBO9781139173933.

[85] Anuj K. Shah, Eldar Shafir, and Sendhil Mullainathan. 2015. Scarcity frames value. *Psychological Science*, 26, 4, 402–412.

[86] 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.

[87] 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.

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

[89] Herbert A. Simon. 1996. *The Sciences of the Artificial*. (3rd ed ed.). MIT Press, Cambridge, Mass.

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

[91] 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 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.

B Cognitive Demand Tables

Here we provide the full qualitative analysis table.

Table 2. This table lists all the causes participants mentioned as contributing to their Mental Demand. The shaded cells represent the percentage of participants citing each source of mental demand, allowing for comparison within columns. The abbreviations are: ST (Short Text Interface), SI (Short Two-phase Interface), LT (Long Text Interface), and LI (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.

[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

B.1 Mental Demand Table

B.2 Physical Demand Table

Table 3. 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 ST	Version SI	Version LT	Version LI	Experiment Conditions Short	Experiment Conditions Long	Experiment Conditions Text	Experiment Conditions 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

B.3 Performance Table

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

[Performance]	Total	Version ST	Version SI	Version LT	Version LI	Experiment Conditions Short	Experiment Conditions Long	Experiment Conditions Text	Experiment Conditions 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

B.4 Temporal Demand Table

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

B.5 Frustration Table

Table 6. Frustration Sources: needs to be updated with some new terms definitions for some of the columns.

[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