

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

2  
3      ANONYMOUS AUTHOR(S)\*  
4

5      Quadratic Surveys (QS) elicit more accurate individual preferences than traditional surveys, such as Likert-scale surveys. However, the  
6      cognitive load associated with QS has hindered its adoption in digital surveys for collective decision-making. We introduce a two-phase  
7      “organize-then-vote” QS interface based on decision-making and preference construction theories designed to lessen the cognitive load.  
8      Since interface design significantly impacts survey results and accuracy, our design scaffolds survey takers’ decision-making while  
9      managing the cognitive load imposed by QS. In a 2x2 between-subject in-lab study on public resource allotment, we compared our  
10     interface with a traditional text interface across QS with 6 (short) and 24 (long) options. Our interface reduced satisficing behaviors  
11     arising from cognitive overload in long QS conditions. Participants using our interface in the long QSs shifted their cognitive effort  
12     from mechanical operations to constructing more comprehensive preferences. This research clarifies how human-centered design  
13     improves preference elicitation tools for collective decision-making.  
14

15  
16     CCS Concepts: • Human-centered computing → Collaborative and social computing systems and tools; Collaborative and  
17     social computing design and evaluation methods; User studies; HCI design and evaluation methods; Interactive systems  
18     and tools; Empirical studies in interaction design.

19  
20     Additional Key Words and Phrases: Quadratic Survey; Survey Response Format; User Interface; Preference Construction; Cognitive  
21     Load

22      **ACM Reference Format:**

23  
24     Anonymous Author(s). 2024. Organize, Then Vote: Exploring Cognitive Load in Quadratic Survey Interfaces. *Proc. ACM Hum.-Comput.*  
25     *Interact.* 1, 1 (September 2024), 42 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

26  
27      **1 Introduction**

28  
29     Designing intuitive survey interfaces is crucial for accurately capturing respondents’ preferences, which directly impact  
30     the quality and reliability of the data collected. For instance, voice assistants have been used to gather high-quality user  
31     feedback [1], and a recent Human-Computer Interaction (HCI) study highlights that certain survey response formats  
32     can increase errors [2]. In this paper, our goal is to introduce an effective interface for **Quadratic Surveys (QS)**, a  
33     survey tool designed to elicit preferences more accurately than traditional methods [3]. Despite the promise of QS,  
34     there has been no research on designing interfaces to support its unique quadratic mechanisms [4], where participants  
35     must rank and rate items – a task that poses significant cognitive challenges. To popularize QS and ensure high-quality  
36     data, this paper addresses the question: *How can we design interfaces to support participants in completing Quadratic*  
37     *Surveys (QS) more effectively?*

38  
39     We envision an effective interface that navigates participants through the complex mechanism and preference  
40     construction process, tailored to the unique challenges of QS. QS improves accuracy in individual preference elicitation  
41     compared to traditional methods like Likert scales by requiring participants to make trade-offs using a fixed budget  
42     of credits, where purchasing  $k$  votes for an option in QS costs  $k^2$  credits [5, 3]. This quadratic cost structure forces

43  
44  
45     Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not  
46     made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components  
47     of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on  
48     servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

49     © 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

50     Manuscript submitted to ACM

51  
52     Manuscript submitted to ACM

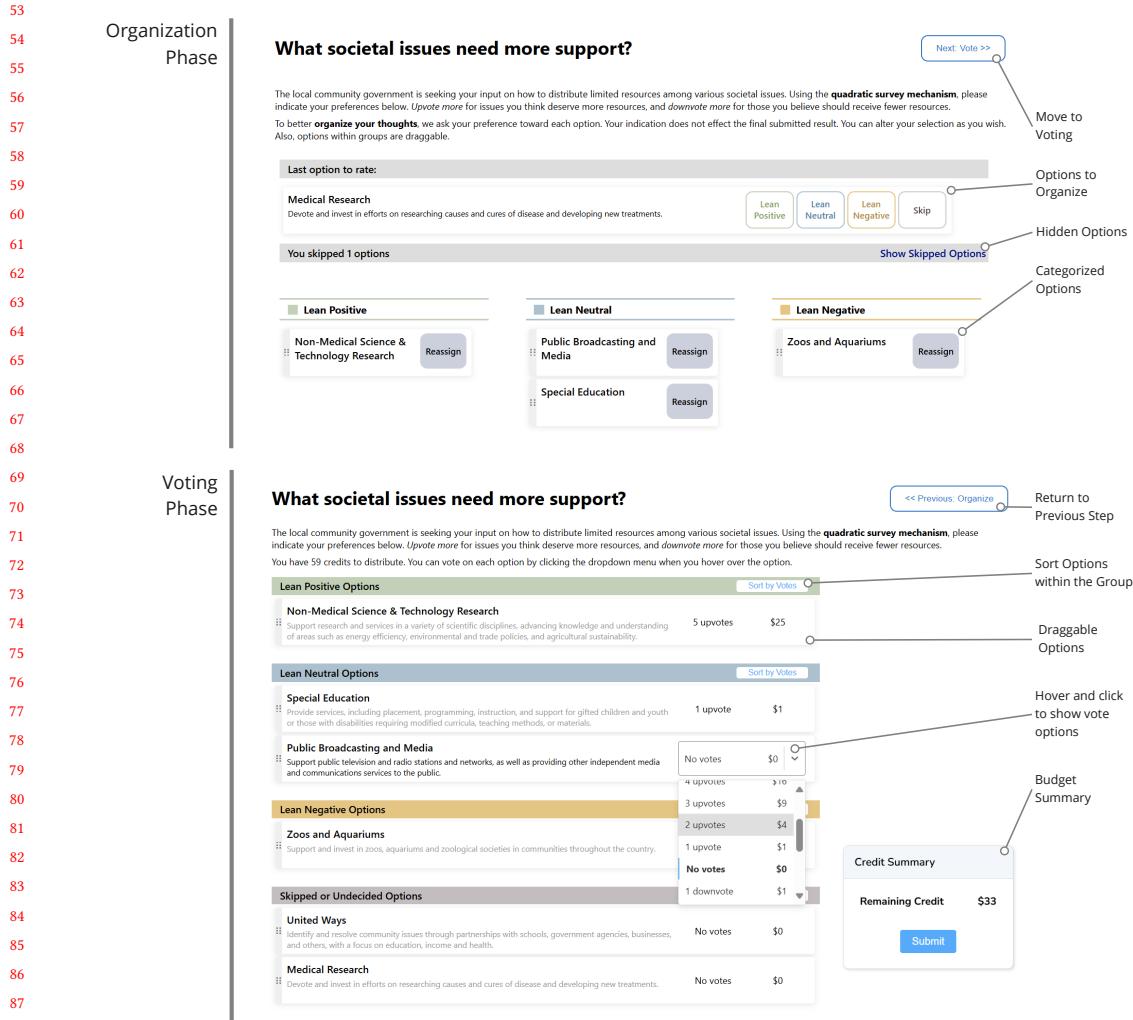


Fig. 1. The Two-Phase Interface: The interface consists of two phases. Survey respondents can navigate between phases using the top right button. In the organization phase, the interface presented one option at a time to the respondents where they choose among four choices: "Lean Positive", "Lean Neutral", "Lean Negative", or "Skip". Skipped options are hidden and can be evaluated later. The chosen options will be listed below. Items can be dragged and dropped across categories or returned to the stack. In the voting phase, options are listed in the order of the four categories. When hovering over each option, respondents can select a vote for that option using the dropdown. Each dropdown contains the cost associated with the vote. A sort button allows ascending sorting within each category. A summary box tracks the remaining credit balance.

respondents to carefully evaluate their preferences, balancing the strength of their support or opposition against the limited budget. While this cost structure forces them to make thoughtful trade-offs, this complexity also increases cognitive load, making it mentally taxing to weigh costs, evaluate options, and construct rankings [6]. Moreover, QS, often referred to as Quadratic Voting (QV) in voting scenarios, can involve hundreds of options [7, 8], increasing the

105 risk of cognitive overload and satisficing [9, 10, 11]. We propose that an effective interface tailored to support users  
106 through these complex trade-offs can reduce cognitive load and facilitate better preference construction.  
107

108 To date, existing quadratic mechanism-powered applications simply present options, allow vote adjustments and  
109 automatically calculate votes, costs, and budget usage. These designs focused heavily on the mechanism, rather than  
110 supporting possible challenges these application users faced. Survey interface literature, while addressing decision-  
111 making and usability, most focus on traditional surveys that do not share the unique option-to-option trade-offs that QS  
112 introduces [12, 13, 14, 15, 16, 2]. The challenges, in particular, cognitive load [17, 18, 15, 19] and scaffolding challenging  
113 tasks [20, 21, 22] like preference construction, are opportunities where user interfaces have shown their promises. As  
114 a result, it remains unclear how different interface designs might support QS in reducing cognitive load and aiding  
115 preference construction.  
116

117 After multiple design iterations, we propose a novel interactive two-phase “organize-then-vote” QS interface (two-  
118 phase interface, Figure 1) that integrates three key elements. First, the interface scaffolds the preference construction  
119 process by having participants initially categorize the survey options into “Lean Positive,” “Lean Neutral,” or “Lean  
120 Negative.” This serves as a cognitive warm-up, easing participants into the more complex QS voting task. Second,  
121 the interface arranges the options according to these categorizations, providing a structured visual layout. Third,  
122 participants can refine the positions of these options using drag-and-drop functionality, giving them greater control  
123 and agency in the preference-construction process. These design features are aligned with preference construction  
124 theory and build upon prior research in interface design to reduce cognitive load and enhance user engagement.  
125

126 To explore how these interface elements mitigate the cognitive load and support preference construction in Quadratic  
127 Surveys, we pose the following research questions:  
128

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

134 We invited 41 participants to a lab study comparing our two-phase interface with a baseline to understand how  
135 different interface designs and option lengths (6 options or 24 options) impact cognitive load. Qualitative findings,  
136 measured using the NASA Task Load Index (NASA-TLX) and semi-structured interviews, revealed that participants  
137 using the two-phase interface experienced cognitive demand more from strategic, holistic thinking compared to personal  
138 relevance and operational tasks, particularly in longer surveys. Quantitative results showed that, although participants  
139 spent more time per option, they made faster decisions during the voting phase, suggesting a more efficient distribution  
140 of cognitive effort. We concluded that the two-phase interface prevented cognitive overload in long QS surveys and  
141 shifted mental load toward more strategic thinking, reducing reliance on mental shortcuts like satisficing [9].  
142

143 *Contributions.* We contribute to the HCI community by proposing the first interface specifically designed for QS  
144 and QV-like applications, aimed at reducing cognitive challenges and scaffolding preference construction through a  
145 two-phase interface with direct manipulation. Before our work, no research had explored QS interfaces, particularly for  
146 long surveys prone to cognitive overload. Few studies in HCI address interfaces for surveys and questionnaires. Our  
147 study demonstrated how user interfaces can facilitate preference construction in situ and reduce satisficing behaviors  
148 by promoting incremental updates and deeper engagement with survey items through interface elements. Additionally,  
149 our interview is also the first in-depth qualitative analysis of user experiences with Quadratic Mechanism applications,  
150

157 identifying key factors contributing to cognitive load. The impact of our contribution extends beyond QS, offering  
 158 design implications for other preference-elicitation tools in complex scenarios. By making QS easier to use and more  
 159 accurate, our design also encourages wider adoption among researchers and practitioners. Finally, our work lays the  
 160 groundwork for future studies on qualitative insights and future interface designs of quadratic mechanisms, supporting  
 161 decision-makers in eliciting accurate respondents' preferences.  
 162

## 163 2 Related Work

164 This research lies at the intersection of three core areas: quadratic surveys, survey and voting interface design, and  
 165 choice overload along with cognitive challenges. In this section, we review the related works in each of these areas.  
 166

### 167 2.1 Quadratic Survey and the Quadratic Mechanism

168 We introduce the term **Quadratic Survey (QS)** to describe surveys that utilize the quadratic mechanism to collect  
 169 individual attitudes. The **quadratic mechanism** is a theoretical framework designed to encourage the truthful revelation  
 170 of individual preferences through a quadratic cost function [4]. This framework gained popularity through **Quadratic**  
 171 **Voting (QV)**, also known as plural voting, which uses a quadratic cost function in a voting framework to facilitate  
 172 collective decision-making [23].

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

$$176 \quad c_k = n_k^2 \quad \text{where} \quad n_k \in \mathbb{Z}$$

177 The total cost of all votes must not exceed the participant's budget:  
 178

$$179 \quad \sum_k c_k \leq B$$

180 Survey results are determined by summing the total votes for each option:  
 181

$$182 \quad \text{Total Votes for Option } k = \sum_{i=1}^S n_{i,k}$$

183 where  $S$  represents the total number of participants, and  $n_{i,k}$  is the number of votes cast by participant  $i$  for option  $k$ .  
 184 Each additional vote for each option increases the marginal cost linearly, encouraging participants to vote proportionally  
 185 to their level of concern for an issue [24].

186 QS adapts these strengths of the quadratic mechanism in *voting* to encourage truthful expression of preferences in  
 187 *surveys*. Unlike traditional surveys that elicit either rankings or ratings, QS allows for *both*, enabling participants to cast  
 188 multiple votes for or against options, incurring a quadratic cost. Cheng et al. [3] showed that this mechanism aligns  
 189 individual preferences with behaviors more accurately than Likert Scale surveys, particularly in resource-constrained  
 190 scenarios like prioritizing user feedback on user experiences.

191 In recent years, empirical studies on QV have expanded into various domains [25, 26]. Applications based on the  
 192 quadratic mechanism have also grown, including Quadratic Funding, which redistributes funds based on outcomes  
 193 from consensus made using the quadratic mechanism [27, 28]. Recent work by South et al. [29] applies the quadratic  
 194 mechanism to networked authority management, later used in Gov4git [30]. Despite the increasing breadth and depth  
 195

of applications utilizing the quadratic mechanism, little attention has been paid to user experience and interface design, which support individuals in expressing their preference intensity. Our work aims to address this by designing interfaces supporting quadratic mechanisms.

## 2.2 Design Implications from Surveys and Voting Interfaces

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

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

These components present options, calculate costs, and allow vote adjustments. However, these interfaces focus purely on mechanics without little understanding of voter's usability needs or offering cognitive support to help them complete the task effectively. In addition, the HCI community conducted few research [31, 32] on survey and questionnaire interfaces components, with more work focusing more on alternative input modalities like bots, voice, and virtual reality [33, 34, 35, 36]. Thus, we turn to marketing and research literature for insights into how digital survey interface elements can impact user behavior and usability.

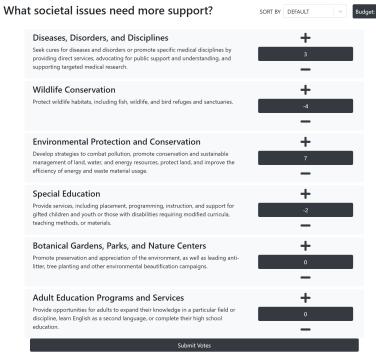
Research in the marketing and research communities focusing on survey and questionnaire design, usability, and interactions examines the influence of presentation styles and 'response format.' Weijters et al. [37] demonstrated that horizontal distances between options are more influential than vertical distances, with the latter recommended for reduced bias. Slider bars, which operate on a drag-and-drop principle, show lower mean scores and higher nonresponse rates compared to buttons, indicating they are more prone to bias and difficult to use. In contrast, visual analog scales that operate on a point-and-click principle perform better [38]. These studies show how even small design changes can have a large impact on usability, highlighting the importance of designing interfaces that prioritize human-centered interaction rather than focusing solely on functionality.

Voting interfaces are a specialized type of survey interface that not only elicit individual choices but often have consequential impacts. For example, the butterfly ballot, an atypical design, may have influenced the outcome of the 2000 U.S. Presidential Election [39]. Research has shown that ballot interfaces can significantly influence democratic processes [12, 40, 41]. Several studies also highlighted how voting interface designs shift voter decisions [12], reduce usability errors [42, 43], or improve interaction [44, 45, 46, 47, 48]. We provide more details to these voting interfaces in the Appendix A.

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

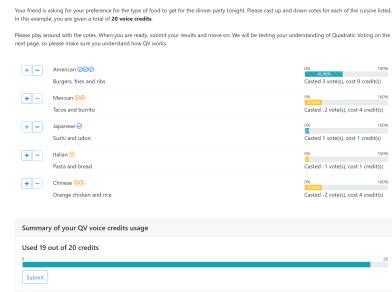
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275

(a) Software by WeDesign, used in the first empirical QV research [5]. Little information is available about the software, except for an image from Posner and Weyl [24]. 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.



276  
277  
278  
279  
280  
281

(c) An open-source QV interface [52] 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.

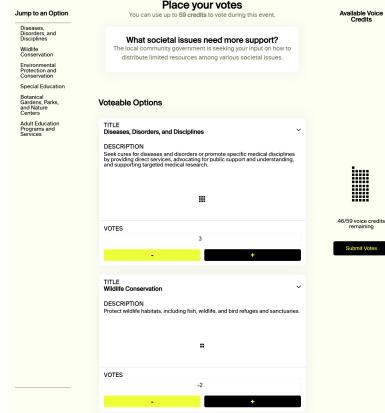


295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312

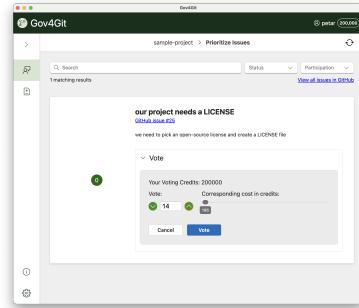
(e) The interface used in the research by Cheng et al. [3] employs the most visual components. Icons depict the current number of votes, with progress bars signifying the current spending.

Manuscript submitted to ACM

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



(b) An open-sourced QV interface [49] forked from GitCoin [50], used by the RadicalxChange community [51]. 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.



(d) The interface designed for gov4git [30] 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.



### 313 2.3 Cognitive Challenges and Choice Overload

314  
 315 The challenge of respondents making difficult decisions using quadratic mechanisms remains unexplored in the  
 316 literature. Lichtenstein and Slovic [6] identified three key elements that make decisions difficult. These elements include  
 317 making decisions in unfamiliar contexts, being forced to make trade-offs due to conflicting choices, and quantifying  
 318 the value of one's opinions. QS fits all three elements: participants may encounter unfamiliar options set by the  
 319 decision-maker, are constrained by budgets that require trade-offs, and cast final votes as numerical values. Thus, we  
 320 believe QS introduces high cognitive load.  
 321

322 According to cognitive load theory, cognitive load refers to the demands placed on a user's working memory  
 323 during the interaction process, which significantly influences the usability of the system [cooper1998research,  
 324 paas2003cognitive]. Cognitive overload can adversely affect performance [drommi2001interface], leading individuals  
 325 to rely on heuristics rather than deliberate, logical decision-making [53]. When presented with excessive information,  
 326 such as too many options, individuals 'satisfice', settling for a 'good enough' solution rather than an optimal one [9,  
 327 10, 11]. Subsequently, too many options can overwhelm individuals, resulting in decision paralysis, demotivation, and  
 328 dissatisfaction [54].  
 329

330 Additionally, Alwin and Krosnick [55] highlighted that the use of ranking techniques in surveys can be time-  
 331 consuming and potentially more costly to administer. These challenges are compounded when ranking numerous items,  
 332 requiring substantial cognitive sophistication and concentration from survey respondents [56].  
 333

334 Notable applications of Quadratic Voting include the 2019 Colorado House, which considered 107 bills [57], and  
 335 the 2019 Taiwan Presidential Hackathon, which featured 136 proposals [58]; both used a single QV question with  
 336 hundreds of options. Psychological and behavioral research highlights the importance of understanding how individuals  
 337 navigate and benefit from new interfaces under long-list QS conditions. These empirical applications of QV suggest  
 338 QS's potential to elicit individual preferences, emphasizing the need to study cognitive load and interface design.  
 339  
 340

## 341 3 Quadratic Survey Interface Design

342 In this section, we present the QS interface. Using components from existing QV interfaces (Fig 2) and insights from  
 343 prior literature, we iterated through paper prototypes and three design pre-tests, detailed in Appendix B. In our initial  
 344 paper prototyping iterations, participants struggled to *rank* relative preferences among options and *rate* the degree  
 345 of trade-offs between them. In this study, we focus on addressing the former challenge, which pertains to preference  
 346 construction.  
 347

### 348 3.1 'Organize-then-Vote': The Two-Phase Interface

349 3.1.1 *Justifying a two-phase approach.* The main objective of the two-phase interface is to facilitate preference con-  
 350 struction and reduce cognitive load. As shown in Figure 1, the interface consists of two steps: an organization phase  
 351 and a voting phase. In both phases, survey respondents can drag and drop options across the presented list.  
 352

353 A *two-phase approach*. Preferences are shaped through a series of decision-making processes [6]. Two major decision-  
 354 making theories informed this two-step interaction interface design: Montgomery [59]'s Search for a Dominance  
 355 Structure Theory (Dominance Theory) and Svenson [60]'s Differentiation and Consolidation Theory (Diff-Con Theory).  
 356 The former suggested that decision-makers prioritize creating dominant choices to minimize cognitive effort by  
 357 focusing on evidently superior options [59]. The latter described a two-phase process where decisions are formed by  
 358 initially *differentiating* among alternatives and then *consolidating* these distinctions to form a stable preference [60].  
 359  
 360

365 Both theories supported the design decision to reduce the dimensions during the initial decision process and help  
 366 emphasize relatively important options to form decisions. Hence, the two-phase design – organize-then-vote – aimed  
 367 to facilitate this cognitive journey explicitly. The first phase focused on differentiating and identifying dominant options,  
 368 enabling survey respondents to preliminarily categorize and prioritize their choices. The second phase presented these  
 369 categorized options in a comparable manner, with drag-and-drop functionality, enhancing one's ability to consolidate  
 370 preferences. This structured approach aimed to construct a clear decision-making procedure that reduced cognitive  
 371 load and enhanced clarity and confidence in the decisions made.  
 372

373  
 374     *Phase 1: Organization Phase.* The goal of the organization phase was to support participants in identifying clearly  
 375 superior options or partitioning choices into distinguishable groups. In this section, we first describe how the interaction  
 376 works, then we detail the reasons for the implemented design decisions.  
 377

378     The organizing interface, depicted on the top half of Figure 1, sequentially presents each survey option. Participants  
 379 select a response among three ordinal categories – “Lean Positive”, “Lean Negative”, or “Lean Neutral”. Once selected,  
 380 the system moves that option to the respective category. Participants can skip the option if they do not want to indicate  
 381 a preference. Options within the groups are draggable and rearrangeable to other groups should the participants wish.  
 382

383     To support preference formation, respondents are shown one option at a time, allowing them to either recall a prior  
 384 judgment or construct a new one based on the presented choices [61]. Limiting the information presented this way also  
 385 helps reduce cognitive load by preventing overload from too many options [62]. This incremental process ensures that  
 386 participants form opinions on individual options, addressing an early prototype issue where the organizing task was  
 387 mistakenly treated as a ranking task.  
 388

389     The three possible options – Lean Positive, Lean Neutral, and Lean Negative – aim to scaffold participants in  
 390 constructing their own choice architecture [63, 64], which strategically segments options into diverse and alternative  
 391 choice presentations while avoiding biases from defaults. We believed that these three categories were sufficient for  
 392 participants to segment the options. We do not limit the number of options one can place in each category to prioritize  
 393 user agency, allowing participants full control over how they organize their preferences [18]. Immediate feedback  
 394 displays the placement of options and allows participants to rearrange them via drag-and-drop, adhering to key interface  
 395 design principles [18]. At the same time, it allows finer-grain control for individuals to surface dominating options and  
 396 create differentiating groups of options.  
 397

400  
 401     *Phase 2: Interactive Voting Phase.* The objective of the voting phase is to facilitate the consolidation of differentiated  
 402 options through interactive elements while reinforcing the differentiation across options constructed by participants in  
 403 the previous phase. This facilitation is achieved by retaining the drag-and-drop functionality for direct manipulation of  
 404 position and enabling sorting within each category.  
 405

406     Options are displayed as they are categorized within each category from the previous step and in the following  
 407 section – Lean Positive, Lean Neutral, Lean Negative, and Skipped or Undecided – as detailed on the bottom half of  
 408 Figure 1. The Skipped or Undecided category contains options left in the organization queue, possibly because survey  
 409 respondents have a pre-existing preference or chose not to organize their thoughts further. The original order within  
 410 these categories is preserved to maintain and reinforce the differentiated options. This ordering sequence mitigated  
 411 early prototype concerns where uncategorized options were left at the top of the voting interface confusing survey  
 412 respondents. Respondents have the flexibility to return to the organization interface at any point during the survey to  
 413 revise their choices.  
 414

417 In the voting interface, options are draggable, allowing participants to modify or reinforce their preference decisions  
 418 as needed. Each category features a sort-by-vote function for reordering within the group, which, although it doesn't  
 419 affect the final outcome, supports information organization and consolidation. Both features aim to group similar  
 420 options automatically and emphasize proximity, reducing cognitive load by following the proximity compatibility  
 421 principle to enhance decision-making [65].

422 While multiple interaction mechanisms exist, drag-and-drop has been extensively explored in rank-based surveys.  
 423 For instance, Krosnick et al. [66] demonstrated that replacing drag-and-drop with traditional number-filling rank-based  
 424 questions improved participants' satisfaction with little trade-off in their time. Similarly, Timbrook [67] found that  
 425 integrating drag-and-drop into the ranking process, despite potentially reducing outcome stability, was justified by the  
 426 increased satisfaction and ease of use reported by respondents. The trade-off was deemed worthwhile as QS did not  
 427 use the final position of options as part of the outcome if it significantly enhanced user satisfaction and usability [68].  
 428 Together, these design decisions led to our belief that a two-phase interface with direct interface manipulation could  
 429 reduce the cognitive load for survey respondents to form preference decisions when completing QS.

430 In addition, we made three aesthetic design decisions. First, we removed visual elements like icons, emojis, progress  
 431 bars, and vote visualizations, as prior research indicated that emojis could influence survey interpretations and reduce  
 432 user satisfaction [69, 15]. While effective visualizations can aid decision-making, this study does not aim to address that  
 433 question. Second, the final interface has all options presented on the screen at the same time, intentionally. Unlike all  
 434 the prototypes and existing interfaces, prior literature emphasized the importance of placing all the options on the same  
 435 digital ballot screen to avoid losing votes. This echoes the proverb "out of sight, out of mind," where individuals might  
 436 be biased toward options that are shown to them, and additional effort is required for individuals to retrieve specific  
 437 information if options are hidden. Last, we decided to use a dropdown positioned to the right of each survey option for  
 438 ease of access to the budget summary when determining the votes. The layout of the votes and cost was inspired by  
 439 online shopping cart checkout interfaces where quantities are supplied next to the itemized costs followed by the total  
 440 checkout amount. After testing two alternative (Figure 3) input methods—click-based buttons, which required multiple  
 441 clicks, and a wheel-based design, which offered intuitive control but was unfamiliar to some participants—we opted for  
 442 a more accessible dropdown menu for vote selection.

	Option	Rating	Cost
453	Voting Item Item description will be placed here	- +3 rating +	\$9
454	Voting Item Item description will be placed here	+3 rating  +4	\$4 \$9 \$16
455	Voting Item Item description will be placed here	+3 rating	\$9

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

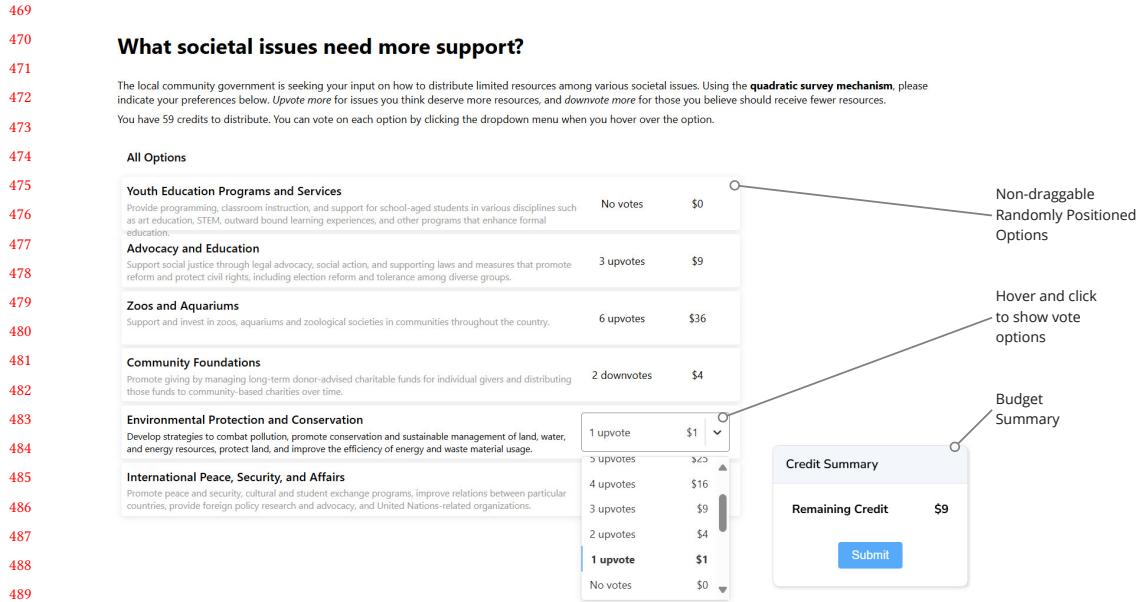


Fig. 4. The text-based interface: This interface is based on the interactive version but does not include the two-phase interactive support and lacks the drag-and-drop functionality. Options are randomly positioned.

### 3.2 Baseline Interface: Text-based Interface

We designed the text-based interface as our control condition to compare how the interactive components influenced participants' cognitive load and behavior. The text-based interface includes all the functionalities of the two-phase interface except for the two-phase interactive design and drag-and-drop feature (Figure 4.) The interface displays the question prompt at the top of the screen. The options appear in a list underneath the prompt. Survey can update their votes by selecting from a dropdown that lists all possible voting options and costs, based on the number of available credits. A small summary box on the right of the interface displays the current total cost and the remaining credits for the respondent. The interface randomly presents options to prevent ordering bias [70, 71].

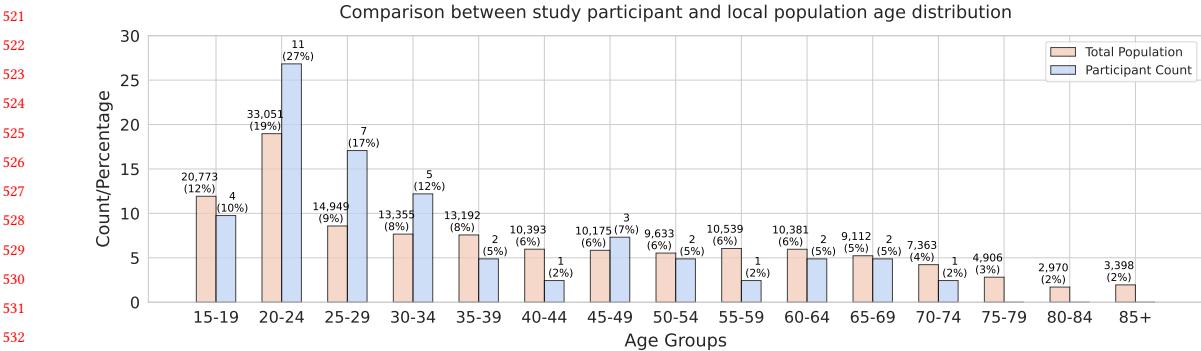
Both experimental interfaces were developed with a ReactJS frontend and a NextJS backend powered by MongoDB. We open-source both interfaces.<sup>1</sup>

## 4 Experiment Design

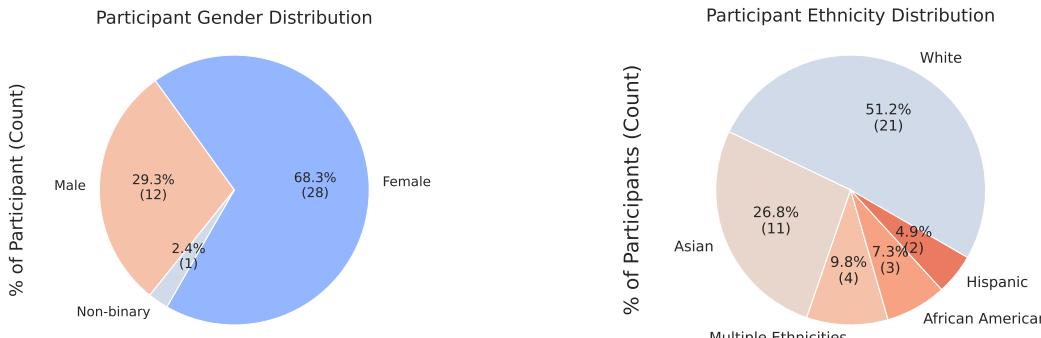
We recruited 41 participants, with one excluded<sup>2</sup> due to data quality concerns, from a United States college town using online ads, digital bulletins, social media posts, online newsletters, and physical flyers in public spaces beyond campus. To ensure participant diversity, we prioritized non-students by selectively accepting them as we monitored demographics. Study participants' mean age was 34.63 years old, with an age distribution similar to the county's demographic profile (Figure 5) albeit a slightly higher representation of younger adults. Gender and race demographics

<sup>1</sup>link-to-github

<sup>2</sup>The participant reported not completing the survey seriously because they believed the experiment was fake.



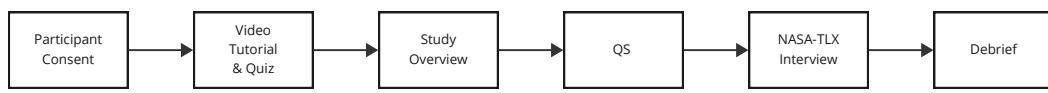
534 Fig. 5. Age distribution of the study participants were similar to the locale's demographic profile.  
535



550 Fig. 6. Gender distribution of our participants skewed towards female participants.  
551

552 Fig. 7. Ethnicity distribution remains diverse with fewer  
553 Hispanic and African American participants.  
554

555 Fig. 8. Demographic distributions: Age, Gender, and Ethnicity  
556



562 Fig. 9. Study protocol: Participants are asked to learn about the mechanism of QS after consenting to the study. The researcher  
563 explained the study overview and asked participants to complete the QS. A NASA-TLX survey followed by interviews to understand  
564 participants' cognitive load. We debriefed participants after the study.  
565

566 are aggregated in Figure 6 and 7. The study was framed as focusing on societal attitudes to avoid response bias. The  
567 university's Institutional Review Board reviewed and approved this study.  
568

569 Figure 9 visually represents the study protocol. Participants completed the study in the lab to control for external  
570 influences. Participants used a 32-inch vertical monitor displaying all options. After consenting, participants watched a  
571

573 video explaining the quadratic mechanism without hints of interface operation followed by a quiz to ensure under-  
574 standing. Participants rewatched the video or consulted the researcher until they could select the correct answers. The  
575 participant's screen was captured throughout the study. The researcher primed the participant that the study aimed to  
576 help local community organizers understand preferences on societal issues to better allocate resources. Participants  
577 were randomly assigned to one of four groups:  
578

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

583 Participants completed the survey independently, without the researcher's presence. They then contacted the  
584 researcher for the NASA-TLX survey, followed by a short audio-recorded semi-structured interview. The session  
585 concluded with a debriefing and a \$15 cash compensation, during which participants were informed of the study goal  
586 on cognitive load and interface design.  
587

588 We made several experimental design choices. First, we selected a between-subject design to minimize study fatigue,  
589 considering the complexity of QS, and avoiding the learning effect that could influence how participants evaluated the  
590 options. Second, we chose the context of public resource allotment, where participants expressed their preferences  
591 regarding their preference across 6 or 24 societal issue options, following the methodology of Cheng et al. [3]. These  
592 issues are relevant to all citizens and effectively demonstrate the need to prioritize limited public resources. We curated  
593 26 societal issues used by Charity Navigator [72] which evaluates over 20,000 charities in the United States. The  
594 interface randomly presents options from this list to participants. Appendix D contains the full list.  
595

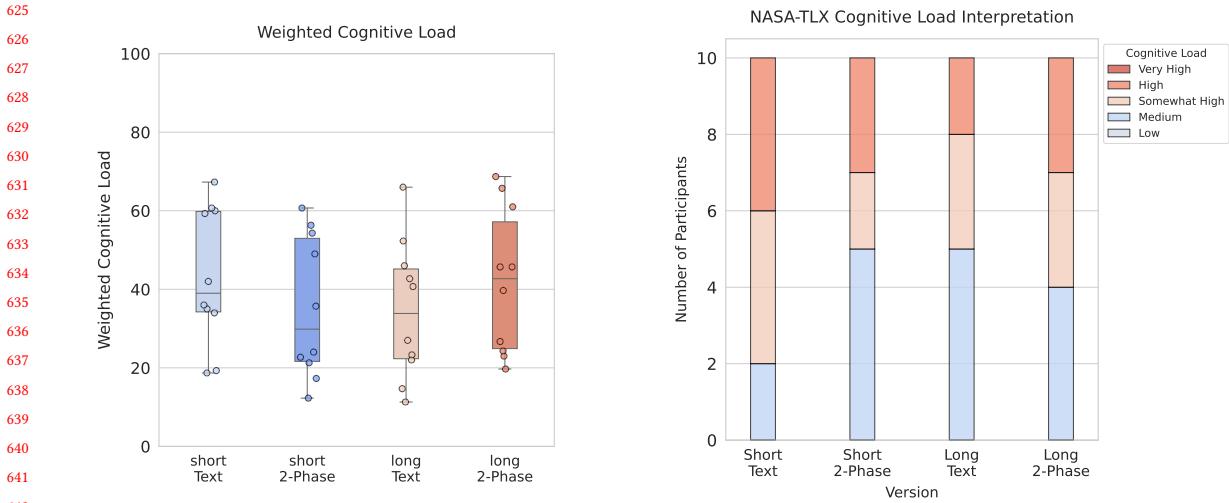
596 We decided to test 6 and 24 options, representing short and long lists, as identifying the 'breaking point' for cognitive  
597 overload would require impractical time and resource commitments. Constant sum surveys and the Analytic Hierarchy  
598 Process (AHP) recommend fewer than ten and seven options, respectively [73, 74, 75]. Miller [76]'s classic work on  
599 cognitive processing capacity and Saaty and Ozdemir [77]'s theoretical proof supported the use of  $7 \pm 2$  items. A  
600 meta-analysis by Chernev et al. [78] identified 6 and 24 as common values for short and long lists in choice overload  
601 studies, rooted in the original experiment by Iyengar and Lepper [54].

602 Finally, we deployed self-report subjective surveys and analytical measures (i.e., time and clickstream data). We  
603 adopted the paper-based weighted NASA Task Load Index (NASA TLX), a widely used multidimensional tool that  
604 averages six subscale scores to represent overall workload after completing a task [79, 80, 81]. NASA-TLX is favored for  
605 its low cost and ease of administration [82], with less variability compared to one-dimensional workload scores [83],  
606 making it suitable for our study. Given the extended nature of QS, we did not choose to measured cognitive load using  
607 performance measures, psychophysiological measures, subjective measures, and analytical measures [82].  
608

## 609 5 Cognitive Load and Sources across Experiment Conditions

610 This section presents cognitive load across experiment groups and the sources contributing to each cognitive load  
611 dimension. Due to the smaller sample size, we focus on descriptive statistics and qualitative assessments. We present  
612 quantitative data from survey tasks and qualitative insights from post-survey interviews.  
613

614 The first author conducted an inductive thematic analysis [84] after transcribing the interviews. Snippets were coded  
615 based on research questions and topics of interest, and similar codes were merged to form themes. When differences  
616



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

(b) NASA-TLX Cognitive Interpretation: More participants in the Short Text Interface, totaling 8, reported a somewhat high or above cognitive load, which is significantly higher compared to the 5 participants who reported similarly for the Short Two-Phase Interface. However, the Long Two-Phase Interface saw slightly more participants, 6 in total, reporting somewhat high or above cognitive load compared to the Long Text Interface.

Fig. 10. This figure shows the box plot results for weighted NASA-TLX scores across experiment groups and participant counts based on individual score interpretations. In 10a, 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 10b, these trends are clearer when NASA-TLX scores are grouped into five tiers.

emerged across experiment conditions and hypotheses were formed, the first author applied a deductive coding process to the relevant text snippets. We begin by presenting an overview of the cognitive load findings.

## 5.1 Overall Cognitive Load

To answer our research question on how the number of options in QS (**RQ1**) and the interface (**RQ2a**) impacted cognitive load, we derive the weighted NASA-TLX scores across the four experiment conditions. We show these results in Figure 10. Weighted NASA-TLX uses a continuous 0-100 score, with higher values indicating greater cognitive load. We use predefined mappings of NASA-TLX scores to cognitive levels: low, medium, somewhat high, high, and very high, as listed by Hart and Staveland [79]. We show value interpretations in Figure 10b.

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

677 We acknowledge that these results may not fully reflect actual cognitive load due to potential noise from factors such  
 678 as small sample size, task nature, or participants' interpretation of the scale. To explore the similarities and sources of  
 679 cognitive load across interfaces (**RQ2b**), we turn to qualitative insights from post-task interviews based on NASA-TLX  
 680 cognitive load dimensions. Among the six NASA-TLX dimensions, we focus on mental demand, temporal demand, and  
 681 frustration while summarizing key findings regarding physical demand, performance, and effort for brevity. Appendix C  
 682 provides further details on all six dimensions.  
 683  
 684

## 685 5.2 Sources of Mental Demand

686 We began by examining mental demand, which refers to the amount of mental and perceptual activity required to  
 687 complete a task. Interview results indicated that the primary drivers of participants' mental demand were *Budget*  
 688 *management* and *Preference construction*.  
 689  
 690

691 5.2.1 *Mental Demand Source #1: Budget management.* 14 participants expressed demand from budgeting within limited  
 692 credit ( S032 *Q [...] for certain societal issues you had to ... take away from other societal issues that you could support.* ,  
 693  $N = 5$  ), tracking remaining credits ( S006 *Q [...] looking at the remaining credits, I'm trying to mentally divide that up*  
 694 *before I start allocating* ,  $N = 10$  ), and maximizing credit use ( S032 *Q [...] I used all the credit that I had available* ,  
 695  $N = 8$  ).  
 696  
 697

698 We categorized budget management-induced mental demand as either operational (single interface-level action,  
 699 e.g., using the last remaining credit) or strategic (higher-level goal, e.g., evenly distributing credits across options).  
 700 Participants who completed longer surveys more frequently reported operational causes, suggesting that an increase in  
 701 survey options induced short-term thinking.  
 702

703 5.2.2 *Mental Demand Source #2: Preference construction.* All but one participant reported increased mental demand  
 704 due to preference construction. We further break it down into three sources: comparative preference evaluation (i.e.,  
 705 evaluating relative importance between options; S002 *Q Figuring out ... how much I prioritize option 1 over option 2* ,  
 706  $N = 16$  ), resource-constraint prioritization (i.e., trading off between options due to resource constraints, S005 *Q [...]*  
 707 *very hard to take decisions ... because I felt that multiple options deserve equal amount of credit ... but you have given very*  
 708 *limited amount of credit* ,  $N = 17$  ), and deciding the exact number of votes ( S023 *Q [...] having to pick how many*  
 709 *upvotes would go to each one* ,  $N = 30$  ).  
 710  
 711

712 Almost all participants mentioned preference construction as a source of mental demand, supporting our claim that  
 713 preference construction when completing QS is a difficult and mentally demanding task. Notably, more participants  
 714 using the text interface reported mental demand from deciding the exact number of votes compared to the two-phase  
 715 interface (18 vs. 12). We conjecture that the first pass on the survey items and organizing them helped participants  
 716 construct preliminary preferences and reduced their mental demands when they started allocating votes.  
 717  
 718

719 In addition, when categorizing preference construction-induced mental demand, participants ( $N = 8$ ) in the long text  
 720 interface tend to consider a smaller scope that focuses on personal relevance. Conversely, participants ( $N = 9$ ) in the  
 721 long two-phase interface considered the broader societal impact and evaluated options more comprehensively. Compare  
 722 the following two quotes, where one focused on adjusting credits between two options and the other reflecting across  
 723 broader societal values:  
 724

725 *Trying to figure out what upvotes I should give [...] went back and forth between those two. [...] it was very mentally tasking for me.*  
 726 *Q S015 (LT)*  
 727

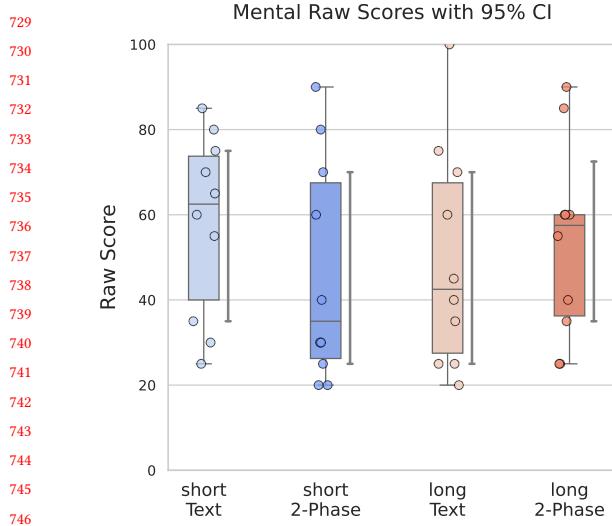


Fig. 11. Mental Demand Raw Score: Across all four experiment groups, participants' reported mental demand is spread across a wide range with many participants experiencing high mental demand.

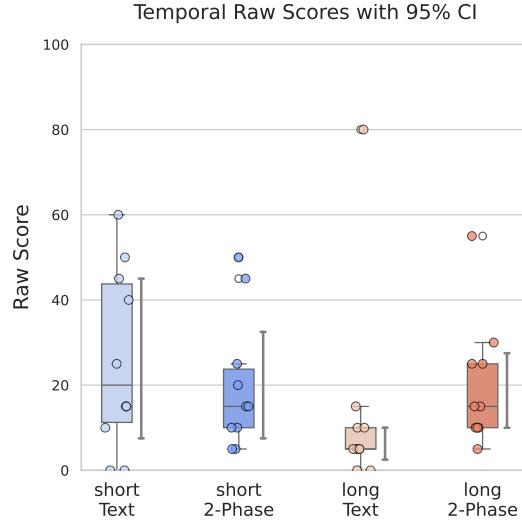


Fig. 12. Temporal Demand Raw Score: The short text interface results in the highest temporal demand, while the long text interface is the lowest. Two-phase interfaces show moderate temporal demand, suggesting that interactive elements allowed participants to pace themselves better.

[...] 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, we did not notice significant differences in mental demand raw values (Figure 11) across the four experiment groups, especially between the interfaces across long QS. However, the experiment groups' mental demands came from different sources.

### 5.3 Sources of Temporal Demand

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

**5.3.1 Two-phase Interface Reduced Temporal Demand on Short QS.** The raw NASA-TLX values in Fig 12 show that participants in the Short Text Interface reported the highest temporal demand. Five participants expressed concerns about time spent on decision-making, feeling themselves invested more time and effort than expected, prompting them to rush. However, the two-phase interface reduced this, with only one participant in the short survey group reporting similar concerns.

**5.3.2 Long QS on Text Interface Showed the Lowest Temporal Demand.** Surprisingly, participants in the long text interface exhibited the lowest temporal demand (Fig. 12) despite making more decisions and operations compared to the short text group. Two possible explanations might explain this counter-intuitive result. First, more participants in

781 the short survey group ( $N = 7$ ) expressed a desire to complete the task efficiently, compared to just one participant  
 782 ( $N = 1$ ) in the long survey group, saying things like:  
 783

784 *I wanna get through things in an efficient manner [...] to move through this quickly and efficiently.*

785 S032 (ST)

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

789 *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.*

790 S035 (LT)

791 We will discuss this possibility further in Section 7.1.  
 792

793 5.3.3 *Two-phase Interface Increased Temporal Demand in Long QS.* Despite the unexpectedly low temporal demand  
 794 in long QS with a text interface, two-phase interface participants have indifferent temporal demand across survey  
 795 lengths. All five participants who mentioned a feeling of time pressure on decision-making in the long two-phase group  
 796 described the pressure affirmatively. This means their pressure stemmed from having too many remaining decisions  
 797 to make ( S022 *So it didn't take too much time, but obviously there were a lot of things to consider, so there was some*  
 798 *temporal demand.* ), not from the time they have already spent (i.e., framed negatively) as that in the short text group.  
 799

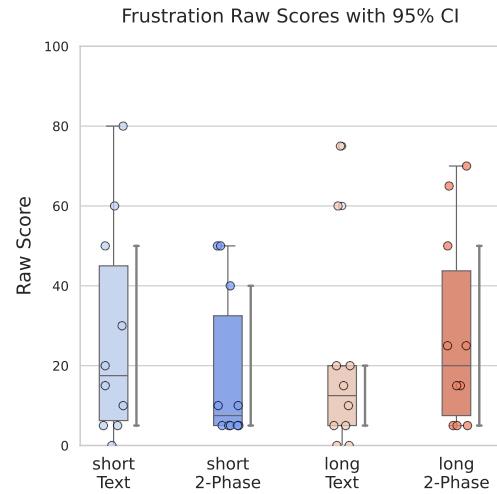
#### 800 5.4 Source of Frustration

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

805 In general, the frustration derived from societal concerns  
 806 did not seem strongly affected by any of the experimental conditions. We saw some discrepancies with respect to operational  
 807 action-driven frustration. The long text interface condition had  
 808 the fewest participants expressing operational frustration, with  
 809 half expressing no frustration, mirroring the trends in the actual  
 810 scores (Figure 13). Similar to the finding that the long text group  
 811 has the lowest temporal demand, this is counter-intuitive as  
 812 more options and dense text are known to lead to more frustration  
 813 in interface design [85]. Participants engaging in sacrificing  
 814 behaviors in the long text interface may explain this phenomenon – prior literature [86, 87] indicates that satisficers tend  
 815 to be less frustrated and happier than maximizers.

#### 816 5.5 Physical Demand, Effort and Performance

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



820 Fig. 13. Frustration Raw Score: Participants other than  
 821 the long text interface highlighted several operational  
 822 tasks that led to frustration. All groups share causes  
 823 from strategic planning.

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

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

## 5.6 Summary across all cognitive load dimensions

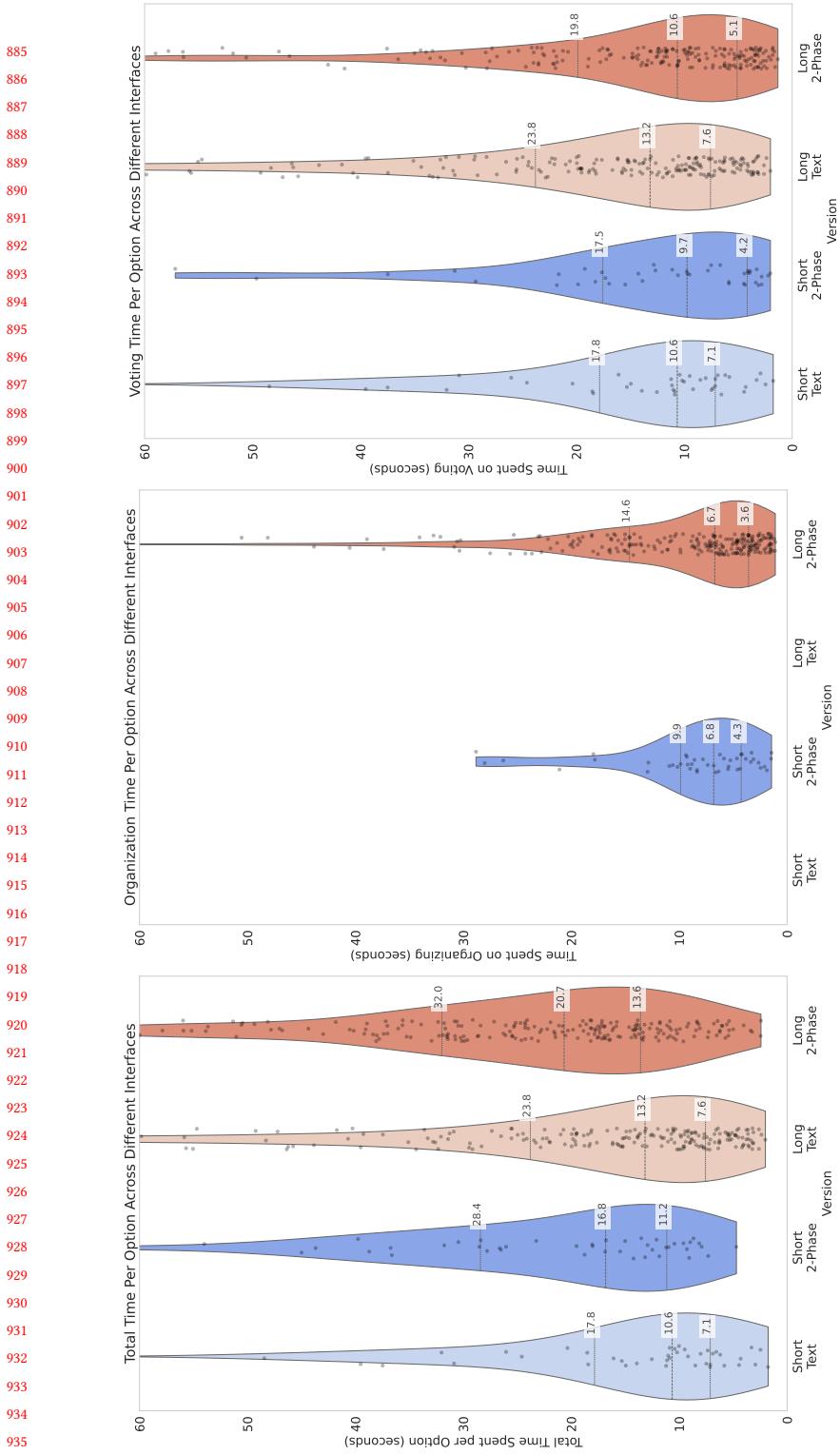
Overall, participants using the two-phase interface tended to think more comprehensively and critically, while those using the text interface focused more on operational tasks. This is reflected by *mental demand* findings where participants using the text interface reported greater difficulty in determining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

## 6 Behavior Results

To understand how interfaces and options influenced survey response behaviors (RQ3), we investigate time-to-action and remaining credit differences across experiment conditions. Time-to-action is a widely used metric in decision sciences, where longer decision time often indicates more complex cognitive processing [88]. Additionally, resource allocation strongly influences decision-making. Cheng et al. [3] showed that the number of given credits influences the validity of QV. Decision science studies like Shah et al. [89] and [90] showed how scarcity influences decisions, increases risk aversion, and adds cognitive load. These measures serve as proxies for participant behavior, and all analyzed data is publicly available<sup>3</sup> for transparency and to facilitate further research.

---

<sup>3</sup>link-to-github



- (a) Total Time per option: We identified that the two-phase interface skewed slightly higher than the text interface, as expected. This discrepancy can be attributed to the extra organization step required in the two-phase interface, leading to slightly longer overall completion time per option.
- (b) Organization Time per option: Only the two-phase interface includes an organization phase, hence the other experimental conditions do not exhibit any accumulated organization time. This figure reinforces that organization time was isolated to the two-phase design.
- (c) Voting Time per option: We observe statistically significant faster voting times for the long QS in the two-phase interface. This suggests that the two-step design improves voting efficiency, especially for longer surveys.

Fig. 14. Time per option across all experiment conditions. The violin plots visualize time spent per option across the experimental conditions, with dots representing total time per option and horizontal lines indicating the median and interquartile ranges. While total time was slightly higher for the two-phase interface, the distinct phases (organization and voting) allowed for a more structured approach, particularly benefiting longer surveys.

937 **6.1 Time Spent per Option**

938 Our first analysis focuses on understanding how much time participants spent per option across different stages and  
 939 experiment conditions. Based on the QS system log, we extracted the following detail: *the option* involved in the  
 940 interaction, *the type of interaction* (such as updating a certain number of votes), and *the time* between this interaction  
 941 and the previous one. Each dot on Figure 14 is a specific type of time a participant spent for one option.  
 942

943 *Total time.* Total time is the aggregate of all time spent on each option across both phases. Each dot on Figure 14a  
 944 visualizes the total time. Participants spent slightly more time per option on the two-phase interface than the text  
 945 interface. A non-parametric Mann-Whitney U test showed a small effect size (long QS:  $p < 0.0000001$ , Rank-biserial:  
 946  $-0.304$ , 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  
 947 interface has an additional step of organizing the options.  
 948

949 We further break down the total time spent into organization time and voting time. To minimize noise, we intentionally  
 950 dropped all the time participants spent on the first option in the organization phase or voting phase. The goal is to  
 951 exclude time spent on reading the prompt, forming their preference, or understanding the interface.  
 952

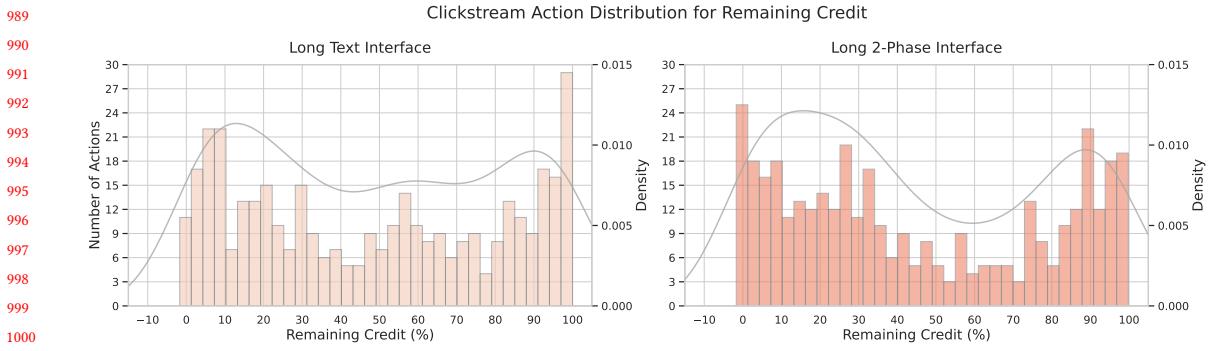
953 *Organization time.* Organization time covers both placing options into categories and the drag-and-drop time during  
 954 the organization phase. Illustrated in Figure 14b, we observed minimal difference in organization time per option  
 955 between short and long surveys, as the interface shows options one at a time for categorization. It also suggests that  
 956 even for longer surveys, the organization functionalities did not significantly impact the time participants needed for  
 957 an option.  
 958

959 *Voting time.* Voting time strictly refers to the time participants took to update vote values for each option. As shown  
 960 in Figure 14c, participants spent significantly less time voting on the two-phase interface than on the text interface  
 961 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  
 962 short survey ( $p > 0.4$ , Power=0.051). This supports our hypothesis that the two-step design in the two-phase interface  
 963 facilitates more efficient decision-making, especially in longer surveys.  
 964

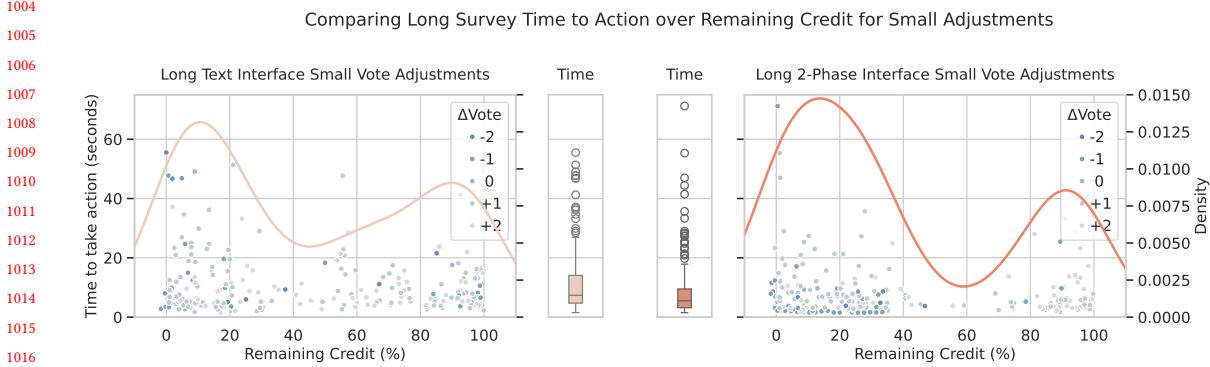
965 **6.2 Budget and Voting Behaviors**

966 We further breakdown and highlight key differences of how long QS participant's voting behaviors when credits changed  
 967 their voting behaviors with detailed analysis in Appendix F. Figure 15a first shows the number of vote adjustments  
 968 at a given remaining budget across the two interfaces. We then plotted the vote adjustments of two or fewer votes,  
 969 which is 10% of the possible values one can choose among the maximum of 21 votes in Figure 15b. A kernel density  
 970 estimate (KDE) plot is provided to visualize the trends and compare how it changed when we plotted only small vote  
 971 adjustments.  
 972

973 In long surveys, participants exhibited more actions both when the budget was abundant and when it began to run  
 974 out. This pattern was more pronounced with the long two-phase interface. In fact, the bimodal distribution is more  
 975 pronounced in the two-phase interface. This suggests that participants make small adjustments both at the beginning  
 976 and toward the end of the QS. However, the two-phase interface shows more frequent and faster edits towards the  
 977 end. Visually, dots are more clustered in the long two-phase interface for small vote adjustments compared to the long  
 978 text interface. The Mann-Whitney U Test on the time spent on small vote adjustments showed significant differences  
 979 ( $U = 13037$ ,  $p < 0.001$ ), with a small effect size (Rank-biserial: 0.227, Cohen's d: 0.195) and a power of 0.381. This  
 980 indicates that participants had a clearer idea of how to distribute their credits across the options.  
 981



989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001 (a) This plot counts the number of voting actions when there are  $x$  percentages of credits remaining. A KDE plot is provided to help  
1002 better understand the action distribution.  
1003



1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017 (b) This plot further separates participants' interaction behavior based on the number of votes participants adjusted. We observed a  
1018 bimodal interaction pattern across long QS when small vote adjustments are made.  
1019

1020 Fig. 15. Comparison of voting actions and participant behavior in different survey interfaces. Subplot (a) shows the overall distribution  
1021 of actions based on the remaining credit, while subplot (b) further differentiates the interaction based on the size of vote adjustments.  
1022  
1023  
1024  
1025

1026 Five participants highlighted how the interface supported their incremental iterative approach during the interview  
1027 whom all used the two-phase interface. As one participant pointed out:

1028 *I like the fact that it remembers everything that you know. [...] that's very important is that it's an iterative process.* Q S019 (LI)

1029  
1030 **In summary**, participants spent more time on the two-phase interface compared to the text interface in both short  
1031 and long surveys. Across the two-phase interfaces, organization time remained consistent. While voting time did not  
1032 differ between interfaces for the short survey, participants voted more quickly on the two-phase interface in the long  
1033 survey, confirming the hypothesis that the two-step design enhances decision-making efficiency. Voting behaviors  
1034 indicated more frequent actions when the budget was abundant and nearly exhausted, particularly in the long two-phase  
1035 interface. Additionally, the analysis revealed more frequent and faster small vote adjustments towards the end of the  
1036 QS in the two-phase interface, demonstrating an iterative and incremental approach.  
1037  
1038  
1039

## 1041 7 Discussion and Future Works

1042 In the discussion section, we interpret results related to cognitive load and survey respondent behaviors. Specifically,  
1043 we aim to understand why results in Section 5 revealed that long text interface participants did not experience higher  
1044 cognitive load, contrary to expectations referencing prior literature and insights from Section 6.

1045 We focus on three key topics: what two-phase interface elements influenced behavior, how these elements supported  
1046 preference construction, and the remaining design challenges. Additionally, we provide recommendations for using  
1047 QS and suggest design improvements. We concluded that the two-phase interface prevented satisficing behaviors and  
1048 promoted strategic, holistic thinking, unlike the text-based interface, which leaned toward operational tasks. Behavioral  
1049 analysis showed that long two-phase interface participants made frequent, small updates, shifting their cognitive focus.  
1050

### 1051 7.1 Interpretation of results: Two-phase interface limits satisficing during cognitive overload

1052 Participants using the long two-phase interface reported slightly higher cognitive load compared to those using the  
1053 long text interface. In comparison, the short two-phase interface resulted in a lower cognitive load. This suggests that  
1054 the two-phase interface alone does not inherently increase cognitive load.

1055 One explanation is that the two-phase interface reduces cognitive load in the short QS but not in the longer one. In  
1056 longer surveys, the interactive nature of the two-phase interface may require participants to perform more operations  
1057 without significantly altering their decision-making processes. However, our findings suggest that participants using  
1058 the two-phase interface in longer surveys responded more efficiently. Across both short and long surveys, participants  
1059 using the two-phase interface demonstrated iterative and efficient fine-tuning preferences, indicating deeper survey  
1060 engagement. Therefore, we reject the claim that the two-phase interface increases cognitive load in the long survey.

1061 An alternative explanation is that while the two-phase interface limits satisficing during cognitive overload, the  
1062 complexity of the long QS may have counteracted this effect, preventing a full reduction in cognitive load. This suggests  
1063 that long QS participants resorted to satisficing behaviors to manage their cognitive overload. Qualitative results support  
1064 this explanation: participants using the two-phase interface engaged in broader, more strategic considerations, while  
1065 those using the text interface focused on operational tasks. For instance, participants using the long two-phase interface  
1066 experienced less cognitive burden from precise voting and managed their time better. Overall, the two-phase interface  
1067 participants demonstrated deeper and more critical thinking, as they engaged in more strategic decision-making.

1068 In contrast, participants using the long text interface satisficed due to cognitive overload from having to decide  
1069 and allocate more credits for more options. However, fewer participants in this group reported high cognitive load  
1070 compared to those in the short-text interface. This group also experienced the least temporal demand and frustration  
1071 with operational tasks, despite spending a similar amount of time per option as participants using the text interface.  
1072 These counterintuitive findings pointed to the remaining plausible explanation that long-text interface participants  
1073 used satisficing to manage the overload.

1074 **In summary**, the interactive components of the two-phase interface likely prevented mental shortcuts [53, 9, 10,  
1075 11], resulting in a cognitive *shift* towards deeper reflection and decision-making. Thus, while QS with many options  
1076 might lead to satisficing, the two-phase interface redirects participants' attention, facilitating the decision-making  
1077 process that promotes comprehensive preference construction. In the following section, we examine these specific  
1078 elements that guided participants to achieve this.

## 1093      7.2 Bounded rationality and interface design

1094      Bounded rationality emerged as a core theme in participants' responses, highlighting how cognitive limitations lead to  
 1095      sub-optimal decision-making due to their inability to process all available information [9]. This often led to *satisficing*  
 1096      behaviors, where participants settle for *good enough* but not *optimal* decisions [91].

1097      For instance, S036 described making meaningful choices while acknowledging the limits of their effort:

1098      [...] 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  
 1099      diminishing returns. I tried to think of enough things [...] and then move on. [...] I felt like that (the response) was satisfied, but not  
 1100      perfect. Cause perfect is not a reality.

1102      ↗ S036 (ST)

1103      This illustrates typical satisficing decision-making, where participants settle for suboptimal choices. In long QS, the  
 1104      significant increase in decision points—due to the numerous vote and credit options—was so overwhelming to some  
 1105      that even features like pre-calculated vote-credit values in drop-down menus provided relief, as they helped avoid the  
 1106      cognitively demanding task of searching for bounds.

1107      Participants often relied on *heuristics* [92] and *defaults* [64], behaviors less common among two-phase interface  
 1108      participants. Presenting one option at a time reduced reliance on defaults and encouraged deeper reflection, as illustrated  
 1109      by S013, who highlighted how the organization phase supported their option preferences construction:

1110      [...] it (organization phase) gives you time to just focus on that single thing and rank it based on how you feel at that moment.

1112      ↗ S013 (SI)

1113      Conversely, text interface participants, like S003, described how default placements influenced their decisions:

1114      Honestly, if medical research [...] was the first one I saw, I think it would automatically give it a lot more.

1116      ↗ S003 (ST)

1117      The three key elements of the organization phase—presenting options one at a time, grouping them into categories,  
 1118      and enabling drag-and-drop—worked together to structure participant preferences. These elements aligned with  
 1119      cognitive strategies like *problem decomposition* [93] and *dimension reduction* which reduces cognitive overload. Bounded  
 1120      rationality illuminates the importance of decision-making support interfaces rather than being a critique of human  
 1121      behaviors. One participant explained how the organization phase broke down complex decisions into manageable steps:

1122      [...] being able to have a preliminary categorization of all the topics. First, it introduced me to all the topics, [...] to think about  
 1123      and process [...] being able to digest all the information prior to actually allocating the budget or completing the quadratic survey.

1124      ↗ S009 (LT)

1125      Participants using the two-phase interface, especially in the long version, organized options along dimensions such  
 1126      as topics (e.g., health vs. humanitarian) and preferences (positive vs. negative) before voting. Others expressed that  
 1127      the upfront introduction of all options and the ability to rank and group them helped manage their cognitive load  
 1128      effectively. In contrast, almost half of the participants using the long text interface, like S028, expressed a desire for  
 1129      features that can help reduce the decision space when responding to the QS, further supporting the importance of these  
 1130      organizational design elements:

1131      Because with this many (options), especially when I'm thinking ... Ok, where was (the option)... Where was (the option) you know?  
 1132      Oh, that's right. Maybe I could give another up another upvote to the, you know whatever [...]

1134      ↗ S028 (LT)

1135      This quote reflected participants' need to manually track and revisit options, which occupies cognitive load, without  
 1136      a more structured interface.

1137      **In summary**, individual's bounded rationality encouraged participants to exhibit *satisficing* behaviors, *heuristics*,  
 1138      and *defaults* when responding to QS. Showing all options upfront, one at a time, and repositioning options based on partic-  
 1139      ipants' rough preferences prevented participants from using defaults and heuristics. The two-phase organization actively

1145 scaffolds participants' decision-making process, supporting efficiency *problem decomposition* and *dimension reduction*.  
 1146 Together, these elements in the two-phase interface design prevented *satisficing* behaviors and supported participants  
 1147 in making more informed decisions through a more strategic planning and holistic thinking process.  
 1148

### 1149 7.3 Construction of Preference on Quadratic Survey

1151 Completing QS is a series of difficult decision tasks Lichtenstein and Slovic [6]. Svenson [60]'s differentiation and  
 1152 consolidation theory helps explain how participants process these decisions. The decision process begins with dif-  
 1153 ferentiation, where participants identify differences and eliminate less favorable options, followed by consolidation,  
 1154 which strengthens their commitment to selected choices. This theory aligns with how the two-phase interface helps  
 1155 participants decompose options into categories, effectively reducing decision complexity.  
 1156

1157 Participants started by constructing preferences *in situ*, especially regarding options they hadn't previously consid-  
 1158 ered:  
 1159

1160 [...] 'Oh, there are other aspects that I never care about.' And actually ... some people care <an option>. Sure. Why? Why (should) I  
 1161 spend money on that?  
 1162 ☰ S037 (LI)

1163 Those using the text interface, lacking the interactive tools, found it challenging to facilitate differentiation as S025  
 1164 noted:  
 1165

1166 I would like to be able to like, click and drag the categories themselves so I could maybe reorder them to like my priorities. [...] make  
 1167 myself categories and subcategories out of this list ... If I could organize it.  
 1168 ☰ S025 (LT)

1169 In contrast, the two-phase interface allowed participants to express at least one dimension of differentiation more  
 1170 easily. The drag-and-drop feature helped blend this differentiation into the consolidation phase. Not only does partic-  
 1171 ipants drag-and-drop options post voting to reflect and assure a correct vote allocation, it also enables participants,  
 1172 like S039, to make fine-grain comparisons between options:  
 1173

1174 I think the system was actually really helpful because I could just drag them. [...] I can really compare them, I can drag this one up  
 1175 here, and then compare it to the top one [...]  
 1176 ☰ S039 (SI)

1177 The bi-modal behavior observed in the long interactive interface participants provided in the results aligns with the  
 1178 differentiation and consolidation framework. Participants in the two-phase interface began differentiating options earlier  
 1179 allowing them to later adjust fine-grain votes. The faster and smaller vote updates indicated participants consolidating.  
 1180 The less prominent bi-modal behavior from the long text interface participants implied that the interface guided this  
 1181 decision framework as participant 037 explained:  
 1182

1183 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  
 1184 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  
 1185 second part.  
 1186 ☰ S037 (LI)

1187 These evidences explain how the organization phase and the drag-and-drop features supported differentiation and  
 1188 consolidation, scaffolding a decision-making framework enabling deeper engagement.  
 1189

1190 **In summary**, participants construct their preferences as they complete QS. We observed behaviors and qualitative  
 1191 insights that align with the differentiation and consolidation theory in decision-making. Our interface scaffolded many  
 1192 of the differentiation stages through pre-voting organization and some consolidation phases through drag-and-drop,  
 1193 explaining how the two-phase approach supports preference construction to yield effective QS responses.  
 1194

## 1197 7.4 Opportunities for better budget management

1198 Budget management is a recurring theme in participant interviews. While they appreciated the automatic calculation  
 1199 feature in modern QV interfaces, we identified three challenges for future QS interfaces: *cognitive load, the cold-start*  
 1200 *problem, and navigating between budget, votes, and outcome.*

1203 *7.4.1 Automatic calculation is critical.* Over one-third of participants ( $N = 14$ ) from all four experiment conditions  
 1204 emphasized the importance of automated calculation for deriving costs and tracking expenditures. For example:

1206 *I thought I have [...] (to) do all the numbers or calculations myself as a part of checking my ability of doing mathematics. [...] I said*  
 1207 *that credit summary to be very specific. The credit summary section was really wonderful in doing all the calculations on that end.*

1208 S005 (LT, keeping track of spent)

1210 The quotes marked the importance that QS must be facilitated by computer-supported interfaces.

1212 *7.4.2 The coldstart problem.* We notice from the study that one of the biggest challenges for participants is deciding  
 1213 'how many votes' to start with. This challenge pertains to the initial vote, not the relative vote. Some participants began  
 1214 by equally distributing their credits to all options and then made adjustments. Others established 1, 2, and 3 votes as  
 1215 starting points. A small handful surprisingly used the tutorial's example of 4 upvotes as their anchor.

1217 This arbitrary voting decision echoes discussions in prior literature about the existence of an absolute value for  
 1218 individuals. Coherent arbitrariness [94], similar to the anchoring effect in marketing, refers to participants' willingness  
 1219 to allocate votes, which can be influenced by an arbitrary value. However, the ordinal utility remains intact among the  
 1220 set of preferences.

1223 *7.4.3 Navigating Between Budget, Votes, and Actual Impact.* The third challenge is participants' confusion between  
 1224 budget, votes, and outcomes, despite understanding their definitions. One participant stated:

1226 *[...] 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*  
 1227 *organization or your participant. You have X amount of dollars you need to. You can only distribute X amount of dollars to these*  
 1228 *causes. So you have to figure out which ones get the most, which ones don't get as much. [...]*

1229 Interviewer: [...] Do you feel that the more votes you're giving to a cause you're actually spending more on it?

1230 Yeah.

S003 (ST)

1232 Participants like S003 bypassed the quadratic formulation, directly translating votes to resource allocation. While this  
 1233 does not invalidate the power of the quadratic mechanism, it causes frustration and friction for participants to construct  
 1234 a clear picture of how to make voting decisions. Future interfaces should better communicate these relationships to  
 1235 facilitate respondents' trade-offs.

1237 **In summary**, while the interface supports budget management through automated cost calculation, participants still  
 1238 face cognitive load from managing the budget. The cold-start problem and the confusion between budget, votes, and  
 1239 actual impact are open questions for future research. These challenges highlight the need for better budget management  
 1240 support to complete the QS interface.

## 1243 7.5 Quadratic Survey Usage, Design Recommendations and Future Work

1244 With a deeper understanding of how survey respondents interact with QS and the sources of cognitive load, we  
 1245 recognize that while this current interface may not significantly reduce cognitive load, it represents a crucial step  
 1246 toward constructing better interfaces to support individuals responding to QS. In this subsection, we outline usage

1249 and design recommendations applicable to all applications using the quadratic mechanism and highlight directions for  
1250 future work.  
1251

1252 *7.5.1 Usage Recommendation: QS for Critical Evaluations.* Our study highlighted the complex cognitive challenges  
1253 and in-depth consideration required when ranking and rating options using QS, even in a short survey. Similar to  
1254 survey respondents needing to make trade-offs across options, researchers and agencies seeking additional insights and  
1255 alignment with respondent preferences must ensure that survey respondents have the cognitive capacity to complete  
1256 such surveys rigorously. QS should be designed for critical evaluations, such as investment decisions, or situations  
1257 where participants have ample time to think and process the survey. For instance, revealing the options ahead of time  
1258 can aid in preference construction.  
1259

1260

#### 1261 *7.5.2 Design Recommendations.*

1262

1263 *Use Organization Phases for Quadratic Mechanism Applications.* Our study demonstrated that preference construc-  
1264 tion can shift from operational to strategic and higher-level causes. An additional organizational phase with direct  
1265 manipulation capability allows survey respondents to engage in higher-level critical thinking. We believe this approach  
1266 should extend beyond QS to other ranking-based surveying tools, such as rank-choice voting and constant sum surveys.  
1267 Further research should examine how implementing such functionality alters survey respondents' mental models.  
1268

1269 *Facilitate Differentiation through Categorization, Not Ranking.* Participants in our study were less inclined to provide  
1270 a full rank unless necessary. The final 'rank' of option preferences often emerged as a byproduct of their vote allocation,  
1271 constructed in situ. Therefore, for survey designs to be effective in constructing preferences, it is more important to  
1272 facilitate differentiation than to focus on direct manipulation solely for fine-tuning. Emphasizing categorization can  
1273 better support participants in articulating their preferences.  
1274

1275 *7.5.3 Future Work: Support for Absolute Credit Decision.* Deciding the absolute amount of credits in QS is highly  
1276 demanding. Designing interfaces and interactions that address the cold start challenge and help participants decide the  
1277 absolute vote value while considering ways to limit direct influences remains an open question. Future research should  
1278 explore innovative solutions to support participants in making these complex decisions effectively.  
1279

1280 By implementing these recommendations and pursuing future research directions, we can improve the usability and  
1281 effectiveness of QS and other quadratic mechanism-powered applications, ultimately aiding respondents in making  
1282 more informed and accurate decisions.  
1283

1284

## 1285 **8 Limitations**

1286

1287 Evaluating the QS interface is challenging due to its novelty. During the study, we identified several limitations that  
1288 require further research.  
1289

1290 *Understanding results influence on decision-makers.* Further research is required to understand how the QS interface  
1291 impacts decision-makers and broader societal resource distributions. Since QS is still in its early stages, we prioritize its  
1292 widespread adoption and usage before attempting a comprehensive assessment of its influence on decision-making.  
1293 Future studies will examine how decision-makers interpret and use QS data, as well as its broader implications for  
1294 societal decisions.  
1295

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

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

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

## 9 Conclusion

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

## References

- [1] 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](https://doi.org/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](https://doi.org/10.1145/3613904.3642397).
- [3] 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](https://doi.org/10.1145/3449281).

- 1353 [4] Theodore Groves and John Ledyard. 1977. Optimal Allocation of Public Goods: A Solution to the "Free Rider" Problem. *Econometrica*, 45, 4, 783–809.  
 1354 JSTOR: 1912672. doi: 10.2307/1912672.
- 1355 [5] David Quarfoot, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein, and Ellen Konar. 2017. Quadratic voting in the wild: real  
 1356 people, real votes. *Public Choice*, 172, 1-2, 283–303.
- 1357 [6] Sarah Lichtenstein and Paul Slovic, eds. 2006. *The Construction of Preference*. (1. publ ed.). Cambridge University Press, Cambridge.
- 1358 [7] Adam Rogers. 2019. Colorado Tried a New Way to Vote: Make People Pay—Quadratically | WIRED. *Wired*, (Apr. 2019). Retrieved June 22, 2024  
 1359 from.  
 1360 [8] Internet Team. [n. d.] Taiwan Digital Minister highlights country's use of technology to bolster democracy in FT interview. [https://www.roctaiwan.org/uk\\_en/post/6295.html](https://www.roctaiwan.org/uk_en/post/6295.html). () Retrieved June 13, 2024 from.
- 1361 [9] Herbert A. Simon. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69, 1, 99–118. JSTOR: 1884852. doi:  
 1362 10.2307/1884852.
- 1363 [10] John W. Payne, James R. Bettman, and Eric J. Johnson. 1988. Adaptive strategy selection in decision making. *Journal of Experimental Psychology:  
 1364 Learning, Memory, and Cognition*, 14, 3, (July 1988), 534–552. doi: 10.1037/0278-7393.14.3.534.
- 1365 [11] Amos Tversky and Daniel Kahneman. [n. d.] Judgments of and by Representativeness.
- 1366 [12] Erik J Engstrom and Jason M Roberts. 2020. *The Politics of Ballot Design: How States Shape American Democracy*. Cambridge University Press.
- 1367 [13] Bert Weijters, Elke Cabooter, and Niels Schilwaeert. 2010. The effect of rating scale format on response styles: The number of response categories  
 1368 and response category labels. *International Journal of Research in Marketing*, 27, 3, (Sept. 2010), 236–247. doi: 10.1016/j.ijresmar.2010.02.004.
- 1369 [14] N. D. Kieruj and G. Moors. 2010. Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of  
 1370 Public Opinion Research*, 22, 3, (Sept. 2010), 320–342. doi: 10.1093/ijpor/edq001.
- 1371 [15] Vera Toepoel, Brenda Vermeeren, and Baran Metin. 2019. Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on  
 1372 Substantive Response, Questionnaire Experience and Response Time. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*,  
 1373 142, 1, (Apr. 2019), 57–74. doi: 10.1177/0759106319834665.
- 1374 [16] Habiba Farzand, David Al Baiaty Suarez, Thomas Goodge, Shaun Alexander Macdonald, Karola Marky, Mohamed Khamis, and Paul Cairns. 2024.  
 1375 Beyond Aesthetics: Evaluating Response Widgets for Reliability & Construct Validity of Scale Questionnaires. In *Extended Abstracts of the 2024  
 1376 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.
- 1377 [17] Christian Jilek Paula Gauselmann Yannick Runge and Tobias Tempel. 2023. A relief from mental overload in a digitalized world: How context-  
 1378 sensitive user interfaces can enhance cognitive performance. *International Journal of Human–Computer Interaction*, 39, 1, 140–150. eprint:  
 1379 <https://doi.org/10.1080/10447318.2022.2041882>. doi: 10.1080/10447318.2022.2041882.
- 1380 [18] A Norman Donald. 2013. *The Design of Everyday Things*. MIT Press.
- 1381 [19] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2021. To reuse or not to reuse? A framework and system for evaluating summarized  
 1382 knowledge. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW1, (Apr. 2021). doi: 10.1145/3449240.
- 1383 [20] Benjamin Lafreniere, Andrea Bunt, and Michael Terry. 2014. Task-centric interfaces for feature-rich software. In *Proceedings of the 26th Australian  
 1384 Computer-Human Interaction Conference on Designing Futures: The Future of Design* (OzCHI '14). Association for Computing Machinery, New York,  
 1385 NY, USA, 49–58. doi: 10.1145/2686612.2686620.
- 1386 [21] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator chatbot for deliberative discussion: Effects of discussion structure  
 1387 and discussant facilitation. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW1, (Apr. 2021). doi: 10.1145/3449161.
- 1388 [22] Amy X. Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proc. ACM Hum.-Comput.  
 1389 Interact.*, 2, CSCW, (Nov. 2018). doi: 10.1145/3274465.
- 1390 [23] Steven P Lalley, E Glen Weyl, et al. 2016. Quadratic voting. Available at SSRN.
- 1391 [24] Eric A Posner and E Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.
- 1392 [25] Ryan Naylor et al. 2017. First year student conceptions of success: What really matters? *Student Success*, 8, 2, 9–19.
- 1393 [26] Charlotte Cavaille and Daniel L Chen. [n. d.] Who Cares? Measuring Preference Intensity in a Polarized Environment.
- 1394 [27] Vitalik Buterin, Zoë Hitzig, and E. Glen Weyl. 2019. A Flexible Design for Funding Public Goods. *Management Science*, 65, 11, (Nov. 2019), 5171–5187.  
 1395 doi: 10.1287/mnsc.2019.3337.
- 1396 [28] Luis Mota Freitas and Wilfredo L. Maldonado. 2024. Quadratic funding with incomplete information. *Social Choice and Welfare*, (Feb. 2024). doi:  
 1397 10.1007/s00355-024-01512-7.
- 1398 [29] Tobin South, Leon Erichsen, Shrey Jain, Petar Maymounkov, Scott Moore, and E. Glen Weyl. 2024. Plural Management. SSRN Scholarly Paper.  
 1399 Rochester, NY, (Jan. 2024). doi: 10.2139/ssrn.4688040.
- 1400 [30] 2023. Gov4git: A Decentralized Platform for Community Governance. (Mar. 2023). Retrieved June 13, 2024 from.
- 1401 [31] Syavash Nobarany, Louise Oram, Vasantha Kumar Rajendran, Chi-Hsiang Chen, Joanna McGrenere, and Tamara Munzner. 2012. The design space  
 1402 of opinion measurement interfaces: exploring recall support for rating and ranking. In *Proceedings of the SIGCHI Conference on Human Factors in  
 1403 Computing Systems*, 2035–2044.
- 1404 [32] Paul Van Schaik and Jonathan Ling. 2007. Design parameters of rating scales for web sites. *ACM Transactions on Computer-Human Interaction  
 (TOCHI)*, 14, 1, 4–es.

- [33] Jing Wei, Weiwei Jiang, Chaofan Wang, Difeng Yu, Jorge Goncalves, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding how to administer voice surveys through smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 6, CSCW2, (Nov. 2022). doi: [10.1145/3555606](https://doi.org/10.1145/3555606).
- [34] Aman Khullar et al. 2021. Costs and benefits of conducting voice-based surveys versus keypress-based surveys on interactive voice response systems. In *Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies (Compass '21)*. Association for Computing Machinery, New York, NY, USA, 288–298. doi: [10.1145/3460112.3471963](https://doi.org/10.1145/3460112.3471963).
- [35] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland UK, (May 2019), 1–12. doi: [10.1145/3290605.3300316](https://doi.org/10.1145/3290605.3300316).
- [36] Martin Feick, Niko Kleer, Anthony Tang, and Antonio Krüger. 2020. The virtual reality questionnaire toolkit. In *Adjunct Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 68–69.
- [37] 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](https://doi.org/10.1016/j.ijresmar.2020.04.002).
- [38] 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](https://doi.org/10.1080/08898480.2018.1439245).
- [39] 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](https://doi.org/10.1215/0003055X-95-4-793).
- [40] Dana Chisnell. 2016. Democracy Is a Design Problem. 11, 4.
- [41] 2015. Designing usable ballots | Center for civic design. <https://civicdesign.org/fieldguides/designing-usable-ballots/>. (June 2015). Retrieved June 17, 2024 from.
- [42] Whitney Quesenberry. 2020. Opinion | Good Design Is the Secret to Better Democracy. *The New York Times*, (Oct. 2020). Retrieved June 17, 2024 from.
- [43] 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](https://doi.org/10.1145/1357054.1357195).
- [44] 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](https://doi.org/10.1145/2858036.2858567).
- [45] 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)*.
- [46] 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](https://doi.org/10.1145/1502650.1502727).
- [47] 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](https://doi.org/10.1075/idj.20.3.01gil).
- [48] 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](https://doi.org/10.1016/j.ijhcs.2008.09.010).
- [49] 2024. RadicalxChange/quadratic-voting. RadicalxChange. (May 2024). Retrieved June 17, 2024 from.
- [50] [n. d.] Read the Whitepaper | Gitcoin. <https://www.gitcoin.co/whitepaper/read/>. (). Retrieved June 17, 2024 from.
- [51] [n. d.] About RxC. <https://www.radicalxchange.org/wiki/about/>. (). Retrieved June 17, 2024 from.
- [52] yehjxraymond. 2024. Yehjxraymond/qv-app. (Mar. 2024). Retrieved June 17, 2024 from.
- [53] Kahneman Daniel. 2017. *Thinking, Fast and Slow*.
- [54] 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.
- [55] 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.
- [56] 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](https://doi.org/10.1080/00049537308255849).
- [57] Peter Coy. 2019. A New Way of Voting That Makes Zealousy Expensive - Bloomberg. *Bloomberg*, (May 2019). Retrieved Dec. 16, 2023 from.
- [58] 2022. Quadratic Voting Frontend. Public Digital Innovation Space. (Jan. 2022). Retrieved Dec. 16, 2023 from.
- [59] 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](https://doi.org/10.1016/S0166-4115(08)62243-8).

- [60] 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](https://doi.org/10.1016/0001-6918(92)90044-E).
- [61] 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](https://doi.org/10.1007/978-1-4612-4798-2_7).
- [62] John Sweller. 2011. Cognitive Load Theory. In *Psychology of Learning and Motivation*. Vol. 55. Elsevier, 37–76. doi: [10.1016/B978-0-12-387691-1.0002-8](https://doi.org/10.1016/B978-0-12-387691-1.0002-8).
- [63] 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](https://doi.org/10.1002/bdm.1897).
- [64] 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.
- [65] 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.
- [66] Jon A Krosnick, Charles M Judd, and Bernd Wittenbrink. 2018. The measurement of attitudes. In *The Handbook of Attitudes*. Routledge, 45–105.
- [67] Jerry P Timbrook. 2013. *A Comparison of a Traditional Ranking Format to a Drag-and-Drop Format with Stacking*. PhD thesis. University of Dayton.
- [68] 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.
- [69] 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](https://doi.org/10.1145/3375629).
- [70] Robert Ferber. 1952. Order Bias in a Mail Survey. *Journal of Marketing*, 17, 2, 171–178. JSTOR: [1248043](https://doi.org/10.2307/1248043). doi: [10.2307/1248043](https://doi.org/10.2307/1248043).
- [71] M. P. Couper. 2001. Web survey design and administration. *Public Opinion Quarterly*, 65, 2, 230–253. doi: [10.1086/322199](https://doi.org/10.1086/322199).
- [72] 2023. Charity Navigator. <https://www.charitynavigator.org/index.cfm?bay=search.categories>. (May 2023). Retrieved Dec. 16, 2023 from.
- [73] 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).
- [74] Thomas L. Saaty and Kirti Peniwati. 2013. *Group Decision Making: Drawing Out and Reconciling Differences*. RWS Publications, (Nov. 2013).
- [75] 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](https://doi.org/10.1007/978-3-642-86679-1_3).
- [76] 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](https://doi.org/10.1037/h0043158).
- [77] 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.
- [78] 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](https://doi.org/10.1016/j.jcps.2014.08.002).
- [79] 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.
- [80] 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: [10.1177/15419312060500909](https://doi.org/10.1177/15419312060500909).
- [81] Brad Cain. 2007. A review of the mental workload literature. *DTIC Document*.
- [82] 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](https://doi.org/10.1080/00140139.2013.790483).
- [83] 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](https://doi.org/10.1111/j.1464-0597.2004.00161.x).
- [84] 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](https://doi.org/10.1007/978-1-4939-0378-8).
- [85] Jakob Nielsen. 1997. How users read on the web.
- [86] 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](https://doi.org/10.1002/bdm.647).
- [87] 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](https://doi.org/10.1037/0022-3514.83.5.1178).
- [88] John W. Payne, James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge University Press, Cambridge. doi: [10.1017/CBO9781139173933](https://doi.org/10.1017/CBO9781139173933).
- [89] Anuj K. Shah, Eldar Shafir, and Sendhil Mullainathan. 2015. Scarcity frames value. *Psychological Science*, 26, 4, 402–412.
- [90] 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](https://doi.org/10.1007/s11238-021-09802-7).
- [91] 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](https://doi.org/10.1037/0033-295X.103.4.650).
- [92] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 4157, 1124–1131. Retrieved June 21, 2024 from JSTOR: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124).

- 1509 [93] Herbert A. Simon. 1996. *The Sciences of the Artificial*. (3rd ed ed.). MIT Press, Cambridge, Mass.  
 1510 [94] Dan Ariely, George Loewenstein, and Drazen Prelec. 2003. “Coherent Arbitrariness”: Stable Demand Curves Without Stable Preferences\*. *The  
 1511 Quarterly Journal of Economics*, 118, 1, (Feb. 2003), 73–106. doi: [10.1162/00335530360535153](https://doi.org/10.1162/00335530360535153).

## 1512 **A Voting Interface Breakdown**

1513 Compared to digital survey interfaces, there exist rich literature on voting interfaces, which we argue is a special type  
 1514 of survey interfaces. We categorize these related work into three main categories detaile below:

1515 *Designs that shifted voter decisions:* For example, states without straight-party ticket voting (where voters can select  
 1516 all candidates from one party through a single choice) exhibited higher rates of split-ticket voting [12]. Another example  
 1517 from the Australian ballot showing incumbency advantages is where candidates are listed by the office they are running  
 1518 for, with no party labels or boxes.

1519 *Designs that influenced errors:* Butterfly ballots increased voter errors because voters could not correctly identify the  
 1520 punch hole on the ballot. Splitting contestants across columns increases the chance for voters to overvote [42]. On  
 1521 the other hand, Everett et al. [43] showed the use of incorporating physical voting behaviors, like lever voting, into  
 1522 graphical user interfaces.

1523 *Designs that incorporated technologies:* Other projects like the Caltech-MIT Voting Technology Project have sparked  
 1524 research to address accessibility challenges, resulting in innovations like EZ Ballot [44], Anywhere Ballot [45], and  
 1525 Prime III [46]. In addition, Gilbert et al. [47] investigated optimal touchpoints on voting interfaces, and Conrad et al.  
 1526 [48] examined zoomable voting interfaces.

## 1527 **B Interface design process**

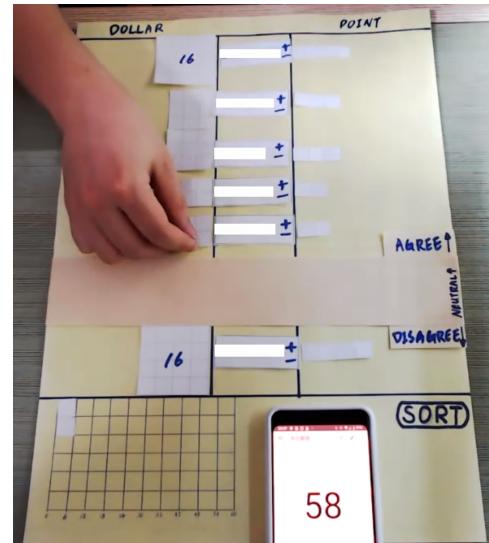
1528 In this section, we outline the design process leading to our final interface. As mentioned in the paper, our design  
 1529 iteration began from existing QV applications in the wild.

### 1530 **B.1 Prototype 1: Ranking-Vote**

1531 Considering that relative preference is often through ranking items, we tested whether ranking options before voting  
 1532 would help establish an individual’s relative preference in our prototype 1. This prototype allowed respondents to  
 1533 reposition options before voting. Pretests revealed that respondents rarely moved the options and questioned the  
 1534 necessity of full ranking, as it did not influence their QS submission. Additionally, many were unaware that options  
 1535 were draggable until shown. This insight indicates that full ranking is unnecessary for establishing relative preferences.  
 1536 Therefore, we decided to ask respondents to select a subset of options instead of requiring a full rank among all options.

### 1537 **B.2 Prototype 2: Select-then-Vote**

1538 Based on feedback from Prototype 1, instead of *allowing* individuals to rank options, Prototype 2 implemented a  
 1539 two-phase process that *intentionally* asks respondents to select options to express opinions before voting. As shown in  
 1540 Figure 18, survey respondents selected their preferred options (Figure 18a), and the interface positioned these options at  
 1541 the top of the list for voting (Figure 18b). We identified several issues during the prototype 2 pretest: many respondents  
 1542 marked most options as ‘options they care about,’ which undermined the design’s purpose. Additionally, the lack of  
 1543 clear distinction between selected and unselected options confused respondents about the necessity of Step 1. Thus, we  
 1544 need a clearer distinction and connection between the two phases to effectively construct relative preferences.



(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. 16. Initial paper prototypes designed for QS interface

### B.3 Prototype 3: Organize-then-Vote

Figure 19 shows the last prototype where we built on the previous takeaway by providing finer-grain groupings and creating a clear connection between option organization and voting position. Specifically, we provided three categories: Lean Positive, Lean Negative, and Lean Neutral. Initially, respondents see all options under the section labeled 'I don't know,' which includes only the option descriptions. We ask respondents to move these options into the categories 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.

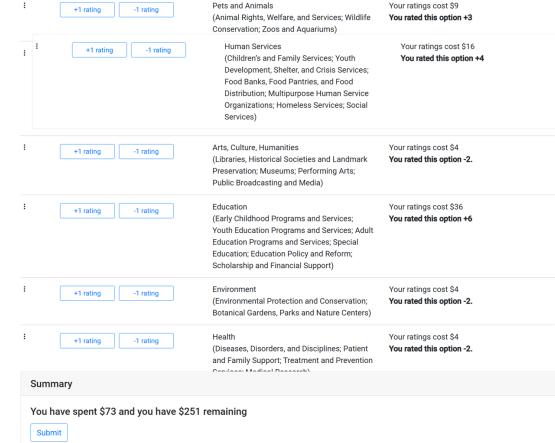
Feedback indicated that survey respondents are comfortable with the two-phase organize-then-vote design, demonstrating it as a central strategy for our interface development. However, several areas for enhancement were identified: First, the dragging and dropping mechanism in the organization phase is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

1613 **What societal issues need more support?**

1614 Please express your opinion using this survey mechanism as described above. You have a total of \$324 for the following 9 issues. You do not need to use up all your budget, but you cannot exceed \$324.

1615 If you think that an issue needs more support, you can rate the issue higher. Vice versa, you can rate the issue lower if you think it requires less support.

1616

1617 

1618

1619

1620

1621

1622

1623

1624

1625

1626

1627

1628

1629

1630

1631

1632

1633

1634 Fig. 17. A Ranking-Vote Prototype: The goal of this prototype is to test whether ranking options prior to voting help establish an individual's relative preferences and reduce effort when voting. Each option is draggable to position in a specific location amongst the full list of options. Votes can be updated using the buttons to the right of the interface with vote count and costs to the right of the interface. A summary box is placed sticky to the bottom of the screen.

1635

1636

1637

1638

1639

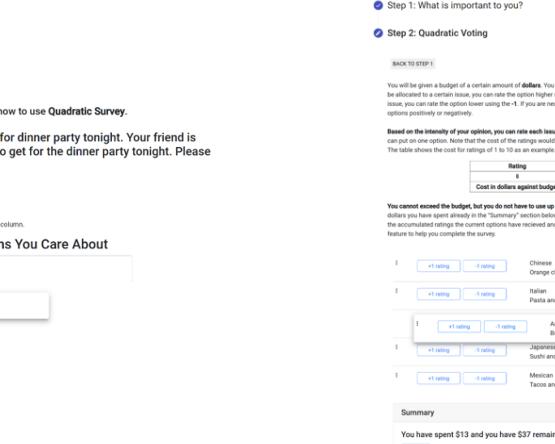
1640

1641

1642 This is a playground designed to help you understand how to use Quadratic Survey.

1643 There is a limited budget to purchase the food for dinner party tonight. Your friend is asking for your preference of the type of food to get for the dinner party tonight. Please complete the survey below.

1644

1645 

1646

1647

1648

1649

1650

1651

1652

1653

1654 (a) Options are dragged and dropped to the 'Option You Care About' box.

1655

1656

1657

1658

1659

1660

1661

1662

1663

1664

#### Step 1: What is important to you?

#### Step 2: Quadratic Voting

BACK TO STEP 1

You will be given a budget of a certain amount of dollars. You will use the available money to rate the options. If you think more resources should be allocated to a certain issue, you can rate the option higher using the +1. If you believe that less resources should be allocated on a certain issue, you can rate the option lower using the -1. If you are neutral on an issue, you can choose to not rate the option. You are allowed to rate options multiple times.

Based on the intensity of your opinion, you can rate each issue positively and negatively. The stronger your opinion is, the higher the rating you can put on one option. Note that the cost of the ratings would increase quadratically in other words, rating of X will cost  $X^2$  (square of X) dollars.

The table shows the cost for ratings of 1 to 10 as an example. You can rate higher than 10 or lower than -10 if the budget allows you to do so.

Rating	1	2	3	4	5	6	7	8	9	10
Cost in dollars against budget	1	4	9	16	25	36	49	64	81	100

You cannot exceed the budget, but you do not have to use up all the budget either. You can see your total budget you have and the amount of dollars you have spent already in the "Summary" section below. The interface will provide real-time calculation of the remaining budget you have, the accumulated ratings you have given, and the total spent on each option. This interface also provides a drag-and-drop feature to help you complete the survey.



Summary

You have spent \$13 and you have \$37 remaining

Submit

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

1665 Fig. 18. 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. 18a, the first step involves selecting options for further consideration. Important options are placed at the top of the list for voting shown in Fig. 18b, but options can be placed anywhere on the list if desired. The rest of the controls remain the same as the previous prototype.

1665  
 1666 **What societal issues need more support?**  
 1667 Please express your opinion using this survey mechanism as described above. You have a total of \$324 for the following 9 issues. You do not need to use up all your budget, but you cannot exceed \$324.  
 1668 If you think that an issue needs more support, you can rate the issue higher. Vice versa, you can rate the issue lower if you think it requires less support.  
 1669 **I don't know**  
 1670 Pets and Animals (Animal Rights, Welfare, and Services; Wildlife Conservation; Zoos and Aquariums)  
 1671 Arts, Culture, Humanities (Libraries, Historical Societies and Landmark Preservation; Museums; Performing Arts; Public Broadcasting and Media)  
 1672 Health (Healthcare, Disorders, and Diseases; Patient and Family Support; Treatment and Prevention Services; Medical Research)  
 1673 Veterans (Religious Activities; Religious Media and Broadcasting)  
 1674 Veterans (Wounded Troops Services; Military Social Services; Military Family Support)  
 1675 Positive  
 1676 Education (Community Programs and Services; Youth Education Programs and Services; Adult Education Programs and Services; Special Education; Education Policy and Reform; Scholarship and Financial Support)  
 1677 Negative  
 1678 Environment (Environmental Protection and Conservation; Botanical Gardens, Parks and Nature Centers)  
 1679 Neutral  
 1680 International (Development and Relief Services; International Peace, Security, and Affairs; Humanitarian Relief Supplies)  
 1681 Homeless (Children's and Family Services; Youth Development, Shelter, and Crisis Services; Food Banks, Food Pantries, and Food Distribution; Multipurpose-Human Service Organizations; Homeless Services; Social Services)  
 1682 Next

What societal issues need more support?

Please express your opinion using this survey mechanism as described above. You have a total of \$324 for the following 9 issues. You do not need to use up all your budget, but you cannot exceed \$324.

If you think that an issue needs more support, you can rate the issue higher. Vice versa, you can rate the issue lower if you think it requires less support.

I don't know

Pets and Animals (Animal Rights, Welfare, and Services; Wildlife Conservation; Zoos and Aquariums)	Your ratings cost \$4 You rated this option +2
Arts, Culture, Humanities (Libraries, Historical Societies and Landmark Preservation; Museums; Performing Arts; Public Broadcasting and Media)	Your ratings cost \$4 You rated this option -2
Health (Diseases, Disorders, and Diseases; Patient and Family Support; Treatment and Prevention Services; Medical Research)	Your ratings cost \$9 You rated this option +7
Veterans (Wounded Troops Services; Military Social Services; Military Family Support)	Your ratings cost \$4 You rated this option -2
Positive	
Education (Early Childhood Programs and Services; Youth Education Programs and Services; Adult Education Programs and Services; Special Education; Education Policy and Reform; Scholarship and Financial Support)	Your ratings cost \$0 You rated this option 0
Negative	
Environment (Environmental Protection and Conservation; Botanical Gardens, Parks and Nature Centers)	Your ratings cost \$36 You rated this option +6
Neutral	
International (Development and Relief Services; International Peace, Security, and Affairs; Humanitarian Relief Supplies)	Your ratings cost \$4 You rated this option -2
Human Services (Children's and Family Services; Youth Development, Shelter, and Crisis Services; Food Banks, Food Pantries, and Food Distribution; Multipurpose-Human Service Organizations; Homeless Services; Social Services)	Your ratings cost \$0 You rated this option 0

Summary

You have spent \$117 and you have \$207 remaining

Submit

(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.

1686 (a) The Organization Interface: Options are shown initially in the first bin labeled as 'I don't know.' Survey respondents can then drag and drop these options into the latter bins: Lean Positive, Lean Neutral, or Lean Negative. Only the details of each option are shown on this interface.

1691 Fig. 19. Organize-then-Vote Prototype: The goal of this prototype is to encourage participants at deriving finer grain categories  
 1692 among options before voting. Survey respondents first organize their thoughts into categories, then vote on the options.

## C Cognitive Load remaining

### C.1 Sources of Physical Demand

1704 **Key Differences:** Two-phase interface experienced higher physical demand from increased mouse usage.

1708 Physical demand refers to the physical effort required to complete a task, such as physical exertion or movement. Most participants reported minimal physical demand ( $N = 32$ ), reflected in the low NASA-TLX physical demand scores (Figure 20). 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 20), 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.

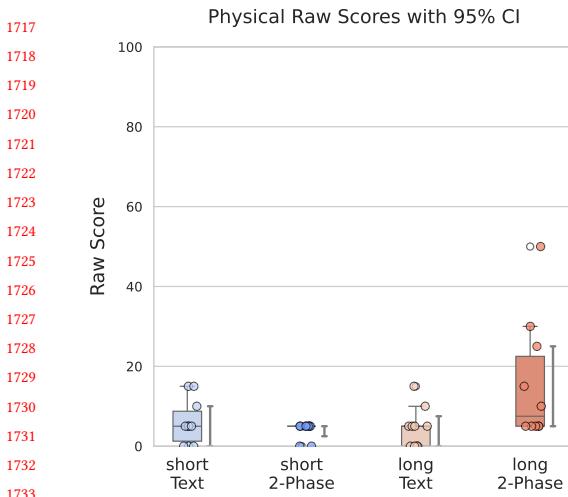


Fig. 20. Physical Demand Raw Score: Participants other than the long two-phase interface reported minimal physical demand. The long two-phase interface had the highest physical demand, likely due to increased mouse clicks and extended time spent looking at the vertical screen.

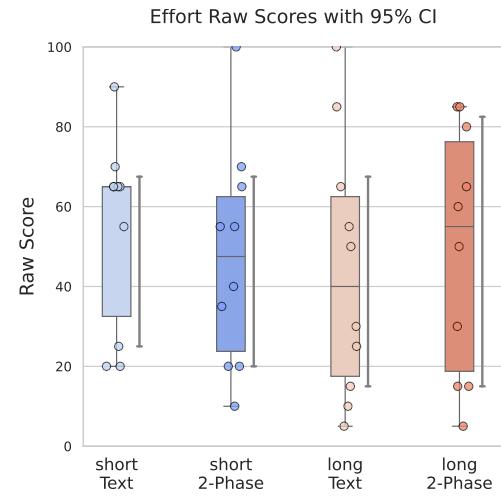


Fig. 21. Effort Raw Score: Effort scores shows indifference across groups.

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
<b>Operational</b>	21	6	5	8	2	11	10	14	7
<b>Strategic</b>	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
<b>None/Little/a bit</b>	9	2	1	3	3	3	6	5	4

## C.2 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. Since effort includes both mental and physical resource intensity, refer to Section 5.2 and Appendix C.1 for definitions.

1769 While the raw NASA-TLX effort scores (Figure 21) showed a similar spread across experiment groups, the qualitative  
 1770 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  
 1771 5.2).

1772 *C.2.1 Effort Source #1: Operational Tasks.* 14 of the 20 participants using the text interface mentioned Operational  
 1773 Tasks as effort sources, compared to 7 using the two-phase interface, with the lowest mention by the long two-phase  
 1774 interface group ( $N = 2$ ).

1775 *C.2.2 Effort Source #2: Strategic Planning.* Different from Operational Tasks, 11 participants in the text interface  
 1776 compared to 17 participants described strategic planning as sources of effort, with almost all participants ( $N = 9$ ) from  
 1777 the long two-phase interface. We further categorize strategic planning into *narrow* and *broad* scopes as we did for  
 1778 mental demand section 5.2. Participants using the two-phase interface ( $N = 7$ ) had nearly mentioned double ( $N = 4$ )  
 1779 times regarding global strategies.

### 1780 C.3 Source of Performance

1781 **Q Key Differences:** Participants who used a two-phase interface were generally more positive about their final  
 1782 outcome – they were twice as likely to report "feeling good" about their final results

1783 Performance refers to a person's perception of their success in completing a task. Lower values mean good perceived  
 1784 performance; higher values mean poor perceived performance. We found minimal qualitative differences between  
 1785 experiment groups regarding factors influencing perceived performance. Two influencing factors emerged: *Operational*  
 1786 *Actions* and *Social Responsibility*<sup>4</sup>. Despite most participants reporting positively on their performance, nuances exist  
 1787 in how different groups interpret their performance.

1788 *C.3.1 Operational Actions.* Operational actions, like the theme presented in temporal demand, refer to specific,  
 1789 executable procedures participants perform in the survey. This could involve: pressure to spend all credits or stay  
 1790 within budget ( $N = 6$ ), fears that final vote choices did not reflect true preferences ( $N = 5$ ), or concerns that they had  
 1791 finished the task inefficiently ( $N = 6$ ).

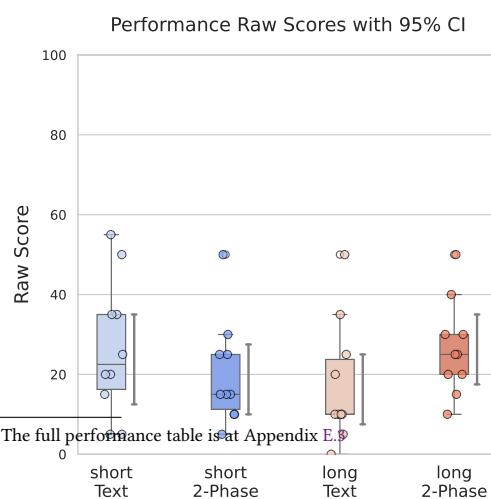


Fig. 22. Performance Demand Raw Score: Participants showed indifferent performance raw scores across experiment conditions, all trending toward satisfactory.

1804  
 1805 *C.3.2 Social Responsibility.* Social responsibility-based con-  
 1806 cerns around performance came up when participants reflected  
 1807 on how their final vote counts would be perceived by others ( S041 Q I don't want people to think that I just like don't care  
 1808 about <ethnicity> people at all ) or influence real-world decision-  
 1809 making ( S027 Q Some of these things might ... have outcomes  
 1810 that I didn't foresee ).

1811 All groups cited social responsibility as source to evaluate  
 1812 effort. Raw NASA-TLX scores (Figure 22) show participants  
 1813 had indistinguishable performance scores. This aligns with the  
 1814 interview results where most participants felt positive about  
 1815 their final submission.

1821  
 1822  
 1823  
 1824  
 1825  
 1826  
 1827  
 1828  
 1829  
 1830 more positive about their final outcome – they were twice as likely to report "feeling good" about their final re-  
 1831 sults ( $N = 11$  v.s.  $N = 6$ ).  
 1832  
 1833  
 1834 **D List of Options**  
 1835 We provide the full list of options presented on the survey.  
 1836

- 1837 • **Animal Rights, Welfare, and Services:** Protect animals from cruelty, exploitation and other abuses, provide  
 1838 veterinary services and train guide dogs.
- 1839 • **Wildlife Conservation:** Protect wildlife habitats, including fish, wildlife, and bird refuges and sanctuaries.
- 1840 • **Zoos and Aquariums:** Support and invest in zoos, aquariums and zoological societies in communities through-  
 1841 out the country.
- 1842 • **Libraries, Historical Societies and Landmark Preservation:** Support and invest public and specialized  
 1843 libraries, historical societies, historical preservation programs, and historical estates.
- 1844 • **Museums:** Support and invest in maintaining collections and provide training to practitioners in traditional  
 1845 arts, science, technology, and natural history.
- 1846 • **Performing Arts:** Support symphonies, orchestras, and other musical groups; ballets and operas; theater  
 1847 groups; arts festivals; and performance halls and cultural centers.
- 1848 • **Public Broadcasting and Media:** Support public television and radio stations and networks, as well as  
 1849 providing other independent media and communications services to the public.
- 1850 • **Community Foundations:** Promote giving by managing long-term donor-advised charitable funds for indi-  
 1851 vidual givers and distributing those funds to community-based charities over time.
- 1852 • **Housing and Neighborhood Development:** Lead and finance development projects that invest in and  
 1853 improve communities by providing utility assistance, small business support programs, and other revitalization  
 1854 projects.
- 1855 • **Jewish Federations:** Focus on a specific geographic region and primarily support Jewish-oriented programs,  
 1856 organizations and activities through grantmaking efforts
- 1857 • **United Ways:** Identify and resolve community issues through partnerships with schools, government agencies,  
 1858 businesses, and others, with a focus on education, income and health.
- 1859 • **Adult Education Programs and Services:** Provide opportunities for adults to expand their knowledge in a  
 1860 particular field or discipline, learn English as a second language, or complete their high school education.
- 1861 • **Early Childhood Programs and Services:** Provide foundation-level learning and literacy for children prior  
 1862 to entering the formal school setting.
- 1863 • **Education Policy and Reform:** Promote and provide research, policy, and reform of the management of  
 1864 educational institutions, educational systems, and education policy.

- 1873 • **Scholarship and Financial Support:** Support and enable students to obtain the financial assistance they  
1874 require to meet their educational and living expenses while in school.
- 1875 • **Special Education:** Provide services, including placement, programming, instruction, and support for gifted  
1876 children and youth or those with disabilities requiring modified curricula, teaching methods, or materials.
- 1877 • **Youth Education Programs and Services:** Provide programming, classroom instruction, and support for  
1878 school-aged students in various disciplines such as art education, STEM, outward bound learning experiences,  
1879 and other programs that enhance formal education.
- 1880 • **Botanical Gardens, Parks, and Nature Centers:** Promote preservation and appreciation of the environment,  
1881 as well as leading anti-litter, tree planting and other environmental beautification campaigns.
- 1882 • **Environmental Protection and Conservation:** Develop strategies to combat pollution, promote conservation  
1883 and sustainable management of land, water, and energy resources, protect land, and improve the efficiency of  
1884 energy and waste material usage.
- 1885 • **Diseases, Disorders, and Disciplines:** Seek cures for diseases and disorders or promote specific medical  
1886 disciplines by providing direct services, advocating for public support and understanding, and supporting  
1887 targeted medical research.
- 1888 • **Medical Research:** Devote and invest in efforts on researching causes and cures of disease and developing  
1889 new treatments.
- 1890 • **Patient and Family Support:** Support programs and services for family members and patients that are  
1891 diagnosed with a serious illness, including wish granting programs, camping programs, housing or travel  
1892 assistance.
- 1893 • **Treatment and Prevention Services:** Provide direct medical services and educate the public on ways to  
1894 prevent diseases and reduce health risks.
- 1895 • **Advocacy and Education:** Support social justice through legal advocacy, social action, and supporting laws  
1896 and measures that promote reform and protect civil rights, including election reform and tolerance among  
1897 diverse groups.
- 1898 • **Development and Relief Services:** Provide medical care and other human services as well as economic,  
1899 educational, and agricultural development services to people around the world.
- 1900 • **Humanitarian Relief Supplies:** Specialize in collecting donated medical, food, agriculture, and other supplies  
1901 and distributing them overseas to those in need.
- 1902 • **International Peace, Security, and Affairs:** Promote peace and security, cultural and student exchange  
1903 programs, improve relations between particular countries, provide foreign policy research and advocacy, and  
1904 United Nations-related organizations.
- 1905 • **Religious Activities:** Support and promote various faiths.
- 1906 • **Religious Media and Broadcasting:** Support organizations of all faiths that produce and distribute religious  
1907 programming, literature, and other communications.
- 1908 • **Non-Medical Science & Technology Research:** Support research and services in a variety of scientific  
1909 disciplines, advancing knowledge and understanding of areas such as energy efficiency, environmental and  
1910 trade policies, and agricultural sustainability.
- 1911 • **Social and Public Policy Research:** Support economic and social issues impacting our country today, educate  
1912 the public, and influence policy regarding healthcare, employment rights, taxation, and other civic ventures.

## 1925 E Cognitive Demand Tables

1926 Here we provide the full qualitative analysis table.

### 1929 E.1 Mental Demand Table

1930  
1931  
1932  
1933 Table 2. This table lists all the causes participants mentioned as contributing to their Mental Demand. The shaded cells represent  
1934 the percentage of participants citing each source of mental demand, allowing for comparison within columns. The abbreviations  
1935 are: ST (Short Text Interface), SI (Short Two-phase Interface), LT (Long Text Interface), and LI (Long Two-phase Interface). Short and  
1936 Long refer to the sum across both interfaces; Text and Inter refer to the sum across both survey lengths. We include Sparklines for  
1937 comparisons across these experiment groups.

1938 [ Mental Demand ]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
<b>1940 Budget Management</b>	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
<b>1950 Preference Construction</b>	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
<b>1960 Demand from Experiment Setup</b>	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
<b>1966 Justification or Reflection on response</b>	8	2	2	1	3	4	4	3	5
<b>1967 External Factors</b>	12	3	1	4	4	4	8	7	5
<b>1968 Demand due to Interface</b>	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

## 1977 E.2 Physical Demand Table

1978  
1979  
1980  
1981  
1982  
1983 Table 3. Physical Demand Causes: Most participants expressed little or no physical demand. Results reflected that participants in the  
1984 long two-phase interface required more actions, hence the higher mention of mouse usage as a source.

1986 [ Physical ]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
1988 <b>Reading</b>	4	0	2	1	1	2	2	1	3
1989 <b>Mouse</b>	16	3	5	2	6	8	8	5	11
1990 <b>Vertical Screen</b>	4	1	0	1	2	1	3	2	2
1991 <b>None/Little</b>	32	8	9	8	7	17	15	16	16

## 1993 E.3 Performance Table

1994  
1995  
1996  
1997  
1998 Table 4. Performance Causes: Most causes are shared across experiment conditions. We provided qualitative interpretations of their  
1999 own performance assessments.

2007 [ Performance ]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
2009 <b>Operational Action</b>	13	2	3	3	5	5	8	5	8
2010 Budget Control	6	1	1	2	2	2	4	3	3
2011 Preference Reflection	6	1	1	2	2	2	4	3	3
2012 Limited Resources	5	1	2	1	1	3	2	2	3
2013 <b>Social Responsibility</b>	8	2	2	2	2	4	4	4	4
2014 Decision maker	7	1	2	2	2	3	4	3	4
2015 Outcome Uncertainty	7	1	2	2	2	3	4	3	4
2016 <b>Performance Assessment</b>	8	2	1	3	2	3	5	5	3
2017 Did their best	17	3	5	3	6	8	9	6	11
2018 Feel Good	10	2	2	3	3	4	6	5	5
2019 Good Enough	10	2	2	3	3	4	6	5	5

**E.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
<b>Budget Management</b>	4	0	1	1	2	1	3	1	3
<b>Decision Making</b>	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
<b>Operational</b>	16	5	6	3	2	11	5	8	8
Task completion	8	2	2	3	1	4	4	5	3
Being efficient	8	3	4	0	1	7	1	3	5

**E.5 Frustration Table**

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

[ Frustration ]	Total	Version				Experiment Conditions			
		ST	SI	LT	LI	Short	Long	Text	Inter
<b>Strategic</b>	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 broader society and common values	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
<b>Operational</b>	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
<b>None/Little</b>	16	4	5	5	2	9	7	9	7

**F Additional behavioral results**

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

Manuscript submitted to ACM

outlined, credit scarcity might influence decision-making. Figure 23 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 kernel density estimate (KDE) plot is provided to visualize the trends. We did not follow Quarfoot et al. [5] in counting accumulated votes over time due to varying total times across individuals.

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.

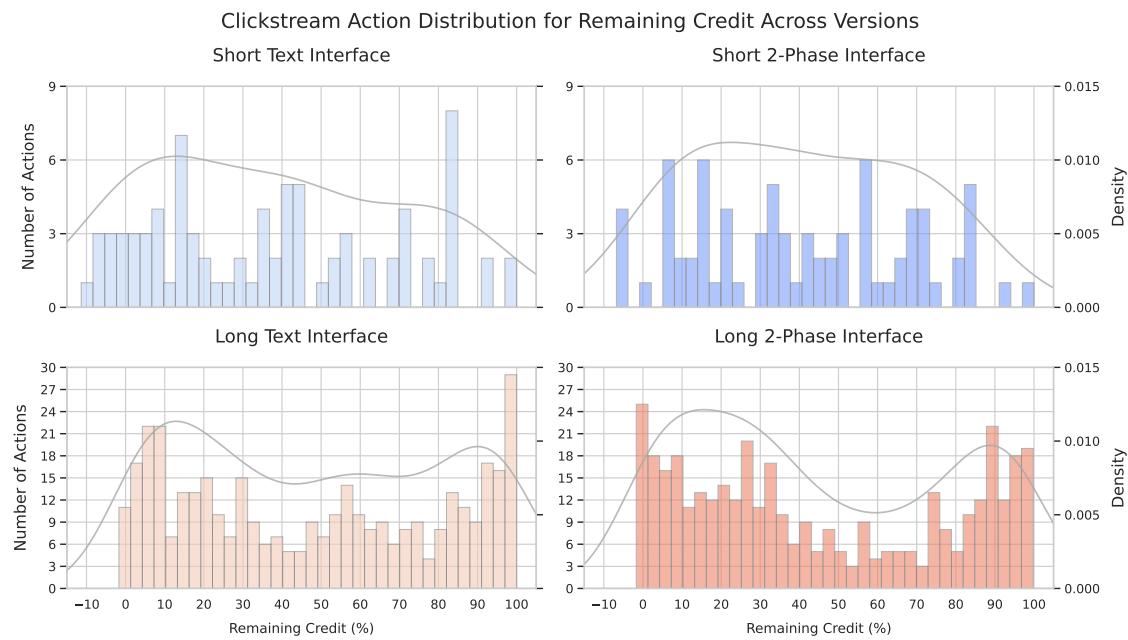


Fig. 23. This plot counts the number of voting actions when there are  $x$  percentages of credits remaining. A KDE plot is provided to help better understand the action distribution.

In Figure 23, 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 24, 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.

We plotted all actions against the time to complete them. Revisiting the KDE curve in the second row in Figure 23 and the curve of the second row in Figure 24 show a stronger bimodal distribution for small vote adjustments across interfaces. In fact, the bimodal distribution is more pronounced in the two-phase interface. This suggests that participants make small adjustments both at the beginning and toward the end of the QS. However, the two-phase interface shows

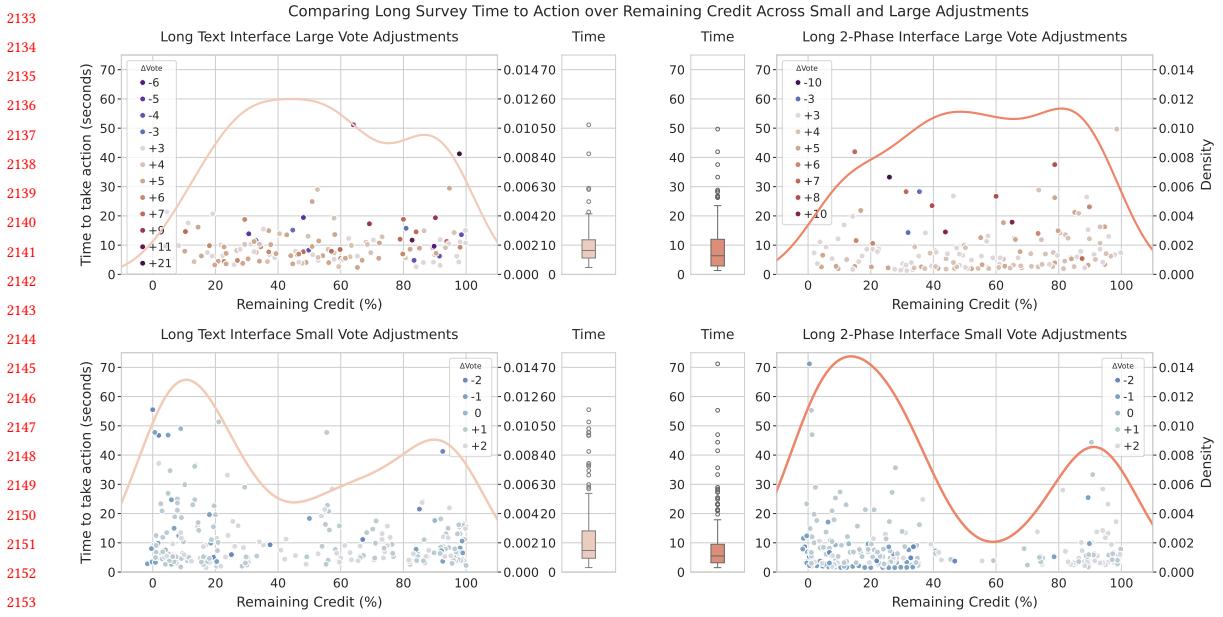


Fig. 24. This plot further separates participants' interaction behavior based on the number of votes participants adjusted. We observed a bimodal interaction pattern across long QS when small vote adjustments are made.

more frequent and faster edits towards the end. Visually, dots are more clustered in the long two-phase interface for small vote adjustments compared to the long text interface. The Mann-Whitney U Test on the time spent on small vote adjustments showed significant differences ( $U = 13037, p < 0.001$ ), with a small effect size (Rank-biserial: 0.227, Cohen's  $d$ : 0.195) and a power of 0.381. Based on the KDE plots in the first row of Figure 24, 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. [...] that's very important is that it's an iterative process.* S019 (LI)