

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

2
3 ANONYMOUS AUTHOR(S)*
4

5 Quadratic Surveys (QS) elicit more accurate preferences than traditional methods like Likert-scale surveys. However, the cognitive
6 load associated with QS has hindered their adoption in digital surveys for collective decision-making. We introduce a two-phase
7 “organize-then-vote” QS to reduce cognitive load. As interface design significantly impacts survey results and accuracy, our design
8 scaffolds survey takers’ decision-making while managing the cognitive load imposed by QS. In a 2x2 between-subject in-lab study
9 on public resource allotment, we compared our interface with a traditional text interface across QS with 6 (short) and 24 (long)
10 options. Two-phase interface participants spent more time per option and exhibited shorter voting edit distances. We qualitatively
11 observed shifts in cognitive effort from mechanical operations to constructing more comprehensive preferences. We conclude that this
12 interface promoted deeper engagement, potentially reducing satisficing behaviors caused by cognitive overload in longer QS. This
13 research clarifies how human-centered design 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 (December 2024), 54 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. Recent Human-Computer Interaction (HCI) studies highlight how
31 certain survey response formats can increase errors [1, 2] and influence survey effectiveness [3]. In this paper, our
32 goal is to introduce an effective interface for **Quadratic Surveys (QS)**, a survey tool designed to elicit preferences
33 more accurately than traditional methods [4]. Despite the promise of QS, there has been no research on designing
34 interfaces to support its unique quadratic mechanisms [5], where participants must rank and rate items — a task that
35 poses significant cognitive challenges. To popularize QS and ensure high-quality data, this paper addresses the question:
36 *How can we design interfaces to support participants in completing Quadratic Surveys (QS) more effectively?*

37
38 We envision an effective interface that navigates participants through the complex mechanism and preference
39 construction process, **tailored to QS**. QS improves accuracy in individual preference elicitation compared to traditional
40 methods like Likert scales by requiring participants to make trade-offs using a fixed budget of credits, where purchasing k
41 votes for an option in QS costs k^2 credits [6, 4]. This quadratic cost structure forces respondents to carefully evaluate their
42 preferences, balancing the strength of their support or opposition against the limited budget. As individual preferences

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.

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

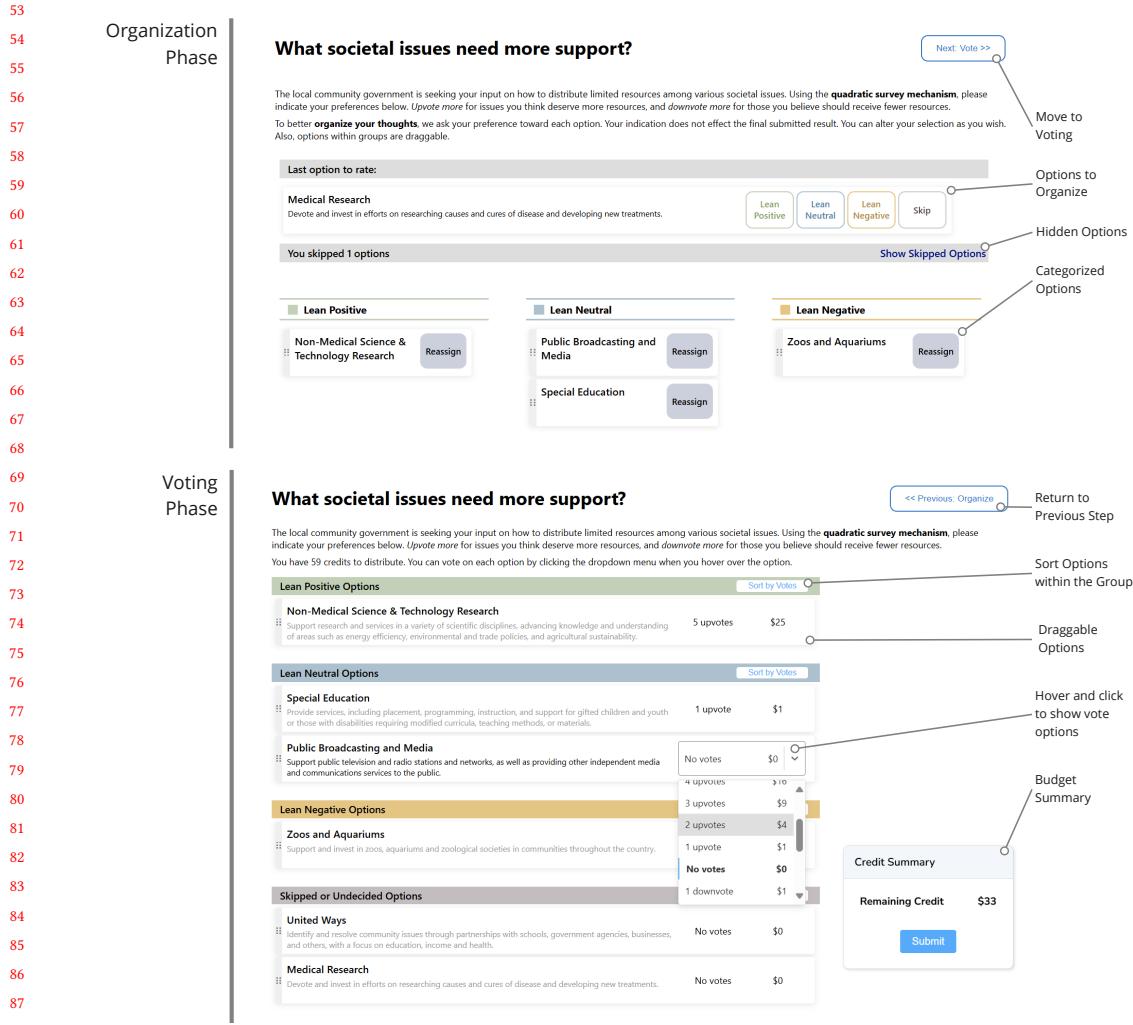


Fig. 1. The Two-Phase Interface: The interface consists of two phases. Survey respondents can navigate between phases using the top right button. In the organization phase, the interface presented one option at a time to the respondents, and they chose four choices: “Lean Positive”, “Lean Neutral”, “Lean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

are often constructed when given the options, even though this cost structure forces participants to make thoughtful trade-offs, the construction process increases cognitive load, making it mentally taxing to weigh costs, evaluate options, and construct rankings [7]. Moreover, QS, often referred to as Quadratic Voting (QV) in voting scenarios, can involve hundreds of options [8, 9], increasing the risk of cognitive overload and taking mental shortcuts [10, 11, 12].

To date, existing quadratic mechanism-powered applications simply present options, allow vote adjustments and automatically calculate votes, costs, and budget usage. These designs focused heavily on the mechanics operating the tool, rather than supporting possible challenges these application users faced. Survey interface literature, while addressing decision-making and usability, most focus on traditional surveys that do not share the unique option-to-option trade-offs that QS introduces [13, 14, 15, 16, 17, 1]. Prior research in HCI and beyond explored techniques to managing cognitive load [18, 19, 16, 20, 21] and scaffolding challenging tasks [22, 23, 24, 25] showing promise in supporting preference construction under QS. Thus, this study aims to bridge this gap.

We propose a novel interactive two-phase “organize-then-vote” QS interface (referred to as the two-phase interface for short, Figure 1) after multiple iterations. It aims to facilitate preference construction and reduce cognitive load when making trade-offs through three key elements. First, the interface scaffolds the preference construction process by having participants initially categorize the survey options into “Lean Positive,” “Lean Neutral,” or “Lean Negative.” This serves as a cognitive warm-up, easing participants into the more complex QS voting task. Second, the interface arranges the options according to these categorizations, providing a structured visual layout. Third, participants can refine the positions of these options using drag-and-drop functionality, giving them greater control and agency in the preference-construction process.

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

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

We invited 41 participants to a lab study comparing our two-phase interface with a baseline to understand how different interface designs and option lengths (6 options or 24 options) impact cognitive load.

Self-reported cognitive load using the NASA Task Load Index (NASA-TLX) and semi-structured interviews identified common challenges in Quadratic Surveys (QS), such as preference construction and budget management, while highlighting differences between text and two-phase interfaces. The two-phase interface fostered more strategic engagement with survey options considering broader impacts in the long QS, reduced time pressure in the short QS, and participants expressed greater affirmative satisfaction (e.g., ‘feeling good’). Quantitative results showed that the organizing phase in the two-phase interface led participants, particularly in long surveys, to traverse the option list less often without reducing edits and spend more time per option, signifying deeper engagement and a shift toward more strategic thinking when constructing their preferences.

Contributions. We contribute to the HCI community by proposing the first interface specifically designed for QS and QV-like applications, aimed at reducing cognitive challenges and scaffolding preference construction through a two-phase interface with direct manipulation. Before our work, no research had explored QS interfaces, particularly for long QS prone to cognitive overload. Few studies in HCI address interfaces for surveys and questionnaires. Our study demonstrated how user interfaces can facilitate preference construction in situ and promote deeper engagement with survey options through interface elements. Additionally, this paper offers the first in-depth qualitative analysis of user experiences among Quadratic Mechanism applications, identifying usability challenges and key factors contributing to

cognitive load. The impact of our contribution extends beyond QS, offering design implications for other preference-elicitation tools in **multi-option scenarios**. By making QS easier to use and more accurate, our design also encourages wider adoption among researchers and practitioners. Finally, our work lays the groundwork for future quadratic mechanisms interface design to better facilitate individuals in communicating their preferences.

2 Related Work

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

2.1 Quadratic Survey and the Quadratic Mechanism

We introduce the term **Quadratic Survey (QS)** to describe surveys that utilize the quadratic mechanism to collect individual attitudes. The **quadratic mechanism** is a theoretical framework designed to encourage the truthful revelation of individual preferences through a quadratic cost function [5]. This framework gained popularity through **Quadratic Voting (QV)**, also known as plural voting, which uses a quadratic cost function in a voting framework to facilitate collective decision-making [26].

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

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

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

$$\sum_k c_k \leq B$$

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

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

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

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

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

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

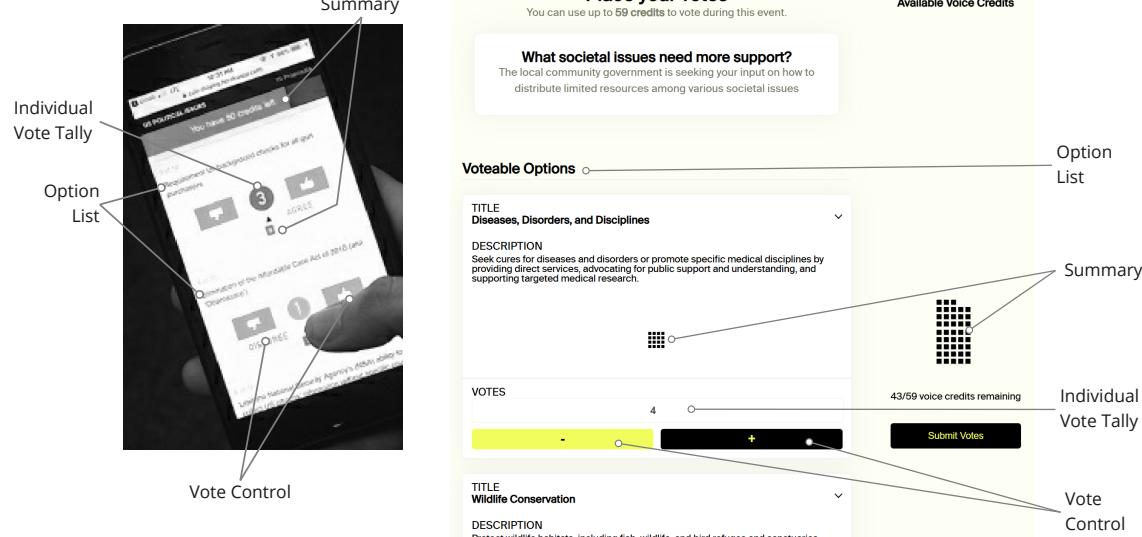


Fig. 2. A selection of two QV interfaces. The interface on the left was used in the first empirical QV research [6]. Little information is available about the software, except for an image from Posner and Weyl [27]. The interface on the right is an open-sourced QV interface [34] forked from GitCoin [35], used by the RadicalxChange community [36]. Both interfaces share the common elements with different visual representations.

2.2 Existing QV Interfaces

Given QS shares the same mechanism with QV, we conducted a snowball sampling process to identify publicly available Quadratic Voting (QV) applications from known news reports and academic publications. No widely adopted QV interfaces have been developed by a single vendor or platform to date. Fig. 2 shows two variations of existing interfaces¹, with all QV interfaces employing a single-step approach with different visual representations of common elements. 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 allow individuals to operate QV, focusing purely on mechanics without little understanding of voters' usability needs nor offering cognitive support to help them complete the task. In addition, the HCI community conducted few research [37, 38] on survey and questionnaire interfaces components, with more work focusing more on alternative input modalities like bots, voice, and virtual reality [39, 40, 2, 41].

¹Appendix C lists the other interface we surveyed

261 2.3 Cognitive Challenges and Choice Overload

262 The challenge of respondents making difficult decisions using quadratic mechanisms remains unexplored in the
 263 literature. Lichtenstein and Slovic [7] identified three key elements that make decisions difficult. These elements
 264 include making decisions in unfamiliar contexts, quantifying the value of one's opinions, and being forced to make
 265 trade-offs due to conflicting choices. QS fits at least two of the three elements: participants may encounter a selection
 266 of unfamiliar options by the survey designer; they are asked to quantify the difference between option preferences
 267 through a numerical vote; and the budget constraint enforces trade-offs under a non-linear function, which means that
 268 a vote decrease for one option is not necessarily equivalent to an increase for another, making iterative adjustment and
 269 evaluating tradeoffs difficult. Thus, we believe QS introduces a high cognitive load.

270 Cognitive load refers to the demands placed on a user's working memory during the interaction process, which
 271 significantly influences the usability of the system [42, 43]. Cognitive overload can adversely affect performance [44],
 272 leading individuals to rely on heuristics rather than deliberate, logical decision-making [45]. When presented with
 273 excessive information, such as too many options, individuals 'satisfice', settling for a 'good enough' solution rather than
 274 an optimal one [10, 11, 12]. Subsequently, too many options can overwhelm individuals, resulting in decision paralysis,
 275 demotivation, and dissatisfaction [46].

276 Additionally, Alwin and Krosnick [47] highlighted that the use of ranking techniques in surveys can be time-
 277 consuming and potentially more costly to administer. These challenges are compounded when ranking numerous items,
 278 requiring substantial cognitive sophistication and concentration from survey respondents [48].

279 Notable applications of Quadratic Voting include the 2019 Colorado House, which considered 107 bills [49], and the
 280 2019 Taiwan Presidential Hackathon, which featured 136 proposals [50]; both used a single QV question with hundreds
 281 of options. These empirical applications of QV suggest the importance of understanding QS with many options' impact
 282 on cognitive load and support developing interfaces for practical uses.

283 3 Quadratic Survey Interface Design

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

289 3.1 'Organize-then-Vote': The Two-Phase Interface

290 3.1.1 *Justifying a two-phase approach.* The main objective of the two-phase interface is to facilitate preference con-
 291 struction and reduce cognitive load. As shown in Figure 1, the interface consists of two steps: an organization phase
 292 and a voting phase. In both phases, survey respondents can drag and drop options across the presented list.

293 A *two-phase approach*. Preferences are shaped through a series of decision-making processes [7]. Two major decision-
 294 making theories informed this two-step interaction interface design: Montgomery [51]'s Search for a Dominance
 295 Structure Theory (Dominance Theory) and Svenson [52]'s Differentiation and Consolidation Theory (Diff-Con Theory).
 296 The former suggested that decision-makers prioritize creating dominant choices to minimize cognitive effort by
 297 focusing on evidently superior options [51]. The latter described a two-phase process where decisions are formed by
 298 initially differentiating among alternatives and then consolidating these distinctions to form a stable preference [52].

313 Both theories supported the design decision to reduce the dimensions during the initial decision process and help
314 emphasize relatively important options to form decisions. Hence, the two-phase design – organize-then-vote – aimed
315 to facilitate this cognitive journey explicitly. The first phase focused on differentiating and identifying dominant options,
316 enabling survey respondents to preliminarily categorize and prioritize their choices. The second phase presented these
317 categorized options in a comparable manner, with drag-and-drop functionality, enhancing one's ability to consolidate
318 preferences. This structured approach aimed to construct a clear decision-making procedure that reduced cognitive
319 load and enhanced clarity and confidence in the decisions made.
320

322 *Phase 1: Organization Phase.* The goal of the organization phase was to support participants in identifying clearly
323 superior options or partitioning choices into distinguishable groups. In this section, we first describe how the interaction
324 works, then we detail the reasons for the implemented design decisions.
325

326 The organizing interface, depicted on the top half of Figure 1, sequentially presents each survey option. Participants
327 select a response among three ordinal categories – “Lean Positive”, “Lean Negative”, or “Lean Neutral”. Once selected,
328 the system moves that option to the respective category. Participants can skip the option if they do not want to indicate
329 a preference. Options within the groups are draggable and rearrangeable to other groups should the participants wish.
330

331 To support preference formation, respondents are shown one option at a time, allowing them to either recall a prior
332 judgment or construct a new one based on the presented choices [53]. Limiting the information presented this way also
333 helps reduce cognitive load by preventing overload from too many options [54]. This incremental process ensures that
334 participants form opinions on individual options, addressing an early prototype issue where the organizing task was
335 mistakenly treated as a ranking task.
336

337 The three possible options – Lean Positive, Lean Neutral, and Lean Negative – aim to scaffold participants in
338 constructing their own choice architecture [55, 56], which strategically segments options into diverse and alternative
339 choice presentations while avoiding biases from defaults. We believed that these three categories were sufficient for
340 participants to segment the options. We do not limit the number of options one can place in each category to prioritize
341 user agency, allowing participants full control over how they organize their preferences [57]. Immediate feedback
342 displays the placement of options and allows participants to rearrange them via drag-and-drop, adhering to key interface
343 design principles [57]. At the same time, it allows finer-grain control for individuals to surface dominating options and
344 create differentiating groups of options.
345

346 *Phase 2: Interactive Voting Phase.* The objective of the voting phase is to facilitate the consolidation of differentiated
347 options through interactive elements while reinforcing the differentiation across options constructed by participants in
348 the previous phase. This facilitation is achieved by retaining the drag-and-drop functionality for direct manipulation of
349 position and enabling sorting within each category.
350

351 Options are displayed as they are categorized within each category from the previous step and in the following
352 section – Lean Positive, Lean Neutral, Lean Negative, and Skipped or Undecided – as detailed on the bottom half of
353 Figure 1. The Skipped or Undecided category contains options left in the organization queue, possibly because survey
354 respondents have a pre-existing preference or chose not to organize their thoughts further. The original order within
355 these categories is preserved to maintain and reinforce the differentiated options. This ordering sequence mitigated
356 early prototype concerns where uncategorized options were left at the top of the voting interface confusing survey
357 respondents. Respondents have the flexibility to return to the organization interface at any point during the survey to
358 revise their choices.
359

365 In the voting interface, options are draggable, allowing participants to modify or reinforce their preference decisions
 366 as needed. Each category features a sort-by-vote function for reordering within the group, which, although it doesn't
 367 affect the final outcome, supports information organization and consolidation. Both features aim to group similar
 368 options automatically and emphasize proximity, reducing cognitive load by following the proximity compatibility
 369 principle to enhance decision-making [58].
 370

371 While multiple interaction mechanisms exist, drag-and-drop has been extensively explored in rank-based surveys.
 372 For instance, Krosnick et al. [59] demonstrated that replacing drag-and-drop with traditional number-filling rank-based
 373 questions improved participants' satisfaction with little trade-off in their time. Similarly, Timbrook [60] found that
 374 integrating drag-and-drop into the ranking process, despite potentially reducing outcome stability, was justified by the
 375 increased satisfaction and ease of use reported by respondents. The trade-off was deemed worthwhile as QS did not
 376 use the final position of options as part of the outcome if it significantly enhanced user satisfaction and usability [61].
 377 Together, these design decisions led to our belief that a two-phase interface with direct interface manipulation could
 378 reduce the cognitive load for survey respondents to form preference decisions when completing QS.
 379

380 In addition, we made three aesthetic design decisions considering existing QV-based interfaces. First, we removed
 381 visual elements like icons, emojis, progress bars, and vote visualizations, as prior research indicated that emojis could
 382 influence survey interpretations and reduce user satisfaction [62, 16]. While effective visualizations can aid decision-
 383 making, this study does not aim to address that question. Second, the final interface has all options presented on the
 384 screen at the same time, intentionally. Unlike all the prototypes and existing interfaces, prior literature emphasized
 385 the importance of placing all the options on the same digital ballot screen to avoid losing votes. This echoes the
 386 proverb "out of sight, out of mind," where individuals might be biased toward options that are shown to them, and
 387 additional effort is required for individuals to retrieve specific information if options are hidden. Last, we decided to use
 388 a dropdown positioned to the right of each survey option for ease of access to the budget summary when determining
 389 the votes. The layout of the votes and cost was inspired by online shopping cart checkout interfaces where quantities
 390 are supplied next to the itemized costs followed by the total checkout amount. After testing two alternative (Figure 3)
 391 input methods—click-based buttons, which participants dislike making multiple clicks, and a wheel-based design, which
 392 offered intuitive control but was unfamiliar to some participants—we opted for a more accessible dropdown menu for
 393 vote selection.
 394

	Option	Rating	Cost
401	Voting Item Item description will be placed here	- +3 rating +	\$9
405	Voting Item Item description will be placed here	+3 rating \$4 +4 \$9 \$16	\$4 \$9 \$16
409	Voting Item Item description will be placed here	+3 rating	\$9

413 Fig. 3. Alternative vote control. The click-based design (upper) mirrors traditional vote control used in other QV interfaces, where
 414 each click controls one vote. The wheel-based design (the latter two) allows control through both clicks and mouse wheel rotation.
 415

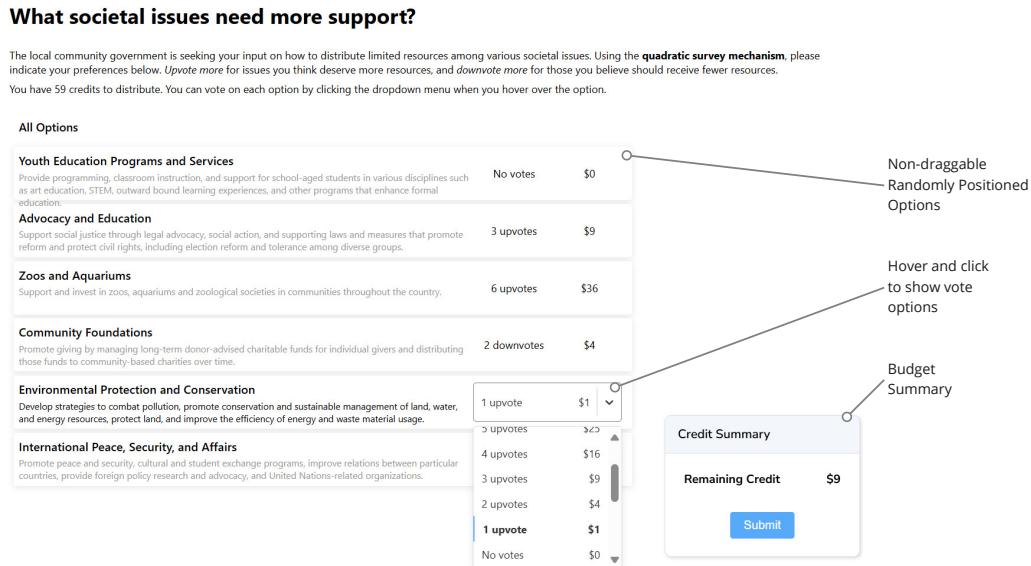


Fig. 4. The text-based interface: This interface is based on the interactive version but does not include the two-phase interactive support and lacks the drag-and-drop functionality. Options are randomly positioned.

3.2 Baseline Interface: Single-Phase Text Interface

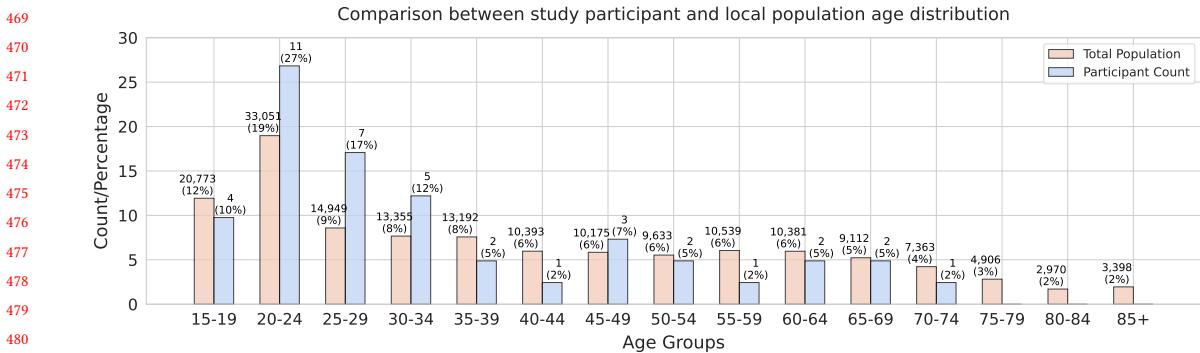
We implemented the single-phase text interface (referred to as text interface for short, Figure 4) as our control condition to compare how the interactive components influenced participants' cognitive load and behavior. The text-based interface, like all existing interfaces, contains a list of static elements, a summary box, and a vote control. We followed the same design considerations, removing visual elements, presenting all options in the same screen, and using the dropdown for vote control, following the two-phase interface interface to provide a more direct comparison. We position the question prompt at the top followed by a randomly ordered option list to prevent ordering bias [63, 64] below. Individual option costs and the remaining credits' summary box are presented to the right of the screen given our interface layout.

Both experimental interfaces were developed with a ReactJS frontend and a NextJS backend powered by MongoDB. We open-source both interfaces.²

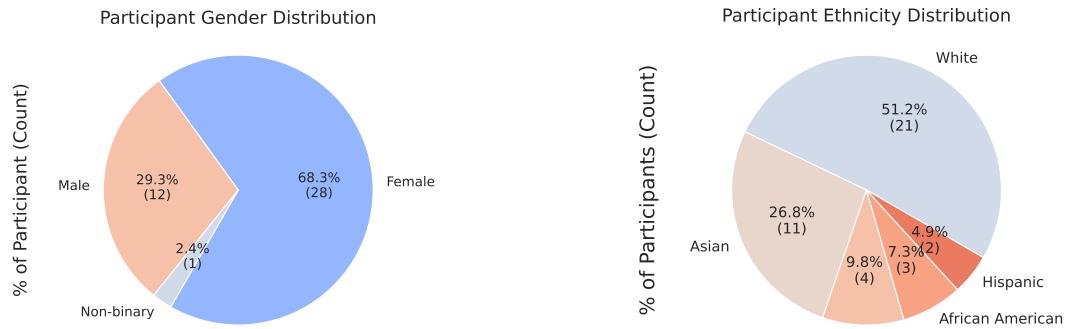
4 Experiment Design

In this section, we describe our experiment design. The study was approved by the university's Institutional Review Board (IRB).

²link-to-github



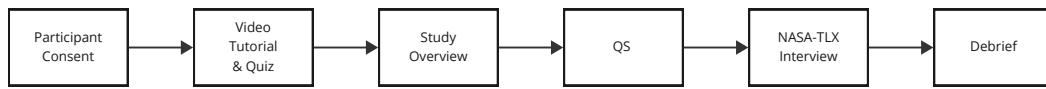
481 (a) Age distribution of the study participants were similar to the locale's demographic profile.



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

499 (c) Ethnicity distribution remains diverse with fewer His-
500 panic and African American participants.

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



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

513 4.1 Recruitment and Participants

515 We recruited 41 participants from a United States college town using online ads, digital bulletins, social media posts,
516 email newsletters, and physical flyers in public spaces beyond campus. We advertised the study as focusing on societal
517 attitudes to mitigate potential response bias. One participant was excluded due to data quality concerns³.

518
519 ³The participant reported not completing the survey seriously, as they believed the experiment was fake.

To ensure diversity, we prioritized non-students by selectively accepting them and monitoring demographic distribution. The mean participant age was 34.63 years, with an age distribution similar to the county's demographic profile (Figure 5a), although there was a slightly higher representation of younger adults. Gender and race demographics are presented in Figures 5b and 5c. Demographic differences between groups were reasonably balanced, although participants using the short text interface skewed slightly younger ($\mu = 32.1$), and those in the long two-phase interface group had a broader age range ($\mu = 38.8$, $\sigma = 19.6$). Full details are provided in Appendix D.

4.2 Experiment Design

We implemented a between-subject design to minimize fatigue, account for the complexity of QS, and avoid learning effects that could influence participants' cognitive load. The experiment focused on public resource allotment, following the methodology of Cheng et al. [4], in which participants expressed preferences across societal issues. These issues are relevant to all citizens and effectively highlight the need to prioritize limited public resources. Participants received a survey with options randomly drawn from the 26 societal topics⁴ evaluated by Charity Navigator [65], an organization that assesses over 20,000 charities in the United States. Randomly selecting the options each participant saw aimed to control for potential systematic content biases introduced by specific voting options across surveys of different lengths.

Participants were randomly assigned to one of four groups:

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

The choice of 6 and 24 options, representing short and long lists, was guided by prior research. Studies recommend fewer than 10 options for constant-sum surveys [66] and fewer than 7 for the Analytic Hierarchy Process [67]. Classic cognitive load research [68, 69] suggests the use of 7 ± 2 items. A meta-analysis by Chernev et al. [70] identified 6 and 24 as common values for short and long lists in choice overload studies, which are rooted in the original experiment by Iyengar and Lepper [46].

4.3 Experiment Procedure

Participant's spent on average 40 minutes (min=27, max=68, $\sigma = 9$) in the lab. Figure 6 visually represents the study protocol detailed in the following subsections.

4.3.1 Consent, Instructions, and Quiz. Participants were invited to the lab to control for external influences and used a 32-inch vertical monitor to display all options. After consenting, participants watched a video explaining the quadratic mechanism without any mention of the interface's operation, followed by a quiz to ensure understanding. Participants rewatched the video or consulted the researcher until they successfully selected the correct answers. Each participant's screen was captured throughout the study.

4.3.2 QS Survey. The researcher informed participants that the study aimed to help local community organizers understand preferences on societal issues to improve resource allocation. Aware that their screens were being recorded, participants completed the survey independently inside a semi-enclosed space in the lab, without the researcher's presence. Once they completed the survey, participants notified the researcher.

⁴See Appendix E for the full list.

573 4.3.3 *NASA-TLX Survey and Interview.* Each participant was provided with a paper-based weighted NASA Task Load
 574 Index (NASA TLX), followed by a semi-structured interview after being informed that the researcher would begin audio
 575 recording. We adopted the paper-based weighted NASA Task Load Index (NASA TLX), a widely used multidimensional
 576 tool that averages six subscale scores to measure overall workload after task completion [71, 72, 73]. NASA-TLX is
 577 favored for its low cost and ease of administration [74], and it exhibits less variability compared to one-dimensional
 578 workload scores [75], making it suitable for our study. While cognitive load can be assessed through performance,
 579 psychophysiological, subjective, and analytical measures [74], the length and complexity of QS make some of these
 580 impractical. Performance and analytical measures require task switching or interruptions, which risk increasing overall
 581 cognitive load and experiment time. Psychophysiological measures, such as pupil size [76] and ECG [77], are costly,
 582 sensitive to external factors, and often require participants to wear additional equipment.
 583

584
 585 4.3.4 *Demographic, Debrief, and Compensation.* After the interview, the researcher collected participant's demographics
 586 and debriefed them, explaining that the study's goal was to understand interface design and cognitive load. Participants
 587 received a \$15 cash compensation.
 588

591 5 Result: Self-Reported Cognitive Load in Quadratic Surveys

592 593 This section presents findings on cognitive load in QS, focusing on how the number of options and different interfaces
 594 influence it (**RQ1**, **RQ2a**). We analyze similarities and differences in cognitive load sources across conditions (**RQ2b**).

595 596 Qualitative findings are based on an inductive thematic analysis [78], conducted after transcribing the interviews.
 597 Snippets were coded according to the research questions and merged into overarching themes. Differences across
 598 599 conditions were refined and validated using a deductive coding process.

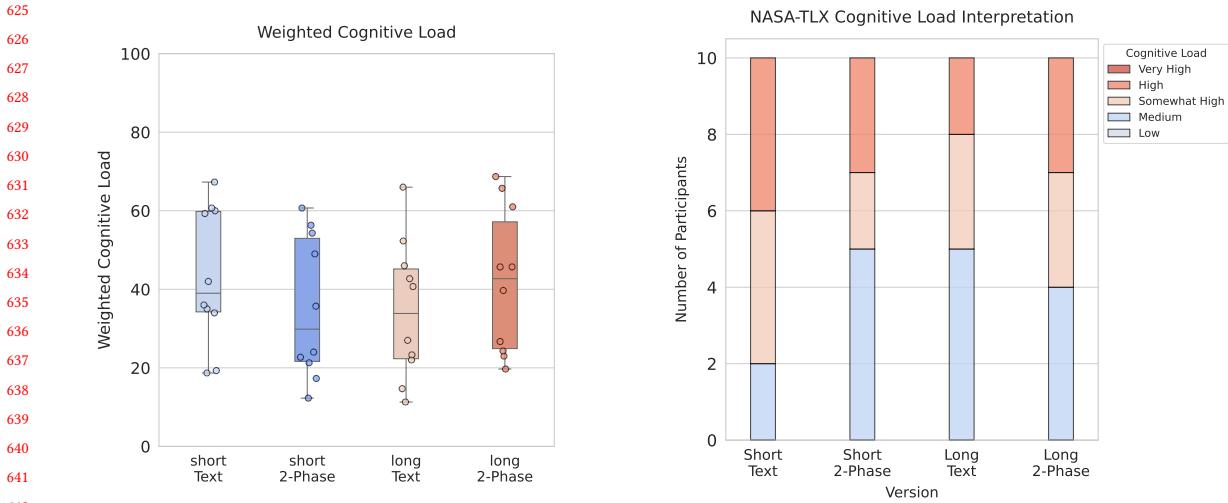
600 601 Quantitative findings are derived from a Bayesian approach, which enhances transparency by interpreting posterior
 602 distributions and moving beyond binary thresholds [79]. Bayesian methods suit various sample sizes, leveraging
 603 maximum entropy priors to ensure conservative and robust inferences [80].
 604

605 5.1 Overall Cognitive Load

606 607 Weighted NASA-TLX uses a continuous 0-100 score, with higher values denoting greater cognitive load. We use
 608 609 predefined mappings of NASA-TLX scores to cognitive levels: low, medium, somewhat high, high, and very high, as
 610 described by Hart and Staveland [71]. Results are shown in Figure 7, with value interpretations presented in Figure 7b.

611 612 Given the sparsity of the data, we modeled the weighted NASA-TLX scores using cognitive levels as ordinal outcome
 613 variables. Then, we developed a hierarchical Bayesian ordinal regression model to analyze ordinal response data. The
 614 model includes length as an ordinal predictor, and interface type as a categorical predictor modeled with hierarchical
 615 priors to allow partial pooling across categories. Interaction effects between length and interface are captured using a
 616 non-centered parameterization constrained by an LKJ prior to account for correlations [80]. We use the same model for
 617 the NASA-TLX subscales. Given that subscales do not have cognitive level interpretations, we constructed weighted bins
 618 to facilitate the ordinal regression model. We present details of this model and additional subscale results in Appendix I.

619 620 In Bayesian analysis, the 94% high-density interval (HDI) represents the range where the true parameter is most
 621 likely to lie. While the results (Figure 8) are not statistically significant because 0 is within this range, the HDI quantifies
 622 probabilistic trends and accounts for uncertainty in a transparent manner.
 623



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

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

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

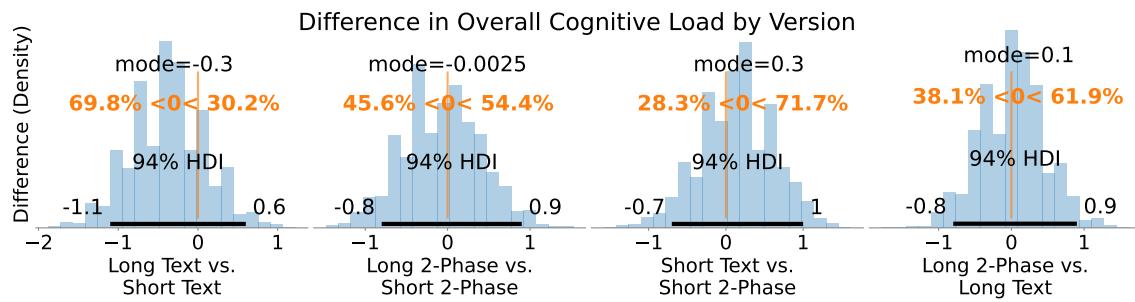


Fig. 8. The figure shows the contrast distribution of the average posterior ordinal category between experimental conditions, highlighting that, while our Bayesian model does not indicate statistically significant differences, longer text interfaces are more likely to reduce cognitive load, and the two-phase interface has a higher probability of lowering cognitive load.

- Increased option length with text interface trends to *reduced* cognitive load with a posterior probability of approximately 69.8%. This reflects a median cognitive load of 33.85 (mean = 34.60, SD = 17.69) compared to a median of 39.00 (mean = 43.23, SD = 17.65).

- Within short QS, the interactive interface trends to *reduced* cognitive load, with a posterior probability of 71.7% supporting the reduction. Participants report a median cognitive load of 29.85 (mean = 35.36, SD = 18.17) under the two-phase interface compared to a median of 39.00 (mean = 43.23, SD = 17.65) under the text interface.
- For the long QS, there trends an *increase* in cognitive load with a posterior probability of 61.9%. The median cognitive load is 42.70 (mean = 42.02, SD = 18.48) under the two-phase interface compared to 33.85 (mean = 34.60, SD = 17.69) in the text interface.

This result contradicts our hypothesis that more options would increase cognitive load and that interfaces can reduce it. Thus, we explore qualitative results to identify possible explanations. To understand the similarities and differences in sources of cognitive load (**RQ2b**), we analyze qualitative results across the six NASA-TLX subscales: mental demand, physical demand, temporal demand, effort, frustration, and performance. Detailed breakdown of each subscale are provided in Appendix F.

5.2 Qualitative Analysis: Common Sources of Cognitive Load

Our analysis reveals several themes across different cognitive load subscales. We identify four themes common to all experimental conditions.

Preference Construction is cited by 97.5% (N=39) of participants as a significant source of mental demand, consistent with prior literature suggesting that preferences are often constructed in context rather than fixed [7]. Specific tasks contributing to this demand include evaluating the relative importance between options (e.g., S002 *Figuring out [...] how much I prioritize option 1 over option 2 , 40% (N = 16)*), making trade-offs due to limited resources (e.g., S005 *[...] very hard to take decisions ... I felt that multiple options deserve equal amounts of credit ... but you have given very limited credit . , 42.5% (N = 17)*), and deciding the exact number of votes (e.g., S023 *[...] having to pick how many upvotes would go to each one , 70% (N = 30)*).

Budget Management emerges as a source of both mental and temporal demand. 25% (N=10) of participants describe the challenge of working with limited credits while trying to maximize their allocation (e.g., S032 *[...] for certain societal issues, you had to ... take away from other issues you could support*). An equal percentage of participants find it mentally taxing to keep track of remaining credits (e.g., S006 *[...] looking at the remaining credits, I'm trying to mentally divide that up before I start allocating*).

Operational Actions refer to reactive efforts addressing immediate, tactical needs. These actions involve direct task execution, responding to constraints without reflection on broader, long-term implications. Examples include adjusting choices to stay within budget (e.g., S003 *I had to alter [...] I kept going under budget*), re-reading options (e.g., S010 *I just had to reread it again*), completing questions efficiently (e.g., S010 *I was trying to be efficient in responding to the question*), and interacting with the survey interface (e.g., S023 *I was trying to be efficient in responding to the question*). 40% (N=16) of participants attribute Operational actions to temporal demand. Additionally, 37.5% (N=15) attribute this cause to frustration, and 32.5% (N=13) attribute it to performance. While this is a commonly cited source across experiment conditions, there are different distributions.

Internal Conflicts and Regretful Trade-offs are cited by 27.5% (N=11) of participants as a source of frustration, particularly when making decisions that conflict with personal values or societal preferences. These findings suggest the potential benefits of Quadratic Surveys (QS) in encouraging participants to balance broader societal considerations and the broader population with their personal preferences.

729 I would have loved to have given more to other groups ...and I felt stressed [...] it's a group that you know is still ...you
 730 know ...important [...] – S020, long text interface
 731

732 733 5.3 Qualitative Analysis: Different Sources of Cognitive Load

734 There are several notable differences between the text and two-phase interfaces.

735 First, regardless of length, when analyzing performance, which refers to a person's perception of their success in
 736 completing a task, participants describe their performances differently. We categorize them into indications of satisficing
 737 behaviors ("good enough"), exhausting their effort (i.e., "done their best"), or feeling positive (i.e., "feeling good") There
 738 are almost twice as many participants using the two-phase interface to report a positive feeling about their final
 739 submission (55% v.s 30% (N=11 vs. 6)).
 740

741 Second, 70% (N=14) of text interface participants attribute operational actions as contributors to effort, double the
 742 percentage observed in the two-phase interface group (35%, N=7). This partially echoes the finding that 90% (N=18) of
 743 text interface participants report mental demand from deciding the exact number of votes, compared to 60% (N=12) in
 744 the two-phase interface group.
 745

746 The distinction between the text and two-phase interfaces becomes more pronounced in the context of the long
 747 survey. 80% of the long text interface participants (N=8) attribute operational actions to effort, compared to only 20%
 748 (N=2) in the long two-phase interfaces. Conversely, 90% of long two-phase interface participants (N=8) attribute effort
 749 to strategic actions, compared to 50% (N=5) in the text interface. **Strategic considerations** refer to reflective decisions
 750 oriented toward long-term goals. They focus on determining priorities, considering broader implications, and aligning
 751 actions with overarching objectives. Mental demand shows similar patterns. 80% of participants (N = 8) in the long text
 752 interface focused on a narrower scope, emphasizing personal relevance and comparing fewer options. In contrast, 90%
 753 of participants (N = 9) in the long two-phase interface consider broader societal impacts and evaluate more options
 754 simultaneously, compared to 30% (N = 3) in the text interface. The following quotes highlight these differences:
 755

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

758 [...] really having to think, especially with so many different societal issues. How do I personally prioritize them? And to what extent
 759 do I prioritize them? Q S009 (L2P)

760 These qualitative differences highlight the variation in **levels of engagement** with the survey content. Participants
 761 using the two-phase interface report higher satisfaction qualitatively about their performance. For the long survey,
 762 they considered broader aspects across different options and how to strategically allocate their credits.
 763

764 765 5.4 Qualitative Analysis: Instances of Satisficing

766 When individuals cannot process all available information, prior research has found that people exhibit *satisficing*
 767 behaviors, which refers to settling for *good enough* rather than *optimal* decisions [81]. While we did not explicitly
 768 ask participants if they 'satisficed,' nor did we measure it quantitatively, we identified satisficing behaviors based on
 769 participants' explanations of how they completed the survey. For example,

770 [...] 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
 771 diminishing returns. I tried to think of enough things [...] and then move on. [...] Q S032 (ST)
 772 I felt like that (the response) was satisfied, but not perfect. Cause perfect is not a reality. Q S036 (ST)

This quote illustrates satisficing decision-making, where participants chose to settle for suboptimal outcomes. Satisficing was observed primarily at the beginning and end of the survey, where participants allocated large amounts of credit initially and then managed the remaining credits to confirm their final vote allocations. For instance,

[...] Because that (the credit) was what was left. [Laughter] I probably wouldn't use that on <optionA> instead of <optionB>. [...]

Q S015 (LT)

I tried to use them [...] it went negative, and then I just settled for just \$6 remaining. [...] I don't think it's perfect. But I think I'm satisfied. Yeah, I'm satisfied.

Q S033 (LT)

[...] when I had first started like looking at the first few, I was just doing it kinda like willy nilly, I'm not really paying that much attention to necessarily how many credits I had, or how many categories there were.

Q S041 (LT)

Participants also exhibited satisficing behaviors regarding *defaults*, particularly when constructing their preferences. For example, participant S003, described how default placements influenced their final decisions:

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

Q S003 (ST)

Our qualitative analysis found that 60% of short-text participants ($N = 6$) and 50% of long-text participants ($N = 5$) expressed instances of satisficing behaviors when describing how they completed the survey, compared to none of the short two-phase participants and 30% of long-text participants ($N = 3$). These qualitative results highlighted potential satisficing behavior from QS participants.

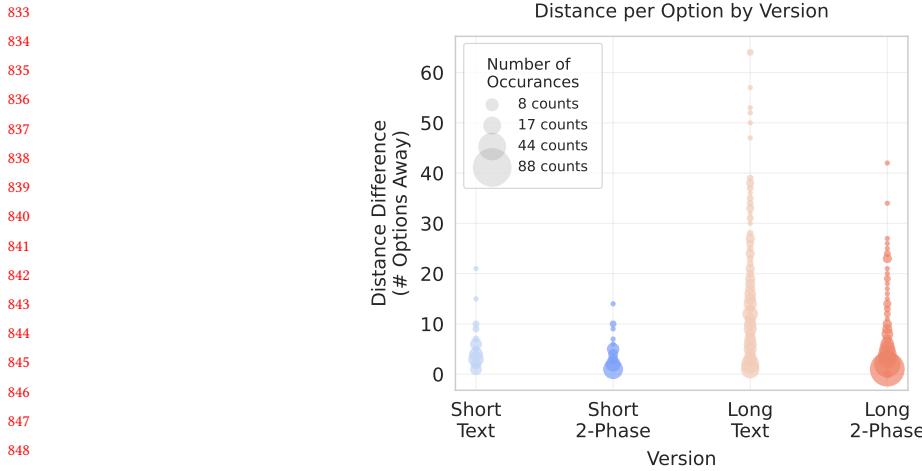
6 Clickstream data: Interface reduces edit distance in long surveys

Following our findings on cognitive load, we analyze voting behaviors to identify differences in how participants cope with survey lengths, how interfaces influence their behavior, and why the long text interface might exhibit lower cognitive load. All data are publicly available⁵ to ensure transparency and support further research. This measure reveals trends in participants' navigation and engagement with survey options. We examine three dimensions of this measure: edit distance per option, edit distance per action, and cumulative edit distance throughout the survey.

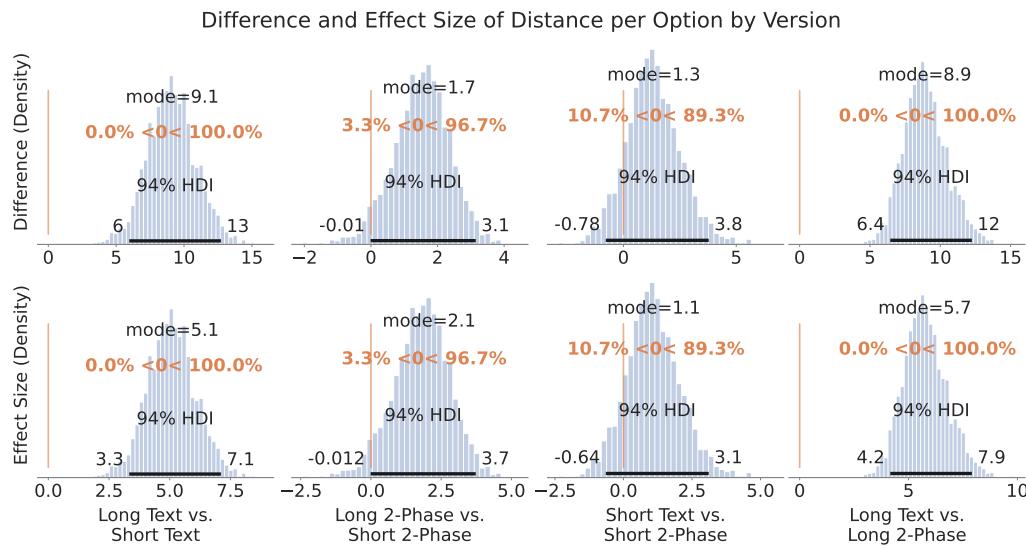
Edit distance per option: We sum up all the distances a participant moves while adjusting values for a single option. Each of these totals is referred to as the edit distance per option. Figure 9 illustrates differences across the four experimental conditions, with the long text interface showing the largest variance in the distance traveled and the highest mean. We implement a hierarchical Bayesian framework to model edit distance differences across experimental conditions. The observed distance differences are modeled using an exponential distribution, where the scale parameter is linked to survey length (treated as an ordinal variable), interface type (treated as a categorical variable), interaction effects between length and interface, and controlling for individual user variability. The linear predictor includes a global intercept and slope for length, random effects for each interface condition with an LKJ prior that captures the correlations among interface categories, and user-specific random effects to account for individual heterogeneity. Detailed mathematical formulations of the model are provided in Appendix K.1.

Figure 10 illustrates the pairwise posterior distributions for differences in edit distances across experimental conditions. For example, the difference in edit distances between the short and long static interfaces has a mode of 9.1, with a 94% highest density interval (HDI) of [6, 13]. This indicates that participants in the long text interface move approximately 9.1 steps more than those in the short text interface, with a high degree of confidence. The effect size is large (mode = 5.1, 94% HDI = [3.3, 7.1]), suggesting a statistically significant difference, which is expected due to the greater number of options in the long text interface.

⁵link-to-github



850 Fig. 9. Edit Distance Per Option: We sum the total number of edit distances for each option, with the figure using the radius to
851 indicate how often a specific edit distance occurred within an experimental condition. Interpretation: Participants in the two-phase
852 interface completed their votes for more options with fewer edit distances, whereas the Long Text interface shows a long tail of
853 options requiring a wider range of edit distances.



875 Fig. 10. The figure shows the contrast distributions of the mean edit distance per option between pairwise experimental conditions,
876 with the first row representing absolute differences and the second row depicting effect sizes. The main finding is that participants in
877 the long text estimated more edit distance per option compared to those in the short text and the long two-phase condition. Notably,
878 the long two-phase interface required estimated only slightly more edit distances despite the longer survey length.

879
880 Similarly, participants using the two-phase interface make approximately 8.9 fewer steps per option (mode = 8.9, 94%
881 HDI = [6.4, 12]) compared to those in the long text interface, with a large effect size (mode = 5.7, 94% HDI = [4.2, 7.9]).
882 Comparatively, the increase in edit distances between the short and long two-phase interfaces is substantially smaller
883

(mode = 1.7, 94% HDI = [-0.01, 3.1]) compared to their static counterparts discussed above. The comparison between the short text and short two-phase interfaces shows weak evidence for a difference, with a mode of 1.3 and a 94% HDI of [-0.78, 3.8]. While the interval includes zero, the posterior distribution slightly favors (with 89.3% probability) the two-phase interface requiring fewer steps. Results from this model suggest that the organization phase in the two-phase interface reduces participants' edit distance per option on average, especially for the long QS.

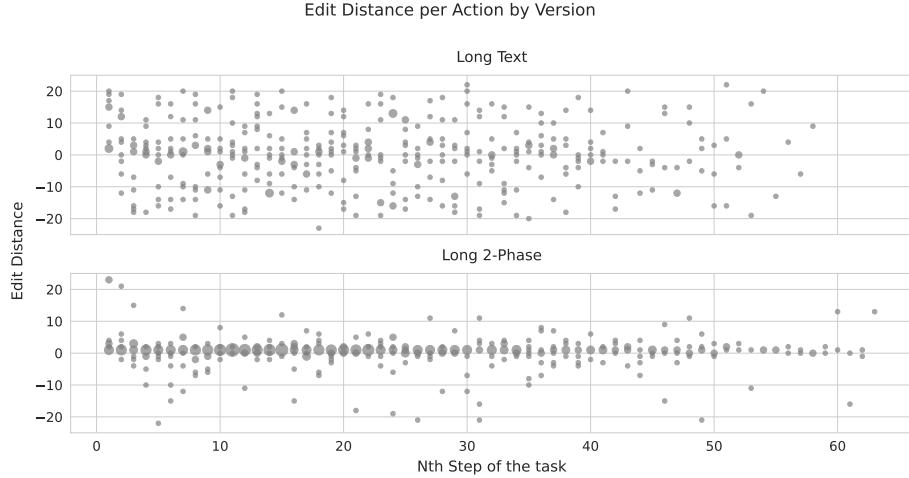


Fig. 11. Edit Distance Per Action: This plot shows the frequency of specific edit distances at each step across the text interface and two-phase interface. Interpretation: Participants in the long two-phase interface tend to make adjustments closer to their previous actions, resulting in visually less variance in edit distances throughout the entire survey.

Edit distance per action: Building on the statistical disparities observed in the previous analysis and the unique patterns exhibited by long text interface participants, we present analyses focusing on edit distance per action and cumulative edit distance throughout the survey between the long text and long two-phase interfaces. Edit distance per action measures how far participants move during each adjustment while completing the survey. Figure 11 illustrates how, at each step, the number of participants moving a given distance (represented by the size of the dots) varies across experimental conditions. Visually, participants move less on average per option within the two-phase interface, with lower variance at smaller scales. This indicates that participants are making local edits, meaning their adjustments tend to occur near their previous edits in terms of edit distance. This also highlights that the organization phase effectively adjusts option positions for easier access, despite participants still having the freedom to move across the interface as all options are presented to them.

In contrast to earlier analyses, we use a hierarchical Bayesian model (detailed in Appendix K.2) to jointly estimate the mean and variance of edit distances across experimental conditions. The model assumes that edit distances are continuous and follow a Normal likelihood. This approach accounts for both central tendencies and variability, using separate predictors for the mean and variance. The model includes hierarchical effects for survey length, interface type, interactions between length and interface, and user-level random effects. Non-centered parametrization is used for survey length and interface type to improve convergence, while interaction effects are modeled with an LKJ prior to capture the correlations between factors. User-level random effects reflect individual differences in behavior, incorporating variability into the model.

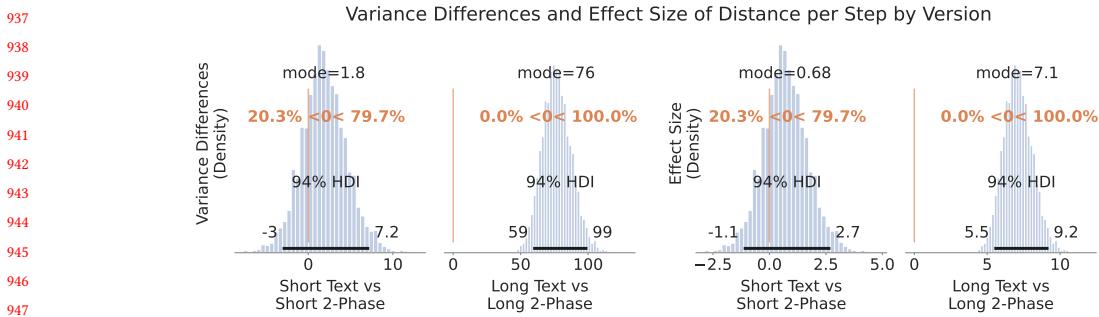


Fig. 12. The figure shows the contrast distributions of the mean edit distance per step between the two-phase interface and text interface for different survey lengths. The left two subplots represent absolute differences, while the right two depict effect sizes. The main finding is that participants in the long text condition exhibited greater variance in edit distance per step compared to those in the long two-phase interface. Similarly, the short text condition showed higher differences, although these were not statistically significant in Bayesian terms.

Figure 12 illustrates the posterior variance distributions, confirming our hypothesis. Participants in the long text interface exhibit greater variance in movement, frequently navigating across the interface, compared to those in the long two-phase interface. This is evidenced by a variance difference mode of 76 (95% HDI = [59, 99]) and a large effect size (mode = 7.1, 95% HDI = [5.5, 9.2]).

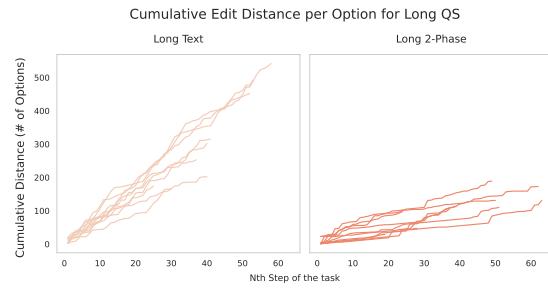


Fig. 13. This plot shows how the cumulative edit distances gained over the course of the survey between long text and long interactive groups. Interpretation: Participants in the long two-phase interface tend to make smaller, more incremental adjustments, resulting in a visually flatter slope compared to the text interface.

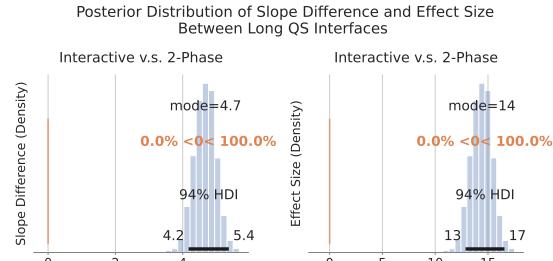


Fig. 14. The figure shows the contrast distributions of slope differences in cumulative edit distance between the two-phase interface and text interface for long QS. The left subplots show absolute differences, while the right depict effect sizes. Main Finding: Participants in the long text interface exhibited a steeper slope, indicating a faster increase in cumulative edit distance compared to the long two-phase interface.

Cumulative edit distance for a participant: This reduction in per action distance due to the two-phase interface's effect on edit distance adds up, as Figure 13 shows the cumulative edit distance over time. Some long text participants traverse double the amount of distance to complete the task compared to the long two-phase participants. We model this growth rate using a hierarchical Bayesian regression model (Detailed in Appendix K.3), with cumulative distance as the predictive variable. The experimental variables include interface type as a categorical variable, individual users modeled with random effects, and steps taken as a continuous variable. The model incorporates a shared global intercept, version-specific intercepts and slopes with partial pooling to balance data across conditions, and user-specific random

989 effects to capture variability. A truncated normal likelihood constrains cumulative distances to positive values and
 990 varies these distances across steps for each participant while masking incomplete data.
 991

992 Figure 14 shows that the slope for the long text interface is approximately 4.7, meaning each step by the text interface
 993 would add 4.7 edit distance (94% HDI = [4.2, 5.4]), compared to the long two-phase interface, which shows a statistically
 994 significant difference with a mode of 1.4 (94% HDI = [1.3, 1.7]). These results explain that the variance in edit distance
 995 per action and the increase in per option edit distance are consistent across participants between the two groups,
 996 showing that the organization phase allows participants to focus on adjusting options within proximity without having
 997 to navigate the interface to locate and make adjustments during the voting phase.
 998

999 **Evidence from qualitative analysis:** Recall the differences in sources of cognitive load between the two exper-
 1000 imental conditions: while two-phase interface participants make adjustments with nearby options, they experience
 1001 cognitive demand from preference construction due to broader considerations involving more options and higher-order
 1002 values. Similarly, the qualitative results highlight that long text interface participants construct narrower preferences,
 1003 yet their edit distance indicates that their movements cover more options.
 1004

1005 Notably, fewer participants (60%, N=6) report precise resource allocation in the long two-phase interface compared
 1006 to 90% in the long text interface. These results make it evident that two-phase interface participants are more focused
 1007 on deliberating preferences than simply completing the survey. Furthermore, the ability to make localized adjustments
 1008 while considering broader decisions suggests that participants construct preliminary preferences during the grouping
 1009 phase, allowing them to focus on deciding their votes.
 1010

1011 These results provide evidence that the initial pass through the survey items, combined with the organizational phase,
 1012 helps participants construct preliminary preferences, thereby reducing the need for large traversals between options.
 1013 This could exemplify that participants in the long text interface are more concerned about operating to 'complete' the
 1014 task (i.e., looking for an option to adjust votes) rather than continuing to stay engaged with the survey options and the
 1015 preference construction task, particularly in the long survey.
 1016

1018 7 Clickstream data: Interface participants' time spent

1019 In addition to distance, participants in the short survey took an average of 2.7 minutes (short-text: $\mu = 2.3$, $\sigma = 1.27$; short
 1020 two-phase: $\mu = 3$, $\sigma = 1.02$), while those in the long survey took 9.7 minutes (long-text: $\mu = 7.5$, $\sigma = 3.45$; long two-phase:
 1021 $\mu = 11.95$, $\sigma = 2.73$). For a fairer comparison of interaction patterns, we analysis total **time-spend-per-option** using
 1022 QS system logs in this section. For participants in the two-phase interface conditions, this includes both organization
 1023 and voting times for that option. The results are visualized in Figure 15.
 1024

1025 Overall, participants spend slightly more time per option in the two-phase interface than in the text interface.
 1026 To quantify these observations, we model the time data as predictive variables of separate Gamma distributions to
 1027 characterize the continuous response times observed under distinct experimental conditions defined by survey length
 1028 and interface type. Each of the four resulting subsets of data is modeled independently, with separate Gamma-distributed
 1029 parameters governing the shape and rate of each group's time distributions.
 1030

1031 We calculated the posterior differences between the two-phase and text interfaces for all pairwise comparisons of
 1032 the four groups. The results in Figure 16 indicate that participants using the two-phase interface consistently spend
 1033 more time per option than those using the text interface, regardless of survey length. For both the short and long QS,
 1034 participants most likely spend 6.1 seconds (94% HDI = [1.0, 11.0]) and 6.7 seconds (94% HDI = [3.7, 9.4]) more per option,
 1035 respectively, with medium effect sizes of $d = 0.49$ (94% HDI = [0.077, 0.89]) and $d = 0.41$ (94% HDI = [0.24, 0.59]). In
 1036 both cases, the intervals lie outside the ROPE of 0 ± 1 , indicating statistical significance. These findings suggest that the
 1037

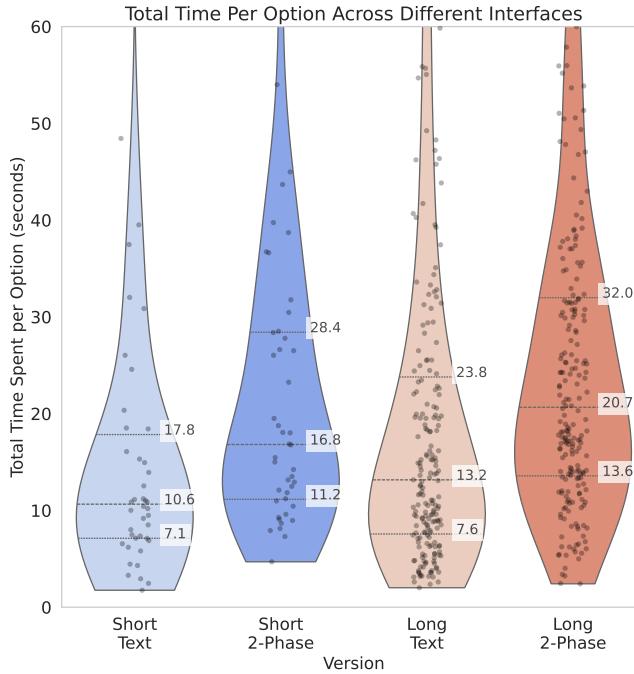


Fig. 15. Total Time per Option: The two-phase interface skewed slightly higher than the text interface, as expected. This discrepancy is due to the additional organization step required in the two-phase interface, resulting in slightly longer overall completion times per option.

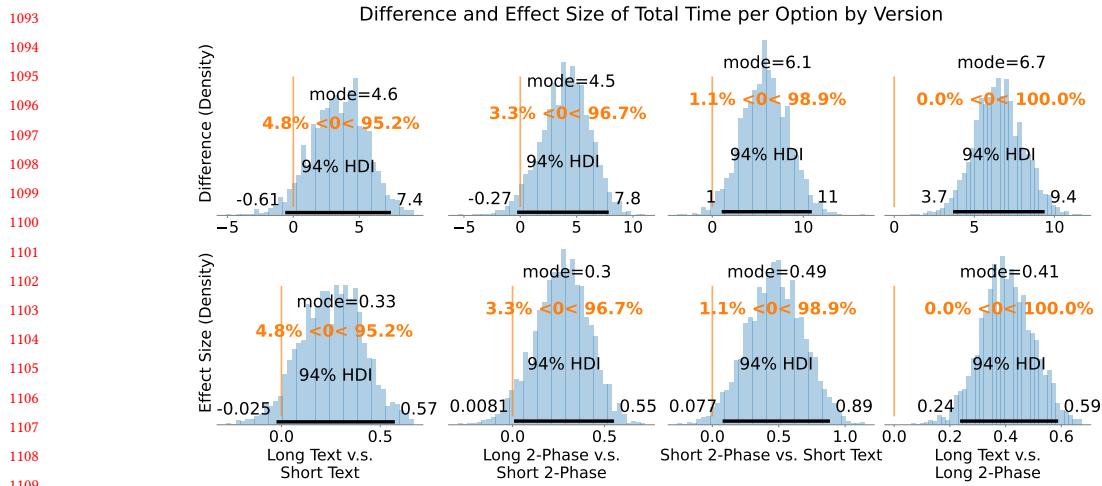
two-phase interface encourages longer deliberation, particularly for longer lists of options. Details of the model are provided in Appendix J.

Some literature points to increased time can lead to cognitive fatigue [82, 83], which can impair decision-making. Other decision science literature suggests that longer decision times can indicate deeper cognitive processing [84, 45]. Our qualitative analysis points to the latter.

Descriptively, participants in the long two-phase condition remained actively engaged during the voting phase, editing their votes an average of 39.3 times per participant ($\sigma = 39.3$, range=19 – 63) compared to 39.1 times ($\sigma = 13.29$, range=15 – 58) in the long text condition. This suggests that the two-phase interface does not reduce engagement despite the additional organization step.

Quantitatively, other than the difference in operational thinking and strategic consideration discussed in Section ??, we find that 37.5% of participants (N=15) who attribute time to *Decision Making* as a source of temporal demand frame such demand differently. We label a participant as *affirmative* if they describe the pressure to make decisions as a source of temporal demand. For example, S022 *So it didn't take too much time, but obviously there were a lot of things to consider, so there was some temporal demand.* is an affirmative statement. Conversely, we label a participant as *negative* if they express concern about the time and effort they have already invested. For example, S024 *maybe I should just hurry up and make a decision.* is a negative statement.

50% of participants (N=5) in the long two-phase group describe the pressure to make decisions affirmatively and none negatively. This suggests that their pressure stems from having too many remaining decisions to make, rather than



1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144

Fig. 16. The figure shows the contrast distributions of the mean time to complete per option between pairwise experimental conditions, with the first row representing absolute differences and the second row depicting effect sizes. The main finding is that participants in the long two-phase condition spent more time per option compared to those in the long text and short two-phase conditions. Additionally, short two-phase participants took longer per option than short text participants.

from the time already invested. This is reflected in their higher average time spent per option and overall time spent ($\mu = 716.86$ seconds, $\sigma = 164.04$ seconds) completing the QS survey compared to the long text group ($\mu = 449.64$ seconds, $\sigma = 206.97$ seconds). We interpret this as evidence that participants are thoughtfully engaged in constructing their preferences and choose to invest additional time, rather than being driven by decision-related pressures or experiencing a sense of urgency.

Conversely, in the short text group, 50% of participants (N=5) express concern about the time and effort they have already invested (S024 *Q: maybe I should just hurry up and make a decision.*) and none frame it affirmatively. Descriptively, participants in the short text group spend comparatively less time than those in the long QS (short text: $\mu = 139.83$ seconds, $\sigma = 76.43$ seconds; short two-phase: $\mu = 178.78$ seconds, $\sigma = 61.07$ seconds). This suggests that participants in the short text group expect themselves to complete the task sooner than they actually do.

Surprisingly, participants in the long text interface exhibit a temporal demand lower than the short text and long 2-phase participants(Figure 17, quantifiable results in Appendix I.1.7), despite spending more time per option and traversing the longest distance (Section 6). Only 30% of participants (N=3) mention the time spent making a decision as a source of temporal demand. One possible explanation is that some participants are satisficing, as we pointed out in Section 5.4.

In summary, we interpret the result that participants in the two-phase interface spend more time per option as a sign of deeper cognitive processing. This is further supported by examining participants' nuanced voting behaviors under budget constraint conditions for the long QS, which we omit for brevity. Notably, two-phase interface participants make more small vote adjustments (i.e., adding or removing at most 2 votes on an option) when they have fewer remaining credits, further supporting our claim that they experience deeper engagement with preference construction, which we elaborate on further in Appendix H.

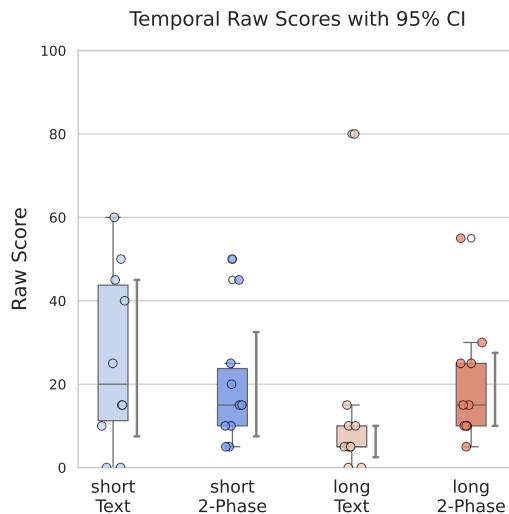


Fig. 17. Temporal Demand Raw Score: The short text interface results in the highest temporal demand, while the long text interface shows the lowest. Two-phase interfaces exhibit moderate temporal demand, suggesting that interactive elements allowed participants to pace themselves better.

8 Discussion and Future Works

In this section, we interpret the findings on cognitive load and respondent behavior in QS. We focus on the rationale and elements supporting the two-phase interface for preference construction and its potential to mitigate satisficing behavior. Additionally, we offer usage and design recommendations for practitioners and outline future directions for improving QS interfaces.

8.1 Two-phase interface: a worthwhile trade-off

Decision makers who deploy surveys aim to elicit thoughtful responses from participants. This means the interface should balance survey usability, respondent satisfaction, and the effort individuals invest in their responses. Results from the study lead us to conclude that the two-phase interface encouraged deeper participant engagement with the options and reduced satisficing behaviors, despite its increased time per option and higher cognitive load for long QS.

8.1.1 Analysis through the lens of cognitive load theory. Cognitive load theory [54], when applied to QS, identifies three components of cognitive load: intrinsic load (the cognitive demand required to understand questions and response options), germane load (associated with deeper processing and preference evaluation), and extraneous load (stemming from navigating and operating the survey interface).

Participants were randomly assigned to experimental conditions, with survey lengths containing options randomly drawn from a common pool to control intrinsic load within the same group.

When QS is short, participants can engage with all options simultaneously. Participants using the two-phase interface traded a slightly longer survey response time for a potential reduction in cognitive load and edit distance. We interpret

1197 this as participants freeing up cognitive demand from extraneous load for germane load, prompting them to better
 1198 construct and express their preferences.
 1199

1200 When QS is long, participants face more options, resulting in a higher intrinsic load at the start of the survey. We
 1201 believe the two-phase interface traded a longer survey response time and a potential increase in cognitive load for
 1202 deeper engagement with the survey. Quantitatively, these participants made shorter traversals while spending more
 1203 time on each option. Qualitatively, they reported experiencing demand more from strategic considerations (germane
 1204 load) than from operational actions (extraneous load), which text interface participants experienced more frequently.
 1205

1206 While some might argue that the additional organizing phase offers participants more opportunities to familiarize
 1207 themselves with the options compared to text interface participants, the greater overall edit distance and high variance
 1208 in edit distance per option suggest that text interface participants traversed the list frequently. This finding is further
 1209 supported by qualitative data, where 70% of long-text participants (N=7) reported scanning the list while voting. This
 1210 behavior suggests that while long-text participants had opportunities to familiarize themselves with the options, the
 1211 explicit organization phase encouraged deeper reflection on their preferences.
 1212

1213 The effect of the two-phase interface shows nuanced differences influencing cognitive load outcomes; however, both
 1214 analyses suggest that the two-phase interface *shifted* participants' cognitive focus when completing QS.
 1215

1216 **8.1.2 Potential in limiting Satisficing.** Qualitative findings (Section 5.4) on potential satisficing behavior highlight the
 1217 importance of careful consideration when deploying long QS. However, the two-phase interface appeared to limit
 1218 satisficing behaviors, as evidenced by fewer observations compared to the long text interface for long QS and none for
 1219 short QS. We believe the potential reasons lie in the design of the two-phase interface, which scaffolds the preference
 1220 construction process.
 1221

1222 The deliberate one-option-at-a-time presentation during the voting task in the two-phase interface reduced re-
 1223 liance on defaults and encouraged deeper reflection using cognitive strategies such as *problem decomposition* [85] and
 1224 *dimension reduction*, both of which are known to reduce cognitive overload.
 1225

1226 When asked about their experience with the interface, four participants highlighted how the organization phase
 1227 supported their preference construction. S013 illustrated how the one-option-at-a-time approach reduced the dimensions
 1228 of decision-making:
 1229

1230 [...] it (organization phase) gives you time to just focus on that single thing and rank it based on how you feel at that moment.
 1231 ☰ S013 (SI)
 1232

1233 This focused mode enabled deeper reflection. When considering relative preferences among QS, S013 described how
 1234 it structurally decomposed the problem:
 1235

1236 [...] to have a preliminary categorization of all the topics [...] (allowed me) to think about and process [...] digest all the information
 1237 prior to actually allocating the budget [...] ☰ S009 (L2P)

1238 This quote highlighted how participants' deliberation occurred during the organization phase, enabling them to focus
 1239 on constructing preferences without worrying about budget management—both of which are cited sources of cognitive
 1240 load. Although direct measurement of satisficing behavior reduction is challenging, qualitative data and participant
 1241 feedback suggest that the two-phase interface has the potential to limit such behaviors. Based on this evidence, we
 1242 advise against using long QS unless paired with a two-phase interface and ample time for participants to complete. We
 1243 suggest future research investigate the mental processes underlying satisficing behaviors in long QS.
 1244

1245 **In summary**, we argue that the trade-off of a longer completion time and potentially higher cognitive load in
 1246 the two-phase interface is justified. Drawing on cognitive load theory, we propose that the interface fosters deeper
 1247

1249 engagement with the options. Additionally, our qualitative findings and participant feedback suggest that the interface
1250 may reduce satisficing, aligning with decision-makers' goals of obtaining thoughtful and deliberate responses from
1251 participants.
1252

1253 8.2 Preference Construction guided by Organize, Then Vote

1254 Completing QS involves a series of in-situ difficult decision tasks Lichtenstein and Slovic [7]. As one participant reflected
1255 when completing the survey with options they had never considered before:
1256

1257 *Oh, there are other aspects that I never care about. [...] Why (should) I spend money on that?*

1258  S037 (L2P)

1259 When processing these unfamiliar options, we believe the two-phase interface supported participants' preference
1260 construction process.
1261

1262 First, 40% of long-text participants (N=3) found it challenging to facilitate differentiation without interactive tools
1263 that would allow grouping or drag-and-drop, as S025 said:

1264 *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
1265 myself categories and subcategories out of this list ... If I could organize it.*

 S025 (LT)

1266 In contrast, 60% (N=6) of long two-phase participants appreciated the upfront introduction of all options, which
1267 enabled them to organize and use drag-and-drop features to facilitate completing QS. Not only did participants use
1268 drag-and-drop options post-voting to reflect and ensure correct vote allocation, but it also enabled participants, like S039,
1269 to make fine-grained comparisons between options:

1270 *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
1271 here, and then compare it to the top one [...]*

 S039 (SI)

1272 This supports our intention of applying Svenson [52]'s differentiation and consolidation theory, where participants
1273 attempt to identify differences and eliminate less favorable options. The organization phase and the drag-and-drop
1274 supported some degree of differentiation process.

1275 *[...] the hardest part deciding in which category of place (prefernece bin) each issue is.*

 S021 (L2P)

1276 This quote by S021 best represents the potential of the organization phase in separating part of the difficult decisions
1277 one needs to make when differentiating their preferences during preference construction. With the selected options, the
1278 shorter edit distance of long two-phase interface participants suggested that they were consolidating their identified
1279 preferences through votes.

1280 8.3 What We Learned: Quadratic Survey Usage and Design Recommendations

1281 This study represents a crucial step toward developing better interfaces to support individuals responding to QS, by
1282 providing a deeper understanding of how survey respondents interact with QS and the sources of cognitive load. In this
1283 subsection, we outline usage and design recommendations applicable to all applications of the quadratic mechanism.

1284 8.3.1 *QS should have Limited Options or for critical evaluations.* We recommend that QS, even with our two-phase
1285 interface, be deployed with limited options or used for critical evaluations, such as eliciting stakeholder preferences
1286 before making investment decisions, as our findings reveal complex cognitive challenges and increased time requirements
1287 when the number of options grows. Even though our two-phase interface scaffolds the decision-making process, we
1288 suggest that practitioners allow ample time for survey participants to deliberate on the options and complete their
1289

1301 responses. When the two-phase QS interface is not available, survey designers should present the options ahead of
 1302 time, allowing participants to familiarize themselves with the choices and deliberate before completing the QS.
 1303

1304 **8.3.2 Facilitate Quadratic Mechanism Applications through Categorization, Not Ranking.** We suggest that future quadratic
 1305 mechanism interface designs focus on categorization rather than ranking. With or without the organization phase,
 1306 participants did not exhibit a ranking process. The final 'rank' of option preferences often emerged as a byproduct
 1307 of vote allocation, constructed in situ. Therefore, for survey designs to effectively construct preferences, it is more
 1308 important to facilitate differentiation than to focus on direct manipulation for fine-tuning.
 1309

1310 We demonstrated this through the organization phase, where participants exhibited deeper engagement with survey
 1311 options and potentially completed the survey more effectively. We believe this approach should extend beyond QS to
 1312 other ranking-based survey tools, such as ranked-choice voting and constant-sum surveys. Further research should
 1313 examine how implementing such functionality influences survey respondents' mental models.
 1314

1315

1316 **8.4 Future work: Opportunities for Better Budget Management**

1317 Budget management emerged as one of the most prominent issues in our study, which the two-phase interface did not
 1318 address. 35% of participants ($N = 14$) emphasized the ability of current quadratic mechanism applications to perform
 1319 automated calculations, but noted that this is not sufficient. We identified three key challenges for future work:
 1320

1321 First, participants struggled to decide on an initial vote allocation. Some distributed credits equally across options,
 1322 while others used 1, 2, or 3 votes as starting points. A few anchored their decisions to the tutorial's example of
 1323 four upvotes. This suggests a need to better understand whether individuals have absolute value preferences among
 1324 options. Second, 12.5% of participants ($N = 5$) expressed confusion about the relationship between budget, votes,
 1325 and outcomes, despite understanding their definitions. They struggled to make trade-offs between votes and budget,
 1326 leading to frustration and hampered decision-making. Third, determining the absolute amount of credits in QS is highly
 1327 demanding. Designing interfaces and interactions to address the cold start challenge and help participants decide on
 1328 the absolute vote value, while also considering ways to limit direct influences, remains an open question.
 1329

1330 We believe that, with the power of computing and a better understanding of how individuals calculate trade-offs can
 1331 provide innovative solutions to help participants more easily express their preferences using QS.
 1332

1333

1334 **9 Limitations**

1335 Evaluating the QS interface is challenging because of its novelty. During the study, we identified several limitations
 1336 that warrant further research.
 1337

1338

1339 ***Individual differences in cognitive capacity.*** Variations in individual cognitive capacity influenced participants' per-
 1340 formance and cognitive scores. For example, participants with greater experience in decision-making may be better
 1341 able to manage multiple options. A within-subject study could clarify shifts in cognitive load, but deconstructing
 1342 established preferences and altering options introduces additional complexity. Therefore, we opted for this in-depth,
 1343 between-subject study, although the small sample size may introduce noise, potentially distorting the measurement of
 1344 cognitive load. Future research should aim to quantify the impact of different QS interfaces on cognitive load at a larger
 1345 scale. Furthermore, participants completed this study in a controlled laboratory environment, with options displayed
 1346 on a large screen. Future work should also investigate how individuals respond to QS on smaller devices and in less
 1347 controlled environments.
 1348

1349

Limited experience with QS. Participants lacked prior experience with the QS interface. After completing a tutorial and quiz, participants proceeded to perform tasks using the QS interface. While participants understood the mechanics of QS, their familiarity with the interface likely influenced their strategies and cognitive load. As quadratic mechanisms become more prevalent, future research could compare the performance of novices and experts.

Duration between clicks and edit distance to represent decision-making. While time and distance are common metrics for quantifying decision-making, it is likely that participants considered other options simultaneously. We acknowledge that these metrics are approximate indicators of decision-making effort. Despite these limitations, this approach provides valuable insights into decision-making within our experimental constraints.

10 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 reduced participant’s edit distance between vote adjustments throughout the survey despite spending more time per option. Qualitative insights highlighted two-phase interface encouraged more iterative and reflective preference construction and its potential at reducing satisficing behaviors even though it did not clearly reduce overall cognitive load for the longer QS. Nonetheless, this design shift promoted deeper engagement and strategic thinking compared to the text-based interface, by distributing cognitive effort more effectively. By integrating the organization and drag-and-drop functions, the interface facilitated both preference differentiation and consolidation, making it easier for respondents to refine their decisions. This two-phase interface design supports the development of future software tools that facilitate preference construction and promote the broader adoption of Quadratic Surveys. Future research should explore how to better support individuals in deciding the allocation of budget and design interfaces for smaller devices.

References

- [1] Martin Pielot and Mario Callegaro. 2024. Did You Misclick? Reversing 5-Point Satisfaction Scales Causes Unintended Responses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, (May 2024), 1–7. doi: [10.1145/3613904.3642397](https://doi.org/10.1145/3613904.3642397).
- [2] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, (May 2019), 1–12. doi: [10.1145/3290605.3300316](https://doi.org/10.1145/3290605.3300316).
- [3] Muhsin Ugur, Dvijesh Shastri, Panagiotis Tsiamyrtzis, Malcolm Dcosta, Allison Kalpakci, Carla Sharp, and Ioannis Pavlidis. 2015. Evaluating smartphone-based user interface designs for a 2d psychological questionnaire. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 275–282.
- [4] Ti-Chung Cheng, Tiffany Li, Yi-Hung Chou, Karrie Karahalios, and Hari Sundaram. 2021. “I can show what I really like.”: Eliciting Preferences via Quadratic Voting. *Proceedings of the ACM on Human-Computer Interaction*, 5, (Apr. 2021), 1–43. doi: [10.1145/3449281](https://doi.org/10.1145/3449281).
- [5] Theodore Groves and John Ledyard. 1977. Optimal Allocation of Public Goods: A Solution to the “Free Rider” Problem. *Econometrica*, 45, 4, 783–809. JSTOR: [1912672](https://doi.org/10.2307/1912672). doi: [10.2307/1912672](https://doi.org/10.2307/1912672).
- [6] David Quarfoot, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein, and Ellen Konar. 2017. Quadratic voting in the wild: real people, real votes. *Public Choice*, 172, 1–2, 283–303.
- [7] Sarah Lichtenstein and Paul Slovic, eds. 2006. *The Construction of Preference*. (1. publ ed.). Cambridge University Press, Cambridge.
- [8] Adam Rogers. 2019. Colorado Tried a New Way to Vote: Make People Pay—Quadratically | WIRED. *Wired*, (Apr. 2019). Retrieved June 22, 2024 from.

- 1405 [9] 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. (). Retrieved June 13, 2024 from.
- 1406 [10] Herbert A. Simon. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69, 1, 99–118. JSTOR: 1884852. doi: 10.2307/1884852.
- 1407 [11] John W. Payne, James R. Bettman, and Eric J. Johnson. 1988. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 3, (July 1988), 534–552. doi: 10.1037/0278-7393.14.3.534.
- 1408 [12] Amos Tversky and Daniel Kahneman. [n. d.] Judgments of and by Representativeness.
- 1409 [13] Erik J Engstrom and Jason M Roberts. 2020. *The Politics of Ballot Design: How States Shape American Democracy*. Cambridge University Press.
- 1410 [14] Bert Weijters, Elke Cabooter, and Niels Schillewaert. 2010. The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 3, (Sept. 2010), 236–247. doi: 10.1016/j.ijresmar.2010.02.004.
- 1411 [15] N. D. Kieruj and G. Moors. 2010. Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of Public Opinion Research*, 22, 3, (Sept. 2010), 320–342. doi: 10.1093/ijpor/edq001.
- 1412 [16] Vera Toepoel, Brenda Vermeeren, and Baran Metin. 2019. Smileys, Stars, Hearts, Buttons, Tiles or Grids: Influence of Response Format on Substantive Response, Questionnaire Experience and Response Time. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 142, 1, (Apr. 2019), 57–74. doi: 10.1177/0759106319834665.
- 1413 [17] Habiba Farzand, David Al Baitay Suarez, Thomas Goodge, Shaun Alexander Macdonald, Karola Marky, Mohamed Khamis, and Paul Cairns. 2024. Beyond Aesthetics: Evaluating Response Widgets for Reliability & Construct Validity of Scale Questionnaires. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems* (CHI EA '24). Association for Computing Machinery, New York, NY, USA, (May 2024), 1–7. doi: 10.1145/3613905.3650751.
- 1414 [18] Christian Jilek Paula Gauselmann Yannick Runge and Tobias Tempel. 2023. A relief from mental overload in a digitalized world: How context-sensitive user interfaces can enhance cognitive performance. *International Journal of Human-Computer Interaction*, 39, 1, 140–150. eprint: <https://doi.org/10.1080/10447318.2022.2041882>. doi: 10.1080/10447318.2022.2041882.
- 1415 [19] Sharon Oviatt. 2006. Human-centered design meets cognitive load theory: designing interfaces that help people think. In *Proceedings of the 14th ACM International Conference on Multimedia*, 871–880.
- 1416 [20] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2021. To reuse or not to reuse? A framework and system for evaluating summarized knowledge. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW1, (Apr. 2021). doi: 10.1145/3449240.
- 1417 [21] Helena M Reis et al. 2012. Towards reducing cognitive load and enhancing usability through a reduced graphical user interface for a dynamic geometry system: An experimental study. In *2012 IEEE International Symposium on Multimedia*. IEEE, 445–450.
- 1418 [22] Benjamin Lafreniere, Andrea Bunt, and Michael Terry. 2014. Task-centric interfaces for feature-rich software. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design* (OzCHI '14). Association for Computing Machinery, New York, NY, USA, 49–58. doi: 10.1145/2686612.2686620.
- 1419 [23] Soomin Kim, Jinsu Eun, Joseph Seering, and Joonhwan Lee. 2021. Moderator chatbot for deliberative discussion: Effects of discussion structure and discussant facilitation. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW1, (Apr. 2021). doi: 10.1145/3449161.
- 1420 [24] Emin İbili. 2019. Effect of augmented reality environments on cognitive load: pedagogical effect, instructional design, motivation and interaction interfaces. *International Journal of Progressive Education*, 15, 5, 42–57.
- 1421 [25] Amy X. Zhang and Justin Cranshaw. 2018. Making sense of group chat through collaborative tagging and summarization. *Proc. ACM Hum.-Comput. Interact.*, 2, CSCW, (Nov. 2018). doi: 10.1145/3274465.
- 1422 [26] Steven P Lalley, E Glen Weyl, et al. 2016. Quadratic voting. Available at SSRN.
- 1423 [27] Eric A Posner and E Glen Weyl. 2018. *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton University Press.
- 1424 [28] Ryan Naylor et al. 2017. First year student conceptions of success: What really matters? *Student Success*, 8, 2, 9–19.
- 1425 [29] Charlotte Cavaille and Daniel L. Chen. [n. d.] Who Cares? Measuring Preference Intensity in a Polarized Environment.
- 1426 [30] Vitalik Buterin, Zoë Hitzig, and E. Glen Weyl. 2019. A Flexible Design for Funding Public Goods. *Management Science*, 65, 11, (Nov. 2019), 5171–5187. doi: 10.1287/mnsc.2019.3337.
- 1427 [31] Luis Mota Freitas and Wilfredo L. Maldonado. 2024. Quadratic funding with incomplete information. *Social Choice and Welfare*, (Feb. 2024). doi: 10.1007/s00355-024-01512-7.
- 1428 [32] Tobin South, Leon Erichsen, Shrey Jain, Petar Maymounkov, Scott Moore, and E. Glen Weyl. 2024. Plural Management. SSRN Scholarly Paper. Rochester, NY, (Jan. 2024). doi: 10.2139/ssrn.4688040.
- 1429 [33] 2023. Gov4git: A Decentralized Platform for Community Governance. (Mar. 2023). Retrieved June 13, 2024 from.
- 1430 [34] 2024. RadicalxChange/quadratic-voting. RadicalxChange. (May 2024). Retrieved June 17, 2024 from.
- 1431 [35] [n. d.] Read the Whitepaper | Gitcoin. <https://www.gitcoin.co/whitepaper/read/>. (). Retrieved June 17, 2024 from.
- 1432 [36] [n. d.] About RxC. <https://www.radicalxchange.org/wiki/about/>. (). Retrieved June 17, 2024 from.
- 1433 [37] Syavash Noabarany, Louise Oram, Vasanth Kumar Rajendran, Chi-Hsiang Chen, Joanna McGrenere, and Tamara Munzner. 2012. The design space of opinion measurement interfaces: exploring recall support for rating and ranking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2035–2044.
- 1434 [38] 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.

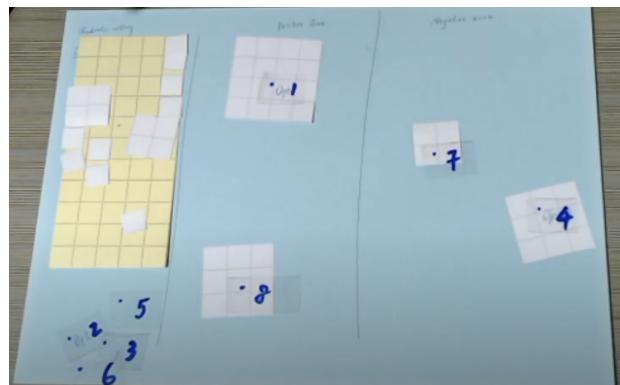
- [39] Jing Wei, Weiwei Jiang, Chaofan Wang, Difeng Yu, Jorge Goncalves, Tilman Dingler, and Vassilis Kostakos. 2022. Understanding how to administer voice surveys through smart speakers. *Proc. ACM Hum.-Comput. Interact.*, 6, CSCW2, (Nov. 2022). doi: [10.1145/3555606](https://doi.org/10.1145/3555606).
- [40] 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).
- [41] 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.
- [42] Graham Cooper. 1998. Research into cognitive load theory and instructional design at UNSW. (1998).
- [43] Stoo Sepp, Steven J. Howard, Sharon Tindall-Ford, Shirley Agostinho, and Fred Paas. 2019. Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory. *Educational Psychology Review*, 31, 2, (June 2019), 293–317. doi: [10.1007/s10648-019-09461-9](https://doi.org/10.1007/s10648-019-09461-9).
- [44] Antonio Drommi, Gregory W Ulferts, and Dan Shoemaker. 2001. Interface design: A focus on cognitive science. In *The Proceedings of ISECON 2001*. Vol. 18.
- [45] Kahneman Daniel. 2017. *Thinking, Fast and Slow*.
- [46] 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.
- [47] 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.
- [48] 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).
- [49] Peter Coy. 2019. A New Way of Voting That Makes Zealotry Expensive - Bloomberg. *Bloomberg*, (May 2019). Retrieved Dec. 16, 2023 from.
- [50] 2022. Quadratic Voting Frontend. Public Digital Innovation Space. (Jan. 2022). Retrieved Dec. 16, 2023 from.
- [51] 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).
- [52] 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).
- [53] 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).
- [54] John Sweller. 2011. Cognitive Load Theory. In *Psychology of Learning and Motivation*. Vol. 55. Elsevier, 37–76. doi: [10.1016/B978-0-12-387691-1.00002-8](https://doi.org/10.1016/B978-0-12-387691-1.00002-8).
- [55] 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).
- [56] 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.
- [57] A Norman Donald. 2013. *The Design of Everyday Things*. MIT Press.
- [58] 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.
- [59] Jon A Krosnick, Charles M Judd, and Bernd Wittenbrink. 2018. The measurement of attitudes. In *The Handbook of Attitudes*. Routledge, 45–105.
- [60] Jerry P Timbrook. 2013. *A Comparison of a Traditional Ranking Format to a Drag-and-Drop Format with Stacking*. PhD thesis. University of Dayton.
- [61] 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.
- [62] 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).
- [63] 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).
- [64] 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).
- [65] 2023. Charity Navigator. <https://www.charitynavigator.org/index.cfm?bay=search.categories>. (May 2023). Retrieved Dec. 16, 2023 from.
- [66] 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).
- [67] 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).
- [68] 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).
- [69] Thomas L Saaty and Muğan S Ozdemir. 2003. Why the magic number seven plus or minus two. *Mathematical and computer modelling*, 38, 3-4, 233–244.
- [70] 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).

- [71] 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.
- [72] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50, 9, (Oct. 2006), 904–908. doi: [10.1177/154193120605000909](https://doi.org/10.1177/154193120605000909).
- [73] Brad Cain. 2007. A review of the mental workload literature. *DTIC Document*.
- [74] Qin Gao, Yang Wang, Fei Song, Zhizhong Li, and Xiaolu Dong. 2013. Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics*, 56, 7, (July 2013), 1070–1085. doi: [10.1080/00140139.2013.790483](https://doi.org/10.1080/00140139.2013.790483).
- [75] 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).
- [76] Oskar Palinko, Andrew L. Kun, Alexander Shyrokov, and Peter Heeman. 2010. Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*. ACM Press, Austin, Texas, 141. doi: [10.1145/1743666.1743701](https://doi.org/10.1145/1743666.1743701).
- [77] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. 2010. Psycho-physiological measures for assessing cognitive load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*. ACM, Copenhagen Denmark, (Sept. 2010), 301–310. doi: [10.1145/1864349.1864395](https://doi.org/10.1145/1864349.1864395).
- [78] 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).
- [79] Matthew Kay, Gregory L Nelson, and Eric B Hekler. 2016. Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4521–4532.
- [80] Richard McElreath. 2018. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Chapman and Hall/CRC.
- [81] 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).
- [82] Thomas Kundinger, Celena Mayr, and Andreas Riener. 2020. Towards a Reliable Ground Truth for Drowsiness: A Complexity Analysis on the Example of Driver Fatigue. *Proceedings of the ACM on Human-Computer Interaction*, 4, EICS, (June 2020), 1–18. doi: [10.1145/3394980](https://doi.org/10.1145/3394980).
- [83] Enamul Karim, Hamza Reza Pavel, Sama Nikanfar, Aref Hebri, Ayon Roy, Harish Ram Nambiappan, Ashish Jaiswal, Glenn R Wylie, and Fillia Makedon. 2024. Examining the landscape of cognitive fatigue detection: a comprehensive survey. *Technologies*, 12, 3, 38.
- [84] 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).
- [85] Herbert A. Simon. 1996. *The Sciences of the Artificial*. (3rd ed ed.). MIT Press, Cambridge, Mass.
- [86] 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).
- [87] 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).
- [88] Dana Chisnell. 2016. Democracy Is a Design Problem. 11, 4.
- [89] 2015. Designing usable ballots Center for civic design. <https://civicdesign.org/fieldguides/designing-usable-ballots/>. (June 2015). Retrieved June 17, 2024 from.
- [90] 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/10.2307/3117714).
- [91] Whitney Quesenberry. 2020. Opinion | Good Design Is the Secret to Better Democracy. *The New York Times*, (Oct. 2020). Retrieved June 17, 2024 from.
- [92] 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).
- [93] 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).
- [94] 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)*.
- [95] 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).
- [96] Juan E. Gilbert, Jerome 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).

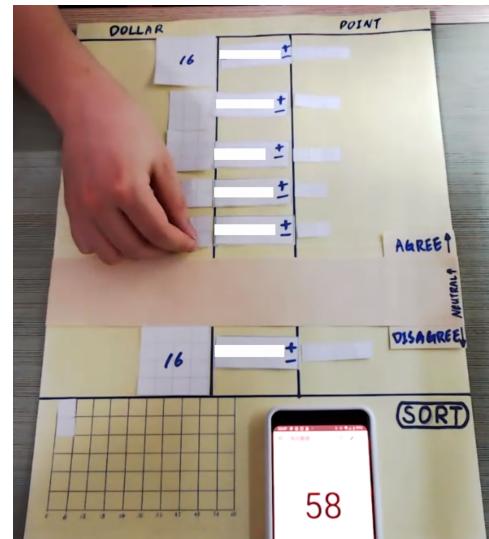
- [97] Frederick G. Conrad, Benjamin B. Bederson, Brian Lewis, Emilia Peytcheva, Michael W. Traugott, Michael J. Hanmer, Paul S. Herrnson, and Richard G. Niemi. 2009. Electronic voting eliminates hanging chads but introduces new usability challenges. *International Journal of Human-Computer Studies*, 67, 1, (Jan. 2009), 111–124. doi: [10.1016/j.ijhcs.2008.09.010](https://doi.org/10.1016/j.ijhcs.2008.09.010).
- [98] yehjxraymond. 2024. Yehjxraymond/qv-app. (Mar. 2024). Retrieved June 17, 2024 from.
- [99] Anuj K. Shah, Eldar Shafir, and Sendhil Mullainathan. 2015. Scarcity frames value. *Psychological Science*, 26, 4, 402–412.
- [100] Eibe Frank and Mark Hall. 2001. A simple approach to ordinal classification. In *Machine Learning: ECML 2001: 12th European Conference on Machine Learning Freiburg, Germany, September 5–7, 2001 Proceedings 12*. Springer, 145–156.

A Interface design process

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



(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. 18. Initial paper prototypes designed for QS interface

A.1 Prototype 1: Ranking-Vote

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

1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664

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.

<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Parks and Recreation (Arts, Culture, and Leisure Services; Wildlife Conservation; Zoos and Aquariums)	Your ratings cost \$9 You rated this option +3
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Human Services (Children and Family Services; Health Services; Nonprofit and Other Services; Food Banks; Food Pantries, and Food Distributors; Multipurpose Human Service Distributors; Homeless Services; Social Services)	Your ratings cost \$16 You rated this option +4
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Arts, Culture, and Leisure Services (Arts, Cultural Activities, and Landmark Preservation; Museums; Performing Arts; Public Broadcasting and Media)	Your ratings cost \$4 You rated this option -2
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Education (Early Childhood Programs and Services; Vocational Education Programs and Services; Adult Education Programs and Services; Special Education; Education Policy and Reform; Scholarship and Financial Support)	Your ratings cost \$34 You rated this option +6
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Environment (Environmental Protection and Conservation; Botanical Gardens, Parks and Nature Centers)	Your ratings cost \$4 You rated this option -2
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Healthcare (Diseases, Disorders, and Disabilities; Patient and Family Support; Treatment and Prevention)	Your ratings cost \$4 You rated this option -2

Summary

You have spent \$73 and you have \$251 remaining

Submit

Fig. 19. 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.

1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664

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

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.

Step 1: What is important to you?

In this step, please elect the options that you care about to the left of the column.

All Options	Options You Care About
American	Ramen
Japanese	Chinese
Mexican	

Step 2: Quadratic Voting

BACK TO STEP 1

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 spend 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 can always spend all the budget. You can always spend the budget you have and the amount of money you have spent already in the "Summary" section below. The interface will provide necessary calculations for the remaining budget you have, the accumulated ratings the current options have received and the dollar spent for each option. The interface also provides a drag and drop feature to help you complete the survey.

<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Chinese Orange chicken and rice	Your ratings cost \$4 You rated this option +2
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Italian Pasta and bread	Your ratings cost \$9 You rated this option -3
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	American Burgers, fries and ribs	Your ratings cost \$0 You rated this option 0
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Japanese Sushi and sashimi	Your ratings cost \$0 You rated this option 0
<input type="button" value="+1 rating"/>	<input type="button" value="-1 rating"/>	Mexican Tacos and burritos	Your ratings cost \$0 You rated this option 0

Summary

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

Submit

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

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

Fig. 20. 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. 20a, the first step involves selecting options for further consideration. Important options are placed at the top of the list for voting shown in Fig. 20b, but options can be placed anywhere on the list if desired. The rest of the controls remain the same as the previous prototype.

A.2 Prototype 2: Select-then-Vote

Based on feedback from Prototype 1, instead of *allowing* individuals to rank options, Prototype 2 implemented a two-phase process that *intentionally* asks respondents to select options to express opinions before voting. As shown in Figure 20, survey respondents selected their preferred options (Figure 20a), and the interface positioned these options at

Manuscript submitted to ACM

the top of the list for voting (Figure 20b). We identified several issues during the prototype 2 pretest: many respondents marked most options as ‘options they care about,’ which undermined the design’s purpose. Additionally, the lack of clear distinction between selected and unselected options confused respondents about the necessity of Step 1. Thus, we need a clearer distinction and connection between the two phases to effectively construct relative preferences.

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)
- Arts, Culture, Humanities
(Books, Media, Literature, and Landmark Preservation; Museums; Performing Arts; Public Broadcasting and Media)
- Health
(Diseases, Disorders, and Disciplines; Patient and Family Support; Treatment and Prevention Services; Medical Research)
- Veterans
(Veteran Support Services; Military Social Services; Military Family Support)
- Positive**
- Education
(Early Childhood Programs and Services; Youth Education; Higher Education; Adult Education Programs and Services; Special Education; Education Policy and Reform; Scholarship and Financial Support)
- Negative**
- Environment
(Environmental Protection and Conservation; Botanical Gardens, Parks and Nature Centers)
- Neutral**
- International
(Development and Relief Services; International Peace, Security, and Affairs; Humanitarian Relief Supplies)
- Human Services
(Child and Family Services; Youth Development; Shelter, and Crisis Services; Food Banks; Food Pantries, and Food Distribution; Multipurpose Human Service Organizations; Homeless Services; Social Services)

Next

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

Fig. 21. Organize-then-Vote Prototype: The goal of this prototype is to encourage participants to derive finer grain categories among options before voting. Survey respondents first organize their thoughts into categories and then vote on the options.

A.3 Prototype 3: Organize-then-Vote

Figure 21 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

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.

Don't know	Positive	Negative	Neutral
Pets and Animals (Animal Rights, Welfare, and Services; Wildlife Conservation; Zoos and Aquariums)	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 (Books, Media, Literature, and Landmark Preservation; Museums; Performing Arts; Public Broadcasting and Media)	Arts, Culture, Humanities (Books, Media, Literature, and Landmark Preservation; Museums; Performing Arts; Public Broadcasting and Media)		Your ratings cost \$4 You rated this option #2
Health (Diseases, Disorders, and Disciplines; Patient and Family Support; Treatment and Prevention Services; Medical Research)	Health (Diseases, Disorders, and Disciplines; Patient and Family Support; Treatment and Prevention Services; Medical Research)		Your ratings cost \$4 You rated this option #2
Veterans (Veteran Support Services; Military Social Services; Military Family Support)	Veterans (Veteran Support Services; Military Social Services; Military Family Support)		Your ratings cost \$4 You rated this option #2
Positive	Educational (Early Childhood Programs and Services; Youth Education; Higher Education; 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 (Child and Family Services; Youth Development; Shelter, and Crisis Services; Food Banks; Food Pantries, and Food Distribution; Multipurpose Human Service Organizations; Homeless Services; Social Services)	Human Services (Child 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

Back

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.

¹⁷¹⁷ the left and right sides of the screen. These insights guided refinements in the final two-phase interface, adhering to the
¹⁷¹⁸ two-phase organize-then-vote design framework.
¹⁷¹⁹

¹⁷²⁰ **B Voting Interface Breakdown**

¹⁷²¹ In this section, we outline additional literature that informed this study. There are two sets of literature that we surveyed:
¹⁷²² Survey response format and voting interfaces.
¹⁷²³

¹⁷²⁴ **B.1 Survey response format**

¹⁷²⁵ Research in the marketing and research communities focusing on survey and questionnaire design, usability, and
¹⁷²⁶ interactions examines the influence of presentation styles and ‘response format.’ Weijters et al. [86] demonstrated that
¹⁷²⁷ horizontal distances between options are more influential than vertical distances, with the latter recommended for
¹⁷²⁸ reduced bias. Slider bars, which operate on a drag-and-drop principle, show lower mean scores and higher nonresponse
¹⁷²⁹ rates compared to buttons, indicating they are more prone to bias and difficult to use. In contrast, visual analog scales
¹⁷³⁰ that operate on a point-and-click principle perform better [87]. These studies show how even small design changes can
¹⁷³¹ have a large impact on usability, highlighting the importance of designing interfaces that prioritize human-centered
¹⁷³² interaction rather than focusing solely on functionality.
¹⁷³³

¹⁷³⁴ **B.2 Voting Interfaces**

¹⁷³⁵ Compared to digital survey interfaces, voting interfaces are a specialized type of survey interface can significantly
¹⁷³⁶ influence democratic processes [13, 88, 89] and often have consequential impacts. Researchers believe that ill-designed
¹⁷³⁷ voting interfaces We categorize these related works into three main categories detailed below:
¹⁷³⁸

¹⁷³⁹ *Designs that shifted voter decisions:* For example, states without straight-party ticket voting (where voters can select
¹⁷⁴⁰ all candidates from one party through a single choice) exhibited higher rates of split-ticket voting [13]. Another example
¹⁷⁴¹ from the Australian ballot showing incumbency advantages is where candidates are listed by the office they are running
¹⁷⁴² for, with no party labels or boxes.
¹⁷⁴³

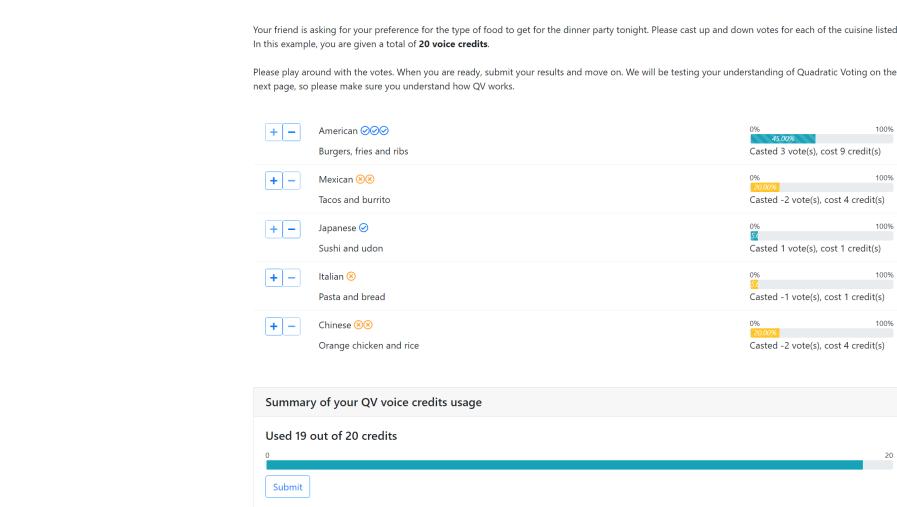
¹⁷⁴⁴ *Designs that influenced errors:* Butterfly ballots, an atypical design, may have influenced the outcome of the 2000 U.S.
¹⁷⁴⁵ Presidential Election [90]. It increased voter errors because voters could not correctly identify the punch hole on the
¹⁷⁴⁶ ballot. Splitting contestants across columns increases the chance for voters to overvote [91]. On the other hand, Everett
¹⁷⁴⁷ et al. [92] showed the use of incorporating physical voting behaviors, like lever voting, into graphical user interfaces.
¹⁷⁴⁸

¹⁷⁴⁹ *Designs that incorporated technologies:* Other projects like the Caltech-MIT Voting Technology Project have sparked
¹⁷⁵⁰ research to address accessibility challenges, resulting in innovations like EZ Ballot [93], Anywhere Ballot [94], and
¹⁷⁵¹ Prime III [95]. In addition, Gilbert et al. [96] investigated optimal touchpoints on voting interfaces, and Conrad et al.
¹⁷⁵² [97] examined zoomable voting interfaces.
¹⁷⁵³

¹⁷⁵⁴ Response format literature and voting interfaces informed how interfaces significantly influence respondent behavior,
¹⁷⁵⁵ decision accuracy, and cognitive load. These burdens are especially problematic for complex systems like QS, where
¹⁷⁵⁶ high cognitive demands may deter researchers and users alike. Developing effective, human-centered interfaces for QS
¹⁷⁵⁷ could enhance usability, reduce cognitive overload, and increase adoption in both research and practical applications.
¹⁷⁵⁸

1769 C Other Existing QV Interfaces

1770 In this section, we list out other QV interfaces we surveyed.



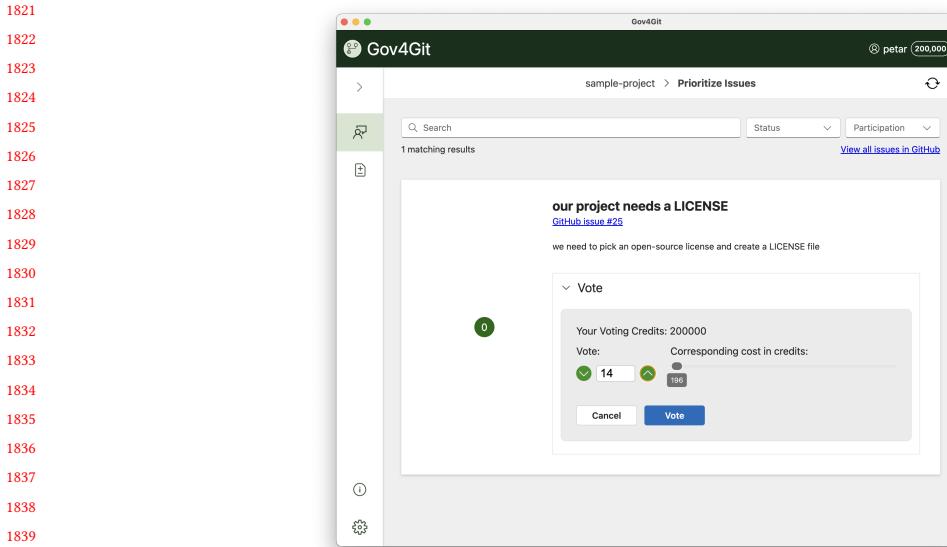


Fig. 24. The interface designed for gov4git [33].

D Demographic Breakdown

We provide the table for a more detail demographic breakdown per group.

Table 1. Participant Age and Gender Distribution by Experimental Condition

Condition	Mean Age	SD	Range	25th	Median	75th	Male	Female	Non-binary
Short Text	31.6	13.7	18–67	23.8	29.5	32.8	4	6	0
Short 2 Phase	32.1	14.0	18–52	20.3	27.0	44.5	4	6	0
Long Text	36.0	14.8	21–61	24.0	33.0	42.8	2	7	1
Long 2 Phase	38.8	19.6	19–71	25.0	28.5	53.0	2	8	0

E List of Options

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

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

- 1873 • **Performing Arts:** Support symphonies, orchestras, and other musical groups; ballets and operas; theater
1874 groups; arts festivals; and performance halls and cultural centers.
- 1875 • **Public Broadcasting and Media:** Support public television and radio stations and networks, as well as
1876 providing other independent media and communications services to the public.
- 1877 • **Community Foundations:** Promote giving by managing long-term donor-advised charitable funds for individual
1878 givers and distributing those funds to community-based charities over time.
- 1879 • **Housing and Neighborhood Development:** Lead and finance development projects that invest in and
1880 improve communities by providing utility assistance, small business support programs, and other revitalization
1881 projects.
- 1882 • **Jewish Federations:** Focus on a specific geographic region and primarily support Jewish-oriented programs,
1883 organizations and activities through grantmaking efforts
- 1884 • **United Ways:** Identify and resolve community issues through partnerships with schools, government agencies,
1885 businesses, and others, with a focus on education, income and health.
- 1886 • **Adult Education Programs and Services:** Provide opportunities for adults to expand their knowledge in a
1887 particular field or discipline, learn English as a second language, or complete their high school education.
- 1888 • **Early Childhood Programs and Services:** Provide foundation-level learning and literacy for children prior
1889 to entering the formal school setting.
- 1890 • **Education Policy and Reform:** Promote and provide research, policy, and reform of the management of
1891 educational institutions, educational systems, and education policy.
- 1892 • **Scholarship and Financial Support:** Support and enable students to obtain the financial assistance they
1893 require to meet their educational and living expenses while in school.
- 1894 • **Special Education:** Provide services, including placement, programming, instruction, and support for gifted
1895 children and youth or those with disabilities requiring modified curricula, teaching methods, or materials.
- 1896 • **Youth Education Programs and Services:** Provide programming, classroom instruction, and support for
1897 school-aged students in various disciplines such as art education, STEM, outward bound learning experiences,
1898 and other programs that enhance formal education.
- 1899 • **Botanical Gardens, Parks, and Nature Centers:** Promote preservation and appreciation of the environment,
1900 as well as leading anti-litter, tree planting and other environmental beautification campaigns.
- 1901 • **Environmental Protection and Conservation:** Develop strategies to combat pollution, promote conservation
1902 and sustainable management of land, water, and energy resources, protect land, and improve the efficiency of
1903 energy and waste material usage.
- 1904 • **Diseases, Disorders, and Disciplines:** Seek cures for diseases and disorders or promote specific medical
1905 disciplines by providing direct services, advocating for public support and understanding, and supporting
1906 targeted medical research.
- 1907 • **Medical Research:** Devote and invest in efforts on researching causes and cures of disease and developing
1908 new treatments.
- 1909 • **Patient and Family Support:** Support programs and services for family members and patients that are
1910 diagnosed with a serious illness, including wish granting programs, camping programs, housing or travel
1911 assistance.
- 1912 • **Treatment and Prevention Services:** Provide direct medical services and educate the public on ways to
1913 prevent diseases and reduce health risks.

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

F Detailed Qualitative Cognitive Load Breakdown

In addition to the discussion on cognitive load sources presented in the main text, we provide additional details on the six cognitive dimensions. Among all dimensions, we also provide the codes representing different types of demand in a table form. The shaded cells represent the percentage of participants citing each source of mental demand, allowing for comparison within columns. The abbreviations in the columns: ST (Short Text Interface), S2P (Short Two-phase Interface), LT (Long Text Interface), and L2P (Long Two-phase Interface). Short and Long refer to the sum across both interfaces; Text and Inter refer to the sum across both survey lengths. We include Sparklines for comparisons across these experiment groups. Future studies can use these as initial codebooks to conduct interface studies on preference construction.

G Sources of Mental Demand

Mental demand refers to the amount of mental and perceptual activity required to complete a task. Table G lists all the mental demand codes. Figure 25 shows the boxplot of participant's subscale response.

G.1 Sources of Physical Demand

Physical demand refers to the physical effort required to complete a task, such as physical exertion or movement. Most participants reported minimal physical demand ($N = 32$), reflected in the low NASA-TLX physical demand scores (Figure 26). 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 26), which show a significant visual difference between short and long two-phase interfaces as well as between text and two-phase interfaces in long surveys. Table 3 presents all the relevant codes across experiment groups.

Manuscript submitted to ACM

Table 2. This table lists all the causes participants mentioned as contributing to their Mental Demand.

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

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

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

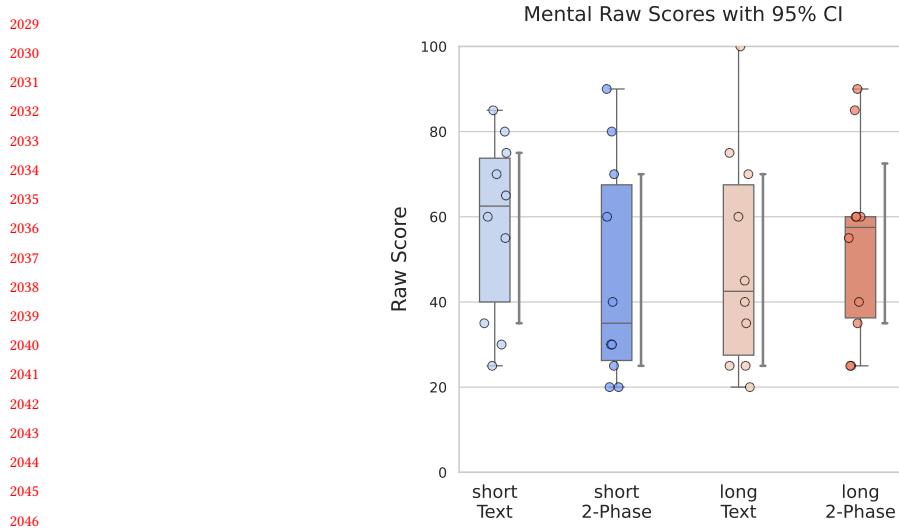


Fig. 25. Mental Demand Raw Score: Across all four experiment groups, participants' reported mental demand is spread across a wide range with many participants experiencing high mental demand.

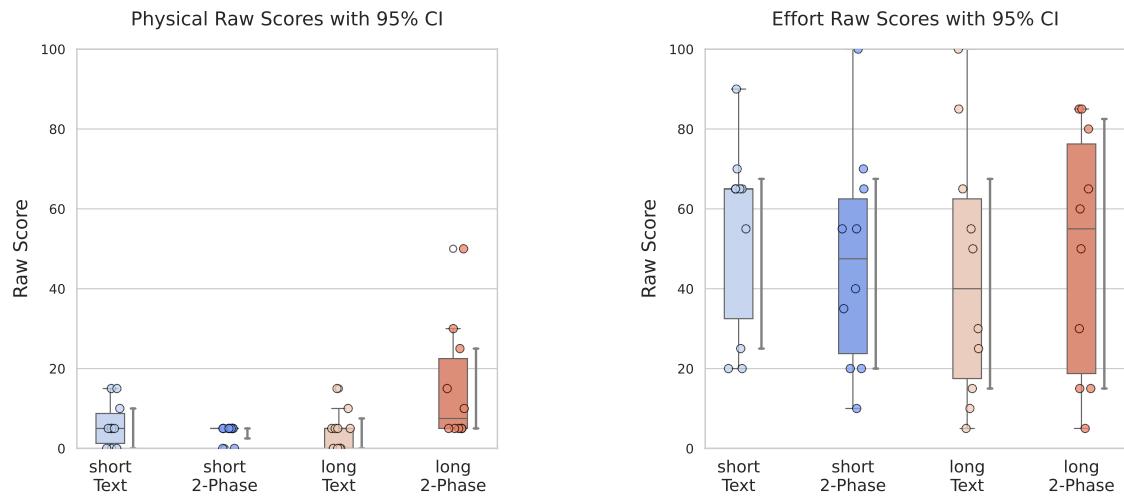


Fig. 26. Physical Demand Raw Score: Participants other than the long two-phase interface reported minimal physical demand. The long two-phase interface had the highest physical demand, likely due to increased mouse clicks and extended time spent looking at the vertical screen.

Fig. 27. Effort Raw Score: Effort scores shows indifference across groups.

G.2 Source of Effort

Effort refers to how hard participants felt they worked to achieve the level of performance they did. Since effort includes both mental and physical resource intensity, refer to ?? and Appendix G.1 for definitions. Raw NASA-TLX Manuscript submitted to ACM

Table 4. Effort Sources: Participants using the text interface focused more on operational tasks, while those using the two-phase interface focused more on strategic planning.

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

effort scores (Figure 27) showed a similar spread across experiment groups, the qualitative analysis showed more distinction that participants using the two-phase interface considered options more comprehensively and felt less effort on completing operational tasks, similar to what we found on mental demands (Section ??). Table 4 contains codes.

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

I wanted to bump up (an option) maybe to 4 or <option> to 5 and realize I couldn't. [...] that would be effort came in of how do I want to really rearrange this to make it (the budget spending) maximize?

– S029, short text interface

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

– S005, long text interface

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

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

G.3 Source from Performance

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

G.3.1 Operational Actions. Operational actions, like the theme presented in temporal demand, refer to specific, executable procedures participants perform in the survey. This could involve: pressure to spend all credits or stay

Table 5. Performance Causes: Most causes are shared across experiment conditions. We provided qualitative interpretations of their own performance assessments.

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

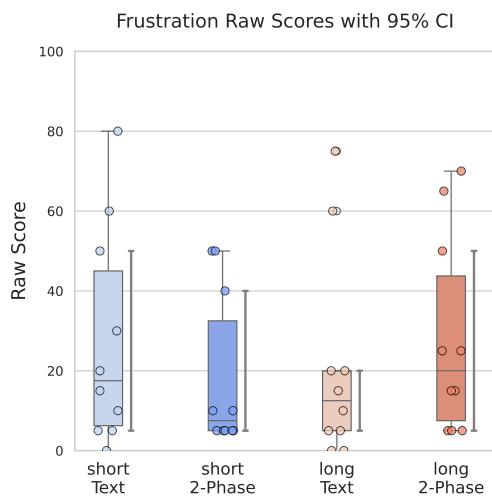


Fig. 28. Frustration Raw Score: Participants other than the long text interface highlighted several operational tasks that led to frustration. All groups share causes from strategic planning.

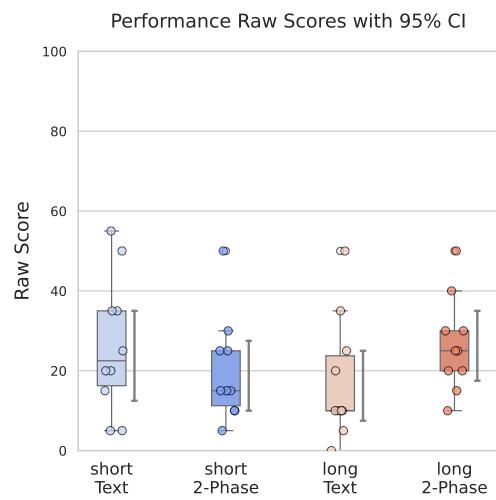


Fig. 29. Performance Demand Raw Score: Participants showed indifferent performance raw scores across experiment conditions, all trending toward satisfactory.

within budget ($N = 6$), fears that final vote choices did not reflect true preferences ($N = 5$), or concerns that they had finished the task inefficiently ($N = 6$).

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

All groups cited social responsibility as source to evaluate effort. Raw NASA-TLX scores (Figure 29) show participants had indistinguishable performance scores. This aligns with the interview results where most participants felt positive about their final submission.

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

G.4 Temporal Demand

Table G.4 lists all the mental demand codes.

Table 6. Temporal Demand Sources: Decision-making and Operational Tasks are the main causes. Participants framed their decision-making sources differently.

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

G.5 Frustration

Table G.5 lists all the mental demand codes.

H Additional voting behavior data

In this section, we describe the additional voting behavior that we observed. The reason why we decided to focus on the percentage of remaining credits comes from prior literature 'scarcity frames value' [99], a driver that makes researchers believe makes quadratic voting more accurate [4]. We did not follow Quarfoot et al. [6] in counting accumulated votes over time due to varying total times across individuals.

We observed the number of vote adjustments given a remaining vote credit percentage. Figure 30 showed all the voting actions over the remaining credit for the four experiment conditions. Here we see two distinct patterns between the short survey and the long survey in terms of participant behaviors. In long surveys, participants exhibited more

Table 7. Frustration Sources: Frustration comes from different levels of strategic operations or operational tasks.

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

actions both when the budget was abundant and when it began to run out. This pattern was more pronounced with the long two-phase interface. This difference is why we further focused on the long QS group.

Figure 31 presents the comparison between when participants make small or large vote adjustments at different budget levels. Revisiting the KDE curve in the second row in Figure 30 and the curve of the second row in Figure 31 show a stronger bimodal distribution for small vote adjustments across interfaces. In fact, the bimodal distribution is more pronounced in the two-phase interface. This suggests that participants make small adjustments both at the beginning and toward the end of the QS. However, the two-phase interface shows more frequent and faster edits towards the end. In comparison, participants also made more large vote adjustments early on that spread more equally compared to the text interface. This indicates that participants had a clearer idea of how to distribute their credits across the options.

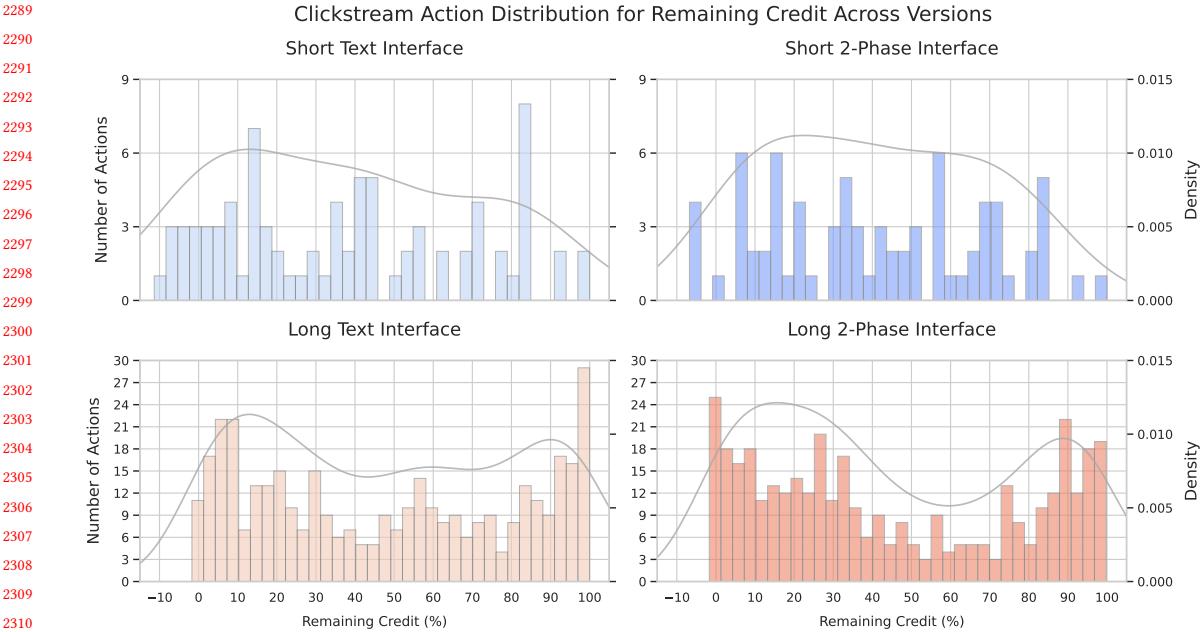


Fig. 30. This plot counts the number of voting actions when there are x percentages of credits remaining. A KDE plot is provided to help better understand the action distribution.

I Modeling NASA-TLX Weighted Scores and Subscales

In this section, we first describe the modeling approach for the NASA-TLX weighted scores and subscales, and then present all subscale results.

I.1 Modeling Approach

We modeled the NASA-TLX weighted scores and subscales using a hierarchical Bayesian ordinal regression model.

I.1.1 Dependent variables.

NASA-TLX weighted scores. are transformed from a continuous 0–100 scale to cognitive levels: low, medium, somewhat high, high, and very high, as described by Hart and Staveland [71]. This transformation helps the model adapt to sparse data. In our study, there were no participants who expressed "low" or "very high"; thus, we modeled the predictive variables as "medium," "somewhat high," and "high."

NASA-TLX subscale ratings. are transformed into ordinal groups using minimum frequency binning [100]. Minimum frequency binning involves grouping adjacent response categories until each bin meets a predefined minimum number of observations. The subscale uses a 21-point Likert scale, with 40 participants, it makes the ordinal data very sparse. Minimum frequency binning mitigates this allowing similar number of participants in each bin. We applied weighted bins across all participants within the same subscale, ensuring that each bin contained at least 10 participants.

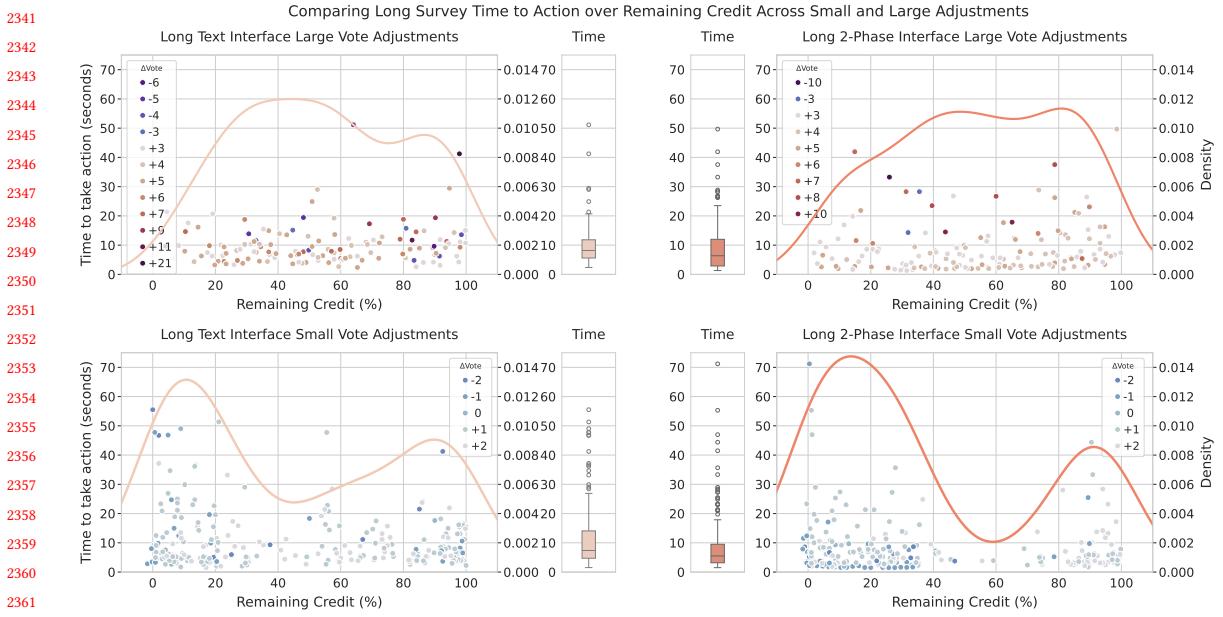


Fig. 31. This plot further separates participants' interaction behavior based on the number of votes participants adjusted. We observed a bimodal interaction pattern across long QS when small vote adjustments are made.

I.1.2 Independent Variables. For this model, we used three independent variables: length (γ_i), interface type (β_I), and the interaction between the two (ϕ_{ij}). Length, categorized as "low" and "short," was modeled as an ordinal variable, as shown in Equation 4. Since there are only two categories, this approach allowed us to model the baseline length effect and the added effect of the longer length. Interface types were set up with hyperpriors, from which the interfaces were drawn. The interaction effect used a non-centered parameterization constrained by an LKJ prior to account for correlations, as described in Equation 5. Weakly informed priors were used for all parameters, as shown in Equations 6, 7, and 8.

I.1.3 Overall Model. We modeled the dependent variables using an Ordered Logistic (Equation 1). The Ordered Logistic model is particularly suited for ordinal outcome variables, where the categories have a natural order but the intervals between them are not necessarily equal. This model has two input parameters: η_i and τ . η_i is the latent predictor derived from a regression equation that incorporates the independent variables, demonstrated as Equation 2. The purpose of it, intuitively, is to model how specific independent variables pushes this latent value towards a higher or lower category. τ as mododed by Equation 3 are the cutpoints that demarcate the boundaries between the ordinal categories. This cutpoint draws from a normal distribution and being transformed to ensure that the thresholds are ordered. The Ordered Logistic model then compares η_i to τ to determine the probability of the observed outcome y_i falling into a specific ordinal category.

2393
 2394
 2395 $y_i \sim \text{OrderedLogistic}(\eta_i, \tau)$ (1)
 2396 $\eta_i = \alpha + \gamma_i + \beta_I[I_i] + \phi_{ij}$ (2)
 2397 $\tau \sim \text{OrderedTransform}(\mathcal{N}(0, 1)^{K-1})$ (3)
 2398 $\gamma_i = \mu_L + \beta_L \cdot L_i$ (4)
 2399 $\phi_{ij} = L_\Omega \cdot (\sigma_\phi \odot z_\phi)$ (5)
 2400
 2401
 2402

2403 *Priors.* We specify priors for all model parameters. The priors are defined as follows:
 2404

2405 $\mu_L, \mu_{\beta_L}, \mu_{\beta_I} \sim \mathcal{N}(0, 1), \quad \sigma_{\beta_L}, \sigma_{\beta_I} \sim \text{Exponential}(1)$ (6)
 2406 $\beta_L \sim \mathcal{N}(\mu_{\beta_L}, \sigma_{\beta_L}), \quad \beta_I \sim \mathcal{N}(\mu_{\beta_I}, \sigma_{\beta_I})$ (7)
 2407 $L_\Omega \sim \text{LKJ}(2), \quad \sigma_\phi \sim \text{Exponential}(1), \quad z_\phi \sim \mathcal{N}(0, 1)$ (8)
 2408
 2409

2410 In Equation 6 and 7 we present the hyperpriors reflecting our belief that the mean effects of length and interface
 2411 are centered around zero with a standard deviation of one. Hyperpriors were used to enable partial pooling where
 2412 information is shared across different levels of the interface type, improving estimation accuracy especially in cases
 2413 with limited data per group. Equation 8 describes the correlation matrix used for the interaction effect. The LKJ prior
 2414 of 2 refers to a moderate correlation without being too restrictive allowing the model to learn appropriate levels of
 2415 interaction terms. σ_ϕ ensuring that the variability of the interaction effects remains positive and allowing the model to
 2416 flexibly adapt to different levels of interaction strength and z_ϕ were assigned to serve as a standardized component
 2417 that, when scaled by σ_ϕ with the correlation matrix L_Ω captures the magnitude and the dependencies of the interaction
 2418 terms effectively.
 2419

2420
 2421 *I.1.4 Model Results.* We conducted the Bayesian analysis using NumPyro, a widely used framework for Bayesian
 2422 inference. We used Markov Chain Monte Carlo (MCMC) sampling, a method commonly applied in Bayesian inference.
 2423 All the models showed that the Gelman-Rubin statistic (\hat{R}) parameters were equal to 1 across two chains, indicating
 2424 that the multiple sampling chains converged. We present each subscale result and provide a short description of these
 2425 results.
 2426

2427
 2428 *I.1.5 Mental Subscale.* Figure 32 shows pairwise Bayesian results from mental demand highlighted 70.4% of posterior
 2429 probability that participants in the long two-phase condition had a higher mental demand compared to the short
 2430 two-phase condition. On the other hand, the short text condition had a 74.5% posterior probability of having a higher
 2431 mental demand compared to the short two-phase condition. This is additional evidence that prompted us to believe that
 2432 the participants in the short two-phase participants benefited from the organization phase. The sheer number of added
 2433 options in the long two-phase condition may have added additional demand to participants, leading to higher mental
 2434 demand.
 2435

2436
 2437 *I.1.6 Physical Subscale.* Figure 33 shows the pairwise comparison of the physical subscale. Noteable results shows that
 2438 there is a 86.1% posterior probability that the long text condition had a lesser physical demand compared to the short
 2439 text condition. This is counter intuitive as the long text participants actually traversed much higher edit distances. We
 2440 are not clear what prompted their self reported value and requires future research.
 2441

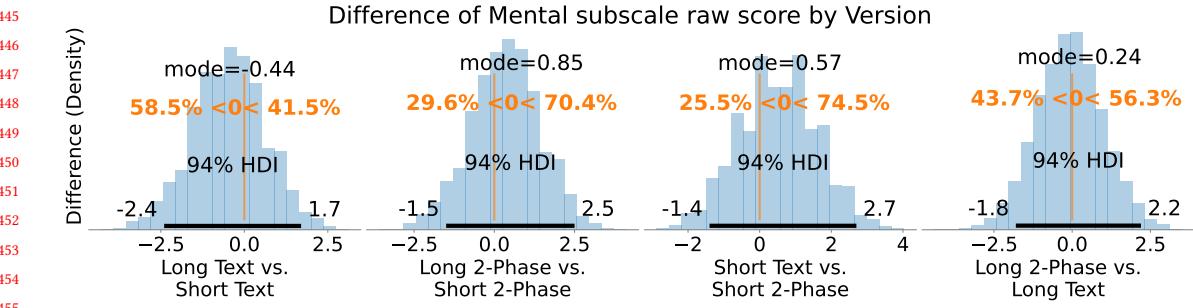


Fig. 32. Differences in the mental subscale scores by version.

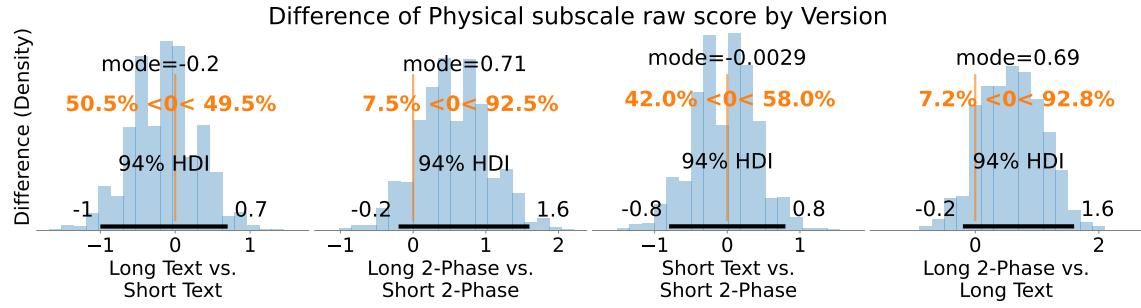


Fig. 33. Differences in the physical subscale scores by version.

I.1.7 *Temporal Subscale*. Figure 34 shows the pairwise comparison of the temporal subscale. The results show that the long two-phase condition once again had a 74.6% posterior probability of having a lower temporal demand compared to the short text condition. Conversely, participants in the long two-phase condition had a 71.1% posterior probability of having a higher temporal demand compared to the short two phase condition, reflecting the longer time they took to complete the survey questions. We believe that the lower temporal demand in the long two-phase condition are potential indicators of participant's satisficing behavior.

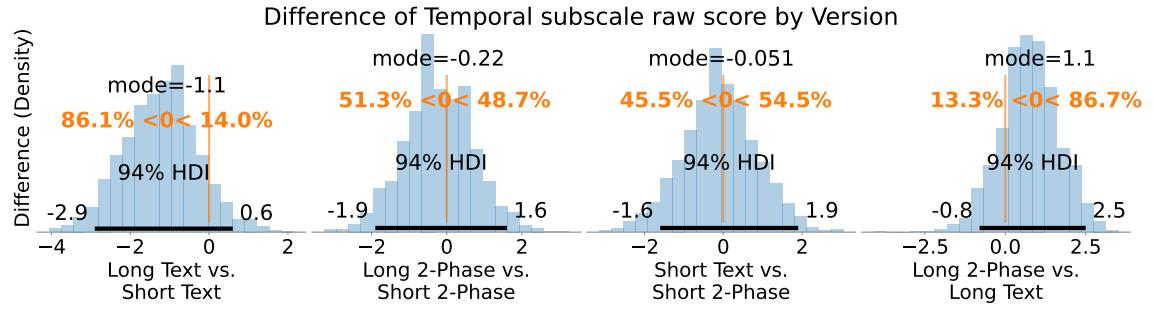


Fig. 34. Differences in the temporal subscale scores by version.

I.1.8 Performance Subscale. We omit the pairwise comparison of the performance subscale due to the mixed signals. We focused on the qualitative results analyzed in the main text.

I.1.9 Effort Subscale. We omit the pairwise comparison of the effort subscale due to its similarity to the mental demand subscale.

I.1.10 Frustration Subscale. Figure 35 shows the pairwise comparison of the frustration subscale. The results show that the long two-phase condition had a 68.3% posterior probability of having a higher frustration compared to the short two-phase condition, likely due to the added number of options to assess.

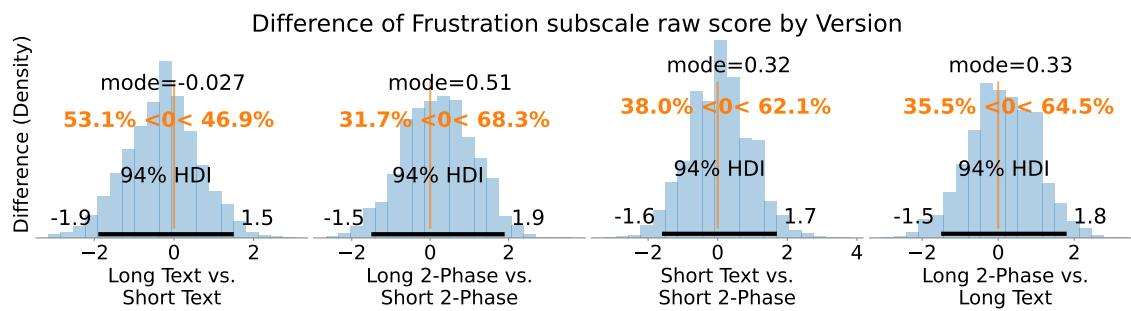


Fig. 35. Differences in the frustration subscale scores by version.

J Modeling Total Time

In this section, we discuss how we modeled the total time per option for each experimental condition.

J.0.1 Dependent Variables. Total time (T_i) refers to the time participants spent on each option, including the time allocated to the organization phase, where participants categorized or reordered options before voting.

J.0.2 Experimental Conditions. We categorize the data into four experimental conditions: Short Text, Short Two-Phase, Long Text, and Long Two-Phase. These conditions are indexed by k , and separate submodels are fit for each condition.

J.1 Modeling Approach

We modeled the total time for each experimental condition using separate Gamma likelihood models. The Gamma distribution is well-suited for modeling positive continuous data, such as time measurements, which are often skewed and strictly positive. Equation 9 shows the model for the total time. The shape parameter α_k and rate parameter β_k were each assigned priors drawn from their own Gamma distributions, as described in Equations 10 and 11.

$$T_i \sim \text{Gamma}(\alpha_k, \beta_k) \quad (9)$$

$$\alpha_k \sim \text{Gamma}(2.0, 0.5) \quad (10)$$

$$\beta_k \sim \text{Gamma}(1.0, 1.0) \quad (11)$$

K Modeling edit distance

In this section, we describe the details for the three models we used to analyze the edit distance data.

K.1 Model 1: Edit Distance per Option

K.1.1 Dependent variables. The dependent variable for this model is the edit total distance accumulated for an option D_i . Distance is a positive continuous variable.

K.1.2 Independent variables. The independent variables for this model are the length of the option L_i , modeled as a ordinal variable (Equation 15); interface type I_i , modeled as a categorical variable; user effect U_i as categorical variables. The ordinal variable L_i consists of a intercept μ_L and added effect β_L , given the interface ordinal value. Since we only have two interfaces, we do not have to worry about the interval between two or more interfaces. Priors are weakly informed in Equation 18. We reparamtereized U_i given the sparser sample from each participant. This is written in Equations 17. Both reparameterization contains an intercept and scaling of the effect due to this user. This will imporse sampling efficiency and help the model converge. Relavent priors are written in Equations 18 and 20. We added an interaction effect between length and interface type ϕ_{ij} described in Equation 16. Similiar to cognitive load model, the interaction effect used a non-centered parameterization constrained by an LKJ prior to account for correlations. Priors for the interaction effect is listed in Equations 19 and 21. Detailed description can be found in Appendix I.

K.1.3 Overall model and Likelihood function. We modeled the dependent variable using an Exponential distribution (Equation 12). Since Exponential distribution takes in a positive value, we transformed it as Equation 13. The observed outcome variable D_i represents the response for the i -th observation parameterized by the latent predictor η_i . η_i is described in Equation 14 as the regression with length, interface, the interaction effect and the interface.

$$D_i \sim \text{Exponential}(\lambda_i) \quad (12)$$

$$\lambda_i = \exp(\eta_i) \quad (13)$$

$$\eta_i = \gamma_i + \beta_I[I_i] + \phi_{ij} + U_i \quad (14)$$

$$\gamma_i = \mu_L + \beta_L \cdot L_i \quad (15)$$

$$\phi_{ij} = L_\Omega \cdot (\sigma_\phi \odot z_\phi) \quad (16)$$

$$U_i = \mu_U + \sigma_U \cdot z_U \quad (17)$$

Priors are defined as:

$$\mu_L, \mu_I, \mu_U, \beta_L, \beta_I, \phi_{ij}, z_U \sim \mathcal{N}(0, 1) \quad (18)$$

$$\sigma_\phi \sim \text{HalfNormal}(0.5) \quad (19)$$

$$\sigma_U \sim \text{Exponential}(0.5) \quad (20)$$

$$L_\Omega \sim \text{LKJ}(3) \quad (21)$$

K.2 Model 2: Edit Distance with Separate Mean and Variance Predictors

K.2.1 Dependent Variables. The dependent variable for this model is the edit distance (with directions) D_i , a positive edit distance refers to participants moving downward. A negaitve edit distance refers to a upward movement.

2601 *K.2.2 Overall Model.* We modeled the dependent variable D_i using a Normal distribution (Equation 22). Since the goal
2602 of this model, unlike some, aims to model the variance since we believe participants in two-phase interface would exhibit
2603 less oscillation than the text interface. Hence, we model independent variables effecting both μ and σ independently for
2604 this analysis to examine this hypothesis.
2605

2606
2607
2608
2609
2610
2611 *K.2.3 Independent Variables.* The independent variables for this model are:
2612
2613
2614
2615
2616
2617
2618

- 2619 • **Length of the option L_i :** Modeled as an ordinal variable. Since we will be modeling both μ_i and σ of a Normal
2620 distribution, Equation 24 and 29 reflects the ordinal variable. Both formula consists of a intercept $\mu_{L,\mu}$, $\mu_{L,\sigma}$ and
2621 added effect $\beta_{L,\mu}$, $\beta_{L,\sigma}$, given the interface ordinal value. Since we only have two interfaces, we do not have
2622 to worry about the interval between two or more interfaces. Priors of both ordinal relationship are weakly
2623 informed in Equation 33 and 34
2624
- 2625 • **Interface type I_i :** Modeled as a categorical variable. Following the previous discussion, they are drawn from a
2626 hyperprior. We reparameterized this independent variable given the added complexity of this model. This is
2627 written in Equations 25 and 30. Both reparameterization contains an intercept and scaling of the effect due to
2628 this interface. Relevant priors are written in Equations 33, 34, and 35.
2629
- 2630 • **User effect U_i :** Users are modeled as categorical variables. Following the interface, it is also reparameterized as
2631 Equations 27 and 32. Priors are defined in Equations 33, 34, and 37
2632
- 2633 • **Interaction between length and interface type ϕ_{ij} :** Similar to the interaction effect for cognitive load, we
2634 used a non-centered parameterization constrained by an LKJ prior to account for correlations. This is described
2635 by Equation 26 and 31. Refer to Appendix I for a more detailed explanation. Relevant priors are described in
2636 Equation 38 and 36. We relaxed the LKJ priors compared to the cognitive load model given the complexity of
2637 the model allowing a lesser belief in correlation among the two variables.
2638

2639
2640
2641
2642
2643
2644 *K.2.4 Likelihood Function.* Given these independent variables, we model both μ and σ as linear regressions. While
2645 we can directly model mu (Equation 23), we need to make sure $sigma$ is strictly positive, we applied a transformation
2646 described in 28. Hence, both μ_i and $\log(\sigma_{obs,i})$ now regresses on the linear combination of length, interface, interaction
2647 effect, and user effect.
2648

2653
 2654
 2655 $D_i \sim \text{Normal}(\mu_i, \sigma_{\text{obs},i})$ (22)
 2656
 2657 $\mu_i = \gamma_{\mu,i} + \beta_{I,\mu}[I_i] + \phi_{\mu,ij} + U_{\mu,i}$ (23)
 2658 $\gamma_{\mu,i} = \mu_{L,\mu} + \beta_{L,\mu} \cdot L_i$ (24)
 2659
 2660 $\beta_{I,\mu}[I_i] = \mu_{I,\mu} + \sigma_{I,\mu} \cdot I_{\mu,I_i}$ (25)
 2661 $\phi_{\mu,ij} = L_{\Omega,\mu} \cdot (\sigma_{\phi,\mu} \odot z_{\phi,\mu})$ (26)
 2662 $U_{\mu,i} = \mu_{U,\mu} + \sigma_{U,\mu} \cdot z_{U,\mu,i}$ (27)
 2663
 2664 $\log(\sigma_{\text{obs},i}) = \gamma_{\sigma,i} + \beta_{I,\sigma}[I_i] + \phi_{\sigma,ij} + U_{\sigma,i}$ (28)
 2665
 2666 $\gamma_{\sigma,i} = \mu_{L,\sigma} + \beta_{L,\sigma} \cdot L_i$ (29)
 2667
 2668 $\beta_{I,\sigma}[I_i] = \mu_{I,\sigma} + \sigma_{I,\sigma} \cdot I_{\sigma,I_i}$ (30)
 2669 $\phi_{\sigma,ij} = L_{\Omega,\sigma} \cdot (\sigma_{\phi,\sigma} \odot z_{\phi,\sigma})$ (31)
 2670 $U_{\sigma,i} = \mu_{U,\sigma} + \sigma_{U,\sigma} \cdot z_{U,\sigma,i}$ (32)

2672 *K.2.5 Priors.* Priors are defined as:

2673
 2674 $\mu_{L,\mu}, \mu_{I,\mu}, \mu_{U,\mu}, \beta_{L,\mu}, \beta_{I,\mu}, z_{\phi,\mu}, z_{U,\mu,i} \sim \mathcal{N}(0, 1)$ (33)
 2675
 2676 $\mu_{L,\sigma}, \mu_{I,\sigma}, \mu_{U,\sigma}, \beta_{L,\sigma}, \beta_{I,\sigma}, z_{\phi,\sigma}, z_{U,\sigma,i} \sim \mathcal{N}(0, 1)$ (34)
 2677 $\sigma_{I,\mu}, \sigma_{I,\sigma} \sim \text{HalfNormal}(0.5)$ (35)
 2678
 2679 $\sigma_{\phi,\mu}, \sigma_{\phi,\sigma} \sim \text{HalfNormal}(0.5)$ (36)
 2680
 2681 $\sigma_{U,\mu}, \sigma_{U,\sigma} \sim \text{Exponential}(0.5)$ (37)
 2682 $L_{\Omega,\mu}, L_{\Omega,\sigma} \sim \text{LKJ}(3)$ (38)

2683 *K.2.6 Model Results.* Here we provide all pairwise comparisons for the variance which the main text only provided the comparison within the same survey length. Figure 36 shows the pairwise comparison of the variance of edit distance in the first row followed by the effect size in the second row. An notable result that we omit from the main text is that if we compare the variance between the long and short text, and the variance between the long and short two-phase, we see that the text group had three times the standard deviation compared to the two-phase group. This indicates that the organization phase minimize the added length of the survey.

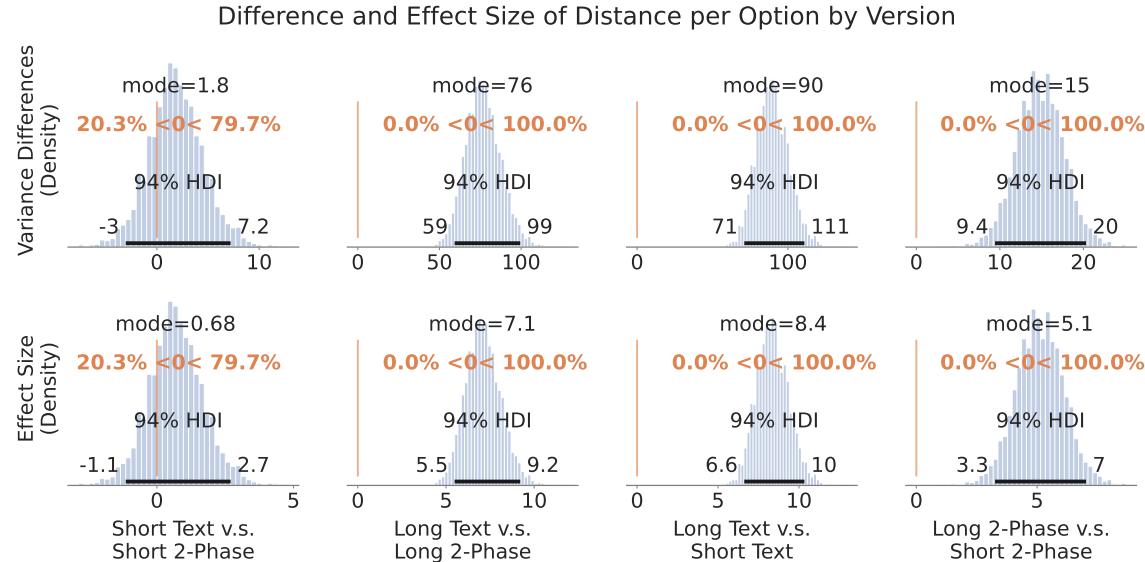


Fig. 36. Differences in the variance of edit distance by version.

K.3 Model 3: Cumulative Edit Distance for long QS

K.3.1 Dependent Variables. The dependent variable for this model is the cumulative edit distance D_i . Cumulative edit distance is a positive continuous variable measured at each step within a version for each user.

K.3.2 Independent Variables. The independent variables for this model involve the following. Steps refers to the n -th step when completing QS (S_i), and interface version refers to the type of interface used (V_i). User-specific effects are included as (U_i). Both interface versions and user-specific effects are modeled with their own hyperpriors to capture variability across these groups.

Equation 46, refers to interface versions, $\beta_v[V_i]$ are drawn from a Normal distribution with hyperparameters defined in Equations 47 and 48 corresponding to the mean and variance of this distribution.

Instead of directly sampling U_i from a hyper distribution, we reparameterize it to account for limited data for each user. This reparameterization is presented in Equation 41. μ_U models the overall mean user effect from users, with σ_U used to capture variability in user effects (Equation 44). A standard normal random variable, Equation 45 introduced individual randomness for each user.

K.3.3 Overall Model and Likelihood Function. We modeled the dependent variable D_i using a Truncated Normal distribution (Equation 39). The observation-specific standard deviation, drawn from a Half-Normal distribution as described in Equation 42. The latent predictors μ_i is modeled as a regression equation (Equation 40). This equation reflects our intuition that the effects from versions and user differences are amplified by steps as the participants complete the survey. The intercept α_{shared} is assigned a prior described in Equation 43. The effect of users σ_U and version $\beta_v[V_i]$ are amplified by the step number S_i .

2757
 2758
 2759 $D_i \sim \text{TruncatedNormal}(\mu_i, \sigma_{\text{obs},i}, \text{lower} = 0)$ (39)
 2760
 2761 $\mu_i = \alpha_{\text{shared}} + \beta_v[V_i] \cdot S_i + U_i \cdot S_i$ (40)
 2762 $U_i = \mu_U + \sigma_U \cdot z_{U,i}$ (41)

2763 Priors used in this model are listed.
 2764

2765 $\sigma_{\text{obs},i} \sim \text{HalfNormal}(0.3)$ (42)
 2766
 2767 $\alpha_{\text{shared}} \sim \mathcal{N}(2.0, 0.5)$ (43)
 2768
 2769 $\mu_U, \sigma_U \sim \mathcal{N}(0, 1), \text{ HalfNormal}(0.1)$ (44)
 2770
 2771 $z_{U,i} \sim \mathcal{N}(0, 1)$ (45)
 2772 $\beta_v[V_i] \sim \mathcal{N}(\mu_\beta, \sigma_\beta)$ (46)
 2773
 2774 $\mu_\beta \sim \mathcal{N}(0.05, 0.05)$ (47)
 2775
 2776 $\sigma_\beta \sim \text{HalfNormal}(0.1)$ (48)