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

ANONYMOUS AUTHOR(S)*

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

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

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

ACM Reference Format:

Anonymous Author(s). 2024. Organize, Then Vote: Exploring Cognitive Load in Quadratic Survey Interfaces. *Proc. ACM Hum.-Comput. Interact.* 1, 1 (December 2024), 30 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

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

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

Manuscript submitted to ACM

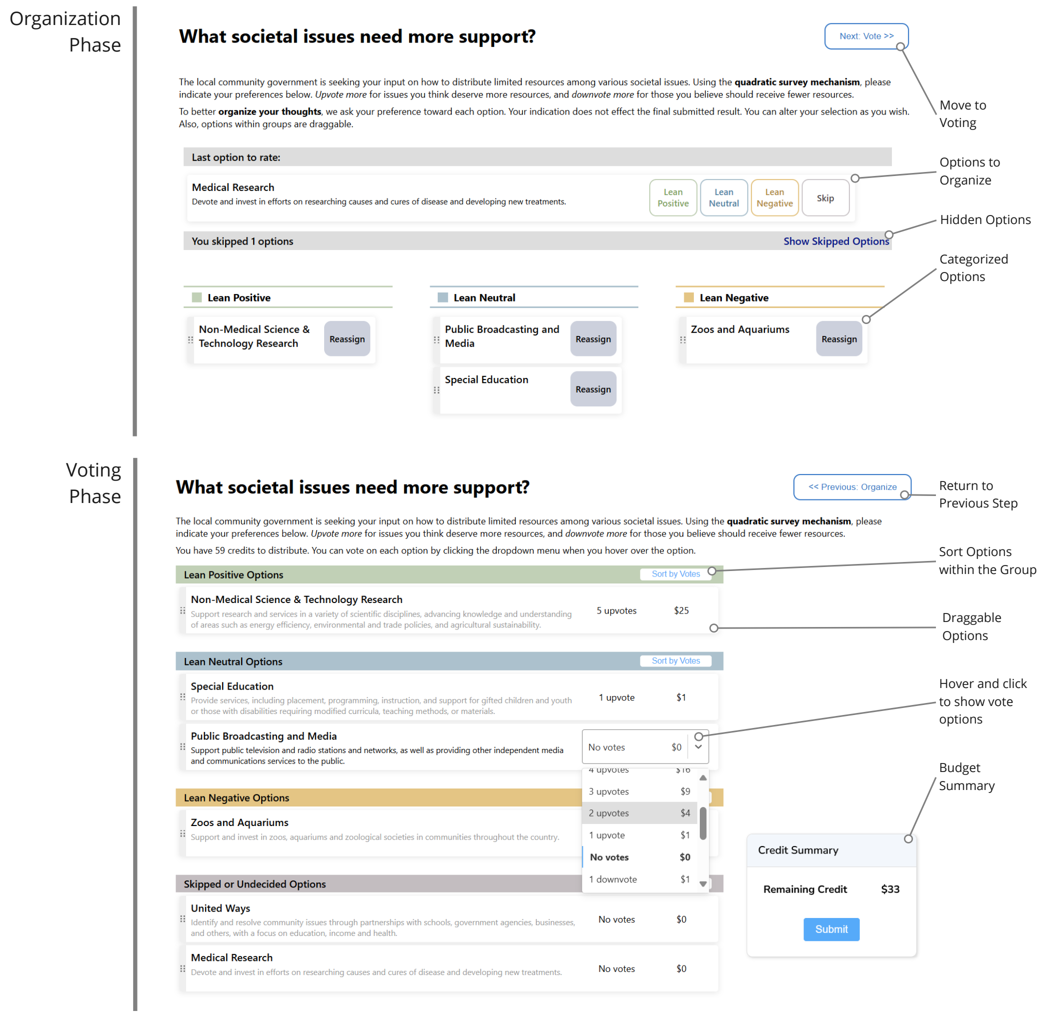


Fig. 1. The Two-Phase Interface: The interface consists of two phases. Survey respondents can navigate between phases using the top right button. In the organization phase, the interface presented one option at a time to the respondents, and they chose four choices: "Lean Positive", "Lean Neutral", "Lean 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.

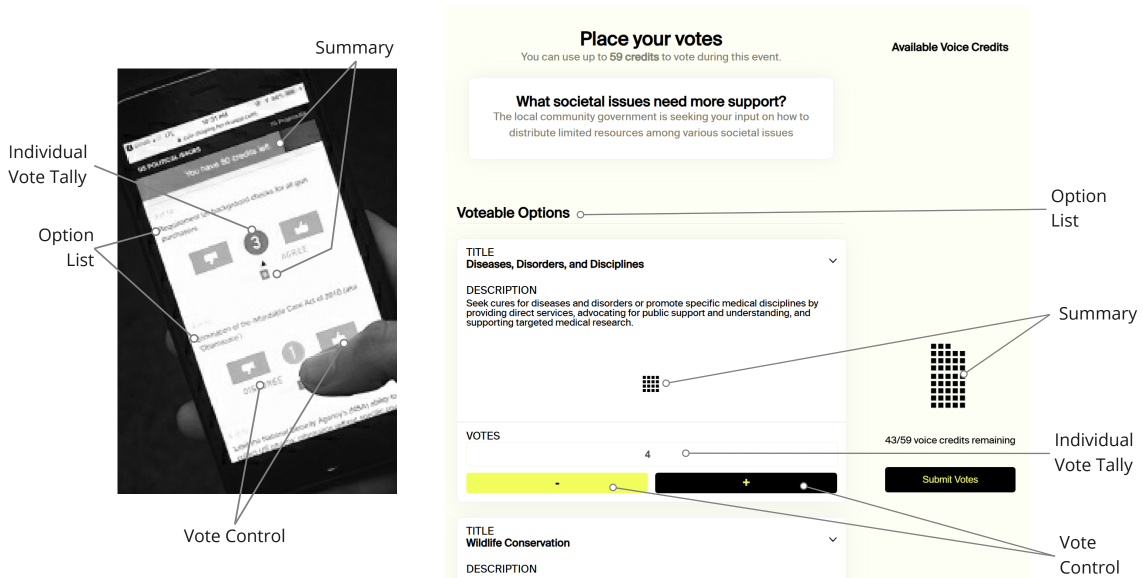


Fig. 2. A selection of two QV interfaces. The interface on the left was used in the first empirical QV research [6]. Little information is available about the software, except for an image from Posner and Weyl [27]. The interface on the right is an open-sourced QV interface [34] forked from GitCoin [35],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

2.2 Design Implications 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 interfaces [1], 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].

[1] Appendix X lists a comprehensive list we surveyed

2.3 Cognitive Challenges and Choice Overload

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

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

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

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

3 Quadratic Survey Interface Design

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

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

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

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

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

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

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

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

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

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

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

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

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

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

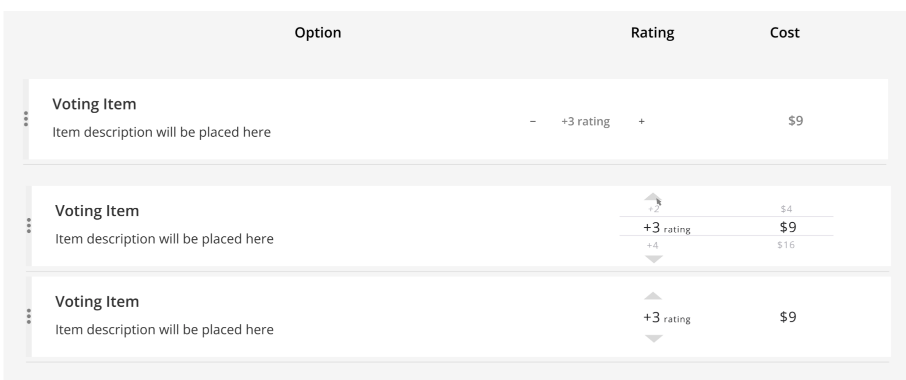


Fig. 3. Alternative vote control. The click-based design (upper) mirrors traditional vote control used in other QV interfaces, where each click controls one vote. The wheel-based design (the latter two) allows control through both clicks and mouse wheel rotation.

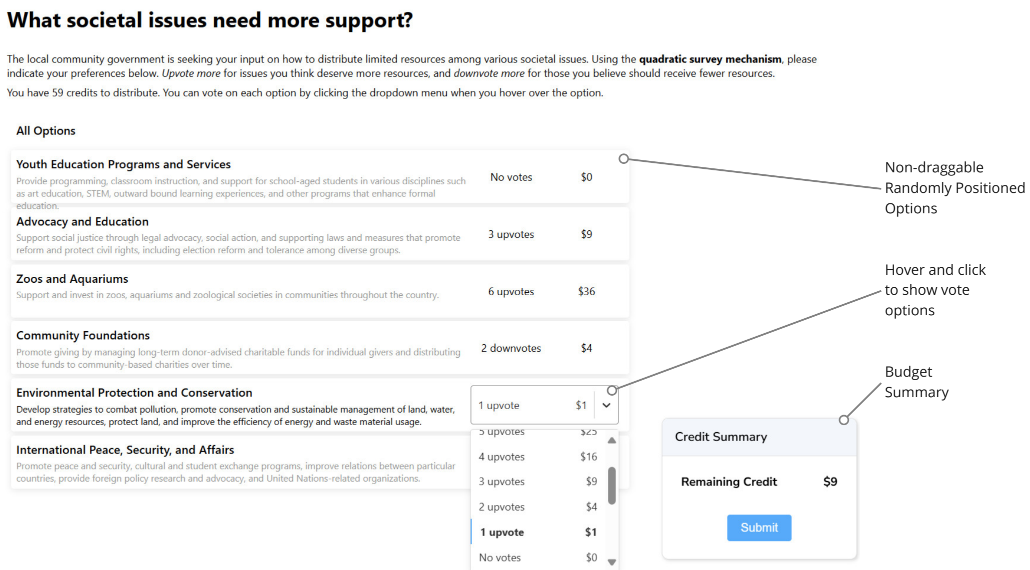


Fig. 4. The text-based interface: This interface is based on the interactive version but does not include the two-phase interactive support and lacks the drag-and-drop functionality. Options are randomly positioned.

3.2 Baseline Interface: Single-Phase Text Interface

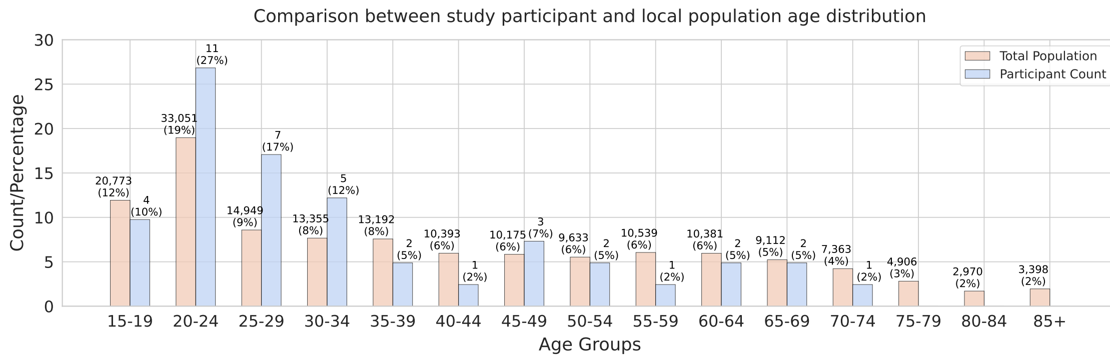
We implemented the single-phase text interface (referred to as text interface for short) as our control condition to compare how the interactive components influenced participants' cognitive load and behavior. The text-based interface, like all existing interfaces, contains a list of static elements, 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]

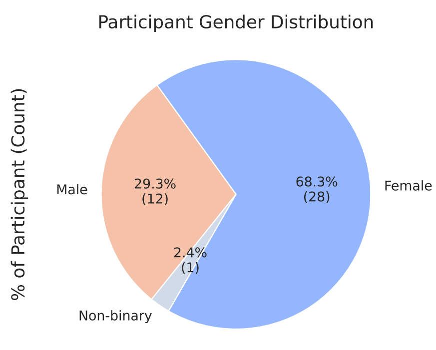
4 Experiment Design

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

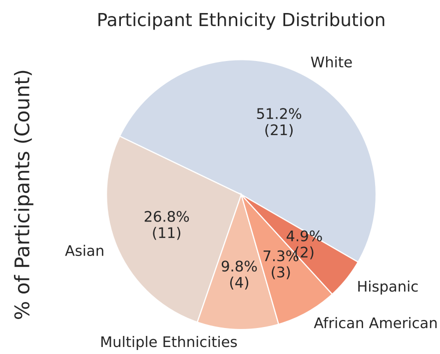
[2] link-to-github



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



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



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

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

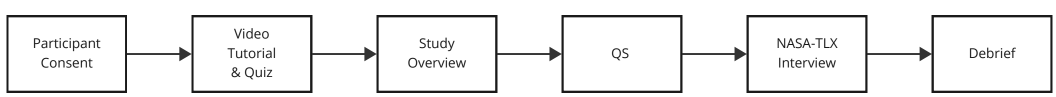


Fig. 6. Study protocol: Participants are asked to learn about the mechanism of QS after consenting to the study. The researcher explained the study overview and asked participants to complete the QS. A NASA-TLX survey followed by interviews to understand participants' cognitive load. We debriefed participants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

4.1 Recruitment and Participants

We recruited 41 participants from a United States college town using online ads, digital bulletins, social media posts, email newsletters, and physical flyers in public spaces beyond campus. We advertised the study as focusing on societal attitudes to mitigate potential response bias. One participant was excluded due to data quality concerns[3].

[3]The 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 **??**.

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 topics[4] 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

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.

[4]See Appendix **??** for the full list.

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

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

5 NASA-TLX Cognitive Load

In this section, we present the results of NASA-TLX cognitive load across experiment groups using descriptive statistics, a Bayesian model, and qualitative findings from post-survey interviews to address how the number of options in Quadratic Surveys (QS) (**RQ1**) and the interface design (**RQ2a**) impact cognitive load.

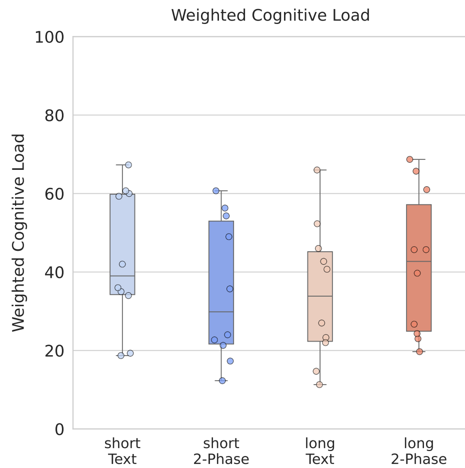
For the qualitative analysis, the first author conducts an inductive thematic analysis [78] after transcribing the interviews. Snippets are initially coded based on the research questions and topics of interest, with similar codes merged to form overarching themes. When differences emerges across experiment conditions and hypotheses are identified, a deductive coding process is applied to refine and validate the findings.

For quantitative analysis, we employ a Bayesian approach to enhance transparency and provide a probabilistic interpretation of results, moving beyond binary significance thresholds [79]. Bayesian methods allow for the interpretation of posterior distributions and are well-suited for both large and small sample sizes due to the use of priors based on maximum entropy distributions, which ensure conservative and robust inferences [80].

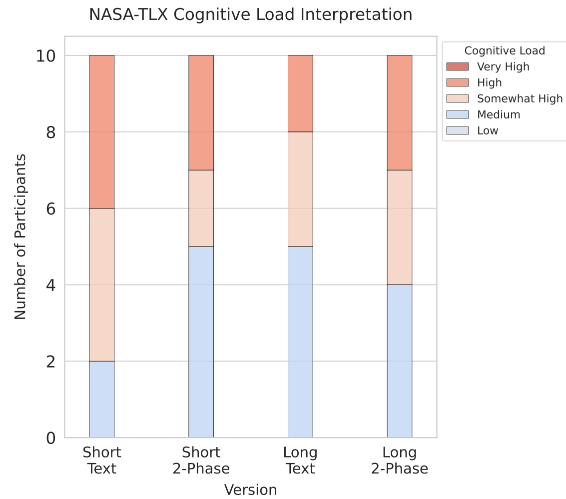
5.1 Overall Cognitive Load

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

Given the sparsity of the data, we modeled the weighted NASA-TLX scores using cognitive levels as ordinal outcomes. Then, we developed a hierarchical Bayesian ordinal regression model to analyze ordinal response data. The model includes length as an ordinal predictor, and interface type as a categorical predictor modeled with hierarchical priors to allow partial pooling across categories. Interaction effects between length and interface are captured using a non-centered parameterization constrained by an LKJ prior to account for correlations. An intercept term establishes the baseline, and ordered cutpoints define the boundaries between ordinal outcome categories. We use the same model for the NASA-TLX subscales. Given that subscales do not have cognitive level interpretations, we constructed weighted bins to facilitate the ordinal regression model. We present details of this model and results in Appendix XX.



(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.

While results (Figure 8) are not statistically significant in Bayesian terms, as 0 does not lie outside the high-density interval, the interval reflects the 94% probability that the true parameter lies within it. This provides quantifiable evidence of trends while transparently accounting for uncertainties:

- Increased option length with text interface trends to *reduced* cognitive load with a posterior probability of approximately 69.8%, corresponding to a medium-to-large effect size ($d = 0.5$). 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 and a small-to-medium effect size ($d = 0.36$). 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% and a small effect size ($d = 0.14$). 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.

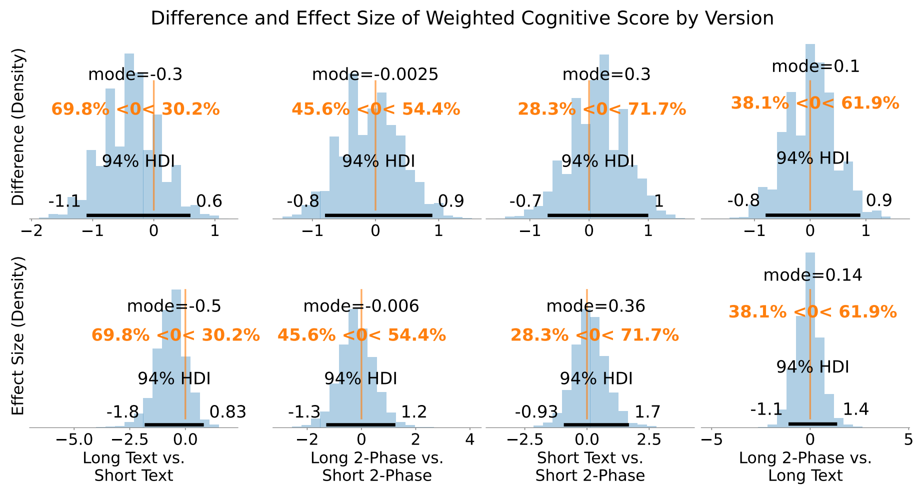


Fig. 8. Edit Distance Per Option

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 analyses are provided in Appendix XX.

5.2 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.

> *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 know . . . important [. . .]* – *S020, long text interface*

5.3 Different Sources of Cognitive Load

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

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

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

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

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

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

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

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 available[5] 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 variabilities. 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 XX.

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.

Similarly, participants using the two-phase interface make approximately 8.9 fewer steps per option (mode = 8.9, 94% 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]). Comparatively, the increase in edit distances between the short and long two-phase interfaces is substantially smaller (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.

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

[5] link-to-github

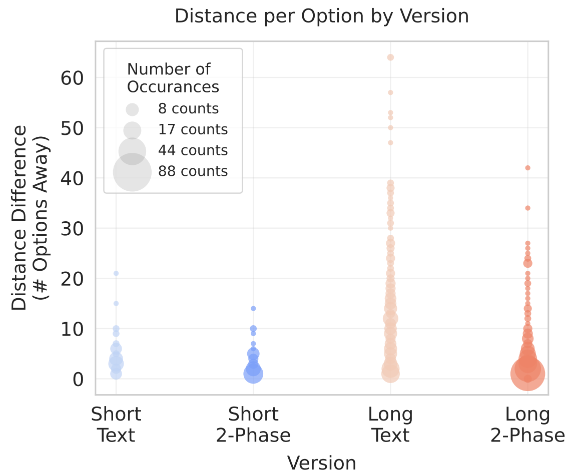


Fig. 9. Edit Distance Per Option

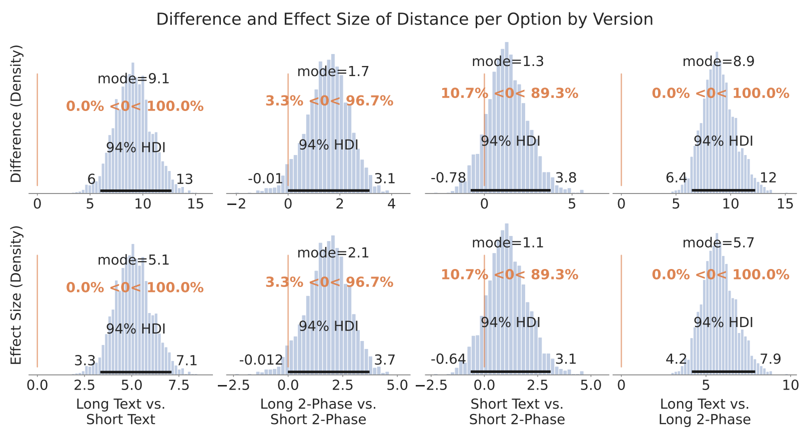


Fig. 10. Step-over Distance

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 XX) 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.

Figure XX 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]).

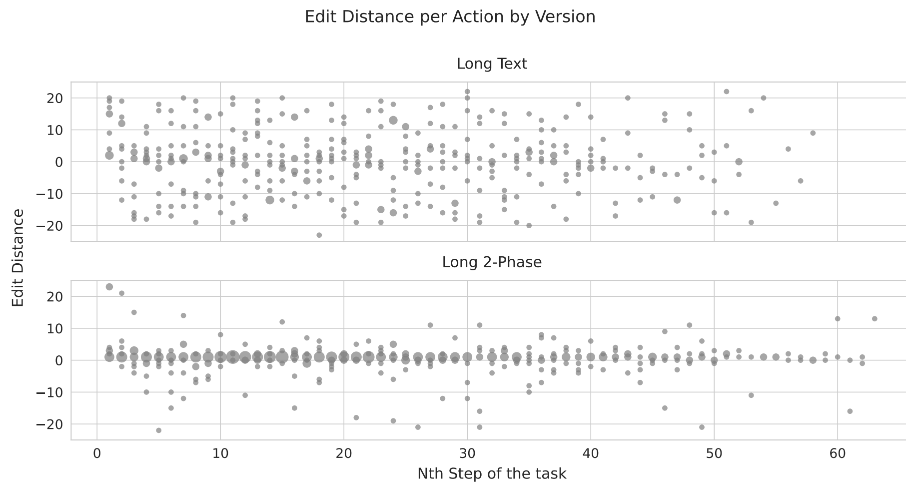


Fig. 11. Step-over Distance

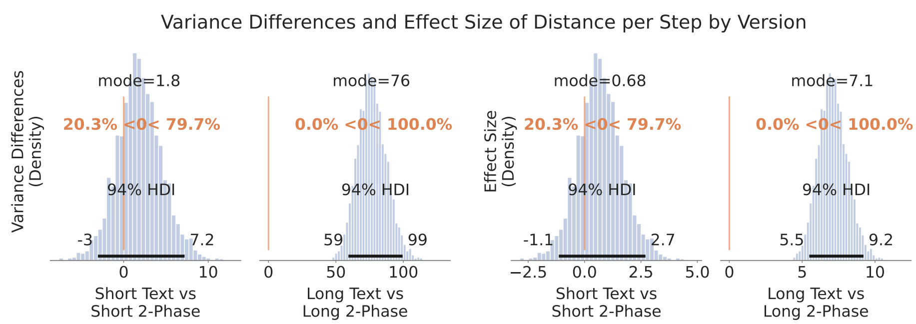
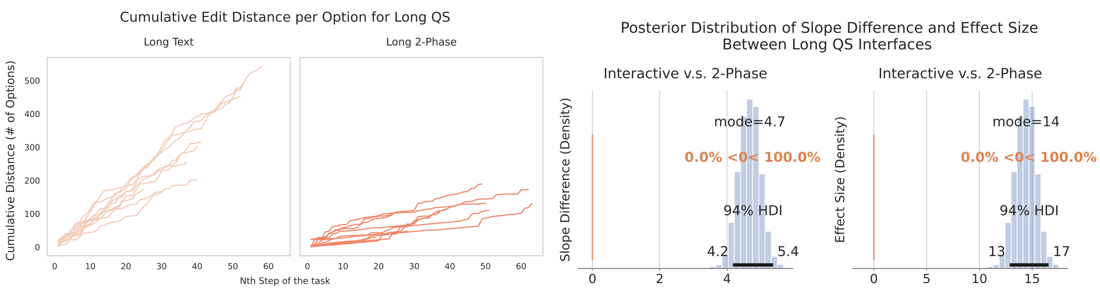


Fig. 12. Step-over Distance

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

Figure 13b shows that the slope for the long text interface is approximately 4.7, meaning each step by the text interface would add 4.7 edit distance (94% HDI = [4.2, 5.4]), compared to the long two-phase interface, which shows a statistically significant difference with a mode of 1.4 (94% HDI = [1.3, 1.7]). These results explain that the variance in edit distance per action and the increase in per option edit distance are consistent across participants between the two groups, showing that the organization phase allows participants to focus on adjusting options within proximity without having to navigate the interface to locate and make adjustments during the voting phase.



(a) Cumulative Edit Distance (Long QS v3/v4)

(b) Slope Difference and Effect Size

Fig. 13. Comparison of Cumulative Edit Distance and Slope Difference/Effect Size.

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

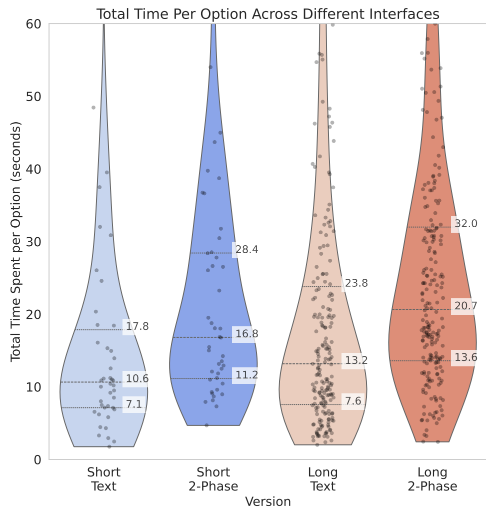
Notably, fewer participants (60%, N=6) report precise resource allocation in the long two-phase interface compared to 90% in the long text interface. These results make it evident that two-phase interface participants are more focused on deliberating preferences than simply completing the survey. Furthermore,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

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

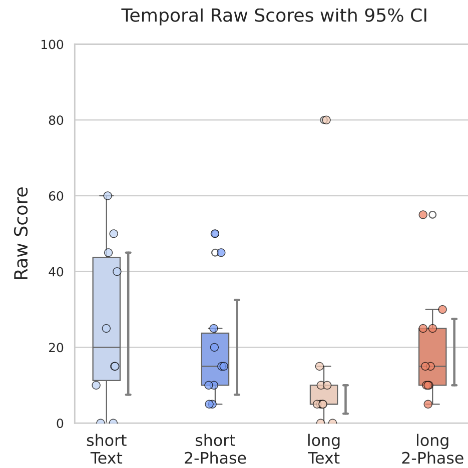
7 Clickstream data: Interface participants' time spent

In addition to distance, we analyze the time participants spend per option. We aggregate the total time participants spend per option using the QS system log. For participants in the two-phase interface conditions, this includes both organization and voting times for that option. The results are visualized in Figure 14a.

Overall, participants spend slightly more time per option in the two-phase interface than in the text interface. To quantify these observations, we model the time data as predictive variables of separate Gamma distributions to characterize the continuous response times observed under distinct experimental conditions defined by survey length



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



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

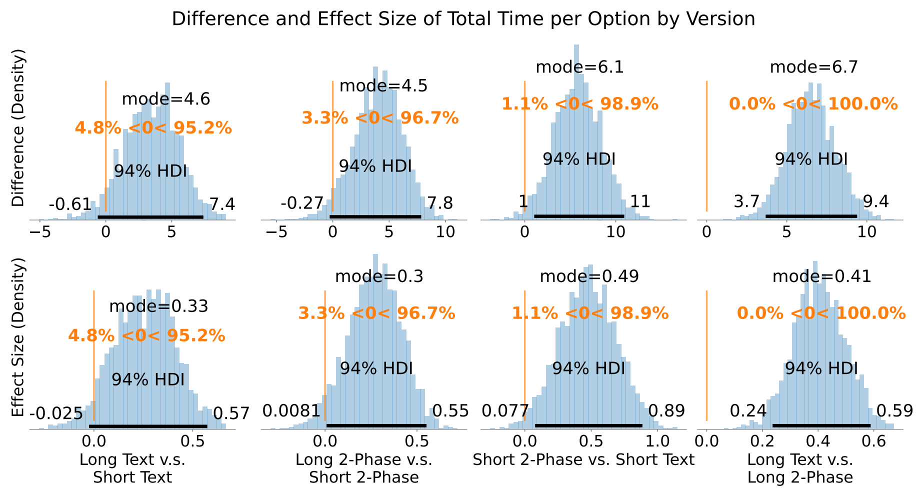


Fig. 15. Bayesian Time Diff

and interface type. Each of the four resulting subsets of data is modeled independently, with separate Gamma-distributed parameters governing the shape and rate of each group's time distributions.

We calculated the posterior differences between the two-phase and text interfaces for all pairwise comparisons of the four groups. The results in Figure 15 indicate that participants using the two-phase interface consistently spend more time per option than those using the text interface, regardless of survey length. For both the short and long QS, 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, 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 both cases, the intervals lie outside the ROPE of 0 ± 1, indicating statistical significance. These findings suggest that the two-phase interface encourages longer deliberation, particularly for longer lists of options. Details of the model are provided in Appendix XX.

Some literature points to increased time leading to time fatigue [**<empty citation>**], which can impair decision-making. Other decision science literature suggests that longer decision times can indicate deeper cognitive processing [81]. Our qualitative analysis points to the latter.

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 from the time already invested. This is reflected in their higher average time spent per option and overall time spent (μ = 716.86 seconds, σ = 164.04 seconds) completing the QS survey compared to the long text group (μ = 449.64 seconds, σ = 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` *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: μ = 139.83 seconds, σ = 76.43 seconds; short two-phase: μ = 178.78 seconds, σ = 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 14b, quantifiable results in Appendix XX), despite spending more time per option and traversing the longest distance (Section **??**). 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, which we will discuss further in Section **??**.

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 **??**.

8 Discussion and Future Works

In this section, we interpret the results related to cognitive load and survey respondent behaviors, emphasizing why the interactive interface did not uniformly reduce cognitive load in the long text interface while providing practical recommendations for practitioners deploying QS.

Our discussion centers on three key topics: elements of the two-phase interface that support preference construction, design recommendations for practitioners, and future challenges. Ultimately, we conclude that the two-phase interface has differential effects on the short and long surveys. While trends suggest a reduction in cognitive load with the two-phase interface compared to the text interface, we observe evidence of deeper engagement with options and enhanced preference construction, particularly in the long survey condition.

8.1 Result Interpretation

8.1.1 Deeper engagements through preference construction in two-phase interfaces. Our main findings indicate that the survey results, qualitative data, and observed behavioral differences reveal shifts in the types of cognitive load experienced by participants, especially for those completing the long survey. Cognitive load theory [54], when applied to the context of QS, identifies the three components of cognitive load: intrinsic load (the cognitive demand required to understand questions and response options), germane load (associated with deeper processing and evaluation of preferences), and extraneous load (stemming from navigating and operating the survey interface).

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

In the short survey condition, participants engage with all options simultaneously. The two-phase interface reduces some extraneous load associated with navigating the interface during voting, though it requires participants to complete the grouping phase. Despite this additional task, participants across both interface types report minimal or no physical demand. The two-phase interface likely facilitates easier engagement with preference construction due to its lower-trended cognitive load, as reflected in the increased likelihood of perceived lower cognitive load.

In the long survey condition, participants cannot engage with all options simultaneously, resulting in a higher intrinsic load at the start of the survey. The organization phase in the two-phase interface shapes participant behavior during the voting phase. While it streamlines the process of locating options, as exemplified by the reduction of edit distance, this benefit may be offset by the additional physical effort required to complete the grouping phase, as reflected by the slightly increased physical demand.

However, qualitative data regarding the voting task suggest that participants maintain their ability to invoke deeper engagement with options. Quantitative data reveal that participants make no fewer overall edits, with a bimodal distribution suggesting continued revision even at low budgets. Additionally, participants strategically consider broader options as they deliberate on nearby ones. These findings indicate a cognitive shift toward germane load, particularly during the voting phase.

In contrast, participants in the long text interface experience higher extraneous load, evident in shallower reflection and shorter overall voting times, despite exhibiting a greater overall edit distance. While some might argue that the additional grouping phase offers participants more opportunities to familiarize themselves with the options, the long edit distance suggests that participants in the text interface traverse the list frequently, providing ample opportunity to adjust their preferences. Qualitative data indicate that 70% of long text participants (N=7) scan the list while voting, with edit distance data reflecting multiple passes across the list.

The deliberate one-option-at-a-time presentation during the voting task in the two-phase interface reduces reliance on defaults and encourages deeper reflection. This is best-illustrated by `S013`, who emphasizes how the organization phase supports their preference construction:

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

Thus, based on this evidence, we argue that a text-based interface is not an optimal solution for long QS where deeper engagement and preference construction are desired. A two-phase interface enables participants to effectively exercise germane load, fostering deeper engagement with the content.

8.1.2 Plausible satisficing behaviors in long QS. In addition, the observed lower overall cognitive load in the long text interface may partly reflect *satisficing behaviors*. Satisficing refers to participants settling for *good enough* rather than *optimal* decisions [82] when unable to process all available information. Interviews reveal that 40% of participants (N=4) in this condition describe using satisficing strategies, while none from the long two-phase interface report such behaviors.These strategies are exemplified by participants prioritizing minimal effort over thorough evaluation, as illustrated by:

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

This quote illustrates satisficing decision-making, where participants settle for suboptimal choices. Additional participants describe similar strategies when deciding on votes:

> *[. . .] Because that was what was left. [Laughter] I probably wouldn't use that on <optionA> instead of <optionB>. [. . .]* `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.* `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.* `S041 (LT)`

These quotes highlight how participants in the long text interface adjust to external constraints rather than carefully weighing internal preferences. This behavior suggests that cognitive overload may lead participants to adopt less effortful strategies. However, further research is needed to fully understand the prevalence and impact of satisficing in long QS surveys.

In summary, the two-phase interface likely reduces extraneous load, particularly in the long survey condition, facilitating a cognitive *shift* toward deeper reflection and more deliberate decision-making. While the extent to which long QS surveys induce cognitive overload or satisficing remains unclear, the interactive interface shows promise in promoting deeper engagement with options and supporting comprehensive preference construction. The following section explores the specific elements that guide participants toward these outcomes.

8.2 Construction of Preference on Quadratic Survey

Completing QS involves a series of difficult decision tasks Lichtenstein and Slovic [7]. Svenson [52]'s differentiation and consolidation theory help explain how participants process these decisions. The decision process begins with differentiation, where participants identify differences and eliminate less favorable options, followed by consolidation, which strengthens their commitment to selected choices. This theory aligns with how the two-phase interface helps participants decompose options into categories, effectively reducing decision complexity.

Participants start by constructing preferences in situ, especially regarding options they have not previously considered:

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

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

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

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

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

The bi-modal behavior observed in the long interactive interface participants aligns with the differentiation and consolidation framework, as described in the results. Participants in the two-phase interface begin differentiating options earlier, allowing them to later adjust fine-grain votes. The faster and smaller vote updates indicate participants are consolidating. The less prominent bi-modal behavior from the long text interface participants implies that the interface guides this decision framework, as participant 037 explains:

> *I only start from the positive one [...] I finish everything ... and then I move to the second part (the neutral box). [...] I want to focus on these and make sure that resources are at least they get the attention they want. And if there's surplus and they can move to the second part.* S037 (LI)

In addition, the three key elements of the organization phase—presenting options one at a time, grouping them into categories, and enabling drag-and-drop—work together to structure participant preferences. These elements align with cognitive strategies like *problem decomposition* [83] and *dimension reduction*, which reduce cognitive overload. Bounded rationality highlights how cognitive limitations lead to sub-optimal decision-making due to the inability to process all available information [10]. It illuminates the importance of decision-making support interfaces rather than serving as a critique of human behaviors. One participant explains how the organization phase breaks down complex decisions into manageable steps:

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

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

> *Because with this many (options), especially when I'm thinking ... Ok, where was (the option) ... Where was (the option) you know? Oh, that's right. Maybe I could give another upvote to the, you know, whatever [...]* S028 (LT)

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

These evidence explain how the organization phase and the drag-and-drop features support differentiation and consolidation, and scaffold a decision-making framework that enables deeper engagement.

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

8.3 Opportunities for Better Budget Management

The current QS interface did not fully address the cognitive challenges of budget management outlined in Section 5.2. We identified three key issues:

First, 35% of participants ($N = 14$) emphasized the importance of automated calculation for deriving costs and tracking expenditures, highlighting the necessity of computer-supported interfaces for QS.

Second, participants struggled with deciding on an initial vote allocation. Some distributed credits equally across options, while others used 1, 2, or 3 votes as starting points. A few anchored their decisions on the tutorial's example of four upvotes. This suggests a need for better understanding of whether individuals possess absolute value preferences among options. Research on coherent arbitrariness [84] indicates that such preferences can be influenced by arbitrary anchors, similar to marketing's anchoring effects.

Finally, 12.5% of participants ($N = 5$) expressed confusion about the relationship between budget, votes, and outcomes, despite understanding their definitions. One participant stated:

> *[...] get rid of the upvote column or just get rid of the word upvote and just really focus on the money column. [...]* `S003 (ST)`

Participants like `S003` bypassed the quadratic mechanism, directly equating votes to resource allocation. While this does not invalidate the mechanism's power, it creates frustration and hampers decision-making. Future interfaces must better communicate these relationships to facilitate respondents' trade-offs.

8.4 Quadratic Survey Usage, Design Recommendations, and Future Work

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

8.4.1 Usage Recommendation: QS for Critical Evaluations. Our study highlighted the complex cognitive challenges and in-depth consideration required when ranking and rating options using QS, even in a short survey. Similar to survey respondents needing to make trade-offs across options, researchers and agencies seeking additional insights and alignment with respondent preferences must ensure that survey respondents have the cognitive capacity to complete such surveys rigorously. QS should be designed for critical evaluations, such as investment decisions, or situations where participants have ample time to think and process the survey. Practitioners should also caution against the use of long QS. If long QS is not avoidable, consider allowing participants to deliberate on each option prior to deploying QS without the organizing phase. For instance, revealing the options ahead of time can aid in preference construction.

8.4.2 Design Recommendations.

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

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

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

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

9 Limitations

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

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

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

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

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

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 decreased reliance on satisficing behaviors and encouraged more iterative and reflective preference construction, even though it did not clearly reduce overall cognitive load. Nonetheless, this design shift promoted deeper engagement and strategic thinking compared to the text-based interface, especially in longer surveys, by distributing cognitive effort more effectively. Quantitative results confirmed that participants, particularly those responding to the longer survey, exhibited more frequent fine-tuning of their votes, reflecting the iterative nature fostered by the interface. By integrating the organization and drag-and-drop functions, the interface facilitated both preference differentiation and consolidation, making it easier for respondents to refine their decisions. This two-phase interface design supports the development of future software tools that facilitate preference construction and promote the broader adoption of Quadratic Surveys. Future research should explore how to better support individuals in deciding the allocation of budget and design interfaces for smaller devices.

References

[1] Martin Pielot and Mario Callegaro. 2024. Did You Misclick? Reversing 5-Point Satisfaction Scales Causes Unintended Responses. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, (May 2024), 1–7. DOI: 10.1145/3613904.3642397.

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

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

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

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

[6] David Quarfoot, Douglas von Kohorn, Kevin Slavin, Rory Sutherland, David Goldstein, and Ellen Konar. 2017. Quadratic voting in the wild: real people, real votes. *Public Choice*, 172, 1-2, 283–303.

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

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

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

[10] Herbert A. Simon. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69, 1, 99–118. JSTOR: 1884852. DOI: 10.2307/1884852.

[11] John W. Payne, James R. Bettman, and Eric J. Johnson. 1988. Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 3, (July 1988), 534–552. doi: 10.1037/0278-7393.14.3.534.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[32] Tobin South, Leon Erichsen, Shrey Jain, Petar Maymounkov, Scott Moore, and E. Glen Weyl. 2024. Plural Management. SSRN Scholarly Paper. Rochester, NY, (Jan. 2024). doi: 10.2139/ssrn.4688040.

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

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

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

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

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

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

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

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

[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.

[69] Thomas L Saaty and Mujgan S Ozdemir. 2003. Why the magic number seven plus or minus two. *Mathematical and computer modelling*, 38, 3-4, 233–244.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

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

[82] 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.

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

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