# Participatory Budgeting Design for the Real World

**Gerdus Benadè,[1] Roy Fairstein,[2] Kobi Gal[2,3]**

[1] Boston University, USA
[2] Ben-Gurion University of the Negev, Israel
[3] University of Edinburgh, UK

### Abstract

Participatory budgeting engages the public in the process of allocating public money to different types of projects. PB designs differ in how voters are asked to express their preferences over candidate projects and how these preferences are aggregated to determine which projects to fund. This paper studies two fundamental questions in PB design. Which voting format and aggregation method to use, and how to evaluate the outcomes of these design decisions? We conduct an extensive empirical study in which 1 800 participants vote in four participatory budgeting elections in a controlled setting to evaluate the practical effects of the choice of voting format and aggregation rule. We find that $k$-approval leads to the best user experience. With respect to the aggregation rule, greedy aggregation leads to outcomes that are highly sensitive to the input format used and the fraction of the population that participates. The method of equal shares, in contrast, leads to outcomes that are not sensitive to the type of voting format used, and these outcomes are remarkably stable even when the majority of the population does not participate in the election. These results carry valuable insights for PB practitioners and social choice researchers.

## Introduction

Participatory budgeting (PB) is a type of direct democracy that allows a community to take part in deciding how to allocate a budget to different type of projects. It is growing in popularity and has been used around the world including in Madrid, Rome, Paris (Sintomer, Herzberg, and Röcke 2008) and New York (Su 2017). PB allows voters to affect local policies that matter to them, and it can increase community satisfaction when it leads to better sets of projects being funded.

Participatory budgeting elections consist of several phases. First, organizers elicit proposals for projects. These proposals are developed further (expanded, merged, costed, etc.) and finally filtered to a set of candidate projects. Citizens are then asked to cast a vote, and the votes are aggregated to determine which projects are funded. This paper assumes that a shortlist of candidate projects has already been fixed and studies the latter two components of PB design: preference elicitation, or how citizens vote for their preferred projects, and how votes are aggregated to determine the funded projects (aggregation).

PB design is a complex task that requires carefully balancing many factors. From the perspective of preference elicitation, voters should be allowed to express complex preferences over projects. For example, a voter may greatly prefer exactly one library being built rather than zero or two libraries, while being indifferent between the two proposed locations. At the same time, voting should be accessible to everyone in the community no matter their educational or socio-economic background, and there is a real risk that requiring too much information from voters may deter participation. Very simple voting formats come with their own problems and can lead to frustration and disillusionment. In practice, participation rates in PB are often low (in extreme cases as low as $0.1\%$ in Germany (Zepic, Dapp, and Krcmar 2017) and 1-3% in Chicago in 2012 and 2014 (Stewart et al. 2014; Carroll et al. 2016)). Organising bodies may also have secondary objectives, for example, designing an input format which exposes voters to budgetary constraints similar to what the city council faces.

The vast majority of real-world PB elections use $k$-approval voting (Aziz and Shah 2021), in which a voter approves their most preferred $k$ projects, though we are we aware of little formal justification for this choice. Other formats that have been proposed in the literature or used in practice include knapsack voting (voters approve their most preferred budget-feasible set) (Goel et al. 2019), ranking projects by value or value-for-money (Aziz and Lee 2020; Benade et al. 2020) and reporting utilities (Peters, Pierczyński, and Skowron 2020).

When it comes to aggregation, the funded projects should be justifiable given the observed votes and represent the preferences of as large a part of the population as possible. Real-world elections almost exclusively use some form of greedy aggregation in which projects are selected in decreasing order of the number of approvals received. Alternatives including the Method of Equal Shares (ES) (Peters, Pierczyński, and Skowron 2020) , Cumulative Single Transferable Vote (Skowron et al. 2020) and variations on the greedy approach (Talmon and Faliszewski 2019) have been proposed but are yet to be of practical significance.

This paper performs a comprehensive study of the effect of PB design on both user experience and the outcome of the election towards answering the fundamental design question of *what input format and aggregation method should be used* *in participatory budgeting elections?*

We recruit more than 1 800 participants on Amazon Mechanical Turk, present each with one of four different PB elections and ask them to vote in one of six input formats.

We track several aspects of the voter experience, including the time it takes to learn and use each format and voters' self-reported belief that an input format accurately captures their preferences. We find that $k$-approval voting leads to the best voter experience overall. Voters using $k$-approval spent the least time learning the format and casting their votes and found the format easiest to use. Somewhat surprisingly, voters also felt that it was the format that allowed them to express their preferences best, despite the fact that $k$-approval captures strictly less information about voter preferences than, for example, rankings.

We compare greedy aggregation to ES. Although the two methods lead to comparable social welfare, we identify two advantages of ES over greedy aggregation. First, it is much less sensitive to the input format used: no matter the input format used, the set of projects that are funded remains almost identical. This makes ES very versatile, as long as ES is used for aggregation the designer can choose almost any reasonable input format to satisfy whatever secondary objectives are present without fear that this will distort the outcome of the election. Second, we find that ES is remarkably stable as participation decreases. In fact, sampling as few as 25-50% of the voters and aggregating only the sampled votes rarely leads to a change in the set of projects being funded. In light of low levels of real-world participation, we believe this is a particularly attractive property.

Our study provides a structured comparison of PB design choices with valuable insights for PB practitioners. Our dataset and code will be made publicly available.

## Related Work

We briefly remark on the most closely related work; for a thorough review of the participatory budgeting literature see Aziz and Shah (2021).

Our work is closest to Benade et al. (2018), who study the role of input formats in lab controlled PB settings. They asked participants to vote over items that may aid their survival in a desert island. Each item has an associated weight and participants are informed that they are only able to carry items up to a specified weight limit. Benade et al. (2018) report on the time it takes to vote, the consistency of preferences across format, and the distortion and welfare of different aggregation methods (under the assumption that the utility of an alternative is equal to the points assigned to it).

In their work voters face only a single set of alternatives, and the desert island framing is far removed from PB. It is hard to argue that their findings are general (for example, there may just be an obvious budget-feasible set of projects that best aids survival). Our experiment much more closely mimics real-world participatory budgeting elections by using projects from real elections and locating projects on a city map. We attempt to check the robustness of our findings by replicating the experiment across four elections with different sizes and projects.

Goel et al. (2019) propose knapsack votes and empirically compare value-for-money comparisons, knapsack and $k$-approval votes in several ways, including through a user survey. It is found that knapsack and $k$-approval votes take a similar amount of time, while value-for-money comparisons induce less cognitive load. We find the opposite, voters consistently take significantly longer to form value-for-money rankings. One possible explanation is that Goel et al. (2019) performed a pair-wise comparison between projects, while we ask the voters for a full ranking over all projects. Goel et al. (2019) show some theoretical properties of knapsack under overlap utilities.

Benade et al. (2020) study the worst-case guarantees of different input formats. They also conduct simulations, using votes from real PB elections, to evaluate the social welfare that result from different input formats under a particular aggregation rule. They found that threshold approval had both the strongest theoretical guarantees and highest welfare in simulations. It does not stand out in our results.

A large literature studies the axiomatic properties of aggregation rules. Analysis most often focus maximizing the social welfare of the outcome (Benade et al. 2020; Goel et al. 2019; Jain, Sornat, and Talmon 2020; Hershkowitz et al. 2021) or satisfying some version of proportionality, which (roughly) requires that a group of voters with similar opinions should have impact on the outcome proportional to the size of the group (Fain, Goel, and Munagala 2016; Aziz et al. 2017; Sánchez-Fernández et al. 2017; Fain, Munagala, and Shah 2018; Aziz, Lee, and Talmon 2018; Skowron et al. 2020; Peters, Pierczyński, and Skowron 2020) or both (Fairstein et al. 2022; Michorzewski, Peters, and Skowron 2020). Notably, ES satisfies a proportionality condition called extended justified representation that guarantees a degree of representation to minority groups with common interests (Peters and Skowron 2020).

We briefly look at welfare and representation but focus on evaluating the stability of greedy aggregation and ES, which concerns the degree to which the outcome changes as one varies either the input format used in the election, or the degree of participation. To the best of our knowledge this has not been studied before.

# Preliminaries

For $k \in \mathbb{N}$, let $[k] := \{1, \ldots, k\}$. A PB instance is defined by a set of n voters $N = [n]$ that express their preferences over a set of $m$ projects $P = [m]$ and a budget $B$. Each project $p \in P$ has cost $c(p)$. The purpose of a participatory budgeting process is to select a subset of projects $S \subseteq P$ to fund which satisfies the budget constraint, so $c(S) = \sum_{p \in S} c(p) \leq B$.

We assume that voter $i$ gains utility $v_i(p)$ from project $p$ being funded. The value that voter $i$ has for a set of projects $S \subseteq P$ is $v_i(S) = \sum_{p \in S} v_i(p)$. The total value of project $p$ is $v(p) = \sum_{i \in N} v_i(p)$ and the social welfare of outcome $S$ is $\mathrm{SW}(S) = \sum_{i \in N} v_i(S) = \sum_{p \in S} v(p)$.

In reality, we do not have access to voters' true utilities. Instead, we receive a vote cast in some input format as proxy for these utilities. We study the following input formats.

1. A *utilities* vote (referred to as POINTS) asks a voter to divide 100 points between projects. Voter $i$ assigns $u_i(p) \geq 0$ to each project $p \in P$ so that $\sum_{p \in P} u_i(p) = 100$.

2. A *$k$-approval* vote (KAPP) asks a voter to approve (up to) their $k$ most preferred projects. The $k$-approval vote of voter $i$ is represented as a binary vector $\alpha_i \in \{0,1\}^m$ with $\sum_{p \in P} \alpha_i(p) \leq k$.

3. A *threshold approval* vote (TAPP) with threshold $t$ asks a voter to approve projects for which they have (normalized) utility at least $t$. Voter $i$'s preferences are represented by a binary vector $\alpha_i \in \{0,1\}^m$.

4. A *knapsack* vote (KNAP) asks each voter to select a subset of projects to maximize their utility subject to the budget. Voter $i$'s vote is a binary vector $\alpha_i \in \{0,1\}^m$ satisfying $\sum_{p \in P} \alpha_i(p) \cdot c(p) \leq B$.

5. A *ranking by value* (RANK) is a strict total order over the projects. Voter $i$'s ranking is denoted $\sigma_i$, where $\sigma_i(p)$ is the position of project $p$ in $i$'s ranking. Observing $\sigma_i(p_i) < \sigma_i(p_j)$ implies $v_i(p_i) \geq v_i(p_j)$.

6. A *ranking by value-for-money* (VFM) is a ranking of projects by the ratio of utility to cost, or 'bang-for-buck'. Voter $i$'s value-for-money ranking is denoted $\sigma_i$. Observing $\sigma_i(p_i) < \sigma_i(p_j)$ implies $v_i(p_i)/c(p_i) \geq v_i(p_j)/c(p_j)$.

The KNAP and VFM input formats require voters to know the budget and/or project costs. TAPP requires that voter utilities are normalized to the same scale.

Since we do not have access to true utilities we deduce proxy valuations for projects purely from the input profiles as follows: For POINTS, we set $v_i(p) = u_i(p)$. For KNAP, KAPP and TAPP, where the vote is a binary vector, we assume $\{0,1\}$ utilities with $v_i(p) = \alpha_i(p)$. For RANK and VFM, we use Borda scores, so $v_i(p) = m - \sigma_i(p)$.

After voters express their preferences, an aggregation method maps the input profile to a budget-feasible set of projects to fund. The first aggregation method that we consider is the *greedy method*, which is most commonly used in real-world PB elections. It selects projects in decreasing order of $v(p)$ until the budget is depleted, skipping projects as necessary.

The second method we consider is called equal shares (Peters and Skowron 2020), or ES. ES virtually divides the PB budget equally among all the voters. At each iteration, it identifies the remaining projects which can be fully funded by their supporters using their remaining virtual funds. It then funds from among these the project $p$ with the smallest ratio between $c(p)$ and $v(p)$. The cost of this project is divided among its supporters proportional to their utility. This process terminates when no project can be funded by only its supporters. One quirk of ES is that it may not return a set of projects which exhaust the budget. To handle this case, we fill out the selected subset by perturbing voter values so that each voter has non-zero value for every project, as proposed in Peters, Pierczyński, and Skowron (2020).

## User Study Description

The user study consists of asking voters to vote using one of the six input formats above in one of four different participatory budgeting elections in a hypothetical city. We recruit roughly 75 different participants for each of the 24 configurations (four elections times six input formats) using Amazon Mechanical Turk, in total just over 1 800 participants.

The small elections (SMALL-A and SMALL-B) consisted of 10 candidate projects while the larger elections (LARGE-A, LARGE-B) had 20 projects. Project descriptions and costs were taken from real participatory budgeting elections from across the world. Each project was assigned a location on a map of the city and categorized in one of five categories (education; streets and transportation; culture and community; facilities and recreation; environment, public health and safety). Full details on every election including the project titles, descriptions, costs, locations, etc. may be found in the appendix.

Participants were first presented with a written and video description of the PB voting task. They had to pass a simple quiz about the task in order to proceed. Next, participants carried out the voting task in their allocated PB configuration. Each participant was assigned a (random) location on the city map and shown the description and location of the projects. We expected that a voter's location relative to the project may affect their preferences (e.g., people may prefer to vote for a library close to their current location).

Figure 1 shows the interface presented to participants who were asked to cast knapsack votes. The left-most column shows the instructions and an indication of what fraction of the budget remains to be allocated. The center column shows the category and title of each proposed project, and a participant can click on a project to see a more detailed description. The right-most panel shows a map of the city on which the locations of the voter and the projects are marked (hovering over a project highlights its location). The other interfaces used in the study may be viewed in the appendix.

After submitting their PB vote, participants were required to answer several consistency questions about their vote that were designed to identify voters who vote randomly or carelessly. Finally, participants were asked to complete a short survey about their subjective experience of voting in the assigned format. They were asked to rate (on a scale of 1 to 5) how easy they found the task, how much they liked the user interface, how expressive they found the input format, and how much the project categories and their location on the map affected their decisions.

Participants were rewarded a fixed sum for participation and received a 75% bonus for passing the consistency questions. IRB approval was obtained from the corresponding institution.

## Effect of Input Format on Voter Experience

We investigate the practical effects of using different input formats by comparing the experiences of voters using the respective input formats.

This comparison has several components. Objectively, we record the time that it takes a voter to complete each of the

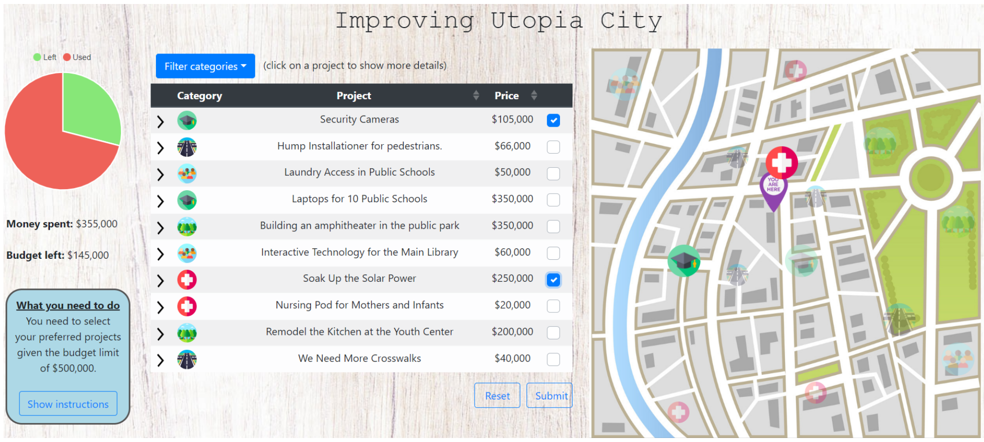



Figure 1: The GUI shown to participants using the KNAP input format (center), showing the city map (right) and budget (left)

first three stages of the task (completing the tutorial, answering the post-tutorial quiz, and casting a vote). We also report the number of participants who fail the post-task consistency test and argue that a participant's inability to recall simple information about the vote they just cast is correlated with the participant finding the task taxing, confusing or tiresome. More subjectively, we examine participants' self-reported scores from the post-completion survey.

## Response time

The time it takes to complete a task is a recognized proxy for the cognitive burden or difficulty of the task (Rauterberg 1992). The average time to complete each stage of the experiment is visually represented in Figure 2. Results in this section are averaged across elections and the conclusions do not change when looking at any individual election.

Participants voting in the VFM format are consistently slowest to complete each stage of the experiment, suggesting that voters find VFM very hard to use. Based on the time it takes to cast a vote, POINTS and TAPP are the next hardest formats to use. In the former case, we believe this supports the idea that asking voters for cardinal utilities is generally not feasible; in the latter, it highlights the complications that sprout from TAPP requiring cross-voter normalization (as well as the fact that TAPP is a relatively niche format which voters are unlikely to have encountered before).

Voters found KAPP, KNAP and RANK easiest to use, in this order.

## Consistency checks

Of the roughly 75 participants recruited for each of the 24 configurations of the experiment, on average 58 passed the consistency test.

The consistency test is designed to check whether the participant can recall simple information about the recently completed task. There are several reasons why a voter may fail the test. For example, they may answer randomly in order to complete task quickly, become distracted if the task is too tiresome, or confused if the input format is hard to un-

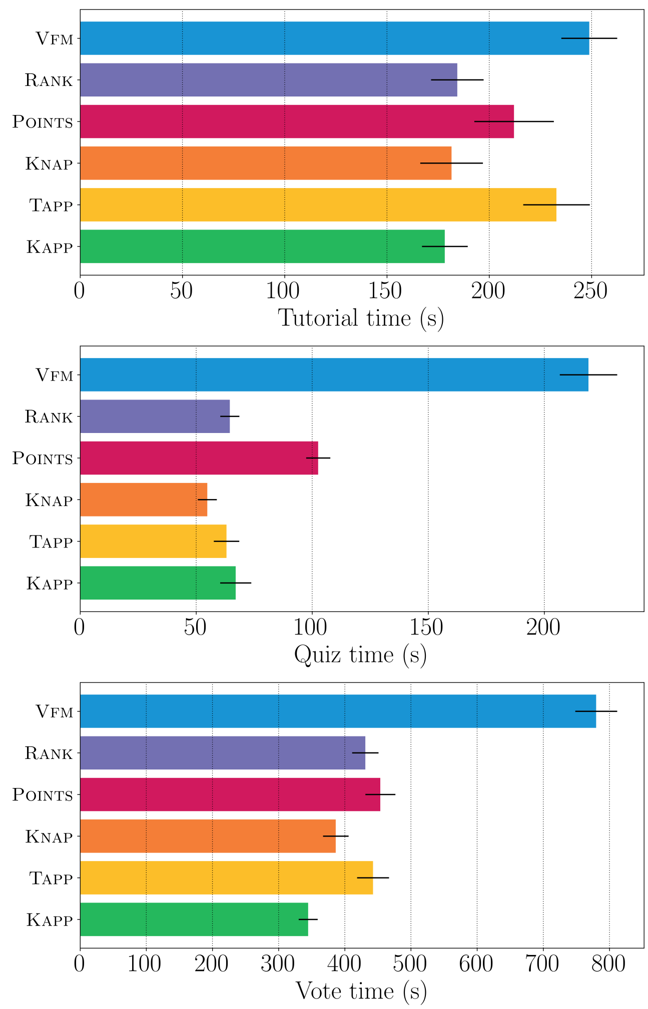



Figure 2: Average time ($s$) to complete each stage.

derstand. As such, we treat this as another signal about the difficulty of expressing your preferences in a given format.

We compare the rate of passing the consistency test across input formats. Participants using the ranking-based formats were most consistent, followed by the approval-based formats. We speculate that the high consistency of VFM is, at least in part, thanks to the inordinately long amount of time voters spend considering their votes. Users of POINTS were comfortably the least consistent, which supports earlier evidence that providing cardinal utilities is a challenging task.

There was virtually no variation in the rate of passing the consistency test across elections. A complete breakdown of the results may be found in the appendix.

### Self-reported voting experience

In addition to the time it took participants to complete the task and how much they could recall about their submitted preferences, we also expressly ask them to rate various aspects of the experience on a scale of 1 to 5. Survey results are summarised in Figures 3 and 4. Results for the questions omitted from this discussion may be found in the appendix. In order to find statistical significance, we first used Kruskal–Wallis test to show the input formats give different results, followed by the Dunn test for post hoc comparison. All of the survey questions gave $p < 0.004$ for the Kruskal test. The Dunn test evaluated statistically significance at the $p < 0.05$ level.

**Ease of use** We asked the voters how easy they found the voting task. Unsurprisingly, participants found VFM significantly harder to use than any other format. KAPP was rated as significantly easier to use than all other formats except POINTS.

**Perceived expressiveness** We asked the participants how well they felt their vote captured their preferences. Participants found KAPP to be most expressive and, in particular, significantly more expressive than TAPP, VFM, and POINTS. It is somewhat surprising that KAPP was rated as more expressive than POINTS, since you can infer a KAPP-vote from your POINTS-vote; we speculate that the relative complexity of POINTS explains part of this phenomenon. VFM was perceived to be least expressive by some margin. Again, the difficulty of using VFM and the fact that voters are forced to explicitly consider project costs may have contributed to the perception that it is less expressive than RANK, which also asks for a ranking of alternatives.

**Effect of additional information** Projects were assigned to one of five categories (e.g. transport or education) and given locations on a city map. Two survey questions asked about the extent that project categories and the map influenced participants' preferences. These results are summarised in Figure 4

Participants using KAPP were influenced more by project locations than any other group. Similarly, KAPP-voters paid significantly more attention to project categories than all other groups except POINTS-voters. It may be that the KAPP format was easy enough to understand and use that it freed participants up to consider additional, non-core information about the projects and election in their decisions.

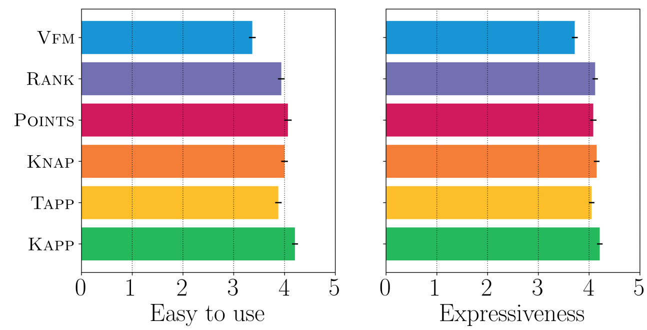



Figure 3: The average feedback for each input format.

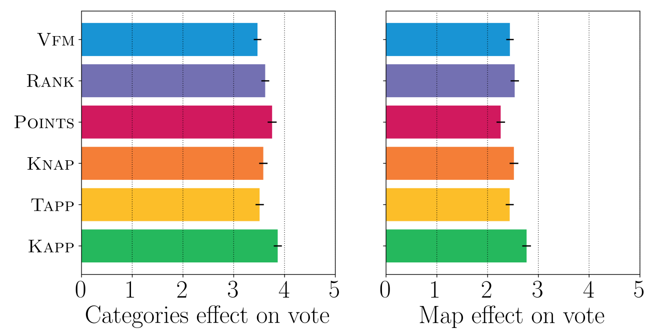



Figure 4: The average effect of the categories and the map.

## Effect of Aggregation Method on Outcome

In this section we analyze the effect of aggregation methods on the outcome of the elections.

### Stability

We first study the stability of the different aggregation methods. We say that an aggregation method is stable when the outcome of the election is:

1. Robust to the choice of input format. This allows the organizers to freely select an input format without fear of affecting the outcome of the election (or default to a format voters find easy to use).
2. Robust to partial participation. Voter turnout may be low in real PB elections, ideally the outcome of the election is not drastically affected by the (non-)participation of a small group of voters.

In this experiment, we first repeatedly sample $n' = 40$ participants per configuration, and aggregate the resulting vote profiles using greedy aggragtion and ES. This process is repeated 200 times.

The fraction of repetitions in which each project is funded in each configuration is summarized in a heatmap in Figure 5. The top row of results corresponds to greedy aggregation and the bottom row to ES. Within each panel projects are ordered in order of increasing cost, and the cells range from black (meaning the project is never funded) to white (always funded).

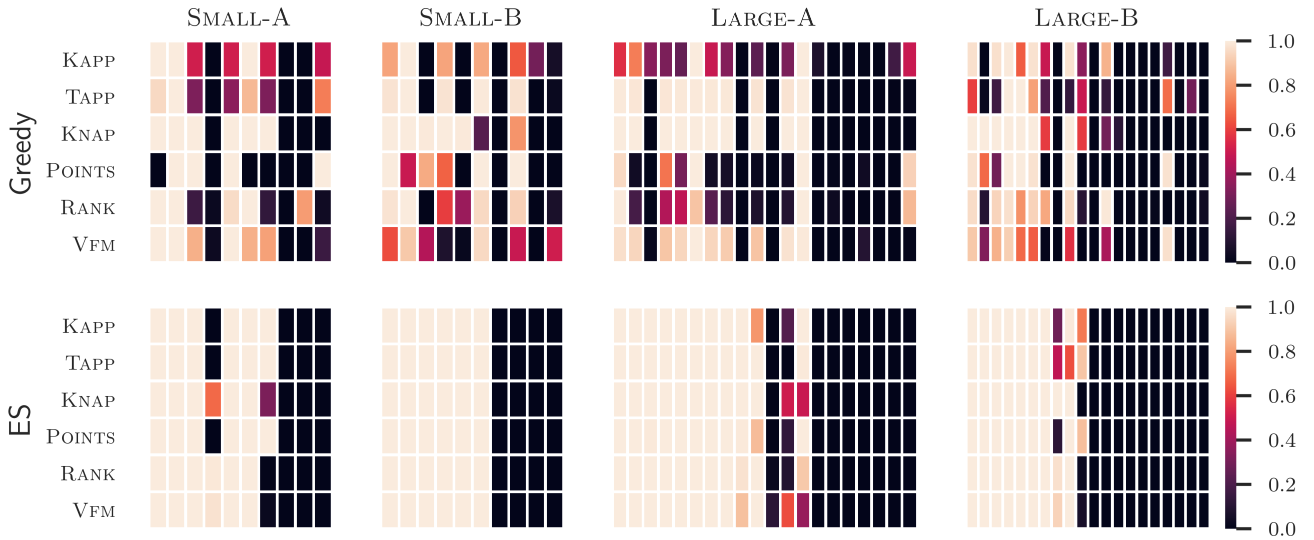



Figure 5: Stability heatmaps for greedy aggregation (top) and ES (bottom) in each election. Within each panel each row represents an input format and each column a project (ordered in increasing cost). The intensity of a cell indicates the fraction of instances in which the project was funded.

Strikingly, the outcome of the election is almost entirely unaffected by the choice of input format when using ES — the majority of the projects are either always funded under every input format or never under any. Greedy aggregation exhibits this property only in rare instances, for example, the outcome of Large-A does not change if the input format is changed from Knap to Tapp. This robustness to the choice of input format remains present when the number of sampled voters is varied $n' \in \{10, 20, 30\}$.

We also observe in Figure 5 that ES is robust to partial participation, it is very rare that the decision to fund a project changes across repetitions when sampling voters at random, i.e. under a uniform model of voter participation/abstention. Greedy aggregation does not exhibit this property.

Of course, since we are sampling 40 votes from (on average) 58 consistent voters the sampled profiles are correlated. We repeat the experiment sampling $n' \in \{10, 20, 30\}$ voters each time to investigate what happens when the correlation in sampled profiles decreases. To quantify stability under partial participation, we compute the entropy of the outcome for each configuration. Suppose that project $p \in P$ is funded with probability $f_{V,A}(p)$ when aggregating votes cast in voting format $V$ using aggregation method $A$, then entropy$(V, A) = -\sum_{p \in P} f_{V,A}(p) \cdot \log_2(f_{V,A}(p))$.

Figure 6 shows the entropy for election Small-A across input formats for greedy aggregation and ES aggregation when varying the degree of participation (i.e. the number of sampled voters). Unsurprisingly, the correlation between vote profiles decreases and entropy increases as fewer voters are sampled. The entropy of ES is consistently significantly lower than that of greedy aggregation across input formats. We conclude that ES is significantly more robust to partial participation than greedy aggregation. One exception is worth highlighting, the entropy of greedy aggregation with Knap votes is fairly competitive with that of ES. This suggests that if the organizers of a PB election are dead-set on using greedy aggregation and also worried the effect of partial participation, Knap may be an attractive option. For the other three elections, ES aggregation comfortably outperforms greedy aggregation across all input formats and sample sizes. Full results may be found in the appendix.

We conclude this section with an observation from Figure 5 unrelated to stability. It appears to be the case that more expensive projects are rarely funded when using ES. We believe this may (at least for the binary input formats) be an artefact of assuming $v_i(p) \in \{0, 1\}$, which makes it very hard for an expensive project to have a ratio of utility to cost that justifies funding it. It remains to be seen whether this trend persists, for example, when assuming $v_i(p) \in \{0, c(p)\}$.

## Social Welfare

An obstacle to comparing outcomes based on social welfare is participants are assigned only one format, so there is no common basis of comparison when aggregating vote profiles from different formats. To address this, we make the assumption that, since voters are randomly assigned to input formats, the distribution of voter preferences are the same across formats. As such, we can deduce voter valuation functions from the full vote profile in *any* format and treat these aggregated values as proxy for the welfare of all voters, regardless which format they used.

We report in Figure 7 the social welfare (averaged over the four elections) as measured when treating the full profile of Points or Knap voters as representative.

Neither aggregation method strictly outperforms the

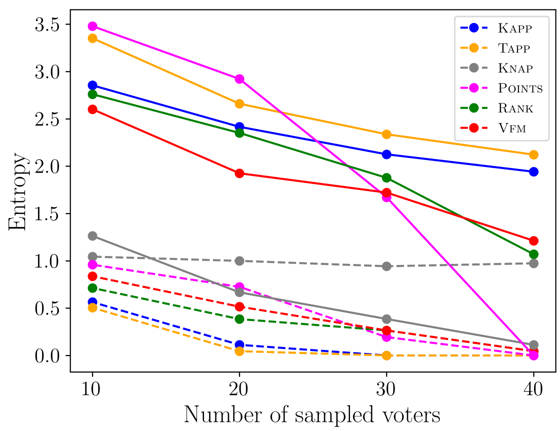



Figure 6: Entropy for each input format under greedy aggregation (solid lines) and ES (dotted lines) as the number of sampled voters changes.

other. However, we observe that, when using greedy aggregation, the choice of input format can have a large effect on welfare. This is seen most clearly when using KNAP-voters to compute welfare. Unsurprisingly, ES is less sensitive to the choice of input format and achieves very similar welfare across input formats. Results when using the other formats to measure welfare are less extreme than the results shown here and may be found in the appendix.

## Discussion

These results highlight practical insights for the design of real participatory budgeting elections. When choosing an input format, we find no reason to deviate from the current standard practice of using KAPP voting: Voters find the format easy to understand, easy to use, and believe it allows them to express their preferences accurately.

Greedy aggregation is used almost universally, however, our experiments find that ES has two big advantages over greedy aggregation. First, the outcomes are largely stable and unchanged when only a subset of the population participates. In light of low levels of real-world participation, we believe this is a particularly attractive property. Second, ES appears to be robust to the choice of input format, in other words, it affords the city officials responsible for administering the participatory budgeting election a great deal of freedom to choose an input format which meets whatever secondary objectives are deemed to be important.

One argument that remains in favour of greedy aggregation is its transparency, it is arguably easier to explain to voters than ES. If greedy aggregation is used and stability to partial participation is a primary concern, then KNAP emerges as a fairly user-friendly format overall and a clear front-runner in terms of stability.

These recommendations should be taken in the context of the limitations of this study. First, the behavior and preferences of participants recruited on Amazon Mechanical Turk may not be sufficiently similar to those of real voters.

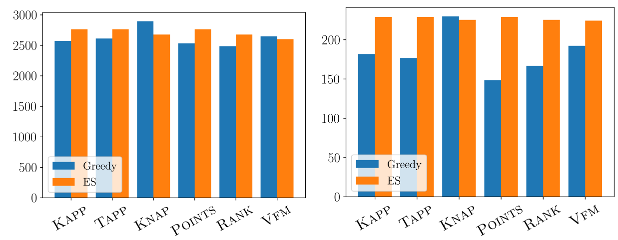



Figure 7: Average social welfare (across elections) using the POINTS- (left) and KNAP-voters (right) to compute welfare.

Though we attempt to ensure the findings are robust across the parameters of different elections by varying the number and type of projects, there is no mimicking the richness, variety and local idiosyncracies of real-world instances. Similarly, real voter utilities likely exhibit complementarities and externalities — a far cry from our utility proxies. Second, it is possible that, despite our best efforts, the voter experience was affected by the interface design choices we made.

We conclude with two directions for future study. When it comes to designing participatory budgeting elections, the proof, to a large extent, should be in the pudding. One can imagine a variant of the experiment presented in this paper in which votes from different PB designs are aggregated and voters are directly asked to compare different elections and outcomes on the transparency of the process, personal satisfaction, perceived fairness of the outcome, etc. Directly asking people their opinion about PB outcomes may yield new insights that can help inform design decisions while avoiding assumptions like additive utilities.

Democratic innovations live and die by the extent to which voters' voices are heard. But this requires representative participation. As a fledgling part of the democratic process, participatory budgeting appears to be particularly susceptible to poor participation. It is important to understand how design choices around voting formats and the transparency of aggregation rules affects the participation (now and in the future) of voters from a wide range of socio-economic backgrounds. Our stability experiments are a step in this direction but assume participation is uniform across the population, which is unlikely to be the case in reality.

## References

Aziz, H.; Brill, M.; Conitzer, V.; Elkind, E.; Freeman, R.; and Walsh, T. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare*, 48(2): 461–485.

Aziz, H.; and Lee, B. E. 2020. The expanding approvals rule: improving proportional representation and monotonicity. *Social Choice and Welfare*, 54(1): 1–45.

Aziz, H.; Lee, B. E.; and Talmon, N. 2018. Proportionally Representative Participatory Budgeting: Axioms and Algorithms. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 23–31.

Aziz, H.; and Shah, N. 2021. Participatory budgeting: Models and approaches. In *Pathways Between Social Science and Computational Social Science*, 215–236. Springer.

Benade, G.; Itzhak, N.; Shah, N.; Procaccia, A. D.; and Gal, Y. 2018. Efficiency and usability of participatory budgeting methods.

Benade, G.; Nath, S.; Procaccia, A. D.; and Shah, N. 2020. Preference elicitation for participatory budgeting. *Management Science*.

Carroll, C.; Crum, T.; Gaete, C.; Hadden, M.; and Weber, R. 2016. Democratizing tax increment financing funds through participatory budgeting. *Chicago, IL: Blocks Together, Great Cities Institute, and Participatory Budgeting Project*.

Fain, B.; Goel, A.; and Munagala, K. 2016. The core of the participatory budgeting problem. In *International Conference on Web Information Engineering (ICWIE)*, 384–399. Springer.

Fain, B.; Munagala, K.; and Shah, N. 2018. Fair allocation of indivisible public goods. In *Proceedings of the 2018 ACM Conference on Economics and Computation (ACM-EC-2018)*, 575–592.

Fairstein, R.; Reshef, M.; Vilenchik, D.; and Gal, K. 2022. Welfare vs. Representation in Participatory Budgeting. In *Proceedings of the 22th AAMAS*, 409–417.

Goel, A.; Krishnaswamy, A. K.; Sakshuwong, S.; and Aitamurto, T. 2019. Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation (TEAC)*, 7(2): 1–27.

Hershkowitz, D. E.; Kahng, A.; Peters, D.; and Procaccia, A. D. 2021. District-fair participatory budgeting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 5464–5471.

Jain, P.; Sornat, K.; and Talmon, N. 2020. Participatory Budgeting with Project Interactions. In *IJCAI*, 386–392.

Michorzewski, M.; Peters, D.; and Skowron, P. 2020. Price of fairness in budget division and probabilistic social choice. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2184–2191.

Peters, D.; Pierczyński, G.; and Skowron, P. 2020. Proportional Participatory Budgeting with Cardinal Utilities. *arXiv preprint arXiv:2008.13276*.

Peters, D.; and Skowron, P. 2020. Proportionality and the limits of welfarism. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 793–794.

Rauterberg, M. 1992. A method of a quantitative measurement of cognitive complexity. *Human-computer interaction: Tasks and organisation*, 295–307.

Sánchez-Fernández, L.; Elkind, E.; Lackner, M.; Fernández, N.; Fisteus, J.; Val, P. B.; and Skowron, P. 2017. Proportional justified representation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Sintomer, Y.; Herzberg, C.; and Röcke, A. 2008. Participatory budgeting in Europe: Potentials and challenges. *International journal of urban and regional research*, 32(1): 164–178.

Skowron, P.; Slinko, A.; Szufa, S.; and Talmon, N. 2020. Participatory Budgeting with Cumulative Votes. *arXiv preprint arXiv:2009.02690*.

Stewart, L. M.; Miller, S. A.; Hildreth, R.; and Wright-Phillips, M. V. 2014. Participatory budgeting in the United States: a preliminary analysis of Chicago's 49th Ward experiment. *New Political Science*, 36(2): 193–218.

Su, C. 2017. From Porto Alegre to New York City: Participatory Budgeting and Democracy. *New Political Science*, 39(1): 67–75.

Talmon, N.; and Faliszewski, P. 2019. A framework for approval-based budgeting methods. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2181–2188.

Zepic, R.; Dapp, M.; and Krcmar, H. 2017. Participatory budgeting without participants: Identifying barriers on accessibility and usage of German participatory budgeting. In *2017 Conference for E-Democracy and Open Government (CeDEM)*, 26–35. IEEE.

## Details omitted from Comparing input formats

### Response time

Figure 8 Time for consistency.

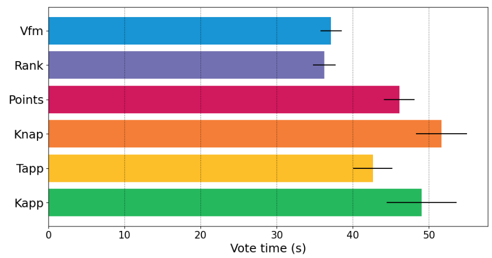



Figure 8: How long it took the participants to do the consistency quiz (in seconds) on average.

### Consistency

Figures 9,10 how many participants were consistent across different formats and across different project sets.

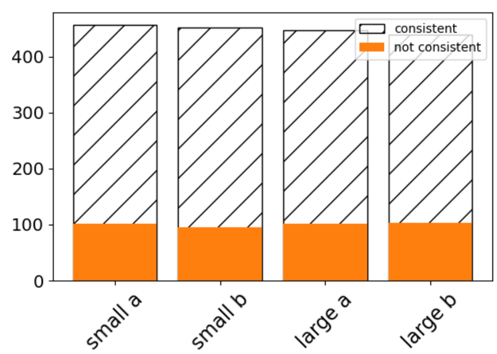



Figure 9: Total and consistent participants for each election.

## Welfare from all formats voters

Figures 11a-11f shows the welfare (averaged across elections), when using each of the input formats voters to calculate the welfare.

## Entropy

Figures 12a-12d shows the entropy for all of the project sets, all input formats when using greedy and ES, given different sample size.

## Interface For Voting

Figure 13a-13f shows the interface for each of the input formats.

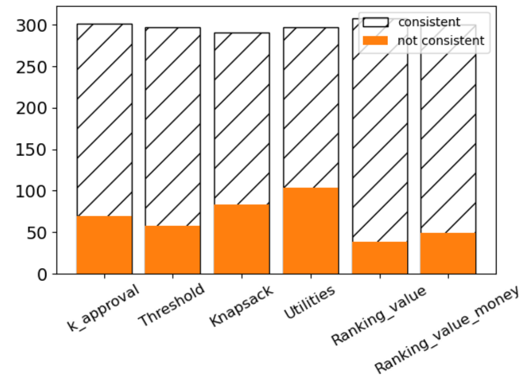


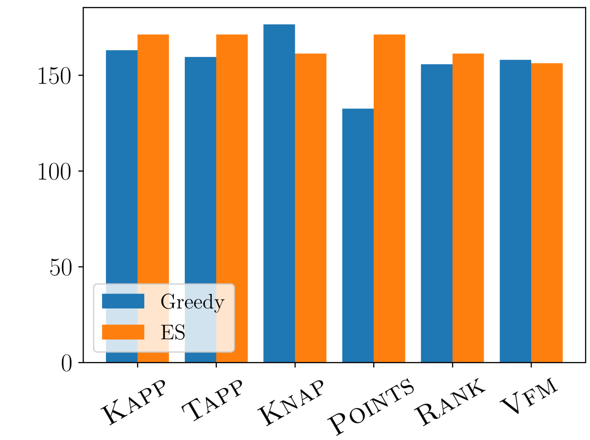

Figure 10: The total number of participants who vote in each format, as well as the number that fail the consistency test. Numbers are aggregated across elections.

## Experiment Configurations

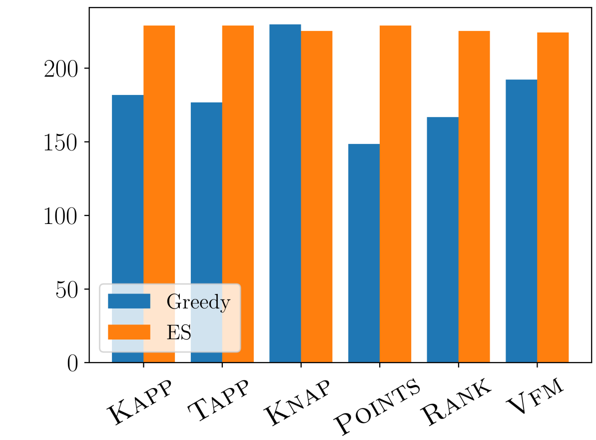
In the experiments each participant encountered one of six input formats and one of 4 possible elections, each with different set of projects, either 10 projects or 20, such that each project belong to one of 5 categories:

- Culture and community.
- Streets, Sidewalks and Transit.
- Environment, public health and safety.
- Facilities, parks and recreation.
- Education.

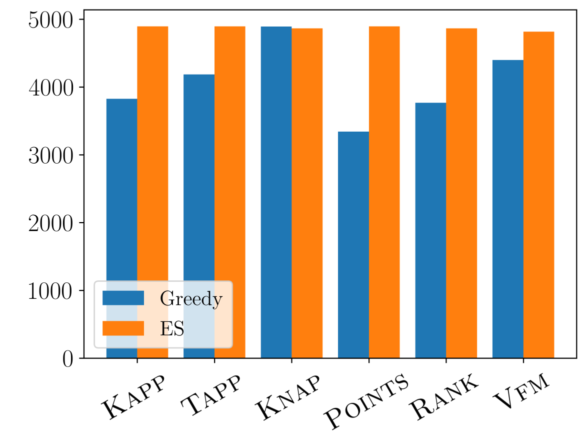
Tables 1-4 show the four different elections the participants might encounter in the experiment, a short description, and to which category it belongs.



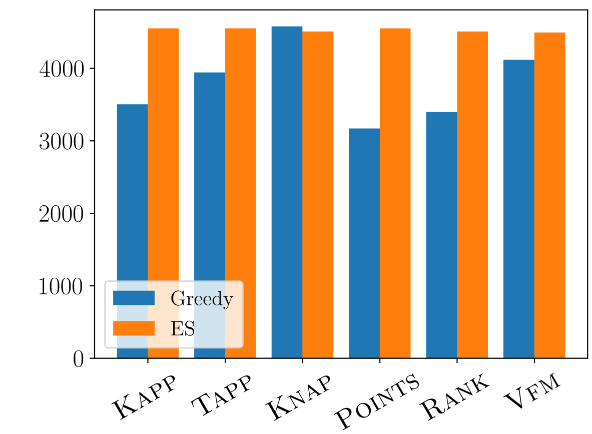

(a) Average social welfare (averaged across elections) using KAPP-voters to compute welfare.



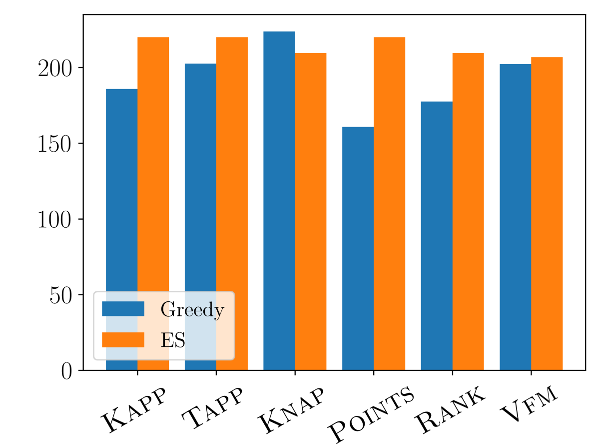

(b) Average social welfare (averaged across elections) using KNAP-voters to compute welfare.



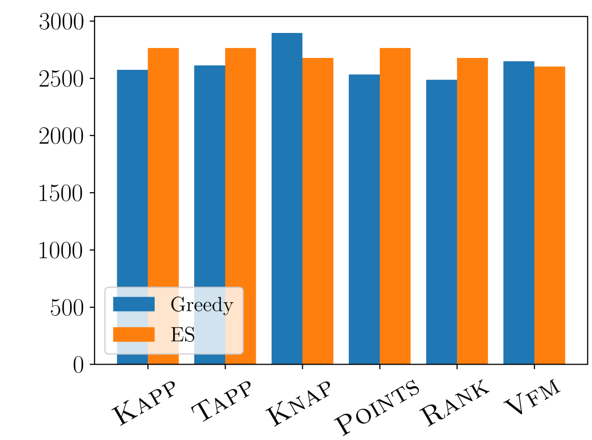

(c) Average social welfare (averaged across elections) using RANK-voters to compute welfare.




(d) Average social welfare (averaged across elections) using VFM-voters to compute welfare.

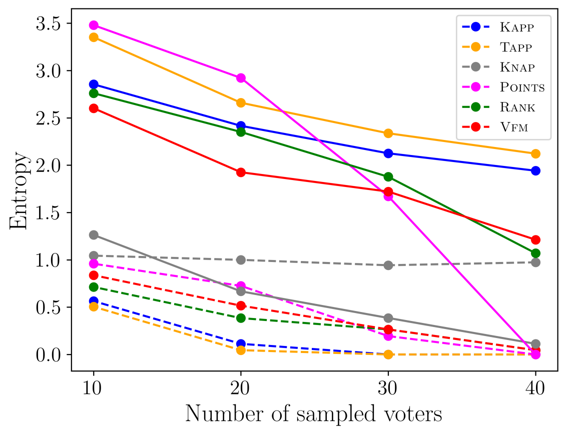



(e) Average social welfare (averaged across elections) using TAPP-voters to compute welfare.

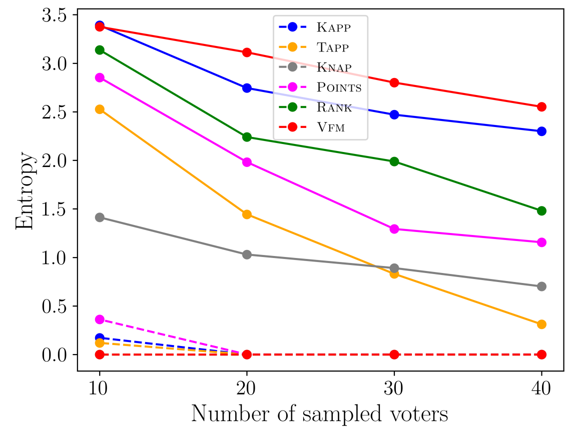



(f) Average social welfare (averaged across elections) using POINTS-voters to compute welfare.

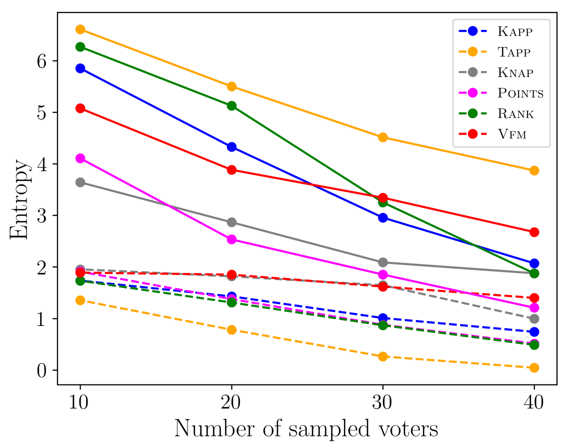
Figure 11: Welfare.



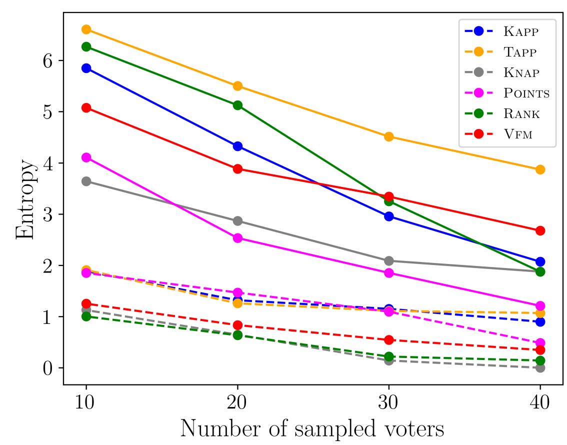

(a) Greedy (full lines) and ES(dotted lines) entropy for each input format (sample format get same color) given different sample size in project set SMALL-A




(b) Greedy (full lines) and ES(dotted lines) entropy for each input format (sample format get same color) given different sample size in project set SMALL-B

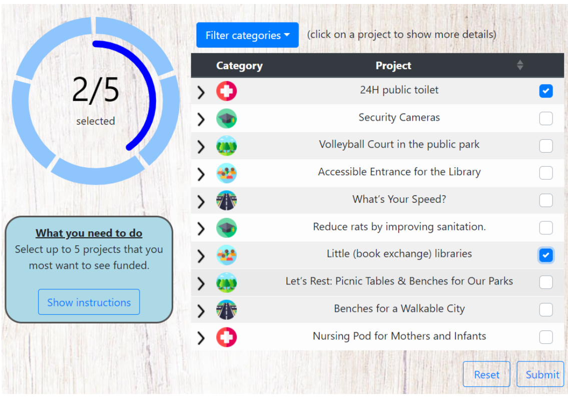



(c) Greedy (full lines) and ES(dotted lines) entropy for each input format (sample format get same color) given different sample size in project set LARGE-A

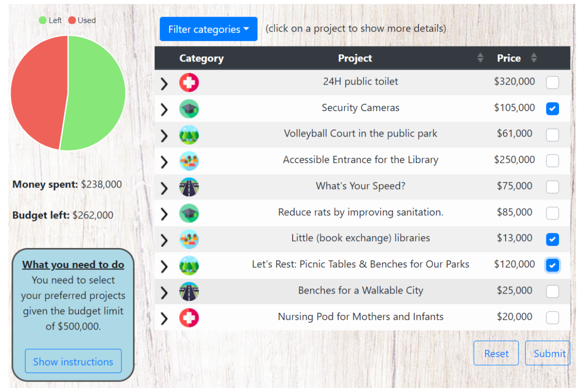



(d) Greedy (full lines) and ES(dotted lines) entropy for each input format (sample format get same color) given different sample size in project set LARGE-B

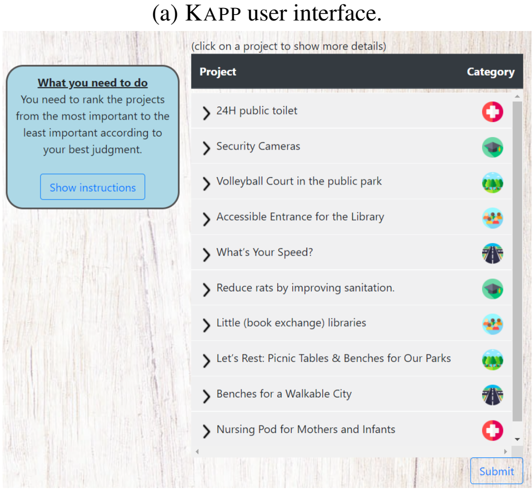
Figure 12: Entropy for each of the elections.



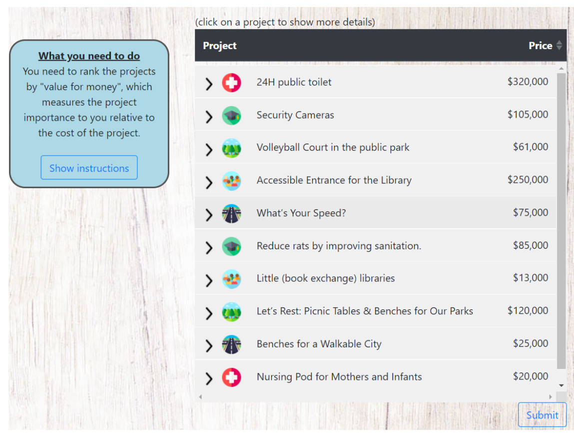

(a) KAPP user interface.



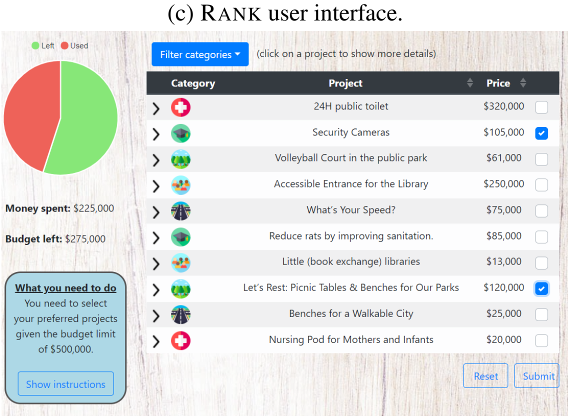

(b) KNAP user interface.



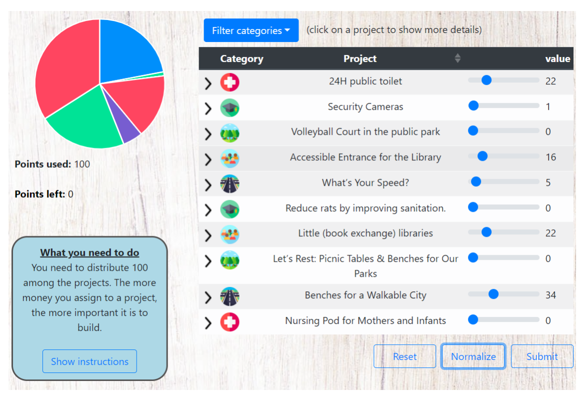

(c) RANK user interface.




(d) VFM user interface.




(e) TAPP user interface.




(f) POINTS user interface.

Figure 13: Input formats interface.

| Project | Description | Category | Price |
|---|---|---|---|
| Computers for the community learning center | Funding 20 laptops including mice and keyboards, giving students a place to study | Culture and community | 27K |
| Laundry Access in Public Schools | Renovate a space in a Cambridge Public School and install washers and dryers for students who do not have easy access to laundry services at home, to use for their clothing and necessities | Culture and community | 50K |
| Real-Time Bus Arrival Monitors in bus stations | Real-time bus arrival monitors bus stops will inform travelers when the next bus will arrive, so they can adjust their plans if needed | Streets, Sidewalks and Transit | 24K |
| Sheltered Bike Parking at the Main Library | The Main Library needs more bicycle parking. A glass pavilion, protecting bikes from the weather, landscaped with paths and trees, will be an attractive and functional addition to the library grounds | Streets, Sidewalks and Transit | 90K |
| 24H public toilet | 24-hour access public toilet near Central Square | Environment, public health and safety | 320K |
| The Sustainable Energy Pilot | Install energy conversion devices on gym equipment and a rapid electric vehicle charging station | Environment, public health and safety | 90K |
| Dog Park | Building a dog park | Facilities, parks and recreation | 250K |
| Let's Rest: Picnic Tables and Benches for Our Parks | Benches and picnic tables bring our community together. Installing new benches and picnic tables in up to 10 of our park will allow people of all ages and abilities to enjoy them for resting, talking, reading, people watching and being outdoors | Facilities, parks and recreation | 120K |
| Installing Lights at the school Basketball Court | Install lighting to extend safe playing hours for basketball courts. Increases safety for community members while expanding healthy alternatives for youth and access to public space | Education | 250K |
| Security Cameras | Install security cameras in public schools | Education | 105K |

Table 1: First small election

.

| Project | Description | Category | Price |
|---|---|---|---|
| Interactive Technology for the Main Library | This project will fund an iPad lending kiosk and 16 iPads, as well as a permanent interactive screen in the Children's Room of the Main Library | Culture and community | 60K |
| Laundry Access in Public Schools | Renovate a space in a Cambridge Public School and install washers and dryers for students who do not have easy access to laundry services at home, to use for their clothing and necessities | Culture and community | 50K |
| We Need More Crosswalks | To enhance pedestrian safety, this project will add a minimum of five new crosswalks to major streets | Streets, Sidewalks and Transit | 40K |
| Hump Installationer for pedestrians | Speed humps create a safer environment by helping slow traffic on streets that students and families cross frequently. When a car hits a pedestrian at a high rate of speed, the collision is more likely to result in a pedestrian fatality. Speed humps slow vehicles and give drivers increased response time and distance for stopping. This makes streets safer for pedestrians | Streets, Sidewalks and Transit | 66K |
| Nursing Pod for Mothers and Infants | Provide an attractive private space where working mothers and community members can breastfeed or pump during the work day | Environment, public health and safety | 20K |
| Soak Up the Solar Power | Free, clean, renewable energy! Let's add solar panels to the Youth Center to reduce greenhouse gas emissions and save money on energy | Environment, public health and safety | 250K |
| Building an amphitheater in the public park | Build an amphitheater in the public park for outdoor performances, music, stories, and other cultural events that the whole community can enjoy | Facilities, parks and recreation | 350K |
| Remodel the Kitchen at the Youth Center | The kitchen area in the Youth Center is in dire need of renovating. Replace the stove, dishwasher, cabinets, and countertops in the Youth Center kitchen | Facilities, parks and recreation | 200K |
| Laptops for 10 Public Schools | Purchasing laptop carts for ten public schools | Education | 350K |
| Security Cameras | Install security cameras in public schools | Education | 105K |

Table 2: Second small election

.

| Project | Description | Category | Price |
|---|---|---|---|
| Little (book exchange) libraries | Installing 13 little free libraries across town | Culture and community | 13K |
| Computers for the community learning center | Funding 20 laptops including mice and keyboards, giving students a place to study | Culture and community | 27K |
| Digital Sign at City Hall in Multiple Languages | Digital sign that will scroll announcements in multiple languages and welcome people to town | Culture and community | 75K |
| Meeting Room Upgrade for libraries | Upgrades will allow for the latest in technology be available for public use in the meeting room | Culture and community | 250K |
| Separate Bike Lanes from Traffic | Improve safety for drivers and bikers by moving bike lanes to be between street parking spots and the sidewalk, reducing car-bike interactions and potential collisions | Streets, Sidewalks and Transit | 50K |
| Real-Time Bus Arrival Monitors in bus stations | Real-time bus arrival monitors bus stops will inform travelers when the next bus will arrive, so they can adjust their plans if needed | Streets, Sidewalks and Transit | 24K |
| Benches for a Walkable City | Install 12 benches across town so that people of all ages and abilities can enjoy benches for resting, talking, tinkering with electronic devices, people watching, and being outdoors | Streets, Sidewalks and Transit | 25K |
| Urban Bicycle Wash Stations | Bicycle owners need to clean and care for their bikes, ideally monthly. But for apartment dwellers, this is really hard! These centrally located bicycle wash stations would allow bicycle owners to wash off dirt, grime, and salt from their bikes | Streets, Sidewalks and Transit | 20K |
| Planting trees in the city | Street trees cool the city, absorb pollution, and make our neighborhoods more livable! planting 100 new trees and building tree wells in the areas that need them most | Environment, public health and safety | 119.4K |
| 24H public toilet | 24-hour access public toilet near Central Square | Environment, public health and safety | 320K |
| 5 Water Bottle Refill Stations | At a water bottle refill station you get a healthy drink for free | Environment, public health and safety | 40K |
| Fire Hydrant Markers | Install high-visibility markers on fire hydrants around town to increase safety for all residents and reduce response time of the fire department by improving ease of locating hydrants in emergencies, at night, and in the snow | Environment, public health and safety | 8K |
| Outdoor Fitness Equipment in the public park | Install outdoor body-weight fitness equipment for stretching, strength building, and plyometric exercises | Facilities, parks and recreation | 65K |
| Volleyball Court in the public park | Creating an outdoor volleyball court would be an exciting addition to the city. The court would have sand and a sturdy net for three-season usage | Facilities, parks and recreation | 61K |
| Inclusive Playground for All Kids | This Universal Design playground would include equipment that is designed to be usable by everyone without special adaptations or retrofitting | Facilities, parks and recreation | 305K |
| Protect the Health and Safety of our Firefighters | This proposal will purchase and install six gear drying units to shorten wait time for clean gear ($55,000), and eleven sets of wireless headsets to protect hearing and improve communication ($55,000). Let's protect those who protect us | Facilities, parks and recreation | 110K |
| New Chairs for Public Schools | New Chairs for Public Schools | Education | 190K |
| Invention and Production of Music | Install music studios and equipment at the Youth Centers to inspire creativity, enable pre-teens and teens to express their skills and passions, and provide youth with another recreational outlet | Education | 150K |
| Upgraded Water Fountains for Public Schools | Project would install 35 new water bottle refilling fountains in public schools | Education | 200K |

| | | | |
|---|---|---|---|
| New Playground for public school | Playground should include a jungle gym, and other equipment for kids to play different games | Education | 200K |

Table 3: First big election

| Project | Description | Category | Price |
|---|---|---|---|
| Little (book exchange) libraries | Installing 13 little free libraries across town | Culture and community | 13K |
| Computers for the community learning center | Funding 20 laptops including mice and keyboards, giving students a place to study | Culture and community | 27K |
| Meeting Room Upgrade for libraries | Upgrades will allow for the latest in technology be available for public use in the meeting room | Culture and community | 250K |
| Accessible Entrance for the Library | Add automatic sliding doors, fix driveway and, if needed, remove steps to benefit seniors and people with disabilities | Culture and community | 250K |
| Bike repair stations | Install 8 bike repair stations with tools and bike pumps across the city for cyclists to quickly, easily, and freely fix routine bike problems | Streets, Sidewalks and Transit | 12K |
| Separate Bike Lanes from Traffic | Improve safety for drivers and bikers by moving bike lanes to be between street parking spots and the sidewalk, reducing car-bike interactions and potential collisions | Streets, Sidewalks and Transit | 50K |
| Sheltered Bike Parking at the Main Library | The Main Library needs more bicycle parking. A glass pavilion, protecting bikes from the weather, landscaped with paths and trees, will be an attractive and functional addition to the library grounds | Streets, Sidewalks and Transit | 90K |
| What's Your Speed? | Remind drivers to slow down by deploying live speed displays on busiest streets | Streets, Sidewalks and Transit | 75K |
| 24H public toilet | 24-hour access public toilet near Central Square | Environment, public health and safety | 320K |
| Flashing Crosswalks for Safer Streets | This project would fund rapid flashing beacons at 10 high pedestrian risk crosswalks. These beacons increase the visibility of pedestrians, especially at night. They can alert drivers to crossing pedestrians, thereby preventing crashes | Environment, public health and safety | 176K |
| Soak Up the Solar Power | Free, clean, renewable energy! Let's add solar panels to the Youth Center to reduce greenhouse gas emissions and save money on energy | Environment, public health and safety | 250K |
| Fire Hydrant Markers | Install high-visibility markers on fire hydrants around town to increase safety for all residents and reduce response time of the fire department by improving ease of locating hydrants in emergencies, at night, and in the snow | Environment, public health and safety | 8K |
| Free Wifi in 6 Outdoor Public Spaces | Install special outdoor wifi access points to offer free public wifi in the public space | Facilities, parks and recreation | 42K |
| Inclusive Playground for All Kids | This Universal Design playground would include equipment that is designed to be usable by everyone without special adaptations or retrofitting | Facilities, parks and recreation | 305K |
| Shade and Wet Weather Canopies for Playgrounds | Installing canopies over playgrounds that do not have protection from the elements will reduce weather-related safety concerns and increase playground availability and use | Facilities, parks and recreation | 146K |
| Let's Rest: Picnic Tables and Benches for Our Parks | Benches and picnic tables bring our community together. Installing new benches and picnic tables in up to 10 of our park will allow people of all ages and abilities to enjoy them for resting, talking, reading, people watching and being outdoors | Facilities, parks and recreation | 120K |

| | | | |
|---|---|---|---|
| New Chairs for Public Schools | New Chairs for Public Schools | Education | 190K |
| Installing Lights at the school Basketball Court | Install lighting to extend safe playing hours for basketball courts. Increases safety for community members while expanding healthy alternatives for youth and access to public space | Education | 250K |
| Upgraded Water Fountains for Public Schools | Project would install 35 new water bottle refilling fountains in public schools | Education | 200K |
| Security Cameras | Install security cameras in public schools | Education | 105K |

.

Table 4: Second big election