

Rank, Pack, or Approve: Voting Methods in Participatory Budgeting

Lodewijk Gelauff, Ashish Goel

Stanford University
 {lodewijk, ashishg}@stanford.edu

Abstract

Participatory budgeting is a popular method to engage residents in budgeting decisions by local governments. The Stanford Participatory Budgeting platform is an online platform that has been used to engage residents in more than 150 budgeting processes. We present a data set with anonymized budget opinions from these processes with K-approval, K-ranking or knapsack primary ballots. For a subset of the voters, it includes paired votes with a different elicitation method in the same process. This presents a unique data set, as the voters, projects and setting are all related to real-world decisions that the voters have an actual interest in. With data from primary ballots we find that while ballot complexity (number of projects to choose from, number of projects to select and ballot length) is correlated with a higher median time spent by voters, it is not correlated with a higher abandonment rate. We use vote pairs with different voting methods to analyze the effect of voting methods on the cost of selected projects, more comprehensively than was previously possible. In most elections, voters selected significantly more expensive projects using K-approval than using knapsack, although we also find a small number of examples with a significant effect in the opposite direction. This effect happens at the aggregate level as well as for individual voters, and is influenced both by the implicit constraints of the voting method and the explicit constraints of the voting interface. Finally, we validate the use of K-ranking elicitation to offer a paper alternative for knapsack voting.

Introduction

Engaging residents/citizens in decision-making is a long-standing challenge (Langton 1979; Ebdon and Franklin 2006), resulting in stakeholder influence ranging from merely being informed to gaining direct control (Arnstein 1969; Fung 2006). In the United States, citizen participation has been mandated and/or commonplace for many decades (Callahan 2007b; Langton 1979) and many cities report some level of involvement by their citizens in the budgeting process (Wang 2001). Citizen participation in budgeting processes has however been criticized as often lacking depth (Callahan 2007a; van Dijk 2012). As society is embracing online technology, administrators are presented with an opportunity to better engage their residents.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Participatory Budgeting (PB) stands out as an empowering mechanism to engage residents and is globally increasingly adopted (Bartocci et al. 2022). PB is associated with many different definitions, which have in common that they allocate a budget across budget items (projects) with stakeholder (citizen) participation in that decision (Williams, Denny, and Bristow 2017; Miller, Hildreth, and Stewart 2019; Sintomer, Herzberg, and Röcke 2008). In many cases in the United States, this will include a proposal, deliberation, and voting phase (Aziz and Shah 2021).

Theoretical research into vote aggregation depends on distribution assumptions, analytical work depends on the availability of clean and realistic data sets and practitioners depend on findings from both. With this paper, we publish a dataset that we believe can be valuable for research into different elicitation and aggregation methods. While such data could be collected from crowdworker platforms on mimicked budgeting elections, this would require assuming that real voters deciding on real issues would behave in the same way as people who have no real stake.

We will focus our paper on the voting phase of Participatory Budgeting and especially focus on aspects that are relevant to PB organizers to design their ballot and make choices with regards to their voting method. We will specifically discuss some findings that we believe to be directly relevant to practitioners. While the data that we use in this paper is gathered on a fully functional platform, there is no reason why participatory budgeting can not be done via simple widgets on platforms such as X and Facebook. Some of the metrics that we study (such as completion time and completion rates) are perhaps even more relevant in these attention-poor social media settings.

Definitions

We will describe the setup of a voting process in a participatory budgeting *election* on the Stanford Participatory Budgeting platform. The election organizer determines the settings of the election including the voting method, available projects and their description, eligibility, and authentication. A voter $i \in \{1, \dots, N\}$ is authenticated, and presented with a *primary ballot* with M projects j ; typically defined by a title, description and cost c_j . The voter is invited to submit a *vote* with approval values for each project x_{ij} , subject to constraints imposed by the *elicitation method*. After submit-

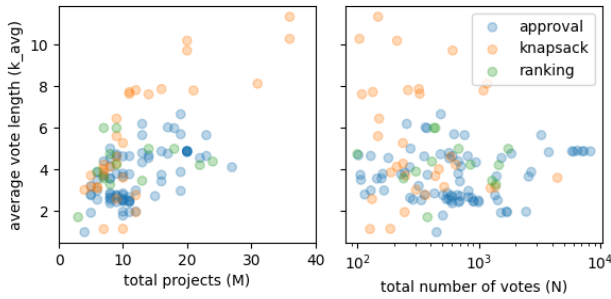


Figure 1: Elections in the dataset
(1 election out of range)

ting the primary ballot, a voter is sometimes given the possibility to submit a *secondary ballot* which is used for research purposes only. Finally, they are often asked to submit a *demographic survey* off-platform, which is not connected to their votes. The organizer can post-authenticate voters based on the authentication information they entered, and void any voters that are not meeting the criteria (e.g. residency in the city). An allocation is determined by the election organizer using an *aggregation method* based on the remaining primary votes.

While the ‘voting method’ is a single choice for the election organizer, it should be noted there are in fact three different choices combined: elicitation, consideration, and aggregation. Elicitation defines how votes are elicited from the voter, the consideration determines what information is exposed (and how prominently), and aggregation defines how the submitted votes result in an aggregate allocation. While consideration is a worthwhile endeavor to investigate in its own right, we will for simplicity consider it associated with the elicitation method.

On this platform, three main elicitation methods are available with a default aggregation method associated with it. The available elicitation methods are:

K-approval voting Each voter selects (approves of) up to K projects and submits a vote where $x_{ij} \in \{0, 1\}$ and $\sum_j x_{ij} = k_i \leq K$. This is the traditional method used in American PB elections and 1-approval is identical to many regular elections in the United States. The advantage of this method is that it is intuitive for the voter. The main downside is that the voter can’t easily trade-off projects of different cost: do they prefer one new playground of \$100,000 or rather upgrade two existing playgrounds for \$50,000 each?

K-ranking voting Each voter selects up to K projects and ranks them in order of value for money. The advantage is that this is still intuitive, voters are increasingly familiar with ranked methods (the description is often simplified to ‘preference’) and it allows voters to express more detailed preferences. The K is often set to a (much) lower value than the number of projects: a full ranking is considered burdensome.

Knapsack voting Each voter selects as many projects as they want, as long as the total cost of their selection does not exceed an available budget B (fits in their ‘knapsack’).

Their ballot is constrained by $\sum_j c_j x_{ij} \leq B$. An advantage of this method is that this is the most natural way for voters to make trade-offs between projects and consider their cost. They make the same choice as a policymaker would. Online, this is an easy task, but on paper we can not implement this the same way. We will later discuss how we could indirectly elicit a knapsack vote through K-ranking, which is easier on paper.

The platform also supports **pairwise comparison voting** (a sequence of project pairs is presented to the voter, and they are asked to indicate their preference between the two) and **K-token voting** (each voter gets K tokens to be distributed among the projects; an implementation of *cumulative voting*) but these methods are primarily research methods and currently not actively used in elections. We will not consider these methods in this overview.

Votes have to be aggregated to result in an allocation, and we will describe how this is implemented on the platform. For K-approval and knapsack elicited ballots, if the projects can only be selected integrally, the projects are ranked in decreasing order of received votes. The budget is distributed in that order. If projects can also be approved partially by a voter, projects are split into single-dollar projects and assigned one point for each voter that approved of them.

Ties can be resolved arbitrarily, and only matter if they concern the last project to be allocated when the budget runs out. How remaining funds are distributed, depends on the context of the election organizer. For example, we can assume a total budget of \$100 and three projects in an aggregate ranking: A (\$70), B (\$50), C (\$30). Partial allocation to the next project is theoretically straight forward and in our example we would allocate \$70 to project A and \$30 to project B. If partial funding is not an option, the organizer could choose to continue down the ranking until a project is found that can be fully funded with the remaining funds (\$70 for project A and \$30 for project C) or the organizer could decide to not spend the remaining funds, but use those to, for example, increase the budget for the next cycle (\$70 for project A and \$30 remaining funds).

For votes with ranking elicitation, it is less obvious what the appropriate aggregation method would be. A few families of aggregation methods for ranked ballots are discussed in this paper:

- **Inferred K-approval:** Assign score 1 to all projects that receive up to rank K , and score 0 to all other projects.
- **Borda count:** Assign score $M - k$ to a project with rank k (with $0.5(M - K_i - 1)$ score for unranked projects, where K_i is the number of ranked projects by voter i) (de Borda 1784). We also consider some variations on the traditional Borda count. *Borda-MK* does not assign any score to unranked projects, assuming that voters don’t have value for them. *Borda-K1* assigns score K_i to the top-ranked project, score 1 to the last-ranked project and no score for unranked projects. For each of these variations, we can replace K_i with K , assuming that each voter fills out their complete ballot. The aggregation code for our platform uses a full Borda count with K at the election level.

- **Inferred Knapsack:** Sequentially assign budget to a project in order of the rank for that voter until the budget is exhausted (Goel et al. 2019). Either use the remaining budget to partially fund the next project (ranking-partial), or skip a project that cannot be funded in favor of a possibly less expensive project that is ranked lower by the voter (ranking-skip). This choice depends on the assumption whether projects can be partially funded or not.

This will then result in an aggregate list of the projects in decreasing order of received votes and budgets are then assigned as above.

Related Work

Participatory Budgeting has a rich literature with case studies and evaluations both by academics and practitioners, offering insights into its potential and effects (e.g. (Su 2017; Falanga and Lüchmann 2020)). This literature is primarily discussed in two fields: in the Public Administration and Political Sciences, PB is approached as a democratic innovation or administrative process to engage stakeholders (citizens/residents), encompassing idea-collection, project shaping, voting and implementation (Bartocci et al. 2022; Hagelskamp et al. 2016). In the social choice literature, PB serves as an important use case for aggregating budgeting opinions (Aziz and Shah 2021; Rey and Maly 2023).

PB spread across the globe as a democratic innovation since it was introduced in 1988 in Porto Alegre, in many different flavors (Sintomer, Herzberg, and Röcke 2008), with a comprehensive survey of its history in (Bartocci et al. 2022). For North America, where most of our data originates, a comparative study of PB evaluations is available in (Hagelskamp et al. 2016). The goals of PB vary widely and in the US context they can explicitly include equity (Lerner and Pape 2020). Engagement with PB has been found to increase political participation of traditionally underrepresented demographics in New York (Johnson, Carlson, and Reynolds 2021) and engagement in budget feedback exercises can enhance understanding of budgetary issues (Kim et al. 2016). As PB increasingly moves online, various tools including the Stanford Participatory Budgeting platform¹ (Goel et al. 2019), Consul² (Arana-Catania et al. 2021; Pina et al. 2022), AppCivist-PB³ (Holston, Issarny, and Parra 2016) and Decidim (Serramia et al. 2019) become available. It would go beyond the scope of this paper to attempt to make a complete inventory or make an extensive comparison, but some tools have been compared in (Society 2016; Cantador and Cortés-Cediel 2018).

Knapsack voting and PB have gained interest in the algorithms and social choice community over the last decade, with PB as an extension to (continuous or discrete) multi-winner elections. Detailed surveys into voting methods are available in (Rey and Maly 2023; Aziz and Shah 2021). Beyond voting methods, there has been an interest in how votes can be aggregated once elicited, for example through

a method that mirrors more deliberative approaches (Goyal et al. 2023), considering fairness measures such as the core (Fain, Goel, and Munagala 2016; Munagala, Shen, and Wang 2022), and justified representation such as the Method of Equal Shares (Peters, Pierczyński, and Skowron 2021).

Vote elicitation in a budgeting context has been explored with theoretical, experimental, and data-driven approaches. In (Benadè et al. 2017) 4 elicitation methods are discussed: K-approval, K-ranking (in two variants), and knapsack voting, concluding among others that knapsack voting might be too burdensome for voters and does not have as many theoretical guarantees. However, (Goel et al. 2019) suggests that knapsack voting was less burdensome in practice and desirable strategic properties. In (Garg et al. 2019) learning rates were proposed to identify optimal ballot design, concluding that the threshold K was historically set too low in K -approval voting for optimal learning. A study into the effect of K-approval, threshold-approval, K-ranking (in two variations), knapsack and K-token voting with recruited crowdworkers evaluated the effect on cognitive load and voters' ability to recall their stated preferences (Fairstein, Benade, and Gal 2023), suggesting Value-for-Money ranking as cognitively the hardest on voters. In contrast to their work, our findings are based on real world PB processes.

Unfortunately, voter preference data from PB elections is scarce, and empirical work spanning multiple elections is limited. A notable exception is the Pabulib, an open library with vote sets from over 800 PB elections, mostly from Poland (Stolicki, Szufa, and Talmon 2020) with mostly data from K -approval ballots and K -ranking ballots with low K . We believe that our data will be a valuable addition, as we report the time spent on the ballot by the voter, have different voting methods available and have votes available from secondary voting methods. This provides both vote pairs (two votes from the same voter on the same projects with different voting methods) and ballot pairs (where voters were randomly assigned to a voting method for their secondary ballot).

Platform and Data Collection

The Stanford Participatory Budgeting platform has been previously described in (Goel et al. 2019) and is one of several platforms that cities in North America have available to organize an online voting phase in their PB election. The software is available under a free license on GitHub and could be installed by an election organizer on their own servers. One instance of the platform can support many ballots, with admin permissions at the election level. This way the Crowdsourced Democracy Team has been able to provide the system to multiple organizers, giving each the possibility to manage their own election and configure their ballot.

Our data set only contains data from elections that were hosted on the instance that is provided by the Crowdsourced Democracy Team. A ballot instance is provided at no cost, and the organizer can then set up the ballot in their desired configuration. Some of the available configuration options are:

- Voting method: K-approval, K-ranking, knapsack or K-

¹<https://pbstanford.org>

²<https://consulproject.org>

³<https://pb.appcivist.org>

token

- Language: Currently, the ballot can be made available in: English, Amharic, Arabic, Bengali, Chinese, Finnish, French, German, Haitian Creole, Hindi, Hmong Daw, Khmer, Polish, Portuguese, Spanish, Tagalog. Select a subset of the languages, with one language as default. The voter can then vote in any of the selected languages.
- Voter validation: SMS confirmation, personal information, generated codes, free-form (no verification).
- Voter registration: Questions that the voter has to answer and that can be used for post-voting verification
- Voting phases: voting phases that can be presented
- Available budget
- Voting constraints (value of K), as appropriate for the voting method
- Project appearance: whether to show the cost, project numbers and maps
- Survey link for demographic survey off-platform

An election is always configured with a primary voting method (the official ballot) and the authors have in some elections (when approved by the organizer) also set up a secondary voting method (the research ballot). If a secondary voting method is set up, the voter is first presented with the primary ballot, before being presented with a consent text and a secondary voting method (with the same projects). The secondary method can be set up as an array of voting methods, in which case the voter is randomly presented with one of the voting methods in the array.

The organizer also configures the content of several city-specific fields, including the landing page text, the contact information, in-person voting date, and location. The organizer also enters the voting dates, projects (title, description, cost (discrete or with cost steps), and optionally location, category, image, video URL, project details, and coordinates.

Data

We present a data set of anonymized ballots from the Stanford Participatory Budgeting platform described above.⁴ Since 2014 we have partnered on 150+ ballots with mostly local governments and foundations in the United States, who used the platform to organize a Participatory Budgeting election. The organizer (partner organization) could configure their election ballot to their own specifications, with their own projects and in the languages that they desired. The organizer was also in charge of recruitment among their stakeholder population and implementing the results.

From the voter's perspective, they arrive on a landing page, authenticate, access the primary ballot, and optionally access a secondary ballot and/or a survey. The voter could exit at any point and may have the option to skip a phase.

⁴The data set is available through the Stanford Digital Repository: <https://doi.org/10.25740/db709zg9088> (Gelauff and Goel 2024)

We do not have access to the information from the demographic survey that was provided by the city and would not be able to connect this back to the voting data.

For this data set, we provide tables *elections* (and a version enriched with summary statistics), *projects*, *voters*, *vote_approvals*, *vote_knapsacks*, *vote_rankings*, and *vote_tokens* as of 22 May 2023. We have added some additional columns and tables with derived data for easier analysis. Elections, voters, projects and votes have been cleaned to ensure quality, and data was removed when that was necessary for anonymization. We describe and justify these steps in the data set documentation.

We present data from 124 primary ballots (79 approval, 32 knapsack, 13 ranking) and 38 secondary ballots (2 approval, 18 knapsack, and 18 ranking) with at least 80 votes per election. Some elections have data from two secondary ballots. Aggregations of this data may be different from published results, due to more aggressive data cleaning for privacy and data quality purposes, as the purpose of this data set is to understand voting behavior. The following tables are included:

- elections: election ID, election name and available budget, text on the ballot, remarks
- elections_rich: the same as elections, but with additional details about ballot settings and summary statistics
- projects: project ID, election ID, category ID, ordering, cost and geo coordinates (when available)
- voters: voter ID, election ID, authentication method, last stage entered, last day updated
- votes: voter ID, election ID, project ID, allocated budget, rank (ranking only), tokens (token only), last day updated, fraction of voters that completed the vote in less time
- inferred votes: inferred knapsack votes when a ranking vote would be used to fill out a knapsack ballot.
- voter utility stats: statistics describing the utility overlap between vote pairs

The voter ID has been altered and the timestamp information is rounded or removed for anonymization purposes. In the data set documentation ("the documentation") we describe the tables and the data cleaning process in more detail.

In Figure 1 we visualize the primary ballots in our dataset, with the number of projects that voters were able to choose from (M), the number of voters that cast a vote (N) and the average number of selected projects on that ballot (k_{avg}).

The data set may prove useful as a resource for realistic voting data under different budget-related voting methods and to understand vote distributions under different conditions. Unique aspects of this data are the relative amount of time spent (as quantile of the population) and the fact that vote pairs under different voting methods from the same voter on the same election are present. Users of the data should be aware that the voters usually only represent a small portion of the eligible population, although the exact size of this portion has not been estimated by the authors. Recruitment methods and other factors can therefore have a significant effect on composition of the set of voters (Gelauff et al. 2020).

Comparing Vote Pairs

We will establish that secondary ballots meaningfully describe the opinion that was expressed by the same voter on the primary ballot. This makes the comparison of vote pairs possible, but also allows us to consider secondary ballots in the first place, which is especially helpful as they contain K -ranking ballots with a relatively high K . Next, we will establish that K-approval and knapsack ballots can be inferred from sufficiently long K-ranking ballots.

We compare the primary and secondary votes to better understand the voting behavior with different voting methods at a granular level. We use inferred knapsack and K -approval votes to compare different aggregation methods, and how their constraints affect the individual choices of a voter and the eventual allocation after aggregation.

Secondary Ballots

We evaluate how similar the primary ballot of voter i (x_i) is to their secondary ballot (y_i) compared to the secondary ballot of all voters i' ($y_{i'}$). The budget overlap utility of a voter is the overlap between a vote and some other set of projects (e.g. the final allocation) in dollars. We calculate the budget overlap utility between any possible vote pair (irrespective of the voting method used) by finding for each project the minimum cost approved between the two votes, and taking the sum. We rank the utility u_{ii} among all $u_{ii'}$ and find its percentile (inclusively: proportion of utilities that are equal or lower) and z-score (the number of standard deviations that u_{ii} is from the average $u_{ii'}$). We calculate for each election the median percentile and median z-score, and report for each combination of voting methods the median, mean, and minimum across all per-election scores as well as the median, mean, and standard deviation per-voter across elections. Finally, we do the same for pairs of knapsack votes and inferred knapsack votes from K-ranking votes (Table 1).

We observe that the overlap percentile, as well as per-election median percentile, is in most elections 1 (perfect overlap between primary and secondary ballots). For each combination, the per-election median percentile is 0.88 or higher. We also observe that the average percentile is 0.88 or higher for each ballot combination. We observe that the median z-score is in most elections 1 or higher (where a positive z-score means higher than average overlap), and the mean z-score is around 1 for each voting method combination.

Because these are different voting methods, we would expect some differences to appear between different ballots of the same person – even when the voters have perfect memory and attention span. These numbers indicate that there is a signal in the secondary ballots and that they are not randomly entered, and in subsequent analysis we found no counter indications that there is a substantive difference between primary and secondary votes. Going forward, we will for brevity sake consider primary and secondary ballots equally, unless otherwise mentioned.

Different Aggregation Methods for K-Ranking Ballots

In our dataset, we have 31 elections with ranking ballots. With some natural assumptions we can use ranking votes to infer how voters would have responded to different voting methods, if they would have followed that ranking. This allows us to explore the effect of the implicit constraints that a voting method imposes on a voter, as this approach excludes any effect from the information available on the ballot or strategic behavior.

For K-approval, we make the assumption that if the voter had been asked to select $K' \leq K$ projects, they would have submitted the K' highest-ranked projects on their K-ranking ballot. For knapsack aggregation, we assume that the voter would have gone down their ranking, adding projects one at a time as the budget constraint permitted. We extract this way what we will call inferred approval and knapsack votes, and then aggregate them as described above for K-approval and knapsack aggregation. While this is a useful approximation, this assumption might not always hold: a voter might for example want to consider interaction effects between projects, or the available ranking might not be long enough.

For each of these elections with ranking ballots, we aggregate the ballots with K-approval ($K = 1, \dots, 5$), 6 variations of Borda count, knapsack-partial and knapsack-skip aggregation methods. For each aggregation method, we create a vector of the portion of the available budget that was allocated to each project with that aggregation method and calculate the cosine similarity between these vectors (see Table 2). Cosine similarity scores range between 0 and 1, where a score of 0 indicates that the vectors are orthogonal, and the larger the score is, the smaller the angle between the vectors.

This is a crude comparison but effectively shows that across all available elections, the aggregation differences between the different members of the Borda family are small with a cosine similarity of 0.92 or higher. We observe that the allocations can be rather different for different choices of K under K-approval voting.

Knapsack Aggregation from Ranked Ballots

Knapsack voting is particularly hard to achieve on paper, where it would be unrealistic to expect voters to constantly calculate whether their selection of projects fits within the allowed budget. In previous work (Goel et al. 2019) it was argued that organizers could effectively provide a paper knapsack ballot, by eliciting a ranked ballot with a large K and infer the knapsack ballot for the voters. The cognitive load that input formats would place on voters has been raised as a concern and it was suggested that experiments would be needed (Benadè et al. 2017). With our data we can attempt to answer this concern.

We have 8 elections with a knapsack and K-ranking ballot pair for the same voters. We calculate the utility overlap between these votes and report the same statistics as before (Table 1). We observe that the percentile scores for utility overlap between the knapsack ballot and the inferred knapsack ballot are high, and that using the ‘skip’ method results

Prim.	Sec.	n	per-election median percentile			per-voter percentile			per-election median z-score			per-voter z-score		
			med	mean	min	med	mean	std	med	mean	min	med	mean	std
app	knap	16	1.00	0.99	0.90	1.00	0.92	0.17	1.37	1.44	1.07	1.41	1.44	0.99
app	rank	11	1.00	1.00	0.98	1.00	0.93	0.17	1.11	1.25	0.80	1.21	1.16	0.80
app	tok	1	1.00	1.00	1.00	1.00	0.93	0.16	1.25	1.25	1.25	1.25	1.00	0.69
knap	app	1	0.88	0.88	0.88	0.88	0.76	0.27	1.12	1.12	1.12	1.12	0.93	1.15
knap	rank	7	1.00	0.97	0.89	0.99	0.82	0.25	1.06	1.11	0.68	0.95	0.88	0.84
rank	knap	1	1.00	1.00	1.00	1.00	0.92	0.16	2.25	2.25	2.25	2.25	2.11	1.25
all	all	26	1.00	0.99	0.88	1.00	0.91	0.19	1.26	1.30	0.68	1.28	1.31	0.98
knap	i-kn _p	7	0.89	0.91	0.81	0.92	0.82	0.24	1.28	1.26	0.93	1.27	1.09	1.02
knap	i-kn _s	7	0.97	0.96	0.88	0.96	0.82	0.24	1.30	1.39	1.22	1.36	1.11	1.08
i-kn _p	knap	1	0.99	0.99	0.99	0.99	0.92	0.16	2.27	2.27	2.27	2.27	2.12	1.26
i-kn _s	knap	1	0.99	0.99	0.99	0.99	0.92	0.16	2.26	2.26	2.26	2.26	2.09	1.28

Table 1: Utility overlap statistics of primary (prim) / secondary (sec) ballot pairs: K-approval (app), knapsack (knap), K-token (tok), K-ranking (rank), inferred knapsack with partial (i-kn_p) and skip (i-kn_s) allocation. Per-election statistics report the median, mean, and minimum of the per-election median utility overlaps. Per-voter statistics give the median, mean, and standard deviation of all individual utility overlaps. n is the number of elections for that ballot pair.

	app ₁	app ₂	app ₃	app ₄	app ₅	full _v	full _k	MK	K1 _v	K1 _k	kn _p	kn _s
app ₁	1.00	0.91	0.84	0.79	0.76	0.81	0.83	0.83	0.78	0.84	0.83	0.71
app ₂	0.91	1.00	0.93	0.88	0.84	0.90	0.91	0.92	0.87	0.93	0.89	0.79
app ₃	0.84	0.93	1.00	0.93	0.88	0.97	0.98	0.98	0.91	1.00	0.90	0.85
app ₄	0.79	0.88	0.93	1.00	0.93	0.96	0.94	0.94	0.93	0.93	0.87	0.84
app ₅	0.76	0.84	0.88	0.93	1.00	0.91	0.90	0.90	0.89	0.88	0.87	0.81
full _v	0.81	0.90	0.97	0.96	0.91	1.00	0.98	0.98	0.91	0.97	0.90	0.88
full _k	0.83	0.91	0.98	0.94	0.90	0.98	1.00	1.00	0.90	0.98	0.89	0.86
MK	0.83	0.92	0.98	0.94	0.90	0.98	1.00	1.00	0.90	0.98	0.90	0.87
K1 _v	0.78	0.87	0.91	0.93	0.89	0.91	0.90	0.90	1.00	0.92	0.85	0.85
K1 _k	0.84	0.93	1.00	0.93	0.88	0.97	0.98	0.98	0.92	1.00	0.91	0.85
kn _p	0.83	0.89	0.90	0.87	0.87	0.90	0.89	0.90	0.85	0.91	1.00	0.81
kn _s	0.71	0.79	0.85	0.84	0.81	0.88	0.86	0.87	0.85	0.85	0.81	1.00

Table 2: Cosine similarity between vectors of projects with the proportion of the budget allocated under different aggregation methods, inferred from identical K-ranking ballots: K-approval (app_K), full Borda ranking, Borda-MK, Borda-K1 (with voter-_v) and election-specific K 's (_k), knapsack with partial (kn_p) and skip-allocation (kn_s).

in inferred votes closer to the knapsack votes. This is expected, as this is in line with what a voter could do in practice (partial allocation is often not allowed) and going forward we will use the ‘skip’ method unless otherwise specified. The z-scores are on average higher than 1, indicating that inferring a knapsack ballot from a ranking ballot is indeed a valid way of achieving a knapsack ballot when online methods are not available. However, while this is a good substitute, it is not identical: we will discuss below the effect that the interface (knapsack compared to ranking) can have on the selected projects.

Project Cost

Voters may consider project costs differently when confronted with explicit budget constraints. Goel et al. (Goel et al. 2019) found that K-approval voting favored more expensive projects in aggregate vote distributions, resulting in a higher average cost for winning projects compared to knapsack voting. This aligns with the notion that voters

would be more frugal when they encounter explicit budget constraints. This analysis focused on aggregate data and did not distinguish between explicit constraints considered by the voter, and implicit constraints imposed by the voting method. Our data allows for a more nuanced exploration.

We aim to dissect the effect of the voting interface and of the voting method. Using elections with ballot pairs, we compare the cost of selected projects using votes under different voting methods by the same voter. We isolate the effect of the user interface and provided information on the ballot by comparing knapsack votes with inferred knapsack votes from the same voter. Similarly, we isolate the impact of the voting method itself by inferring votes under different voting methods from the same K-ranking ballots.

We report our cost statistics consistently per ballot pair (e.g. knapsack and inferred knapsack). We calculate the cost statistics (most expensive selected project, average cost of the 3 most expensive selected projects and average cost of all selected projects) for each vote and report for each ballot

id	n	top-1	top-3	avg	Prim.
4	146	0.004	-0.016 **	0.001	app
5	141	-0.006	-0.004	0.016 **	app
14	162	0.009	0.004	0.011 *	app
17	940	0.024 **	0.039 **	0.029 **	app
103	231	0.013	0.067 **	0.097 **	app
119	175	0.033 **	0.103 **	0.132 **	app
121	149	0.005	0.064 **	0.073 **	app
122	220	0.007	0.098 **	0.141 **	app
126	280	-0.029 *	-0.0	0.035 **	app
128	234	-0.029 **	-0.027 **	-0.018 *	app
130	517	0.01 **	0.006 **	0.01 **	app
152	204	0.09 **	0.062 **	0.116 **	knap
170	267	-0.051 **	-0.034 **	-0.003	app
172	405	-0.006	-0.016 **	0.033 **	app
173	246	-0.038 **	-0.023 **	-0.018 **	app
174	447	-0.006 **	0.005 **	0.005 **	app
255	472	0.063 **	0.033 **	0.033 **	app
avg	17	0.004	0.021	0.041	

Table 3: Average cost difference between pairs of approval and knapsack votes from the same voter. A positive score implies that the statistic for K-approval was that portion of the election budget higher. The bottom row shows the election count and average election statistics.

ballot 1	ballot 2	n	ballot 1 > ballot 2			ballot 1 < ballot 2		
			top-1	top-3	avg	top-1	top-3	avg
app	knap	17	10	9	16	7	8	1
knap	i-kn _s	8	4	7	1	2	1	7
app	i-app	11	10	8	6	0	1	3
i-app ₄	i-kn _s	31	19	15	23	7	11	2

Table 4: Per ballot pair, the number of elections with more voters that had a higher average cost statistic for method 1 than for method 2, and vice versa. Inferred ballots are indicated with ‘i’.

pair per election the average difference of cost statistics, normalized by the budget available in that election. Significance is established through a percentile-bootstrap (1000 simulations) per election and a two-sided test at 95% (*) and 99% (**) confidence intervals, recognizing that the statistics may not follow a normal distribution.

To understand whether this effect is dominated by a small number of voters that made large shifts, we calculated for each election how many voters had a higher cost statistic for the first (K-approval) than for the second method (knapsack) and counted the number of elections where these voters formed the majority. We did this for each ballot pair comparison in this paper, and reported the numbers in Table 4.

K-Approval and Knapsack Voting

We consider 17 elections with sufficient voters who submitted both an approval and a knapsack vote and consider the cost statistics for their selected projects (see Table 3). Except for one election, the primary voting method was K-approval and the secondary method was knapsack.

id	n	top-1	top-3	avg	Prim.
93	740	0.003 **	0.004 **	-0.014 **	rank
152	226	-0.012	-0.005	-0.09 **	knap
171	91	0.011	0.012	0.016	knap
194	137	0.007	-0.001	-0.019 **	knap
248	106	-0.002	0.0	-0.015	knap
249	352	0.005	0.005	-0.006 *	knap
250	408	0.006	-0.006	-0.022 **	knap
251	127	0.013	0.013 *	-0.014 *	knap
avg	8	0.002	0.002	-0.020	

Table 5: Average cost difference between pairs of knapsack votes and inferred knapsack (skip) votes. A positive score implies that the statistic for knapsack was higher.

In our per-voter analysis we validate Goel et al.’s finding: in most elections, the K-approval ballot resulted in a significantly higher average cost than the knapsack method. However, there are two instances where the knapsack ballot has a significantly higher average cost for selected projects. This dominant effect can be explained by the explicit constraints presented in the knapsack voting interface. As voters select their preferred projects, they will have some budget left over, which they can only allocate to a low-cost project. Under K-approval voting, no such pressure exists. This pressure is however not unnatural: it mirrors constraints faced by decision-makers, and can even be argued to be desirable.

This effect is much smaller when we focus on more expensive projects that the voter selected (top-1 and top-3 average costs). In several elections the shift even reverses direction: the voter selected a more expensive project with knapsack than with K-approval. This suggests that the magnitude and direction of this effect depend on the specifics of the election, especially concerning more expensive projects. This trend is even more pronounced in Table 4: in almost every election, voters with more expensive projects on their K-approval ballot were in the majority – but this is not true when we only look at their most expensive selected projects.

These observations align with the explanation that voters are compelled to be more budget-conscious due to budget constraints after selecting their most expensive preferences. If they would choose only one instead of two of the most expensive projects under knapsack constraints, we would expect to see a strong effect in the top-3 statistics (the average number of selected projects K_i is usually below 5).

Effect of the Voting Interface

In order to isolate the effect of the voting interface and information provided to the voter, we compare knapsack votes (from the knapsack interface) with inferred knapsack votes (from the K-ranking interface). This comparison excludes any effects from implicit constraints posed by the voting method. We report the cost statistics in Tables 5 and 4.

If explicit constraints and the information provided to the voter were responsible for the observed cost differences between K-approval and knapsack, we would expect a similar shift here. Indeed, we observe a consistent trend for the

id	n	top-1	top-3	avg	Prim.
103	247	0.016	0.014	0.014	app
118	118	0.026 **	0.015 **	0.015 **	app
119	152	0.011	0.017 *	0.017 *	app
121	151	0.003	0.007	0.007	app
122	207	0.012	0.016	0.016	app
126	312	0.026 *	0.012 *	-0.014 **	app
130	417	-0.001	-0.001	-0.001	app
172	489	0.023 **	0.025 **	-0.003	app
174	491	0.003	-0.001	-0.001	app
255	525	0.001	-0.003	-0.003	app
267	124	0.015	0.011	0.011	app
avg	11	0.012	0.010	0.005	

Table 6: Average cost difference between pairs of K-approval votes and inferred K-approval votes. A positive score implies that the statistic for K-approval was higher.

overall average cost: the average cost is higher for inferred knapsack votes than for knapsack votes. However, the top-3 statistics in Table 4 give a more nuanced picture: almost all elections had a majority of voters select more expensive top-3 projects with explicit knapsack constraints than without. This effect is opposite of what we would expect.

This suggests that while voters may exhibit greater cost-consciousness when confronted with explicit budget constraints, this effect primarily manifests among less expensive projects on the ballot. Among more expensive projects, we do not observe significantly different average project costs in most elections. In cases where the average top-3 project cost was higher on the knapsack ballot than on the inferred knapsack ballot, it could be hypothesized that voters prefer to ‘upgrade’ more expensive projects while ‘downgrading’ or even splitting less expensive ones.

We can similarly compare K-approval votes with inferred K-approval votes to isolate any effect the question to rank the projects may have. To ensure individual-level comparability, we truncated the K-ranking ballots to match the number of projects selected by the voter at their equivalent K-approval ballot (K_i) to obtain the inferred K-approval ballot. We present the cost statistics in Tables 6 and 4.

We would have expected this effect to be small given the visual similarities in the interface: the K-ranking interface is essentially the same as K-approval, followed by an additional step where the voter is asked for their preferred order among the selected projects. Any effect would be attributed to the additional consideration step: this suggests that the step of asking voters to rank their project results in a different cost trade-off than immediately asking them for a smaller selection of projects.

Implicit Constraints from Voting Methods

In order to isolate the effect of implicit constraints imposed by a voting method on the cost of selected projects, we compare the effect of a number of voting methods on the aggregated allocation. We report the average cost statistics of different ranking aggregation methods across elections in Table 7, as ratio of the knapsack-skip aggregation. This shows

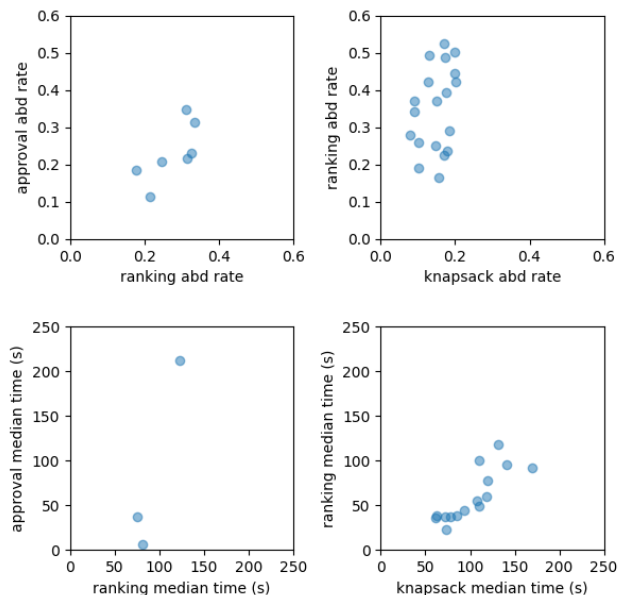


Figure 2: Abandonment (abd) rate and median time for secondary K-approval, K-ranking and knapsack ballots at the same election. Voters were randomly assigned to either secondary voting method.

very clearly that across elections, the standard knapsack aggregation results in lower-cost projects to be selected in aggregate, in line with the findings of Goel et al.

To compare the effect of implicit constraints of the elicitation method (excluding the effect of aggregation), we infer both 4-approval and knapsack ballots from the same ranking votes, and report the cost statistics of this ballot pair in Tables 8 and 4. We observe that the average cost is significantly higher in nearly all elections under 4-approval constraints than under knapsack constraints. This picture is more nuanced when we limit ourselves to the most expensive selected projects per voter, with significantly higher statistics for both ballots. This demonstrates that even if the projects are entirely independent, and if voters would exactly follow the same ranking to inform their approval and knapsack ballots, we would observe differences in the cost statistics.

Abandonment Rate and Median Time Spent

Election organizers indicated in private communications that they choose the K-approval voting method for its perceived ease of use for voters. We gauge voter usability by analyzing the abandonment rate and median time spent during ballot completion. We present in Table 9 the correlation of the main ballot design choices (number of projects to select K , number of projects on the ballot M , text on the ballot, budget available, voting method) with each other and the abandonment rate and median time spent for 121 ballots. In Table 10 we present the results of multiple-linear regressions for both

	app ₁	app ₂	app ₃	app ₄	app ₅	full _v	full _k	MK	K1 _v	K1 _k	kn _p	kn _s
top-1	1.22	1.20	1.19	1.19	1.17	1.16	1.16	1.16	1.13	1.19	1.14	1.0
top-3	1.11	1.11	1.09	1.08	1.08	1.07	1.07	1.07	1.07	1.09	1.11	1.0
avg	1.03	1.04	1.02	1.03	1.03	1.02	1.02	1.02	1.03	1.02	1.07	1.0

Table 7: Average cost ratio of different aggregation methods applied to the same ballot. The ratio is the value divided by the same value for the knapsack (skip) aggregation when considering all selected projects, the 3 most expensive selected projects or the most expensive project. The column headers refer to the various aggregation methods.

id	n	top-1	top-3	avg	Prim.
16	681	-0.011 **	-0.012 **	0.0	rank
93	1460	-0.002 **	-0.007 **	0.001	rank
103	247	0.021 **	0.07 **	0.041 **	app
106	100	-0.008 **	-0.011 **	0.005 **	rank
118	119	0.077 **	0.064 **	0.034 **	app
119	155	0.043 **	0.108 **	0.073 **	app
121	151	0.022 **	0.062 **	0.032 **	app
122	208	0.017 **	0.068 **	0.049 **	app
126	314	0.028 **	0.016 **	0.008 **	app
130	417	0.01 **	0.017 **	-0.004 **	app
152	228	0.073 **	0.065 **	0.02 **	knap
171	103	0.0	0.0	0.0	knap
172	491	0.0	0.001 **	-0.0	app
174	492	0.019 **	0.039 **	0.013 **	app
186	240	0.0	0.0	0.0	rank
194	139	-0.028 *	0.026 **	0.037 **	knap
214	1747	-0.008 **	-0.005 **	0.004 **	rank
225	440	0.0	0.0	0.0	rank
242	378	0.107 **	-0.011	-0.011	rank
243	304	0.056 **	0.098 **	0.042 **	rank
244	1383	0.007 **	-0.0	0.01 **	rank
245	1285	-0.012 **	-0.01 **	0.0	rank
247	854	-0.005 **	-0.01 **	0.0	rank
248	109	0.007	0.041 **	0.029 **	knap
249	356	0.003	-0.007 **	0.007 **	knap
250	411	-0.017 **	0.064 **	0.059 **	knap
251	128	0.044 **	0.046 **	0.031 **	knap
255	525	0.135 **	0.055 **	0.016 **	app
266	411	-0.031 **	-0.01 **	0.001	rank
267	126	0.0	0.0	0.0	app
270	425	0.0	0.0	0.0	rank
avg	29	0.017	0.024	0.016	

Table 8: Average cost difference between pairs of inferred 4-approval votes and inferred knapsack votes. A positive score implies that the statistic for inferred 4-approval was higher.

the median time spent and the abandonment rate⁵ on the ballot design choices.

As anticipated, choices that affect ballot complexity are all significantly predictive for a higher median time spent, but do not predict a higher abandonment rate. The knapsack voting method is the only choice that significantly predicts a higher abandonment rate. However, as it is not possible to have no voting method at all, we should note that the confidence intervals for the different voting methods do overlap.

⁵Abandonment rate is only available in aggregate. Time spent is anonymized at the per-voter level to a percentile score.

Therefore, we can confidently conclude that voters in elections with knapsack voting spend more time and are a little more likely to exit the ballot than voters in elections that use the approval method. However, we can not assert the same when we compare voters in elections using the knapsack and ranking methods – the more relevant comparison, as the knapsack ballot can be inferred from ranking.

These findings only hint at associations, and causality cannot be inferred. For instance, cities choosing knapsack voting might employ a more innovative or extensive recruitment than those choosing approval voting, leading to higher dropout rates. To establish causality, randomized controlled trials would be needed. For some elections, we have effectively established exactly that, by randomly assigning a secondary voting method to voters. Unfortunately, the number of elections where we have sufficient votes available in two randomly assigned voting methods, is limited. In Figure 2 we present the randomized controlled data that we have available in our data set for the approval-ranking and the ranking-knapsack ballot combinations. These data points represent secondary ballots where voters were (after completing the same primary ballot) randomly assigned to different secondary ballots (e.g. after a primary approval ballot, they were assigned to either a knapsack or a 5-ranking secondary ballot). This data shows that the abandonment rate for knapsack ballots is actually lower than K-ranking, even though the median time spent is higher.

This effect could be partly attributed to the fact that the K-ranking and K-approval interfaces are so similar, and that a K-ranking secondary ballot always followed a K-approval primary ballot, and vice versa. This could reinforce a higher median time for knapsack (more familiarity between ranking and approval) and higher abandonment rate for ranking and approval due to voter confusion. Consequently, these findings cannot be readily generalized to primary voting methods without further experimentation.

In summary, the question whether knapsack voting results in a higher abandonment rate remains inconclusive, and further research is warranted.

Discussions and Conclusions

With this analysis, we have uncovered valuable insights that can aid PB election organizers. By examining ballot pairs from the same voters and leveraging long K-ranking votes, we made the following observations regarding voting methods and their effect on the average cost of selected projects:

1. Secondary ballots serve as effective tools for comparing voting methods.

	K	M	text on ballot	budget	approval	knapsack	ranking	abd rate	median time
K	1.00	0.63	0.49	0.43	-0.48	0.49	0.05	0.05	0.58
M	0.63	1.00	0.58	0.27	-0.01	0.05	-0.06	-0.15	0.56
text on ballot	0.49	0.58	1.00	0.13	-0.16	0.05	0.19	-0.00	0.48
budget	0.43	0.27	0.13	1.00	-0.07	0.07	0.01	-0.09	0.22
approval method	-0.48	-0.01	-0.16	-0.07	1.00	-0.79	-0.44	-0.35	-0.18
knapsack method	0.49	0.05	0.05	0.07	-0.79	1.00	-0.20	0.36	0.27
ranking method	0.05	-0.06	0.19	0.01	-0.44	-0.20	1.00	0.04	-0.11
abd rate	0.05	-0.15	-0.00	-0.09	-0.35	0.36	0.04	1.00	-0.02
median time	0.58	0.56	0.48	0.22	-0.18	0.27	-0.11	-0.02	1.00

Table 9: Correlations of election characteristics and abandonment rate, median time. K is defined as the largest number of projects selected by any participant, and M is the number of projects available to choose from.

Variable	Median Time Spent		Abandonment Rate	
	coefficient	95% confidence interval	coefficient	95% confidence interval
K	4.7 *	0.3 – 9.0	-0.0012	-0.005 – 0.002
M	2.6 *	0.3 – 4.9	-0.0010	-0.003 – 0.001
text on the ballot	0.005 *	0.001 – 0.009	0.000	-0.000 – 0.000
budget	0.000	0.000 – 0.000	0.000	-0.000 – 0.000
approval method	20 *	5 – 36	-0.0012	-0.01 – 0.01
knapsack method	33 **	13 – 53	0.043 ***	0.027 – 0.059
ranking method	-10	-34 – 15	0.014	-0.005 – 0.034

Table 10: Multiple Linear Regression (121 elections). K is the largest number of projects selected in any submitted vote. M is the number of available projects on the ballot. Text on the ballot is the cumulative number of characters in project descriptions in English. Approval, knapsack, and ranking are binary indicators of using the respective method.

- For available K-ranking data, the specific Borda count variant chosen has a limited effect on project allocation. The choice of K in K-approval voting has a bigger impact.
- Inferring knapsack votes from sufficiently long K-ranking votes is a valid approximation method when no digital interface is available.

We also provide a more rigorous and nuanced understanding of the following insights in existing literature about the average cost of selected projects under different voting methods:

- In most elections, voters tend to select significantly more expensive projects using the K-approval method, compared to the knapsack method. However, when we limit the analysis to the most expensive projects selected by a voter, they select significantly higher-cost projects under either method, depending on the election.
- In several elections, voters tend to select significantly less expensive projects using the knapsack interface, than when the same constraints are applied to their vote cast with the K-ranking interface. Explicit budget constraints contribute to the lower-cost effect of the knapsack voting method.
- In some K-ranking ballots, voters ranked their most expensive projects lower than would be expected based on their K-approval votes.
- In most elections, voters would select less expensive projects if their choice was budget constrained, than by

number of projects. This excludes effects from the voting interface or information presented. The implicit constraints of the voting method itself contribute to the lower-cost effect of the knapsack voting method.

- Aggregating ranking ballots with a knapsack method results in selecting less expensive projects compared to aggregation with K-approval or Borda Count.

These findings corroborate and expand upon results from (Goel et al. 2019): knapsack voting tends to result in lower-cost selected budgets. We find that it also results in lower-cost individual ballots. This cost effect is influenced by both explicit constraints and presented information, as well as the implicit constraint of the method itself. Notably, this cost effect is consistent across all selected projects, but not for the most expensive selected projects. This suggests that voters prioritize cost considerations for cheaper projects, which should be confirmed in more qualitative user studies.

The conclusion that explicit constraints matter, implies that while knapsack ballots and inferred knapsack ballots are meaningfully similar, and ranked ballot elicitation is a viable alternative to knapsack elicitation (for example, on paper), it is not a perfect substitute and may impact the selected projects. However, since the cost effects align, this does not necessarily have to be a problem. We examined abandonment rates and median time spent for primary ballots, revealing:

- Ballot design can affect the effort that voters have to spend on a ballot. Ballot complexity (e.g. number of

projects to choose from, number of projects to select, and text on the ballot) positively correlated with the median time spent by voters. However, there is no significant correlation with the abandonment rate.

10. Elections using the knapsack voting method correlate with a higher median time spent and higher abandonment rate. Still, the cause remains unclear and the causality remains inconclusive.

A lot more analysis is possible with this data set, and we hope that this will provide a fruitful avenue for other researchers to test various hypotheses.

This study demonstrates the feasibility and value of comparing voting methods in practical settings, although this comes with its own limitations. While it allows a direct evaluation of real voter considerations under different elicitation methods, it does not allow a fully randomized treatment where within the same election. We would encourage more experimental ballots to be included in real PB processes, in order for these findings to be evaluated and understood more thoroughly, and sharing anonymized data of such ballots. Only that way, the potential confounding effect of various factors such as turnout and recruitment method can be adequately addressed.

Future work could also include surveys and user studies with real voters in various processes that help understand to what extent voters are aware of various specifics of the voting process (e.g. the aggregation method selected), and whether they considered this in their voting strategy. This could more definitively inform assumptions in the aggregation method.

To really understand whether knapsack voting is a harder voting method for voters and whether that deters people, we will need to design dedicated experiments with real voters on real issues. Analysis of secondary ballot data (such as the data we are releasing) is a promising avenue to develop such insights, but will require data from more elections for answering some of the questions.

Impact Considerations

The cleaning and anonymization process of the dataset has been executed to maximally preserve reproducibility while assuring privacy of the participants. Cities have not been anonymized as these processes were public. Conclusions were framed to improve understanding for practitioners.

Acknowledgments

The authors acknowledge the many PB organizers that agreed to include a research ballot as part of their election. Part of the data collection efforts was funded through grants from Cisco Systems [unrestricted donation] and from the US Department of the Navy for 'Collaborative Decision Making at Scale: Bridging Theory and Practice' [grant number N00014-15-1-2786]. We thank Sukolsak Sakshuwong for developing and maintaining the Stanford Participatory Budgeting platform for many years, and thank Tanja Aitamurto and Anilesh K. Krishnaswamy for their contributions. We appreciate the helpful discussions with and feedback during the data collection design from Nikhil Garg, the analysis

support from Megan Mou and further feedback on the paper from Mohak Goyal, Irene Lo and anonymous reviewers.

References

- Arana-Catania, M.; Lier, F.-A. V.; Procter, R.; Tkachenko, N.; He, Y.; Zubiaga, A.; and Liakata, M. 2021. Citizen Participation and Machine Learning for a Better Democracy. *Digital Government: Research and Practice*, 2(3): 1–22.
- Arnstein, S. R. 1969. A Ladder Of Citizen Participation. *Journal of the American Institute of Planners*, 35(4): 216–224.
- Aziz, H.; and Shah, N. 2021. Participatory Budgeting: Models and Approaches. In Rudas, T.; and Péli, G., eds., *Paths Between Social Science and Computational Social Science*, 215–236. Cham: Springer International Publishing. ISBN 978-3-030-54935-0 978-3-030-54936-7. Series Title: Computational Social Sciences.
- Bartocci, L.; Grossi, G.; Mauro, S. G.; and Ebdon, C. 2022. The journey of participatory budgeting: a systematic literature review and future research directions. *International Review of Administrative Sciences*, 002085232210789.
- Benadè, G.; Nath, S.; Procaccia, A. D.; and Shah, N. 2017. Preference Elicitation For Participatory Budgeting. *Journal of the ACM*, 1(1): 27.
- Callahan, K. 2007a. Citizen Participation: Models and Methods. *International Journal of Public Administration*, 30(11): 1179–1196.
- Callahan, K., ed. 2007b. *Elements of effective governance: measurement, accountability and participation*. Number 126 in Public administration and public policy. Boca Raton: CRC/Taylor & Francis. ISBN 978-0-8493-7096-0. OCLC: ocm69013806.
- Cantador, I.; and Cortés-Cediel, M. E. 2018. Towards increasing citizen engagement in participatory budgeting digital tools. In *Proceedings of the 19th Annual International Conference on Digital Government Research*, 1–2. ACM Press. ISBN 978-1-4503-6526-0.
- de Borda, J.-C. 1784. Mémoire sur les élections au scrutin. *Histoire de l'Académie royale des sciences*, Année 1781: 31–34.
- Ebdon, C.; and Franklin, A. L. 2006. Citizen Participation in Budgeting Theory. *Public Administration Review*, 66(3): 437–447.
- Fain, B.; Goel, A.; and Munagala, K. 2016. The Core of the Participatory Budgeting Problem. In Cai, Y.; and Vetta, A., eds., *Web and Internet Economics*, volume 10123, 384–399. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-662-54109-8 978-3-662-54110-4. Series Title: Lecture Notes in Computer Science.
- Fairstein, R.; Benade, G.; and Gal, K. 2023. Participatory Budgeting Design for the Real World. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37.
- Falanga, R.; and Lüchmann, L. H. H. 2020. Participatory budgets in Brazil and Portugal: comparing patterns of dissemination. *Policy Studies*, 41(6): 603–622.

- Fung, A. 2006. Varieties of Participation in Complex Governance. *Public Administration Review*, 66(s1): 66–75.
- Garg, N.; Gelauff, L.; Sakshuwong, S.; and Goel, A. 2019. Who is in Your Top Three? In *HCOMP 2019*.
- Gelauff, L.; and Goel, A. 2024. Participatory Budgeting Preferences Data Set.
- Gelauff, L.; Goel, A.; Munagala, K.; and Yandamuri, S. 2020. Advertising for Demographically Fair Outcomes. In *ArXiv, Presented at: 4th Workshop on Mechanism Design for Social Good*. Online. ArXiv: 2006.03983.
- Goel, A.; Krishnaswamy, A. K.; Sakshuwong, S.; and Aitamurto, T. 2019. Knapsack Voting for Participatory Budgeting. *ACM Transactions on Economics and Computation*, 7(2): 1–27.
- Goyal, M.; Sakshuwong, S.; Sarmasarkar, S.; and Goel, A. 2023. Low Sample Complexity Participatory Budgeting. In Etessami, K.; Feige, U.; and Puppis, G., eds., *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 70:1–70:20. Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-278-5.
- Hagelskamp, C.; Rinehart, C.; Silliman, R.; and Schleifer, D. 2016. Public Spending, by the People. Participatory Budgeting in the United States and Canada in 2014–2015. Technical report, Public Agenda.
- Holston, J.; Issarny, V.; and Parra, C. 2016. Engineering software assemblies for participatory democracy: the participatory budgeting use case. In *Proceedings of the 38th International Conference on Software Engineering Companion*, 573–582. Austin Texas: ACM. ISBN 978-1-4503-4205-6.
- Johnson, C.; Carlson, H. J.; and Reynolds, S. 2021. Testing the Participation Hypothesis: Evidence from Participatory Budgeting. *Political Behavior*, 45: 3–32.
- Kim, N. W.; Jung, J.; Ko, E.-Y.; Han, S.; Lee, C. W.; Kim, J.; and Kim, J. 2016. BudgetMap: Engaging Taxpayers in the Issue-Driven Classification of a Government Budget. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1028–1039. San Francisco California USA: ACM. ISBN 978-1-4503-3592-8.
- Langton, S. 1979. American citizen participation: A deep-rooted tradition. *National Civic Review*, 68(8): 403–422.
- Lerner, J.; and Pape, M. 2020. Budgeting for Equity: How Can Participatory Budgeting Advance Equity in the United States? *Journal of Deliberative Democracy*, 12(2).
- Miller, S. A.; Hildreth, R. W.; and Stewart, L. M. 2019. The Modes of Participation: A Revised Frame for Identifying and Analyzing Participatory Budgeting Practices. *Administration & Society*, 51(8): 1254–1281.
- Munagala, K.; Shen, Y.; and Wang, K. 2022. Auditing for Core Stability in Participatory Budgeting. In Hansen, K. A.; Liu, T. X.; and Malekian, A., eds., *Web and Internet Economics*, volume 13778, 292–310. Cham: Springer International Publishing. ISBN 978-3-031-22831-5 978-3-031-22832-2. Series Title: Lecture Notes in Computer Science.
- Peters, D.; Pierczyński, G.; and Skowron, P. 2021. Proportional Participatory Budgeting with Additive Utilities. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, volume 34, 12726–12737. Curran Associates, Inc.
- Pina, V.; Torres, L.; Royo, S.; and Garcia-Rayado, J. 2022. Decide Madrid: A Spanish best practice on e-participation. In Randma-Liiv, T.; and Lember, V., eds., *Engaging Citizens in Policy Making*. Edward Elgar Publishing. ISBN 978-1-80037-436-2 978-1-80037-435-5.
- Rey, S.; and Maly, J. 2023. The (Computational) Social Choice Take on Indivisible Participatory Budgeting. ArXiv:2303.00621.
- Serramia, M.; Lopez-Sanchez, M.; Rodriguez-Aguilar, J. A.; and Escobar, P. 2019. Optimising participatory budget allocation: the Decidim use case. *Artificial Intelligence Research and Development*, 319: 193–202.
- Sintomer, Y.; Herzberg, C.; and Röcke, A. 2008. Participatory Budgeting in Europe: Potentials and Challenges: Participatory budgeting in Europe. *International Journal of Urban and Regional Research*, 32(1): 164–178.
- Society, T. D. 2016. Digital tools and Scotland’s Participatory Budgeting programme. Technical report, The Democratic Society.
- Stolicki, D.; Szufa, S.; and Talmon, N. 2020. Pabulib: A Participatory Budgeting Library. ArXiv:2012.06539.
- Su, C. 2017. From Porto Alegre to New York City: Participatory Budgeting and Democracy. *New Political Science*, 39(1): 67–75.
- van Dijk, J. A. 2012. Digital Democracy: Vision and Reality. In Snellen, I. T. M.; Thaens, M.; and Donk, W. B. H. J. v. d., eds., *Public administration in the information age: revisited*, number volume 19 in Innovation and the public sector, 49–62. Amsterdam ; Washington, DC: IOS Press. ISBN 978-1-61499-136-6.
- Wang, X. 2001. Assessing Public Participation in U.S. Cities. *Public Performance & Management Review*, 24(4): 322.
- Williams, E.; Denny, E. S.; and Bristow, D. 2017. Participatory Budgeting: An Evidence Review. Technical report, Public Policy Institute for Wales.

Paper checklist

1. For most authors...
 - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? Yes
 - (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? Yes
 - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? Yes
 - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? Yes Some additional detail is due to space constraints included in the dataset documentation.
 - (e) Did you describe the limitations of your work? Yes
 - (f) Did you discuss any potential negative societal impacts of your work? Yes. The main potential negative impact relates to voter privacy. We have made every effort to anonymize the data set and described this in detail in the dataset documentation and briefly in the paper.
 - (g) Did you discuss any potential misuse of your work? Yes. We identify that people could try to use the dataset to verify PB election outcomes and warn that this may be different due to data cleaning.
 - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? Yes. The dataset has been anonymized and contains additional documentation. This dataset can be used to reproduce most of our findings (except for abandonment rate).
 - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? Yes
2. Additionally, if your study involves hypotheses testing...
 - (a) Did you clearly state the assumptions underlying all theoretical results? NA
 - (b) Have you provided justifications for all theoretical results? NA
 - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? NA
 - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? Yes
 - (e) Did you address potential biases or limitations in your theoretical framework? Yes
 - (f) Have you related your theoretical results to the existing literature in social science? NA
 - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? NA
3. Additionally, if you are including theoretical proofs...
 - (a) Did you state the full set of assumptions of all theoretical results? NA
 - (b) Did you include complete proofs of all theoretical results? NA
4. Additionally, if you ran machine learning experiments...
 - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? NA
 - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? NA
 - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? NA
 - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? NA
 - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? NA
 - (f) Do you discuss what is "the cost" of misclassification and fault (in)tolerance? NA
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
 - (a) If your work uses existing assets, did you cite the creators? NA
 - (b) Did you mention the license of the assets? NA
 - (c) Did you include any new assets in the supplemental material or as a URL? Yes. The dataset has been made available to the reviewers as zip file and will be linked in the camera ready version.
 - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? Yes The data has been thoroughly anonymized and no identifiable information is left in the dataset. We described this in the dataset documentation as well as to what extent consent was obtained.
 - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? Yes We discuss that the data has been anonymized and how.
 - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR (see ?)? Yes We refer to the dataset documentation. In addition, we intend to make a derivative version of this dataset available for inclusion in the main Participatory Budgeting vote database PaBuLib, but this requires further discussion.
 - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? No. We have attempted to answer the relevant questions in the dataset documentation.
6. Additionally, if you used crowdsourcing or conducted research with human subjects...
 - (a) Did you include the full text of instructions given to participants and screenshots? No Relevant screenshots

are available in another publication and the platform is available online.

- (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? Yes Main risk is anonymity.
- (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? NA
- (d) Did you discuss how data is stored, shared, and de-identified? Yes As far is possible anonymized.