# Proportionality from Low-Dimensional Approval Data

Extended Abstract

Zhiyi Huang
University of Texas at Austin
Austin, United States
zhiyih@cs.utexas.edu

Gregory Kehne
Washington University in St. Louis
St. Louis, United States
kehne@wustl.edu

Chutong Yang
University of Texas at Austin
Austin, United States
cyang98@cs.utexas.edu

## ABSTRACT

Multiwinner voting is seeing increasing application in a wide range of domains, including participatory budgeting, online e-democracy platforms, and reinforcement learning from human feedback (RLHF) for fine-tuning AI models. In virtually all settings, instances can and do exceed the scale for which it is feasible to elicit human agents' input on the full candidate set. Motivated by this, we explore the extent to which proportionality axioms can be satisfied when each agent expresses preferences over only a few of the alternatives.

We consider when only a constant number of queries per voter suffice to identify proportional committees, even as the committee size grows large. We give fine-grained guarantees when voters are one of only finitely many types, and present both query-sparse and query-efficient algorithms. The former proceeds via a complexity metric that captures the difficulty of reconstructing an approval profile from sparse queries, and may be of independent interest.

We also ask when approval queries over mere pairs of candidates are enough. Such structured domains include possibly-single peaked instances, where pairwise queries are enough to both identify a global candidate order and identify proportional committees of arbitrary size.

## CCS CONCEPTS

• **Applied computing** → **Voting / election technologies**; • **Theory of computation** → *Algorithmic game theory*; • **Computing methodologies** → *Multi-agent systems.*

## KEYWORDS

Multiwinner voting, Approval voting, Participatory budgeting

**ACM Reference Format:**
Zhiyi Huang, Gregory Kehne, and Chutong Yang. 2026. Proportionality from Low-Dimensional Approval Data: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/YHKI7695

## 1 INTRODUCTION

How can groups of people effectively express and aggregate their preferences at scale? In voting theory and practice, voters typically express preferences over the collection of candidates, either by ranking them explicitly, as in ranked-choice voting, or by choosing one or more candidates, as in plurality, $k$-approval, or approval voting. Even when voting rules elicit only a small prefix of voters' preference rankings, they are conventionally predicated on voters' reasoned consideration of the full set of candidates in the election.

As the number of candidates becomes large, this quickly becomes cognitively and logistically intractable. For instance, in the ranked-choice voting in New York City's 2023 municipal primary elections, fully 33% of voters ranked at most one candidate in the elections with four or more candidates [14]. In elections with many candidates, voters are also less able to expend the effort to deliver informed assessments of candidates, leading to lower-quality outcomes.

This scalability problem only gains salience as we move from traditional settings to the frontier of modern applications of voting rules and social choice. For example, instances of participatory budgeting (PB)—a generalization of multiwinner voting—regularly feature more than a hundred project proposals (candidates) for voters' evaluation [5, 13]. In this setting many elicitation formats and popular PB methods require only that voters cast approval preferences [16]; but voters may not have efficient means for discerning which projects they support, let alone which they believe to be cost-effective.

Another recent application is online platforms for large-scale deliberation, many of which are approval-based and regularly feature tens of thousands of distinct candidates in the form of statements participants either approve or disapprove of [19]. At this scale, eliciting each participant's input for all candidates is infeasible by multiple orders of magnitude. Instead, voters are presented with only a small subset of the alternatives, and the platform's goal is to choose a small statement set that collectively represents participants' positions or opinions as a whole, or otherwise summarizes the sentiment landscape. Further, these applications of multiwinner voting are not developing in isolation; indeed, e-democracy platforms are promoted for the purpose of eliciting project proposals for participatory budgeting elections [6].

This preference elicitation scalability bottleneck also arises in the application of social choice to AI alignment and the fine-tuning of AI models from human feedback, a subject which has recently generated acute interest within computational social choice [4, 8, 18, 21]. The seminal work on RLHF of Ouyang et al. [15] assumes only a few inputs per agent (voter), and recent distortion-based work on RLHF by Gölz et al. [9] adopts and accommodates this constraint. This motivation from large-scale instances and RLHF is also made explicit by Charikar et al. [3], who study metric distortion within the class of tournament voting rules and low-dimensional generalizations. Within the application of social choice to RLHF, a case is also emerging for multiwinner approaches: Vamplew et al. [20] motivate multi-objective RL from pluralism and the aim of



*Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus.* © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/YHKI7695

representing diverse user preferences, and [12] construct ensembles of reward models in pursuit of a distinct notion of proportionality.

Motivated by the settings of many-candidate *single*-winner elections and online deliberation platforms, Halpern et al. [10] investigate the tractability of computing single-winner voting rules using only sparse queries to voters' *ordinal* preferences. Their point of departure is the approval-based, multiwinner voting model of Halpern et al. [11], who develop approval-based multiwinner voting rules which require only sparse voter queries, and establish an adaptivity gap for the problem of finding *proportional* outcomes: while the number of $k$-sparse queries required by their adaptive rules scales nearly linearly in $m$, non-adaptive multiwinner rules must use $\Omega(m^{11})$ $O(k)$-sparse queries in the worst case.

The proportionality axioms developed in multiwinner approval voting literature [1, 2] and field-tested in participatory budgeting [17] offer a normatively appealing and domain-agnostic standard for designing multiwinner voting rules in restricted settings. Informally, the axiom of justified representation (JR) and its relatives guarantee committee representation to preference-distinct groups of voters that is proportional to group size, and do so *for all groups*. This is sufficiently versatile to provide fairness desiderata in PB, has been proposed as apt means to summarize opinion space for online deliberation platforms, and is a plausible standard for choosing ensembles of objectives or models that provide collective coverage of populations with diverse preferences. More broadly, proportionality is an ideal benchmark against which to measure preference query models: it is always feasible but nontrivial to satisfy, and is not merely a linear function of agents or alternatives.

With these goals in mind, we revisit the setting of Halpern et al. [11] and investigate the power of sparse queries. We ask:

> *When do sparse queries suffice to identify proportional subsets of large spaces of alternatives?*

While we are primarily concerned with when it is possible to identify proportional committees from sparse queries, we also consider the complexity of this identification. This can be measured by computational complexity, or by query complexity, meaning the number of sparse subsets of candidates queried or the number of sparse-query-answering voters necessary. High query complexity generally entails high computational complexity.

## 1.1 Our Contributions

Our contributions are threefold. First, we begin by reinterpreting and then circumventing the sparse-query hard instances of Halpern et al. [11] by considering the **finite types** preference setting. Here we assume the presence of $s$ distinct voter types, and we develop a fine-grained complexity metric and associated profile reconstruction algorithm which enables the identification of proportional committees from $t$-sparse queries for $t$ between $\log s$ and $s$. We show this can be made computationally and query-efficient for $s$-dimensional queries, and conjecture that 2-dimensional queries suffice to identify JR committees when $s \leq k$.

**Theorem 1.** *Let $A$ be an approval profile for which there are at most $s$ types of voters. Then there is a poly-time, $s$-dimensional, $O(m \cdot \log s)$-query algorithm that reconstructs the approval profile.*

Next, we further address the question of when 2-dimensional queries suffice to identify (approximately) proportional committees, even as $k$ grows large. This class of committee selection algorithms forms the approval-based committee voting analog of *tournament rules* in single-winner voting from ranked preferences, in that such rules consider only preferences over pairs of candidates and ignore higher-order correlations between candidate approvals.

In the **candidate interval (CI)** preference setting (and the slight generalization to *CI on a circle* preferences), we demonstrate a distinction between the cases where the ordering permutation of the profile is known versus unknown. When the ordering permutation is known, we propose an appealing rule which satisfies EJR+ using only $O(m)$ 2-dimensional queries. On the other hand, finding proportional committees requires $\Omega(m^2)$ non-adaptive 2-dimensional queries when the permutation is unknown. We match this by demonstrating a (computationally inefficient) algorithm for identifying a common candidate order given only the $O(m^2)$ 2-dimensional queries, and contrast this with the classical poly-time algorithm of Fulkerson and Gross [7] for recovering a common candidate order when the full approval preference profile is known.

**Theorem 2.** *Given an approval profile that is CI (on a circle), there exists a 2-sparse, $O(m^2)$-query $k$-committee-selection algorithm satisfying EJR+.*

In the **sparse approvals** preference setting, we assume every voter approves a small number of candidates, say no more than $q$. We first establish lower bounds on the query complexity of identifying even $2/q$-JR committees. We then develop a matching committee selection rule that uses the $O(m^2)$ 2-dimensional queries to achieve $2/q$-JR.

**Theorem 3.** *There is a 2-sparse, $O(m^2)$-query $k$-committee-selection algorithm that satisfies $2/q$-JR in the sparse approval setting.*

Our third contribution is to generalize this to a single-parameter family of **Thiele rules** implementable via 2-dimensional queries, which can be seen as interpolating between maximizing total approvals (the AV rule) and maximizing total approving voters (the Chamberlin–Courant rule). In order to evaluate our metrics and heuristic rules, we consider Euclidean models and approval profiles from the participatory budgeting instances cataloged on Pabulib [5]; both show that our approaches comfortably outperform the worst case in practice.

# 2 FUTURE DIRECTIONS

First, what other restricted approval domains are amenable to treatment with sparse queries? There is broad spectrum of well-motivated restrictions that can be imposed on the approval profile $A$; identifying which of these allow for effective committee selection within the sparse query model remains largely open. Second, what can be said for the more general (and, arguably, practical) setting of participatory budgeting? More generally, what is the worst-case query sparsity necessary to identify a size-$k$ committee satisfying JR (say)? The worst-case $\Omega(1)$ lower and $O(k)$ upper bounds of Halpern et al. [11] invite improvement. Finally, our conjecture regarding the use of 2-dimensional queries to select a committee satisfying JR in the finite types setting merits further consideration, and presents an intriguing challenge to either prove or refute.

## ACKNOWLEDGMENTS

The authors would like to thank Shuchi Chawla, as well as several anonymous reviewers for their valuable feedback on the framing and presentation of this work.

CY was supported in part by an Amazon AI Fellowship.


## REFERENCES


[1] Haris Aziz, Markus Brill, Vincent Conitzer, Edith Elkind, Rupert Freeman, and Toby Walsh. 2017. Justified representation in approval-based committee voting. *Social Choice and Welfare* 48, 2 (2017), 461–485.

[2] Markus Brill and Jannik Peters. 2023. Robust and Verifiable Proportionality Axioms for Multiwinner Voting. In *Proceedings of the 24th ACM Conference on Economics and Computation, EC 2023*. 301–301.

[3] Moses Charikar, Prasanna Ramakrishnan, Zihan Tan, and Kangning Wang. 2025. Metric Distortion for Tournament Voting and Beyond. In *Proceedings of the 26th ACM Conference on Economics and Computation, EC 2025*. ACM, 790–818.

[4] Vincent Conitzer, Rachel Freedman, Jobst Heitzig, Wesley H. Holliday, Bob M. Jacobs, Nathan Lambert, Milan Mossé, Eric Pacuit, Stuart Russell, Hailey Schoelkopf, Emanuel Tewolde, and William S. Zwicker. 2024. Position: Social Choice Should Guide AI Alignment in Dealing with Diverse Human Feedback. In *Forty-first International Conference on Machine Learning, ICML 2024*.

[5] Piotr Faliszewski, Jaroslaw Flis, Dominik Peters, Grzegorz Pierczyński, Piotr Skowron, Dariusz Stolicki, Stanisław Szufa, and Nimrod Talmon. 2023. Participatory Budgeting: Data, Tools and Analysis. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI*. 2667–2674.

[6] Sören Fillet. 2024. *What is participatory budgeting?* https://www.govocal.com/blog/what-is-participatory-budgeting

[7] Delbert Fulkerson and Oliver Gross. 1965. Incidence Matrices and Interval Graphs. *Pacific journal of mathematics* 15, 3 (1965), 835–855.

[8] Luise Ge, Daniel Halpern, Evi Micha, Ariel D Procaccia, Itai Shapira, Yevgeniy Vorobeychik, and Junlin Wu. 2024. Axioms for AI Alignment from Human Feedback. In *Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024*, Vol. 37. 80439–80465.

[9] Paul Gölz, Nika Haghtalab, and Kunhe Yang. 2025. Distortion of AI Alignment: Does Preference Optimization Optimize for Preferences? arXiv:2505.23749

[10] Daniel Halpern, Safwan Hossain, and Jamie Tucker-Foltz. 2024. Computing voting rules with elicited incomplete votes. In *Proceedings of the 25th ACM Conference on Economics and Computation, EC 2024*. 941–963.

[11] Daniel Halpern, Gregory Kehne, Ariel D. Procaccia, Jamie Tucker-Foltz, and Manuel Wüthrich. 2023. Representation with Incomplete Votes. In *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023*. AAAI Press, 5657–5664.

[12] Daniel Halpern, Evi Micha, Ariel D. Procaccia, and Itai Shapira. 2025. Pairwise Calibrated Rewards for Pluralistic Alignment. *CoRR* abs/2506.06298 (2025).

[13] Annick Laruelle. 2021. Voting to select projects in participatory budgeting. *Eur. J. Oper. Res.* 288, 2 (2021), 598–604.

[14] Deb Otis. 2024. *What We Learned from New York City's Second Ranked Choice Voting Election.* https://fairvote.org/report/rcv-in-nyc-report-2023/

[15] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022*.

[16] Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. 2021. Proportional Participatory Budgeting with Additive Utilities. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021*. 12726–12737.

[17] Dominik Peters and Piotr Skowron. 2025. *Method of Equal Shares for Participatory Budgeting*. https://equalshares.net/

[18] Ariel D. Procaccia, Benjamin Schiffer, and Shirley Zhang. 2025. Clone-Robust AI Alignment. In *Forty-second International Conference on Machine Learning, ICML 2025 (Proceedings of Machine Learning Research, Vol. 267)*.

[19] Christopher Small, Michael Bjorkegren, Timo Erkkilä, Lynette Shaw, and Colin Megill. 2021. Polis: Scaling deliberation by mapping high dimensional opinion spaces. *Recerca: revista de pensament i anàlisi* 26, 2 (2021).

[20] Peter Vamplew, Conor F Hayes, Cameron Foale, Richard Dazeley, and Hadassah Harland. 2024. Multi-objective Reinforcement Learning: A Tool for Pluralistic Alignment. arXiv:2410.11221

[21] Shresth Verma, Niclas Boehmer, Lingkai Kong, and Milind Tambe. 2025. Balancing act: prioritization strategies for llm-designed restless bandit rewards. In *International Conference on Game Theory and AI for Security*. Springer, 376–394.