Strategic Manipulation in Temporal Voting with Undesirable Candidates (Student Abstract)

Tzeh Yuan Neoh[1], **Nicholas Teh**[2]

[1]Agency for Science, Technology and Research, Singapore
[2]University of Oxford, UK
neoh_tzeh_yuan@cfar.a-star.edu.sg, nicholas.teh@cs.ox.ac.uk

Abstract

We study a model of sequential decision-making where voters have dynamic preferences over a set of candidates that are undesirable. This models scenarios such as the implementation of projects that are overall beneficial to society, but impose individual costs on certain affected individuals. We show that while minimizing the sum of agents' disutilities can be done in polynomial time, minimizing the maximum disutility obtained by any agent is computationally intractable, even in restricted cases. We then examine the potential for agents to engage in strategic manipulation in response to these welfare objectives, offering insights into possible misconduct within such decision-making environments.

1 Introduction

The local city council has launched an ambitious multi-year initiative to revitalize the city's tourism industry and enhance its cultural appeal. This initiative aims to boost the local economy while fostering a stronger sense of community through vibrant public projects.

As part of this initiative, residents from across the district are invited to actively participate in a voting process to express their preferences and opinions on a variety of proposed recurring projects. These projects range from establishing a weekly farmer's market and organizing seasonal fireworks displays to hosting large-scale sporting events like marathons and other community-centered activities. The goal is to select projects that will attract visitors, promote local businesses, and enrich the city's cultural landscape.

However, while such projects may bring substantial economic and cultural benefits to the city as a whole, they also raise valid concerns among residents living near the proposed event sites. These residents may be apprehensive about the potential negative impacts on their daily lives, including noise pollution from event setup and activities, increased traffic congestion, difficulties with parking, and overcrowding in their neighborhoods. As a result, some residents might oppose certain projects, even if these projects are broadly supported by the wider community for their perceived economic and cultural value.

Balancing these competing interests—ensuring that local voices are heard while pursuing the greater good—poses a significant challenge for the city council. This dynamic highlights the need for a decision-making framework that takes into account individual residents' preferences, while striving to achieve a *fair* outcome that is also *good* for society.

One possible approach is to study this problem through the lens of *temporal voting* (Elkind, Neoh, and Teh 2024; Zech et al. 2024). However, instead of voters having approval preferences (indicating which projects they would like to support) over a set of projects/candidates, voters now can express *disapproval* preferences (indicating which projects they object to) over a set of projects/candidates.[1] Temporal voting has also been studied under various other names, such as *perpetual voting* (Lackner 2020; Bulteau et al. 2021; Neoh and Teh 2024) and *sequential decision-making* (Chandak, Goel, and Peters 2024). We refer the reader to Elkind, Obraztsova, and Teh (2024a) for a comprehensive survey of other related work in the area.

2 The Temporal Voting Model

For each positive integer z, let $[z] := \{1, \dots, z\}$. Let $N = [n]$ be the set of n *agents*, $P = \{p_1, \dots, p_m\}$ be the set of m *projects* (or *chores*), and $T = [\ell]$ be the set of ℓ *timesteps*. We assume voters have *disapproval preferences*. For each $i \in N$ and $k \in T$, denote $D_{ik} \subseteq P$ as the *disapproval set* of agent i at timestep k. Let the *disapproval vector* of an agent i be $\mathbf{D}_i = (D_{i1}, \dots, D_{i\ell})$. Also, let γ denote the maximum number of disapprovals that each project has at any timestep. An *instance* of our problem is a tuple $(N, P, \ell, (\mathbf{D}_i)_{i \in N})$.

An *outcome* is a vector $\mathbf{o} = (o_1, \dots, o_\ell)$, where $o_k \in P$ for each $k \in [\ell]$. For any $k \in [\ell]$, a *partial outcome* is an outcome $\mathbf{o}^{(k)} = (o_1, \dots, o_k, \varnothing, \dots, \varnothing)$, where $o_{k'} \in P$ for each $k' \in [k]$ and $o_{k'} = \varnothing$ for each $k' \in \{k+1, \dots, \ell\}$.

The *disutility* of an agent $i \in N$ for an outcome $\mathbf{o}$ is given by $d_i(\mathbf{o}) = |\{k \in [\ell] : o_k \in D_{ik}\}|$. Let $\Pi(\mathcal{I})$ denote the space of all possible outcomes for an instance $\mathcal{I}$. A *mechanism* maps an instance $\mathcal{I} = (N, P, \ell, (\mathbf{D}_i)_{i \in N})$ to an outcome in $\Pi(\mathcal{I})$. Note that we require each agent to disapprove of at least one project at *some* timestep, i.e., for each $i \in N$ there exists a $k \in [\ell]$ with $D_{ik} \neq \varnothing$; otherwise we can sim-

[1]Note that the results in the approval framework do not automatically translate to the disapproval framework. This is in contrast to temporal *fair division*, where there might be symmetry observed between positive and negative valuations (Elkind et al. 2024).

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ply delete i as they can never be satisfied. However, we do not require that an agent disapprove of some project at *every* timestep.

3 Our Results

Two commonly studied welfare objectives when candidates are postively-valued are maximizing the *sum* of agents' welfare or maximizing the *minimum* welfare of any agent. We can similarly define analogous welfare objectives for the case of undesirable candidates. For any instance $\mathcal{I}$, we refer to outcomes $\mathbf{o} \in \Pi(\mathcal{I})$ that minimizes the sum of agents' disutilities (i.e., $\sum_{i \in N} d_i(\mathbf{o})$) or minimizes the maximum agents' disutility (i.e., $\max_{i \in N} d_i(\mathbf{o})$) as MIN-SUM and MIN-MAX outcomes, respectively.

To find a MIN-SUM outcome, we can greedily select, at each timestep, the project with the lowest sum of disapprovals at each step. It is easy to observe that this greedy algorithm runs in polynomial time, but it may not be fair. Consider an instance with $2\kappa + 1$ agents $N = \{1, \ldots, 2\kappa + 1\}$, two projects $P = \{p_1, p_2\}$, and ℓ timesteps. Let $\kappa + 1$ agents disapprove of p_2 at each timestep, and the remaining κ agents disapprove of p_1 at each timestep. Then, the MIN-SUM algorithm will select p_1 at every timestep, favoring the $\kappa + 1$ agents and consistently disadvantaging the other κ agents—this is arguably not a fair outcome. As such, we look into another welfare objective—MIN-MAX, which would conceivably provide better fairness guarantees.

Given any instance $\mathcal{I}$, and for $\tau \leq \ell$, let $\mathbf{A}$ denote the set of τ constraints: $\mathbf{A} = \{(t_1, \lambda_1), \ldots, (t_\tau, \lambda_\tau)\}$ where for each pair $(t, \lambda) \in \mathbf{A}$, $t \in T$ and $\lambda \in \mathbb{Z}_{\geq 0}$. Moreover, for $k < k'$, we have that $t_k < t_{k'}$ and $\lambda_k < \lambda_{k'}$.

Then, we let the decision problem of MIN-MAX take as input a problem instance $(N, P, \ell, (\mathbf{D}_i)_{i \in N})$ and a set of constraints $\mathbf{A}$, and answers the question of whether there exists there an outcome $\mathbf{o}$ such that for each $(t, \lambda) \in \mathbf{A}$, $\max_{i \in N} \sum_{j=1}^{t} d_i(o_j) \leq \lambda$. Intuitively, each pair $(t, \lambda) \in \mathbf{A}$ mandates that at timestep t, any agent should have a (cumulative) maximum disutility of λ. This generalizes the standard decision problem studied for maximizing egalitarian welfare in the setting with approval preferences (Elkind, Kraiczy, and Teh 2022; Elkind, Neoh, and Teh 2024).

We show that even in simple settings, the decision problem of MIN-MAX is NP-complete, with the following result.

Theorem 3.1. *The decision problem of* MIN-MAX *is* NP-*complete, even with $m = 2$, $\gamma = 2$, and $\tau = 1$.*

Next, we focus on agents' strategic considerations with respect to each of the considered welfare objectives. One well-studied concept in the social choice literature is *strategyproofness*, which states that no agent should be able to strictly benefit (in this case, decrease their disutility) by misreporting their preferences. For each $i \in N$, let $\mathcal{D}_{-i}$ denote the list of all disapproval vectors except that of agent i: $\mathcal{D}_{-i} = (\mathbf{D}_1, \ldots, \mathbf{D}_{i-1}, \mathbf{D}_{i+1}, \ldots, \mathbf{D}_n)$. Then, strategyproofness is formally defined as follows.

Definition 3.2 (Strategyproofness)**.** A mechanism $\mathcal{M}$ is *strategyproof* (SP) if for every agent $i \in N$ with disapproval vector $\mathbf{D}_i$, and for every possible reported vector $\mathbf{D}'_i$ and every possible report by the other agents $\mathcal{D}_{-i}$, we have that $d_i(\mathcal{M}(\mathbf{D}_i, \mathcal{D}_{-i})) \leq d_i(\mathcal{M}(\mathbf{D}'_i, \mathcal{D}_{-i}))$.

Then, we show the compatibility between the MIN-SUM objective and strategyproofness with the following result.

Theorem 3.3. *The greedy algorithm for computing a* MIN-SUM *outcome is strategyproof.*

Given this, a natural follow-up question would be whether the MIN-SUM objective is compatible with a stronger notion of strategyproofness; one such candidate is the well-studied notion of *group-strategyproofness* (GSP) (Halpern et al. 2020; Suksompong and Teh 2022, 2023). Intuitively, GSP states that no group of agents should be able to misreport their preferences so as to benefit everyone in the group. Observe that GSP implies SP by considering singleton groups.

For each $S \subseteq N$, let $\mathcal{D}_{-S}$ denote the list of all disapproval vectors except that of agents in the set S. Then, GSP is formally defined as follows.

Definition 3.4 (Group-strategyproofness)**.** A mechanism $\mathcal{M}$ is *group-strategyproof (GSP)* if there does not exist a nonempty group of agents $S \subseteq N$ and disapproval vectors $\mathbf{D}_i, \mathbf{D}'_i$ for each $i \in S$ such that $d_i(\mathcal{M}((\mathbf{D}'_i)_{i \in S}, \mathcal{D}_{-S})) < d_i(\mathcal{M}((\mathbf{D}_i)_{i \in S}, \mathcal{D}_{-S}))$ for all $i \in S$.

Unfortunately, MIN-MAX is not compatible with GSP, as we show in the following result.

Proposition 3.5. *Let $\mathcal{M}$ be a mechanism that always returns a* MIN-SUM *outcome. Then $\mathcal{M}$ is not group-strategyproof.*

In contrast, we show that the MIN-MAX objective is incompatible even with strategyproofness. Intuitively, agents have an incentive to misreport their disapproval for projects they do not actually disapprove.

Proposition 3.6. *Let $\mathcal{M}$ be a mechanism that always returns a* MIN-MAX *outcome. Then $\mathcal{M}$ is not strategyproof.*

4 Conclusion and Future Work

In this work, we study a model of temporal voting where agents can express disapprovals over candidates at each round. We investigate the computational complexity of two welfare objectives—MIN-SUM and MIN-MAX—and the strategic implications associated with achieving these objectives. Future directions include looking at more general welfare objectives (e.g., p-mean welfare), proposing suitable *proportionality* concepts (adapted from the well-studied approval setting (Bulteau et al. 2021; Chandak, Goel, and Peters 2024; Elkind, Obraztsova, and Teh 2024b)) for the disapproval setting. Other interesting directions include explore similar questions in settings where it is typical to assume positively-valued candidates, such as the *issue-by-issue voting* setting (Alouf-Heffetz et al. 2022), which can be seen as a restricted setting of temporal voting; or in *portioning* (Elkind, Suksompong, and Teh 2023).

References

Alouf-Heffetz, S.; Bulteau, L.; Elkind, E.; Talmon, N.; and Teh, N. 2022. Better Collective Decisions via Uncertainty Reduction. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, 24–30.

Bulteau, L.; Hazon, N.; Page, R.; Rosenfeld, A.; and Talmon, N. 2021. Justified Representation for Perpetual Voting. *IEEE Access*, 9: 96598–96612.

Chandak, N.; Goel, S.; and Peters, D. 2024. Proportional Aggregation of Preferences for Sequential Decision Making. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 9573–9581.

Elkind, E.; Kraiczy, S.; and Teh, N. 2022. Fairness in Temporal Slot Assignment. In *Proceedings of the 15th International Symposium on Algorithmic Game Theory*, 490–507.

Elkind, E.; Lam, A.; Latifian, M.; Neoh, T. Y.; and Teh, N. 2024. Temporal Fair Division of Indivisible Items. *arXiv preprint arXiv:2410.14593*.

Elkind, E.; Neoh, T. Y.; and Teh, N. 2024. Temporal Elections: Welfare, Strategyproofness, and Proportionality. In *Proceedings of the 27th European Conference on Artificial Intelligence (ECAI)*, 3292–3299.

Elkind, E.; Obraztsova, S.; and Teh, N. 2024a. Temporal Fairness in Multiwinner Voting. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence*, 22633–22640.

Elkind, E.; Obraztsova, S.; and Teh, N. 2024b. Verifying Proportionality in Temporal Multiwinner Voting. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multi-Agent Systems*, 2246–2248.

Elkind, E.; Suksompong, W.; and Teh, N. 2023. Settling the Score: Portioning with Cardinal Preferences. In *Proceedings of the 26th European Conference on Artificial Intelligence (ECAI)*, 621–628.

Halpern, D.; Procaccia, A. D.; Psomas, A.; and Shah, N. 2020. Fair Division with Binary Valuations: One Rule to Rule Them All. In *Proceedings of the 16th Conference on Web and Internet Economics (WINE)*, 370–383.

Lackner, M. 2020. Perpetual Voting: Fairness in Long-Term Decision Making. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, 2103–2110.

Neoh, T. Y.; and Teh, N. 2024. Welfare Maximization in Perpetual Voting (Student Abstract). In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*, 23597–23599.

Suksompong, W.; and Teh, N. 2022. On Maximum Weighted Nash Welfare for Binary Valuations. *Mathematical Social Sciences*, 117: 101–108.

Suksompong, W.; and Teh, N. 2023. Weighted fair division with matroid-rank valuations: Monotonicity and strategyproofness. *Mathematical Social Sciences*, 126: 48–59.

Zech, V.; Boehmer, N.; Elkind, E.; and Teh, N. 2024. Multiwinner Temporal Voting with Aversion to Change. In *Proceedings of the 27th European Conference on Artificial Intelligence (ECAI)*, 3236–3243.