Evaluating Voting Design Vulnerabilities for Retroactive Funding

Jay Yu *
Stanford University
jyu01@stanford.edu

Austin Bennett
Stanford University
gaustinb@stanford.edu

Billy Gao
Stanford University
billygao@stanford.edu

Rebecca Joseph
Stanford University
rjoseph5@stanford.edu

Abstract

Retroactive Public Goods Funding (RetroPGF) rewards blockchain projects based on proven impact rather than future promises. This paper reviews voting mechanisms for Optimism's RetroPGF, where "badgeholders" allocate rewards to valuable projects. We explore Optimism's previous schemes for RetroPGF voting, including quadratic, mean, and median voting. We present a proof-based formal analysis for vulnerabilities in these voting schemes, empirically validate these vulnerabilities using voting simulations, and offer assessments and practical recommendations for future iterations of Optimism's system based on our findings.

1 Introduction

1.1 Background and Motivation

Decentralized Autonomous Organizations (DAOs) have emerged as a novel form of organizational structure that leverages blockchain technology to enable transparent, automated decision-making processes [7]. Many different kinds of DAOs exist today in the blockchain industry, including protocol DAOs such as Uniswap, Optimism, and Arbitrum, collector DAOs such as PleasrDAO and NounsDAO, and investor DAOs. As of writing (November 2024), DAOs today manage over 22 billion USD in treasury value, and with major DAOs such as Optimism holding over 3 billion USD in treasury value [3].

Most DAOs today adopt some form of token-voting, where holders of a publicly-traded blockchain token, such as UNI, ARB, or OP are able to use their token holdings to cast votes on community proposals [4]. Token-based voting enables inclusive, community-driven participation, but practical challenges arise, including the risk of plutocratic capture by wealthy holders, low participation rates, and the need to balance efficiency with diverse interests [2]. Additionally, blockchain's pseudonymity introduces vulnerabilities like Sybil attacks and vote manipulation.

These real-world stakes and relative ease of changing governance processes within DAOs (compared with traditional corporate or political settings) have allowed DAOs to become a fertile testing ground for testing governance mechanism innovation [6].

1.2 Optimism's RetroPGF

Optimism's Retroactive Public Goods Funding (RetroPGF) represents an innovative program to sustaining public goods development in the blockchain ecosystem. Unlike traditional grant programs

*Funded in part through a research grant by the Optimism Foundation

that fund projects based on prospective promises, RetroPGF rewards projects retroactively based on their demonstrated impact. This mechanism aligns incentives by ensuring that funding flows to projects that have already proven their value to the ecosystem [5].

The RetroPGF system operates through rounds of funding, where qualified voters (badgeholders) in Optimism's Citizen House evaluate and allocate rewards to projects that have contributed value to the Optimism ecosystem. The process begins with project teams submitting evidence of their contributions, followed by the careful selection of community members with demonstrated expertise as badgeholders. These badgeholders then engage in a thorough evaluation period, reviewing and assessing project contributions before their votes are aggregated to determine the final funding distribution.

This system presents unique challenges for voting mechanism design. The need to balance the expertise of specialized voters with broad community representation creates an inherent tension in the selection process. The system must prevent collusion while encouraging legitimate collaboration among voters, particularly challenging in a space where coordination is often necessary for effective evaluation. Additionally, the mechanism must enable nuanced evaluation of qualitative impact across diverse project types while maintaining the ability to scale effectively as the number of projects and voters grows.

To address this, Optimism has experimented with different mechanisms in each fo the four rounds of prior RetroPGF funding [5].

- Round 1 used a method of **Quadratic Voting**
- Round 2 used a method of **Mean Voting**
- Round 3 and Round 4 used **Variants of Median Voting**

We will be exploring the voting designs and potential vulnerabilties of these various voting mechanisms in the following sections of this paper.

1.3 Contributions

This paper makes several contributions to the the analysis and development of Retroactive funding processes:

1. We offer a formal model for the voting mechanisms in Optimism's RetroPGF processes
2. We provide proofs and analysis of vulnerabilities within various RetroPGF processes
3. We simulate voting to determine the practical impact of these theoretical vulnerabilities
4. We provide specific recommendations for the implementation of voting mechanisms in future RetroPGF rounds.

The remainder of this paper is organized as follows: Section 2 provides an overarching literature review to situate RetroPGF within a broader context of social choice theory. Section 3 outlines our methodology and introduces a formal model for our proof-based evaluation and simulation framework. Section 4 details our proof-based analysis of vulnerabilities in the target voting mechanisms, and Section 5 assesses their practical impact in a simulation context. Finally, we present our findings in Section 6 and provide recommendations for further iterations of RetroPGF.

2 Literature Review: Social Choice Theory and DAOs

Social choice theory provides a fundamental framework for analyzing collective decision-making processes, particularly in DAOs. This field of study, which examines the aggregation of individual preferences into collective choices, is crucial for understanding the challenges and opportunities presented by voting mechanisms in Optimism's RetroPGF and the broader blockchain-based systems.

At the heart of social choice theory is "collective wisdom," which posits that group decisions can outperform individual judgments under certain conditions. This idea is formalized in mathematical models known as Jury Theorems, which provide a theoretical foundation for democratic decision-making processes.

This underscores the need for effective mechanisms to select and educate badge holders and provide them with high-quality information about project contributions in the context of Optimism's RetroPGF system.

Social choice theory also grapples with the challenges of aggregating diverse preferences, as Arrow's Impossibility Theorem famously demonstrated. This theorem shows that no voting system can simultaneously satisfy all criteria when there are three or more options. This result has profound implications for the design of voting mechanisms in DAOs, highlighting the inherent trade-offs and limitations that must be considered.

Optimism operates with three voting mechanisms to mitigate this inherent tradeoff: Quadratic, Mean, and Median voting for each round before RetroPGF funding. To ensure we operate under the same base assumptions of these 3 mechanisms I will situate each of these voting functions within Optimism and grander social choice theory.

Quadratic voting, used in Round 1, was developed by economists E. Glen Weyl and Steven P. Lalley [10]. This innovative approach builds on the earlier Vickrey-Clarke-Groves (VCG) mechanism, allowing voters to express preference intensity. Simply put, their voting power is determined by the square root of their token holdings. This approach enables participants to express preference intensity while mitigating the risk of power concentration among wealthy holders. In RetroPGF, quadratic voting facilitates a nuanced allocation of funds to projects based on demonstrated impact, aligning incentives with proven contributions to public goods.

Mean voting, utilized in Round 2, while less formally established as a distinct theoretical framework, draws from fundamental statistical concepts developed by mathematicians like Carl Friedrich Gauss and Adolphe Quetelet in the 18th and 19th centuries[9]. These voting mechanisms represent a progression in social choice theory, each offering unique advantages and challenges in aggregating collective preferences. Mean voting involves allocating funding according to the average of all votes received for each project. While straightforward and easy to understand, this method can be vulnerable to manipulation through extreme allocations—particularly in scenarios where adversaries can artificially deflate a project's perceived support. In RetroPGF rounds, mean voting has been used but raised concerns about its robustness against strategic behavior, prompting exploration of alternative methods.

Median voting, implemented in Rounds 3 and 4, is closely tied to the median voter theorem, first articulated by economist Duncan Black, [1] in which Black underlines that median voting determines funding based on the median allocation received by each project, making it less susceptible to outlier influence compared to mean voting. This mechanism aligns well with public goods funding by focusing on the central tendency of voter preferences. However, it also faces vulnerabilities from strategic attacks, such as the "Median Phantom Vote Attack," where adversaries can significantly lower a project's median allocation by introducing minimal votes. In Optimism's RetroPGF implementation, median voting has been utilized in later rounds to enhance resilience against manipulation while ensuring fairer resource distribution among projects.

Social choice theory also grapples with the challenges of aggregating diverse preferences, as famously demonstrated by Arrow's Impossibility Theorem. This theorem shows that no voting system can simultaneously satisfy a set of seemingly reasonable criteria when there are three or more options. This result has profound implications for the design of voting mechanisms in DAOs, highlighting the inherent trade-offs and limitations that must be considered.

The integration of these voting mechanisms within Optimism's RetroPGF system illustrates an ongoing experimentation with social choice principles in decentralized governance.

As DAOs continue to evolve, these frameworks will be critical in shaping effective decision-making processes that promote equitable funding for public goods in blockchain ecosystems.

2.1 Participatory Budgeting

Participatory budgeting, a process that allows citizens to directly decide how to allocate part of a public budget, shares similarities with Optimism's RetroPGF system. Both aim to democratize resource allocation and engage community members in decision-making. However, RetroPGF's focus on retroactive funding for blockchain projects represents a novel application of these principles. Quadratic voting, a mechanism used in Optimism's first round of RetroPGF, has gained attention

in social choice theory for its potential to balance individual preference intensity with collective decision-making. Proposed by Lalley and Weyl, quadratic voting allows participants to express the strength of their preferences by allocating votes, with the cost of votes increasing quadratically [10] . This mechanism aims to prevent the tyranny of the majority while still maintaining democratic principles. In the context of public goods funding, Buterin, Hitzig, and Weyl proposed an extension of quadratic voting called quadratic funding [9]. This mechanism is designed to optimize the provision of public goods in a decentralized ecosystem, addressing the free-rider problem inherent in traditional contributory systems while avoiding the pitfalls of pure one-person-one-vote systems. Quadratic funding works by matching individual contributions to projects using a formula where the total funding is proportional to the square of the sum of the square roots of contributions. This approach aims to incentivize many small contributions and create a more equitable distribution of funds, aligning closely with the goals of Optimism's RetroPGF. The application of these mechanisms in blockchain-based systems like Optimism's RetroPGF represents a significant development in the practical implementation of social choice theory. It offers a real-world testing ground for these theoretical concepts, allowing for empirical evaluation of their effectiveness in promoting fair and efficient allocation of resources for public goods. As we delve deeper into the specific voting mechanisms used in Optimism's RetroPGF rounds, including mean and median voting, it's important to consider how these approaches relate to and differ from the quadratic voting and funding models. This analysis will provide valuable insights into the evolving landscape of decentralized decision-making and public goods funding in blockchain ecosystems.

2.2 Quadratic Voting

Quadratic voting represents a significant innovation in voting mechanism design, aiming to balance individual preference intensity with collective decision-making. Under this system, voting power scales as the square root of token holdings, creating a natural equilibrium between stake and influence. For a participant with n tokens, their effective voting power becomes n, which helps prevent excessive concentration of power while still respecting economic stake[10]. The theoretical underpinnings of quadratic voting suggest that, under certain assumptions regarding token balances and vote determinacy, it can be proven to be uniquely optimal. This optimality stems from its ability to allow individuals to express the intensity of their preferences while maintaining a form of economic efficiency. However, quadratic voting's effectiveness relies on a critical assumption: that everyone's number of voting tokens is known and verifiable. In many blockchain-based systems, this assumption is challenged by the pseudonymous nature of wallet creation, potentially leaving the system vulnerable to Sybil attacks. In these attacks, a user could create multiple identities to gain disproportionate voting power. For Optimism's RetroPGF, this vulnerability is mitigated by the pre-selection and verification of delegates. However, there remains a potential for a direct Sybil attack if colluding delegates decide to average their votes across multiple projects. This strategy comes with significant counterparty risk, which acts as a deterrent. The application of quadratic voting in Optimism's RetroPGF setup presents an interesting case study in the real-world implementation of this theoretical voting mechanism. Its effectiveness in this context will depend on how well it can balance the goals of decentralization, stake-based influence, and resistance to manipulation.

2.3 Median Voting

Median voting, also known as the median voter theorem or the median voting rule, is a well-established concept in social choice theory and political science. This voting mechanism posits that in a majority rule voting system, the outcome of a vote will align with the preferences of the median voter34. The median voting rule has several key properties that make it attractive for collective decision-making: Resistance to Extreme Values: Unlike mean-based voting systems, median voting is less influenced by outliers or extreme votes, potentially making it more robust against manipulation. Condorcet Winner: Under certain conditions, the median alternative is guaranteed to be the Condorcet winner - the option that would win in a head-to-head competition against any other alternative. Strategy-Proofness: In single-peaked preference profiles, the median voting rule is strategy-proof, meaning voters have no incentive to misrepresent their true preferences. However, median voting also has limitations: Applicability: Its effectiveness is most pronounced in single-dimensional issue spaces and may not translate well to multi-dimensional decision problems. Information Loss: By focusing solely on the median, this method may discard valuable information about the distribution of preferences across the entire voting population. In the context of Optimism's RetroPGF, the adoption of median voting

in later rounds represents an interesting shift from earlier mechanisms. This change likely aims to leverage the median vote's robustness against extreme values and potential for strategy-proofness. However, its effectiveness in this specific context will depend on how well the assumptions of median voting align with the realities of the RetroPGF voting environment. The application of median voting in blockchain-based decision-making systems like RetroPGF opens up new avenues for research. It raises questions about how traditional social choice theories translate to decentralized, token-based voting systems, and how they can be adapted to address the unique challenges and opportunities blockchain technology presents.

3 Methodology

As mentioned in Section 1, Optimism has previously conducted RetroPGF using three different mechanisms across their four rounds: Mean Voting, Quadratic Voting, and Median Voting. Our approach will focus on both a proof-based theoretical analysis of these voting mechanisms to expose potential vulnerabilities in each of these models in Section 4. We then empirically verify and assess the impact of these vulnerabilities in simulated voting rounds in Section 5 and discuss our results in Section 6. In this section, we will outline the formal model of Optimism's RetroPGF setup that we will analyze, both in our formal analysis and our simulations.

3.1 Optimism Governance Assumptions

Voting mechanisms are often difficult to evaluate for blockchain-based governance systems, as there are many factors that can affect how we distribute and weigh votes. For example, users with a large number of tokens, high governance reputation score, and control over many nodes may hold varying levels of power in different systems of governance.

Optimism governance incentivizes development by providing retroactive funding to projects based on their success, which is determined by a group of badge holders. In this paper, we explore potential risks under the following assumption model:

1. There is a fixed number of voters in our system.
2. A voter cannot split their funds into multiple voting wallets.
3. There is a fixed number of projects where voters can allocate funds.
4. Voters all have a set number of tokens to allocate.
5. Rational voters will maximize their utility – a function of their allocations to projects.

3.2 Model

Consider a retroactive funding round with $n \in \mathbb{Z}^+$ voters and a total funding allocation of $T \in \mathbb{R}^+$ tokens to be evenly divided among them. This yields a voting power of $\frac{T}{n}$ for each voter. We assume that there are no costs associated with voting. Thus, we define the following model to represent a funding round:

Model Specification for a Single-Shot Optimism Retroactive Funding Vote

- The number of voters within a retroactive funding round: $N \in \mathbb{Z}^+$
- The number of projects available for funding: $P \in \mathbb{Z}^+$
- The number of tokens to be distributed: $T \in \mathbb{R}^+$
- The Allocation Matrix, $A \in \mathbb{R}^{NxP}$, where A_{ij} is equal to voter i's allocation to project j.

Subject to the Following Model Variable Constraints

1. All voters have a maximum of $\frac{T}{N}$ votes: $\sum_j^P A_{ij} \leq \frac{T}{N}, \forall i \in [1, ..., N]$
2. Vote amounts are non-negative: $A \geq 0$

3.3 A Simplified Model of Aggregate Voting Mechanisms

3.3.1 Mean Voting

Given the allocation matrix $A \in \mathbb{R}^{N \times P}$, where A_{ij} represents the allocation of tokens from voter i to project j, the total allocation for each project j is calculated as:

$$T_j = \sum_{i=1}^{N} A_{ij},$$

The allocation is given by:

$$P_j = \frac{T_j}{\sum_{j=1}^{P} T_j}.$$

$$\text{Allocation} = T \cdot P_j$$

3.3.2 Median Voting

Given the allocation matrix $A \in \mathbb{R}^{N \times P}$, where A_{ij} represents the allocation of tokens from voter i to project j, the median allocation for each project j is calculated as:

$$M_j = \text{median}\{A_{ij} \mid A_{ij},\, i = 1, \ldots, N\},$$

The allocation is given by:

$$P_j = \frac{M_j}{\sum_{j=1}^{P} M_j},$$

$$\text{Allocation} = T \cdot P_j$$

3.3.3 Quadratic Voting

Given the allocation matrix $A \in \mathbb{R}^{N \times P}$, where A_{ij} represents the allocation of tokens from voter i to project j, the quadratic voting allocation for each project j is calculated as:

$$Q_j = \sum_{i=1}^{N} \sqrt{A_{ij}},$$

The allocation is given by:

$$P_j = \frac{Q_j}{\sum_{j=1}^{P} Q_j}.$$

$$\text{Allocation} = T \cdot P_j$$

4 Voting Mechanism Analysis

4.1 Mean Voting

Mean voting is a mechanism were we allocate funding according to the mean of the vote amounts for a project. Here, we show the importance of including all wallets under a mean voting system, including those with a vote of zero tokens.

Remark 1 *When all vote allocations are included, then Mean Voting is a Linear Voting Mechanism*

When all votes are included, notice that for a project, j in our model, $j \in [1, ..., P]$, we can represent its allocation total as:

$$\text{Project j's Mean Votes Received} = \frac{1}{N}\sum_{i=1}^{N} A_{ij}$$

Since N is a constant, we see that mean allocation has a linear relationship to the number of tokens allocated to the project. This is proportionally the same as Linear Voting; and, therefore, offers no strategic advantage in resources allocation. It may, however, yield some advantages with respect to ease of determining token allocation. For example, mean voting could allow groups to make decisions about the allocation proportion from a larger pool in an easy-to-view way for voters.

Remark 2 *The Mean Phantom Vote Attack – Where n is the number of non-zero wallets, Mean Voting without zero-value allocations included can lead k adversaries to artificially reduce the vote total by up to $\frac{n}{n+k}$, where $k \leq N - n, k \in \mathbb{Z}^+$.*

To see this, choose some arbitrarily project j with n non-zero allocations and the allocation vector $a_j = (a_{1j}, ..., a_{nj})$. Notice that this is not the same as the allocation matrix, as this is simply the non-zero subset of the column associated with allocations for a project, j. Thus, the mean voting power prior to an attack simply takes the form:

$$\text{Project } j\text{'s Mean Voting Power before Attack} = \frac{1}{n}\sum_{i=1}^{n} a_{ij}$$

Now, say some, k, adversaries don't like this project and want to decrease its mean vote total. This can be for a variety of reasons, though most commonly to reserve funds for another project.

The adversaries can each choose some arbitrarily small $\epsilon_k > 0$ for them to commit to a project, j. This yields, with a slight abuse of notation, the following mean voting power after an attack:

$$\text{Project } j\text{'s Mean Voting Power after Attack} = \lim_{\epsilon_i \to 0^+} \frac{1}{n+k}(\sum_{i=1}^{n} a_{ij} + \sum_{i=1}^{k} \epsilon_i) = \frac{1}{n+k}\sum_{i=1}^{n} a_{ij}$$

From this, we see that the relative mean voting power of a project after the attack compared to before the attack can be simplified as follows:

$$\frac{\text{Project } j\text{'s Mean Voting Power after Attack}}{\text{Project } j\text{'s Mean Voting Power before Attack}} = \frac{\frac{1}{n+k}\sum_{i=1}^{n} a_{ij}}{\frac{1}{n}\sum_{i=1}^{n} a_{ij}} = \frac{n}{n+k}\square$$

This problem arises from the number of voters which we count in the mean vote total. Since an adversary can contribute effectively nothing, they can decrease the average number of vote allocation. If we count all vote allocations, even with zero value, then we have the constant number of participants, N, yielding a scaling term of $\frac{1}{N}$ for all vote allocation totals. This yields the linear system described in Remark 1.

Thus, it is important that we count both zero-votes and non-zero votes alike in a vote total when evaluating mean voting power, otherwise we could have a group of adversaries artificially deflate the mean vote total for a project.

4.2 Median Voting

Median voting is a mechanism were we allocate funding according to the median of the vote amounts for a project. Here, we show how the Phantom Vote attack may be more dangerous for Median voting, as we cannot rely on including all zero votes as we did with Mean voting.

Remark 3 *The Median Phantom Vote Attack – Where n is the number of non-zero wallets, let $a_j^{(i)}$ be the ith order statistic for $i \in [1, ..., n]$. If we have a median allocation a_m, where m is the largest index such that $a_m \leq median(a)$, then Median Voting without zero-value allocations included can lead k adversaries to artificially decrease the voting power to where it is bounded above by $a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}$.*

To see this, invoke a similar strategy as described in the "mean phantom vote" section. More specifically, each adversary, k, can each choose some corresponding arbitrarily small $\epsilon_k > 0$ for them to commit to a project, j, such that $\epsilon_k < a_1$ prior to the attack.

Because the calculation of median depends on the number of elements, we now consider four cases:

Case: n is even, k is even

In this case, we see $n + k$ is even. Prior to the attack, the median value is equal to $\frac{a_j^{(m)} + a_j^{(m+1)}}{2}$. Because we now have k elements below our median, by the definition of median we see that our median value shifts down $\frac{k}{2}$ order statistics and is now $\frac{a_j^{(m-\frac{k}{2})} + a_j^{(m-\frac{k}{2}+1)}}{2}$, as we still have an even number of allocations and need to take the average of the allocations on either side.

This is bounded above by $a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}$, as we have:

$$a_j^{(m-\frac{k}{2})} \leq a_j^{(m-\lceil \frac{k}{2} \rceil + 1)} \implies \frac{a_j^{(m-\frac{k}{2})}}{2} \leq \frac{a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}}{2} \implies \frac{a_j^{(m-\frac{k}{2})} + a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}}{2} \leq a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}$$

Case: n is even, k is odd

In this case, we see $n + k$ is odd. Prior to the attack, the median value is equal to $\frac{a_j^{(m)} + a_j^{(m+1)}}{2}$. Because we now have k elements below our median, by the definition of median we see that our median value shifts down $\frac{k-1}{2}$ order statistics from a_j^m and is now $a_j^{(m-\frac{k-1}{2})}$.

This is bounded above by $a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}$.

Case: n is odd, k is even

In this case, we see $n + k$ is odd. Because we now have k elements below our median, by the definition of median we see that our median value, $a_j^{(m)}$, shifts down $\frac{k}{2}$ order statistics and is now $a_j^{(m-\frac{k}{2})}$.

This is bounded above by $a_j^{(m-\lceil \frac{k}{2} \rceil + 1)}$.

Case: n is odd, k is odd

In this case, we see $n + k$ is even. Because we now have k elements below our median, by the definition of median we see that our median value, $a_j^{(m)}$, shifts down $\frac{k-1}{2}$ order statistics and is now

$\frac{a_j^{(m-\frac{k-1}{2})}+a_j^{(m-\frac{k-1}{2}+1)}}{2}$.

This is bounded above by $a_j^{(m-\lceil\frac{k}{2}\rceil+1)}$.

Thus, we see this bound holds in all cases so the Median Phantom Vote is proven. While perhaps more difficult to quantify, the Phantom Vote attack may be more dangerous in this setting than mean voting. This is because if a large gap between two allocations is identified, it could be targeted by a sufficient number of adversaries to exploit, causing a much larger jump downward than we see with mean voting.

The efficacy of this strategy will ultimately depend on the allocations of votes; however, including zero token votes as a counter measure becomes substantially more difficult under this system. This is because if less than half of the votes are greater than zero, then a median allocation would be zero. For this reason, we may wish to implement alternative measures, such as a minimum allocation threshold in this case.

4.3 Quadratic Voting

First proposed by Lalley and Weyl [8], a Quadratic Voting mechanism reweighs votes by valuing them as the square root of the allocation. In the context of Optimism Governance, the Quadratic Voting Power can be expressed as:

$$\text{Project j's Mean Voting Power} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{A_{ij}}$$

This has many advantages with respect to economic optimality and decentralization; however, Quadratic Voting is difficult to implement in blockchain-based Quadratic Voting systems due to the possibility of a Sybil attack. Through it, an adversary could split their tokens among an arbitrarily large number of wallets and gain an outsize influence on the network.

With respect to Optimism's governance structure, however, this is not a concern as voters cannot split their token votes among some large number of wallets. That being said, Sybil issues still persist should multiple voters decide to collude. For example, two voters with strong preferences for one project who collude can artificially increase their voting power by a factor $\sqrt{2}$. To see this, consider the optimization problem.

Remark 4 *The Quadratic Collusion Attack – For two colluding voters, who care solely about the success of projects p and q, respectively, optimal collusion for both parties under Quadratic Voting allows them to each artificially increase their voting power by a factor of* $\sqrt{2}$.

Consider two colluding voters have a total of T tokens who can contribute (p_1, q_1) and (p_2, q_2) to projects p and q, respectively, yielding total contributions of $P = p_1 + p_2$ and $Q = q_1 + q_2$. Additionally, assume that these voters only care about the number of funds allocated to projects p and q, respectively. Holding all other allocations by voters constant, this yields the simplified utility functions:

- Voter 1 Utility: $u_1(p_1, q_1) = \sqrt{p_1} + \sqrt{p_2}$
- Voter 2 Utility: $u_2(p_2, q_2) = \sqrt{q_1} + \sqrt{q_2}$

To find optimal allocations for both colluding parties, we know that neither will voter will allow the other project to amass more of their collective resources, as this would be suboptimal since voting power increases with votes allocated. Thus, we know that $P = Q = T$.

Now, the task is to find the optimal allocation among p_1, p_2, q_1, and q_2. Our constraints only restrict the relationship between p_1 and p_2 as well as q_1 and q_2, so we can treat these as separate optimization problems and solve for the optimal value for our decision variables.

Starting with p_1 and p_2, we have the following optimization problem:

$$\max_{p_1,p_2} \sqrt{p_1} + \sqrt{p_2}$$

Subject to:

1. $p_1 + p_2 = P$
2. $p_1 \geq 0$
3. $p_2 \geq 0$

The second and third constraints are satisfied by our model definition. Additionally, by the first constraint, we can rewrite $p_2 = P - p_1$. Substituting this into our optimization problem, we obtain a single-variable objective function to maximize. Taking the first order conditions, we can solve for the optimal value as p_1 as follows:

$$\begin{aligned}
\max_{p_1} \sqrt{p_1} + \sqrt{P - p_1} &\implies \frac{\partial}{\partial p_1}(\sqrt{p_1} + \sqrt{P - p_1}) = 0 \\
&\implies (\frac{1}{2\sqrt{p_1}} - \frac{1}{2\sqrt{P - p_1}}) = 0 \\
&\implies \frac{1}{2\sqrt{p_1}} = \frac{1}{2\sqrt{P - p_1}} \\
&\implies \sqrt{p_1} = \sqrt{P - p_1} \\
&\implies p_1 = P - p_1 \\
&\implies 2p_1 = P \\
&\implies p_1 = \frac{P}{2}
\end{aligned}$$

From here, we see that the optimal allocation for $(p_1^*, p_2^*) = (\frac{P}{2}, \frac{P}{2})$, yielding the optimal utility:

$$u_1^*(p_1^*, p_2^*) = 2\sqrt{\frac{P}{2}} = \sqrt{2} \cdot \sqrt{P} = \sqrt{2} \cdot \sqrt{T}$$

Without loss of generality, the same argument can be applied in solving for the optimal $(q_1^*, q_2^*) = (\frac{Q}{2}, \frac{Q}{2})$, yielding the optimal utility:

$$u_2^*(q_1^*, q_2^*) = 2\sqrt{\frac{Q}{2}} = \sqrt{2} \cdot \sqrt{Q} = \sqrt{2} \cdot \sqrt{T}$$

In both cases, we see that the colluding parties, each of whom in theory have a voting power of $\sqrt{T}$, can multiply their voting power by a factor of $\sqrt{2}$ by agreeing to use half of their tokens to vote for the other project. □

5 Voting Mechnanism Simulations

5.1 Experimental Design

This study uses a simulation-based framework to evaluate the performance and robustness of various voting mechanisms in the context of Retroactive Public Goods Funding (RetroPGF) distributions. By simulating different scenarios, the study explores how these mechanisms allocate resources based on voter preferences and how they withstand adversarial behaviors such as collusion and phantom attacks.

Five distinct voting algorithms are implemented, each designed to aggregate voter preferences and distribute funding differently. The simulations examine critical factors such as alignment with voter intentions, resistance to strategic manipulation, and the ability to promote diversity in project funding. To ensure comprehensive analysis, the framework allows the testing of these mechanisms across varying community sizes, diverse voter preference distributions, and targeted attack scenarios, providing insights into their fairness, efficiency, and resilience.

5.2 Preference Matrix Construction

At the core of the simulation framework lies a voter preference matrix, denoted as M, which models the distribution of preferences across voters and projects. Each element of the matrix, $M_{v,p}$ represents the preference of voter v for project p, normalized such that the preferences for each voter sum to 1.

Formally, we can define this by,

$$M_{v,p} \in [0, 1] \quad \forall v, p \tag{1}$$

$$\sum_{p} M_{v,p} = 1 \quad \forall v \tag{2}$$

The matrix can be constructed both from empirical data, utilizing historical voting patterns, or generated synthetically based on specific distributions such as uniform or Gaussian. Each row of the matrix represents the preference distribution of an individual voter across all projects, while each column captures the collective preference for a particular project across all voters. This structure enables a detailed analysis of how voter preferences influence fund allocations under various voting mechanisms and scenarios.

5.3 Voting Mechanisms

The study investigates five distinct voting mechanisms, each employing a unique method for aggregating voter preferences and determining fund allocations. summarized in Table 1.

Table 1: Summary of Voting Mechanisms and Their Allocation Formulas

Mechanism	**Allocation Formula**
Single Selection Maximum	$A_p = \begin{cases} V_{num} & \text{if } p =_p (M_{v,p}) \\ 0 & \text{otherwise} \end{cases}$
Quadratic Voting (Standard)	$A_p = \sqrt{M_{v,p} \cdot V_{num}}$
Quadratic Voting (Collusion)	$A_p = \begin{cases} \sqrt{0.5 \cdot V_{num}} & \text{for } p \in \{_p(M_{v_1,p}),_p (M_{v_2,p})\} \\ 0 & \text{otherwise} \end{cases}$
Mean Voting	$A_p = \frac{\sum_v M_{v,p} \cdot V_{num}}{N}$
Mean Voting under Phantom Attack	$A_p = \begin{cases} \frac{\sum_v M_{v,p} \cdot V_{num}}{N+M} & \text{for attacked project} \\ \frac{\sum_v M_{v,p} \cdot V_{num}}{N} & \text{for other projects} \end{cases}$

Note: V_{num}: standardized voting power (100), N: total voters, M: attackers (10). A_p: voter allocations to project p.

The first mechanism, Single Selection Maximum, allocates all of a voter's voting power to their most preferred project, emphasizing singular project support and concentrating resources on top preferences. Quadratic Voting (Standard) introduces a proportional allocation model, where funds are distributed based on the square root of voter preferences. This mechanism seeks to balance strong individual preferences with a broader equitable allocation. A variation of this approach, Quadratic Voting (Collusion), simulates scenarios where voters strategically collaborate to increase their influence, distributing half of their voting power among colluding projects.

Mean Voting adopts a straightforward approach, averaging voter preferences to allocate funds proportionally across all projects. This mechanism provides simplicity and ease of understanding, making it suitable for general use cases. Finally, Mean Voting under Phantom Attack modifies the standard mean voting mechanism to evaluate the impact of adversarial conditions. It redistributes funds by accounting for phantom voters, artificially inflating support for targeted projects. Each mechanism is designed to address specific challenges in resource allocation, enabling a nuanced analysis of their strengths, weaknesses, and adaptability in dynamic and adversarial environments.

6 Results

6.1 Theoretical Evaluation

Drawing from the proofs of our voting model, we have a few observations:

Observation 1: Median Voting is More Vulnerable to Large-Scale Attacks than Mean Voting

While both measures attempt to find some notion of a "middle" allocation, notice that the upper bounds of our Mean Phantom Vote Attack and Median Phantom Vote Attack are quite different. The former relies on a normalized sum, while the ladder relies on an ordering of the allocations.

For this reason, each allocation will have an equal impact on the value of Mean Voting, while only the center allocations impact the value of a Median Voting, notwithstanding the actual order of the

allocations themselves. Thus, an attacker could cause much massive swings within a median value, especially if there are large gaps in value between the allocations.

To get an intuition for why this difference matters, consider the following set of allocations for a project: $[0, 0, 0, 0, 100, 100, 100, 100]$. Currently, both mechanisms yield a value of 50.

If an adversary contributes 0 to both projects, we now have $[0, 0, 0, 0, 100, 100, 100, 100]$, yielding:

Mean Voting = $\frac{1}{9}(5 \cdot 0 + 4 \cdot 100) = 44.\overline{4}$

Median Voting = 5th Ranked Value = 0

This is an extreme example, as median is unlikely to be swayed this much by a single vote; but in as we've shown in Section 4.2, many adversaries can attack to bring down the median. Additionally, we can't simply count zero-valued votes as we did in Mean Voting, as this would also bring down the median to a point where it could wash out all real allocations.

There have been some clever workarounds proposed to address other issues in Median Voting, such as the use of Capped Median Rule and Moving Phantoms Rules that could provide some sort of stopgap against this attack due to making this drop off less severe. Additionally, raising the threshold of a minimum allocation could make this sort of attack prohibitively expensive; but for now, the threat remains.

Observation 2: Quadratic Voting Exploitation Occurs with Repeated Voters

While the prospect of increasing vote power by a factor $\sqrt{2}$ is quite substantial, this only works so long as two delegates have an incentive to collude.

If the same delegates are consistently used for every vote, over time the benefits of collusion could outweigh the benefits of betraying the effort. Its possible that this repeated collusion never reaches a Nash Equilibrium; but if the delegates are frequently switched, it becomes substantially less likely as the delegates would almost certainly value future collusion less.

At a high-level, if a colluder is going to contribute half of their tokens to the project that someone wants, is it more valuable to give half of their stake to the other project in hopes of future collusion? If future collusion isn't a sure thing, that's probably not going to be a wise choice. Thus, changing out the group of delegates for Quadratic Voting substantially lowers the risk of this collusion.

6.2 Simulation Results

The simulation results reveal distinct patterns in the performance of each voting mechanism, offering valuable insights into their fairness, diversity, and resilience to manipulation. One key metric, the Number of Different Project Winners, measures the diversity of funded projects across the eight voting mechanisms. Mechanisms like Quadratic Voting (Standard) and Mean Voting excelled in promoting diversity, resulting in a broader range of funded projects. This reflects their ability to balance voter preferences across a wider spectrum. In contrast, Single Selection Maximum concentrated resources on a narrow set of highly popular projects, limiting diversity and disproportionately favoring dominant preferences.

The Difference between Quadratic Voting and Voter Attack highlights the vulnerability of Quadratic Voting to collusion among voters. When groups of voters coordinated their votes, the outcomes shifted significantly compared to normal Quadratic Voting, exposing its susceptibility to strategic manipulation. Similarly, the Difference between Quadratic Voting and Project Attack demonstrated that project-level collusion—where multiple projects collaborate to influence results—can substantially distort allocations, reducing the fairness and intended proportionality of this mechanism.

For Mean Voting, the results under the Difference between Mean Voting and Voter Attack revealed a pronounced impact from voter collusion. The averaging process, while straightforward, amplified the influence of coordinated voting, making this mechanism highly sensitive to strategic voter behavior. Furthermore, the Difference between Mean Voting and Project Attack showed that project-level

collusion could significantly skew results, with coordinated efforts among projects exploiting the simplicity of Mean Voting to disproportionately gain from the aggregated preferences.

Overall, the simulation results underscore that while mechanisms such as Quadratic Voting and Mean Voting are effective under normal conditions for promoting fairness and diversity, their performance degrades substantially under adversarial conditions. These findings emphasize the importance of implementing anti-collusion safeguards and designing mechanisms that are inherently resistant to strategic attacks.

6.3 Practical Recommendations

As we have shown by proof and simulation, without proper safeguards delegates could easily manipulate these voting mechanisms. That being said, there are some methods to decrease the efficacy of these attacks, which we summarize below.

Mean Voting

Because Mean Voting Power scales linearly with the amount of tokens committed, the only "fix" needed is to ensure this process is not disrupted. As discussed, this means that we need to include all votes when calculating the mean vote total, even those with a zero allocation.

Median Voting

Given the strong potential for manipulation the Median Phantom Vote Attack, we recommend caution with Median Voting as its exploitation could be quite damaging. It may be that the solution lies in some weighting between Median Voting and other styles, however, safeguards should be put in place if it is to be used on its own.

Quadratic Voting

If Quadratic Voting is the mechanism of choice, ensure that voters are frequently rotated. Otherwise, we risk collusion that can give outsized influence in the mechanism.

7 Further Considerations

Our analysis of voting mechanisms in Optimism's RetroPGF system has revealed several critical areas that warrant deeper investigation. The challenges we identified with median voting, particularly its susceptibility to the Phantom Vote Attack, suggest that it may be difficult to balance. Although Optimism has attempted to use various readjustments in Round 3 and Round 4 to compensate for some of median votings' flaws, these measures run the risk of overcomplicating the voting process and possibly introducing novel ways for adversarial actors to manipulate the voting process. Potential areas of future work would offer more formal analyses of these countermeasures.

Regarding quadratic voting, we want to highlight the critical importance of voter rotation in preventing consistent collusion. Future work can look into practical ways of implementing such rotation systems, balancing potential benefits of anti-collusion against the importance of maintaining expertise knowledge. This balance may be particularly crucial in the case of retroactive funding systems where there requires more sophisticated understanding of specific verticals and projects, and there may be a limited pool of informed voters.

Interestingly, despite its simplicity, linear voting represents a remarkable robust baseline against other mechanisms. While much attention in mechanism design has focused on developing increasingly sophisticated voting systems, our findings suggest that the straightforward nature of linear voting provides several understated advantages. Its resistance to manipulation through phantom voting (when properly implemented to include zero votes), combined with its intuitive understanding by

voters, makes it a more robust choice than initially assumed. This observation challenges the common assumption that more complex mechanisms necessarily yield better results.

The strength of linear voting as a baseline raises important questions about the tradeoffs between mechanism sophistication and practical effectiveness. While quadratic voting provides theoretical benefits in terms of preference intensity expression, and median voting offers potential advantages for outlier resistance, the implementation complexities and attack vectors of these systems must be carefully weighed against the reliable performance of linear approaches. Future research should more thoroughly examine whether the marginal benefits of sophisticated voting mechanisms justify their increased complexity and potential vulnerabilities, particularly in high-stakes environments like retroactive funding.

References

[1] Median voter theorem. Accessed on: December 5, 2024.

[2] James Austgen, Andrés Fábrega, Sarah Allen, Kushal Babel, Mahimna Kelkar, and Ari Juels. DAO Decentralization: Voting-Bloc Entropy, Bribery, and Dark DAOs. *arXiv preprint*, 2023. https://arxiv.org/abs/2311.03530.

[3] DeepDAO. DeepDAO Organizations. https://deepdao.io/organizations. Accessed 2024-11-02.

[4] Compound Docs. Governance. https://docs.compound.finance/v2/governance/. Accessed 2024-10-22.

[5] Optimism Citizen House Docs. How retro funding works. https://community.optimism.io/citizens-house/how-retro-funding-works. Accessed 2024-11-02.

[6] Andrew B. Hall and Eliza R. Oak. What Kinds of Incentives Encourage Participation in Democracy? Evidence from a Massive Online Governance Experiment. 2023. https://andrewbenjaminhall.com/Hall_Oak_Airdrop_Effects_on_Participation.pdf.

[7] Samer Hassan and Primavera De Filippi. Decentralized autonomous organization. *Internet Policy Review*, 10(2):1–10, 2021.

[8] Steven Lalley and E. Glen Weyl. Quadratic voting: How mechanism design can radicalize democracy. *American Economic Association Papers and Proceedings*, 1(1):1–5, 2018.

[9] Eric A. Posner and E. Glen Weyl. Voting squared: Quadratic voting in democratic politics. Technical Report Working Paper No. 657, Coase-Sandor Institute for Law and Economics, University of Chicago, 2014. Accessed on: December 5, 2024.

[10] Eric A. Posner and E. Glen Weyl. Quadratic voting and the public good, 2017. Accessed on: December 5, 2024.