Binary Decisions in DAOs: Accountability and Belief Aggregation via Linear Opinion Pools

Nuno Braz
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa

Miguel Correia
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa

Abstract

We study binary decision-making in governance councils of Decentralized Autonomous Organizations (DAOs), where a group of experts must choose between two alternatives on behalf of the organization. We introduce an information structure model for such councils and formalize desired properties in blockchain governance. Building on these foundations, we propose a decision-making mechanism where we assume the availability of an evaluation tool that ex-post returns a boolean indicating success or failure. Smart contracts allow the implementation of this mechanism in practice.

In our model, experts hold two types of private information: idiosyncratic preferences over the alternatives and subjective beliefs about which alternative is more likely to benefit the organization. The designer's objective is to select the best alternative for the organization, which we assume is given by aggregating the expert beliefs. Framing it as a classification problem rather than a welfare maximization problem.

The mechanism collects the preferences of the experts and computes monetary transfers for each participant accordingly. It then applies other monetary transfers contingent on the boolean returned by the evaluation tool. For aligned experts, whose preferences and beliefs agree, the mechanism is a dominant strategy incentive compatible. For unaligned experts, we prove a Safe Deviation property: no expert can profitably deviate from a known strategy dependent on the expert's private information toward an alternative they individually believe is less likely to succeed.

Our main result decomposes the sum of the reports into idiosyncratic noise and a linearly pooled belief signal whose sign matches the designer's optimal decision. The pooling weights arise endogenously from the equilibrium strategies, and the mechanism achieves correct classification whenever the per-expert budget exceeds a threshold that decreases as experts' beliefs converge.

2012 ACM Subject Classification Theory of computation → Algorithmic mechanism design; Applied computing → Electronic commerce

Keywords and phrases Mechanism Design, Decentralized Autonomous Organizations, Accountability, VCG, Governance Councils, Information Aggregation, Smart Contracts

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

1 Introduction

A Decentralized Autonomous Organization (DAO) is an organization governed by rules encoded as smart contracts on a blockchain, where members coordinate and make collective decisions without centralized authority. In recent years, DAOs have grown into large-scale entities where thousands of members coordinate pseudonymously. Unlike traditional contracts, which merely articulate rules and can only serve as legal proof, smart contracts can execute the rules they encode. The scale of these organizations creates a significant challenge in decision-making. [8, 31, 21] confirms that these organizations suffer from low participation rates and slow execution, validating the need for delegated governance structures. Delegating decision-making to smaller groups or *governance councils* ends up being the solution to this problem. In large blockchains such as Bitcoin and Ethereum, the decision-making is done by the maintainers [10, 3]. Polkadot elects a council [28]. Project Catalyst has a Community

42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:26
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Advisors board and a veteran [29]. Furthermore, even when Governance Councils are not elected, decision power is concentrated in a small number of members [21]. Small councils also help mitigate collusion: while coordinated manipulation remains a concern in any mechanism, a small group is easier to monitor, and heavy penalties can be imposed when collusive behavior is detected.

Restricting power to a small group creates a natural problem: experts often have idiosyncratic preferences that are not aligned with the organization's goal. For example, a council governing a lending DAO might vote to approve a highly risky token as collateral simply because they personally own large amounts of it, exposing users to severe financial risk.

We identify two primary pathways for evolving DAO governance beyond its current limitations. The first, and the one pursued in this work, is the design of decision-making mechanisms for governance councils that ensure their members are directly accountable for the impact of their decisions on the broader community. The second focuses on developing community-wide decision-making mechanisms that enable effective information aggregation to reach better decisions. Notable examples of the latter include decision markets, as implemented by MetaDAO [25], and holographic consensus, as employed by DAOStack [9].

Kiayias et al. defines two essential properties regarding blockchain governance [20]. Although their work focuses on blockchain governance—a subcategory of DAO —these properties are readily extendable to the broader DAO context.

▶ **Definition 1** (Accountability (informal)). *A DAO satisfies the property of Accountability if, whenever participants bring in a change, they are held individually responsible for it in a clearly defined way by the platform.*

▶ **Definition 2** (Sustainable Participation (informal)). *A DAO sustains participation if it incentivizes, via monetary rewards or otherwise, participants who participate in the decision-making process of the platform.*

These properties are in regard to the whole governance protocol, not the council's internal decision-making mechanism. For example, according to the evaluation performed in that study, only Polkadot fully satisfies Accountability at the protocol level, as it does so by forcing users who vote in favor of a proposal to lock tokens until the proposal is deployed. A proposal is deployed if it is approved by the council. An unsolved issue, however, lies in the council itself: even if its members receive some direct compensation, they are not directly accountable for the changes they bring. This is because Polkadot uses majority voting as a decision-making mechanism in the council of experts. In Appendix A, we show that this mechanism does not satisfy Accountability.

These properties are presented as an overview in the original article, but they remain largely informal. Furthermore, they do not consider the decision-making mechanism used by the council. This paper bridges this gap by providing a more formal treatment of these properties regarding the decision-making mechanism in the councils. Sustainable participation is an informal statement of a known property in *mechanism design theory*: strict Individual Rationality. Accountability, by contrast, is seldom addressed within this area.

In an ideal decision-making mechanism, experts whose votes are pivotal—that is, whose votes actually change the outcome—should be rewarded when that outcome is positive and penalized when it is negative. To achieve this, mechanisms such as Decision Markets condition transfers on the observed outcome. This has been done in practice in DAOStack and MetaDAO [9, 25].

1.1 Model Overview and Contributions

We model each expert as possessing two independent, private types of information: *idiosyncratic preferences* —how much they personally stand to gain or lose from each alternative and *subjective beliefs*—their private assessment of which alternative will improve the organization's Key Performance Indicators (KPIs). Unlike standard mechanism design, where the goal is to maximize the sum of idiosyncratic utilities, we frame decision-making as a classification problem: the designer seeks to align the decision with the aggregate expert beliefs, not with the aggregate preferences.

A key distinction arises from the interplay between these two types of private information. An expert is *aligned* when their preferences and beliefs point in the same direction, and *unaligned* when they conflict. This distinction is central to our results, as the mechanism's incentive guarantees differ across the two cases. The full model is presented in Section 3.

Our contributions are:

1. We formalize Accountability and Sustainable Participation as mechanism design properties for governance councils, bridging the gap between the informal governance requirements identified in [20] and classical mechanism design theory. We formalize the designer's objective as a classification problem: align the decision with the aggregate expert beliefs rather than maximize the sum of idiosyncratic utilities.
2. We propose a decision-making mechanism that combines an incentive-compatible transfer rule (Vickrey-Clarke-Groves) with an outcome-contingent reward rule, and prove that it satisfies a series of properties that are desirable for decision-making in DAOs.
3. We show that the mechanism is Dominant Strategy Incentive Compatible (DSIC) for aligned experts. For unaligned experts, we prove a Safe Deviation property: no expert can profitably deviate from a known strategy dependent on the expert's private information toward an alternative they individually believe is less likely to succeed, regardless of other experts' reports.
4. We establish the main result on information aggregation (Theorem 17): the mechanism aggregates experts' information in a way that filters personal preferences, so that the final decision is driven by the experts' beliefs about which alternative benefits the organization.

This work addresses binary decisions exclusively, although the properties are extensible to a setting with more alternatives. The majority of DAO decisions fall within this category: based on the data provided in [36], approximately 66.76% of the analyzed DAO proposals utilized a binary voting pattern. Some scenarios necessitate a wider range of alternatives. An organization does not need to commit to a single decision-making mechanism; rather, it can adapt its approach depending on the specific context. We leave the development of a mechanism that satisfies these properties for cases with more than two alternatives to future work. In this regard, we note that in mechanism design theory, Roberts' Theorem [32] establishes a barrier regarding design possibilities, distinguishing between the binary setting and the settings with three or more alternatives.

The remainder of the paper is organized as follows. Section 2 surveys related work on DAO governance, decision scoring rules, information aggregation, and existing decision-making mechanisms. Section 3 introduces the model: the information structure, the assumptions on expert types and beliefs, the expanded mechanism space with its three components (allocation rule, transfer rule, and ex-post reward rule), a review of the Pivotal Mechanism on which our construction builds, the designer's objective framed as a classification problem, and the formal properties required for the mechanism. Section 4 characterizes the reward rule, presents the complete mechanism construction, and analyzes its computational and on-chain complexity. Section 5 proves the structural and incentive properties, states the main information aggregation result (Theorem 17), and illustrates the mechanism with a worked example. Section 6 discusses limitations and directions for future work.

2 Related Work

This work sits at the intersection of two fields: blockchain governance and mechanism design focusing on information aggregation. The blockchain governance literature identifies desirable properties for DAO governance protocols but lacks mechanisms for decison-making that comply with these properties. Mechanism design provides the formal tools to construct such mechanisms but has not been tailored to the DAO setting, where experts hold both private preferences and private beliefs about organizational outcomes. We organize the related work accordingly: we first review governance in DAOs, then cover Decision Markets and Decision Scoring Rules—the mechanisms most comparable to ours, as they are the only ones that condition transfers on the observed outcome—followed by information aggregation and a survey of standard decision-making mechanisms.

2.1 Governance in DAOs

The research on Decentralized Autonomous Organizations has shifted from simple treasury management [39] toward more complex coordination problems [34]. [39] proposed a complete design of a treasury system for decentralized organizations, focusing on the infrastructure rather than the decision-making mechanism. [34] identifies DAOs as an important application domain for Multi-Agent Systems (MAS) and calls for the MAS community to address their governance challenges. Our work answers this call by bridging informal governance requirements in real DAO use cases with formal mechanism design.

On the analytical side, [21] studies contributors' influence and voting power shifts. [17] provides a formal economic analysis of DAO governance, examining how token-weighted voting creates conflicts between large token holders and minorities. [36] conducts an empirical study on governance dynamics. These analyses collectively confirm that simple token-based voting is insufficient when participants have heterogeneous stakes and information, motivating the need for more sophisticated mechanisms.

Focusing on the second pathway, decision-making mechanisms for the whole community, [13] provides an overview of decision-making mechanisms in DAOs, highlighting the limitations of one-token-one-vote systems. A notable decision-making mechanism currently being adopted in DAOs is Quadratic Voting [22], [5]. It describes an optimal vote pricing rule, where experts buy votes and allocate them to the candidates according to a pricing rule. This pricing rule is, not surprisingly, quadratic. [12] explores the utility of quadratic voting to capture preference intensity, and [37] proposes quadratic voting with staking to deter power concentration and collusion in DAOs. However, QV does not aggregate the participants' beliefs, which makes it suitable only when the community is, in general, aligned with DAO's interests.

Other mechanisms include conviction voting [14], where the influence of a participant's vote grows the longer their tokens remain staked on a specific proposal.

2.2 Decision Markets and Scoring Rules

Decision Scoring Rules (DSRs) and decision markets are the closest related work to our mechanism, as they are the only frameworks that condition transfers on the realized outcome of the chosen action. A DSR consists of a scoring function that rewards experts based on their reports and the actual result, together with a decision rule that selects an action based on those reports. A decision market achieves the same goal through trading: experts trade the possibility of being rewarded by the scoring rule, and the market prices guide the principal's action.

The idea of using markets to guide governance decisions originates with Hanson's futarchy proposal [18], where democratic voting sets values and prediction markets guide policy. MetaDAO [25] implements a version of this on-chain. Holographic consensus used by the DAOstack [9] adds a prediction market layer where participants stake tokens to "boost" specific proposals they believe are high-quality. Once a proposal is boosted, the requirement for a massive community-wide quorum is replaced by a relative majority of active voters. This approach effectively bridges our two identified pathways: it uses the collective intelligence of the market to filter information, while delegating the final decision to a smaller, more efficient group. Holographic Consensus still faces the challenge of ensuring that those making the final decision remain accountable for the long-term outcomes of their votes. An analysis done by [11] noted that the system became self referential. The collective intelligence of the market never showed up and proposals became boosted not because they are deemed relevant, but because their authors wish to see them approved. This rise and fall of the DAOstack reinforces the need to account for idiosyncratic preferences of participants.

The central challenge in this literature is strict properness: ensuring that honest reporting is the unique optimal strategy. Because only the outcome of the chosen action is ever observed, the expert's reward cannot depend on counterfactual outcomes, making it impossible to strictly incentivize truthful reports about alternatives that were not selected under deterministic decision rules.

[27] introduced decision markets, where the outcome to be predicted depends on the action taken, and characterized the set of scoring rules compatible with decision-making. [4] considers a single self-interested expert whose idiosyncratic preferences over decisions create an incentive to misreport. Boutilier introduces *compensation rules* that offset the expert's utility, inducing proper scoring rules despite the conflict of interest. When the principal knows the expert's utility function, a complete characterization of proper compensation rules is provided; under uncertainty, bounds on misreporting incentives are derived. [6] proposed randomized decision rules that restore good incentive properties in market-based settings, overcoming the impossibility of strict properness under deterministic rules. [7] extended this line of work to the joint elicitation of predictions and recommendations, where the principal seeks both forecasts and actionable advice. [26] formalized DSRs as the single-expert analogue of decision markets and established tight conditions for the existence of strictly proper scoring rules in decision problems.

A recent development is [33], which designs incentive-compatible prediction markets that elicit and aggregate beliefs *without* observing the true outcome, paying experts based on agreement with a carefully chosen reference expert. This relaxes the outcome-observation requirement that our mechanism shares with classical DSRs, at the cost of weaker incentive guarantees.

Our mechanism departs from this literature in two key ways. First, the works above are designed for a single expert or for settings where experts interact through a market, and besides [4], do not account for idiosyncratic preferences; our setting involves multiple self-interested experts who simultaneously submit reports that are aggregated via a deterministic allocation rule. Second, none of these mechanisms employ VCG transfers, and therefore they cannot satisfy Accountability — they reward accurate forecasts but do not penalize experts whose decisive actions lead to negative outcomes.

2.3 Information Aggregation

Information aggregation refers to the problem of combining the dispersed private beliefs of multiple experts into a single collective estimate that guides the decision. Our mechanism aggregates experts' beliefs via a weighted linear combination $\sum w_i(p_i^A - p_i^B)$, a form of *linear opinion pooling*. Simple linear averages have been shown to be remarkably robust across empirical settings [38], and are well-calibrated [30]. Aczél and Wagner [1] provided a foundational characterization of weighted arithmetic.

Beyond the choice of aggregation method, a foundational question in committee decision-making is whether voting alone can aggregate dispersed private information. [2] shows that sincere voting is generally not a Nash equilibrium, even with identical preferences, and [15] shows that unanimous rules perform worse than majority rule under strategic voting. When preferences conflict, [24] shows that equilibrium strategies necessarily garble private information under any voting rule without transfers. These results motivate transfer-based mechanisms.

2.4 Decision Making Mechanisms

Several mechanisms have been proposed for aggregating preferences in binary decision settings, each with different trade-offs regarding incentive compatibility, information elicitation, and budget requirements. However, most of these mechanisms do not account for the distinction between experts' idiosyncratic preferences and their beliefs about which alternative benefits the organization—a distinction that is central to governance in DAOs. We briefly describe the most commonly used mechanisms for binary decision-making.

- **Majority Voting (MV):** each expert is asked to report their preference for one of the alternatives. An alternative is selected if it receives more than half of the total votes cast.
- **Quadratic Voting (QV)** each expert can buy multiple votes for their preferred alternative by paying the square of the number of votes purchased i.e, casting n votes costs n^2.
- **Pivotal Mechanism (VCG)** experts are asked to report their valuation of each one of the alternatives and the experts whose specific report flips the outcome are charged a tax equal to the net loss in utility their report caused on others.
- **Decision Scoring Rules (DSR)** mathematical functions used to evaluate the quality of probabilistic forecasts by measuring how closely predicted distributions align with actual outcomes. Experts are rewarded based on the outcome and their prediction.
- **Decision Markets (DM)** use trading to predict which alternative will lead to the best outcome. Experts bet on the outcomes of different alternatives.

Table 1 compares decision-making mechanisms along the following dimensions: whether the mechanism elicits experts' subjective beliefs about the success of each alternative (*Belief Elicitation*); whether it mitigates the influence of experts' private preferences on the collective decision (*Reduce Idiosyncratic Noise*); whether it supports multiple simultaneous participants (*Multi-Agent*); whether it operates in a single round (*One-Shot*); whether it functions without external funding (*No Subsidies*); whether total transfers sum to zero (*Budget Balanced*); and whether it operates without observing the realized outcome (*No Outcome Observation*).

3 Model

In this section, we present our model and the necessary background. We consider a DAO facing a choice between two mutually exclusive alternatives $\{A, B\}$. A set $N = \{1, \ldots, n\}$

Feature	This	MV	QV	VCG	DSR	DM
Belief Elicitation	●	○	○	○	●	●
Reduce Idiosyncratic Noise	●	○	○	○	◐†	◐†
Multi-Agent	●	●	●	●	○	●
One-Shot	●	●	●	●	○	○
No subsidies	○	●	●	●	○	○
Budget Balanced	○	●	○	○	○	○
No Outcome Observation	○	●	●	●	○	○

Table 1 Comparison of decision-making mechanisms in decentralized organizations. ● = satisfied, ◐ = partially satisfied, ○ = not satisfied. †Outcome-contingent subsidies can indirectly mitigate idiosyncratic noise, but these mechanisms were not designed for this purpose.

of n risk-neutral experts are expected utility maximizers in the sense of the Von Neumann–Morgenstern framework [35]. The rules of the mechanism are common knowledge. We focus on environments where the number of experts is insufficient for traditional decentralized prediction markets to function effectively, necessitating a formal mechanism design approach.

3.1 Information Structure

Each expert $i \in N$ is endowed with a private type—a collection of private information that fully characterizes the expert—comprising their idiosyncratic preferences and subjective beliefs.

The expert's idiosyncratic utility is captured by the pair $(\theta_i^A, \theta_i^B) \in \Theta^2$, where θ_i^A and θ_i^B represent the expert's personal utility derived from the implementation of option A and option B, respectively. We assume that at least one of these alternatives yields a non-negative idiosyncratic utility for the expert. Since exactly one of these options will be implemented, the expert can normalize their preference parameters accordingly.

The expert's subjective beliefs are represented by the pair $(p_i^A, p_i^B) \in (0,1)^2$, where p_i^A and p_i^B denote the expert's private assessment of the probability that alternatives A and B will lead to a positive outcome for the organization, respectively. We exclude the cases where the experts are certain.

The expert's type is given by $\tau_i \in \mathcal{T}_i$, where $\tau_i = (\theta_i^A, \theta_i^B, p_i^A, p_i^B)$. We denote the profile of all experts' types by $\tau = (\tau_1, \dots, \tau_n)$ and the belief profile by $\mathbf{p} = (p_1^A, p_1^B, \dots, p_n^A, p_n^B)$. We operate under the Independent Private Values (IPV) model: each expert's type τ_i is drawn independently from a distribution over the type space $\mathcal{T}_i$ with some probability distribution. To clarify this assumption, we contrast our setting with the classical mineral rights auction [19]. There, each bidder receives a private geological signal, but the true value of the tract is a common function of all signals. Learning another bidder's signal would cause a bidder to revise their own valuation. A correlated values model is necessary precisely because no public deliberation stage resolves the common-value uncertainty before bids are placed. In DAO governance the mechanism is preceded by forum discussions, temperature checks, and community calls that generate a common public signal observed by all experts. Each expert absorbs this shared information and updates accordingly. After discussion, an expert's private signal—reflecting domain expertise, privileged data, or personal judgment that the public discussion did not reveal—carries no information about another expert's signal: learning it would not cause a further update.

We do not assume that the two pairs (θ_i^A, θ_i^B) and (p_i^A, p_i^B) are correlated experts tend to benefit from the options they believe will succeed which aligns private incentives with informational contributions and makes the designer's task easier. Independence is therefore a conservative assumption.

We assume the availability of an ex-post evaluation tool deployed via smart contracts. When a DAO is formed, its founding members agree on a set of key performance indicators (KPIs)—for example, treasury growth, active users, or community feedback—and code these metrics into the governance smart contract together with the evaluation tool, which monitors them over a specified time horizon. Following the implementation of a binary decision $a \in \{A, B\}$, this tool deterministically resolves the performance data into a boolean state of the world, $\Delta \in \{-1, 1\}$, where $\Delta = 1$ denotes a positive impact and $\Delta = -1$ denotes a negative impact. For instance, a DeFi protocol voting on a liquidity incentive program may track whether Total Value Locked (TVL) exceeds a target threshold after six months, or a DAO choosing between two grant proposals may resolve Δ via a community satisfaction vote. In practice, DAOs are already implementing such tools: MetaDAO uses the DAO's own token price as the KPI [25], while other DAOs allow the community to express appreciation for a proposal's realized impact. More generally, peer prediction methods [**?**] can be used to elicit the community's assessment of Δ without access to ground truth, and Schelling-point mechanisms such as Kleros [23] already resolve on-chain disputes by rewarding jurors who agree with the majority verdict. This tool enables the execution of contingent transfers, binding the experts' payoffs to the consequences of their actions. The expert's beliefs (p_i^A, p_i^B) are formally defined as the conditional probabilities of observing a positive outcome given the chosen alternative:

$$p_i^A = \mathbb{P}_i(\Delta > 0 \mid a = A)$$

$$p_i^B = \mathbb{P}_i(\Delta > 0 \mid a = B)$$

We assume that the experts' beliefs, when appropriately aggregated, preserve the true ranking of the alternatives. That is, if the aggregate of the experts' private assessments favors alternative A over B, then A is indeed more likely to yield a positive outcome. The assumption does not require individual experts to agree on which alternative is better; experts may hold opposing beliefs. It only requires that the aggregate is informative about the true ranking.

3.2 Mechanism Space

A direct revelation mechanism asks each participant to report their private information and then determines an outcome and monetary transfers as a function of the reported profile. Formally, it consists of two components: an *allocation rule*, which maps reports to an outcome, and a *transfer rule*, which determines the payment each participant makes or receives based on the reports. However, satisfying Accountability requires conditioning payments on the realized outcome, which the standard two-component framework cannot express. We therefore introduce a third component—an ex-post reward rule—that distributes payments after the outcome is observed.

In our setting, the mechanism proceeds in two stages. First, each expert submits a message encoding the intensity of their preferences. The mechanism aggregates these messages, selects an alternative, and charges each expert a transfer that depends only on the submitted messages. Second, after the outcome is observed, the mechanism distributes an ex-post reward that depends on whether the decision led to a positive or negative outcome. The expert's payoff is therefore determined by three factors: which option is chosen, how much they pay upfront, and what reward they receive once the outcome is revealed.

All experts share a common message space $\mathcal{R}$, where each message $m_i \in \mathcal{R}$ is a real number encoding how strongly the expert favors one alternative over the other. A profile of messages submitted by all experts is denoted by $\mathbf{m} = (m_1, \ldots, m_n) \in \mathcal{R}^n$. We write $\mathbf{m}_{-i}$ for the profile of all experts except i, so that $\mathbf{m} = (m_i, \mathbf{m}_{-i})$.

We formally define a mechanism in this setting as a triplet $\mathcal{M} = (x, t, r)$. We use $\mathbf{t}, \mathbf{r}$ to refer to vectors denoting the application of each rule:

- $x : \mathcal{R}^n \to \{A, B\}$ is the deterministic allocation rule representing the collective decision.
- $\mathbf{t} = (t_1, \ldots, t_n)$ is the transfer rule, where $t_i : \mathcal{R}^n \to \mathbb{R}$ represents a monetary transfer to expert i based solely on the submitted messages. A positive transfer means the expert receives money, while a negative transfer means they pay.
- $\mathbf{r} = (r_1, \ldots, r_n)$ is the reward rule, where $r_i : \mathcal{R}^n \times \{-1, 1\} \to \mathbb{R}$ represents the ex-post reward distributed to expert i, contingent on the realized state of the world Δ.

Given a type $\tau_i = (\theta_i^A, \theta_i^B, p_i^A, p_i^B)$ and a profile of submitted messages $\mathbf{m}$, let $a = x(\mathbf{m})$ denote the chosen alternative. The expected utility of expert i reflects the three factors described above:

$$\mathbb{E}[u_i(\mathbf{m} \mid \tau_i)] = \theta_i^a + t_i(\mathbf{m}) + \mathbb{E}_\Delta[r_i(\mathbf{m}, \Delta) \mid p_i^a] \tag{1}$$

Figure 1 illustrates the overall flow: experts submit messages, the mechanism selects an alternative and applies transfers, and after the evaluation period the ex-post reward is distributed based on the realized outcome.

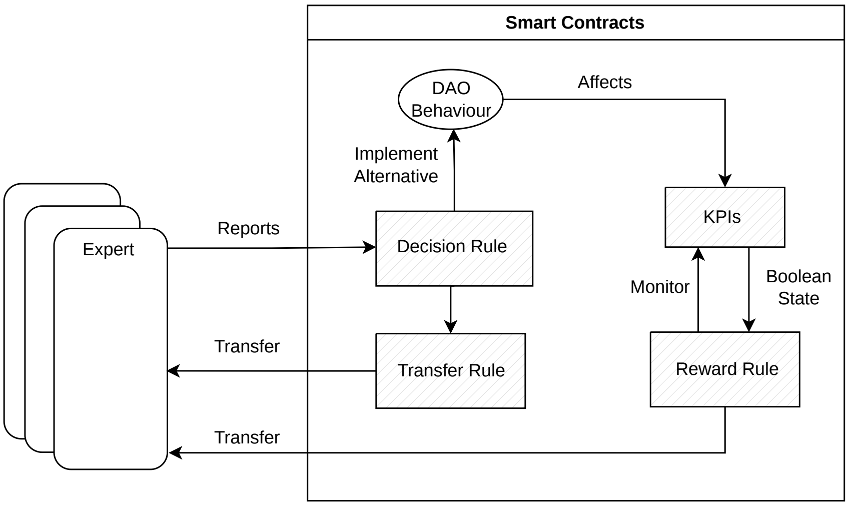

Figure 1 Mechanism Overview

3.3 Background: The Pivotal Mechanism

The Vickrey–Clarke–Groves (VCG) is a class of mechanisms that achieve DSIC. The Pivotal Mechanism [16] is an instance of this class, which achieves DSIC by charging each participant the externality they impose on others, and our mechanism builds on it. We briefly review its construction. Let $v_i(a)$ be the expert's valuation for alternative $a \in \{A, B\}$. The allocation rule x selects the alternative that maximizes the reported utilitarian welfare:

$$x(\mathbf{m}) \in \arg\max_{a \in \{A,B\}} \sum_{i \in N} v_i(a) \tag{2}$$

The Pivotal Mechanism ensures DSIC by charging each expert the externality they impose on others. The transfer t_i—the payment expert i makes based on the submitted messages—is defined as:

$$t_i(\mathbf{m}) = \sum_{j\neq i} v_j(x(\mathbf{m})) - \max_{a\in\{A,B\}} \left(\sum_{j\neq i} v_j(a) \right) \tag{3}$$

One way to implement this allocation rule in a binary setting is by asking the experts to submit a scalar $m_i = \theta_i^A - \theta_i^B$ and selecting Option A if the sum of messages is non-negative, and Option B otherwise:

$$x(\mathbf{m}) = \begin{cases} A & \text{if } \sum_{i=1}^n m_i \geq 0 \\ B & \text{otherwise} \end{cases}$$

The tie-breaking convention $x = A$ when $\sum m_i = 0$ is without loss of generality: relabelling the alternatives yields the symmetric rule, and on a continuous message space the event $\sum m_i = 0$ has probability zero. If an expert is not pivotal, i.e., their reported preferences do not change the chosen outcome $x^*(\mathbf{m})$, then their transfer is exactly 0. Under this allocation rule, the transfer simplifies to:

$$t_i(m) = \begin{cases} -\sum_{j\neq i} m_j & \text{if } -m_i > \sum_{j\neq i} m_j > 0 \\ \sum_{j\neq i} m_j & \text{if } -m_i < \sum_{j\neq i} m_j < 0 \\ 0 & \text{otherwise} \end{cases}$$

The Pivotal Mechanism is insufficient for our environment since it ignores the experts' beliefs (p_i^A, p_i^B), failing to leverage the experts' predictive knowledge regarding the DAO's success metric. To resolve these limitations, our mechanism expands upon this baseline by introducing the ex-post reward function $\mathbf{r}$.

3.4 Designer's Objective

We now define what the mechanism designer is trying to achieve. Unlike standard mechanism design settings that seek to maximize the sum of idiosyncratic utilities, the goal here is to select the alternative most likely to yield a positive outcome for the organization, using the experts' dispersed beliefs while filtering out idiosyncratic noise. Formally, let $\mathbb{P}(\Delta = 1 \mid a, \mathbf{p}^a)$ denote the probability that decision $a \in \{A, B\}$ yields a positive outcome, conditional on the aggregate beliefs of all experts. If the designer had access to all private information, their optimal decision rule $x^*(\mathbf{p})$ would be:

$$x^*(\mathbf{p}) \in \arg\max_{a\in\{A,B\}} \mathbb{P}(\Delta = 1 \mid a, \mathbf{p})$$

By the concordance assumption, the best alternative agrees with the sign of a weighted belief aggregate $\sum_{i=1}^n w_i(p_i^A - p_i^B)$. The optimal rule therefore takes the form:

$$x^*(\mathbf{p}) = \begin{cases} A & \text{if} \quad \sum_{i=1}^n w_i(p_i^A - p_i^B) > 0 \\ B & \text{if} \quad \sum_{i=1}^n w_i(p_i^A - p_i^B) < 0 \end{cases}$$

where $w_i = \frac{1}{1-p_i^{(\cdot)}}$ and $p_i^{(\cdot)}$ denotes the probability of success of expert i's preferred alternative. As we show in Section 5.2, these are precisely the weights induced by the mechanism's equilibrium strategies. The designer does not directly observe the true beliefs $\mathbf{p}$. Therefore, the designer must construct a mechanism $\mathcal{M} = (x, t, r)$ to elicit a message profile $\mathbf{m} \in M$ that minimizes the deviation between the aggregated reports and the true aggregate information for any possible types. Formally, the designer seeks to satisfy:

$$\operatorname{sgn}\left(\sum_{i=1}^{n} m_i\right) = \operatorname{sgn}\left(\sum_{i=1}^{n} w_i(p_i^A - p_i^B)\right), \quad \forall \tau \in \mathcal{T}$$

subject to the constraints that we define in the next section. Unlike classical mechanism design, where the mechanism designer faces a maximization problem—maximizing social welfare—our mechanism frames decision-making as a classification problem. The designer's objective is to align the sign of the aggregated reports with the sign of the aggregate beliefs.

3.5 Formalization of the Properties

We now formalize the desired properties for the decision-making mechanism. We start by defining Accountability: a pivotal participant — one whose report changes the outcome — must be held individually responsible. Formally,

▶ **Definition 3** (Accountability). *Let T_i be the total monetary transfer done by expert i and T_j the total monetary transfer done by expert j. A decision-making mechanism M satisfies Accountability if:*

$$\forall i, j \in N, \forall \tau_i \in \mathcal{T}_i, \forall \tau_j \in \mathcal{T}_j, \forall \mathbf{m} \in \mathcal{R} :$$

$$\Big(x(\mathbf{m}_{-i}) \neq x(\mathbf{m}) \wedge x(\mathbf{m}_{-j}) = x(\mathbf{m})\Big) \implies \begin{cases} T_i > T_j & \text{if } \Delta > 0 \\ T_i < T_j & \text{if } \Delta < 0 \end{cases}$$

Informally, this implies that an expert whose report alters the DAO's decision, must internalize the consequences of their influence. Conditional on their decisive action yielding a positive outcome ($\Delta > 0$), the pivotal expert receives a higher monetary transfer (the sum of the initial transfer and ex-post reward) relative to a non-pivotal participant. Conversely, if the pivotal intervention yields a negative outcome ($\Delta < 0$), the expert incurs a strictly greater monetary penalty relative to a non-pivotal participant.

A mechanism satisfies *Weak Accountability* if the strict reward condition for a positive outcome ($\Delta > 0$) is relaxed to a weak inequality ($T_i \geq T_j$), while the strict penalty for a negative outcome ($\Delta < 0$) is maintained. Informally, weak accountability ensures that a pivotal expert who drives a successful outcome is at least as well off as a non-pivotal expert, though not necessarily strictly better off. They might receive the exact same net payoff as someone whose input did not alter the decision.

Trivially, any mechanism that does not use monetary transfers cannot satisfy Accountability. It turns out that none of the mechanisms from Section 2 satisfies Weak Accountability. We refer the reader to appendix A for the proofs.

Next, we establish the condition under which experts are willing to participate in the DAO's decision-making process, formally known as Individual Rationality (IR). We adopt the notion of interim IR each expert knows their own preferences and beliefs when deciding whether to vote, but does not know what the other experts will report. Additionally, it also exist the notion of ex-ante (before learning one's type) and ex-post (after all messages are submitted).

▶ **Definition 4** (Interim IR). *A decision-making mechanism satisfies Interim Individual Rationality if:*

$$\forall i \in N, \forall \tau_i \in \mathcal{T}_i, \forall \mathbf{m}_{-i} \in \mathcal{R}_{-i} : \quad \mathbb{E}[u_i(m_i^\star, \mathbf{m}_{-i} \mid \tau_i)] \geq \mathbb{E}[u_i(\mathbf{m}_{-i} \mid \tau_i)]$$

for some $m_i^\star \in \mathcal{R}$.

Informally, this condition ensures that no expert is made worse off by participating in the mechanism. Their expected utility from engaging and submitting their optimal message $m_i^\star$ must be at least as great as their expected utility from abstaining, given their private type τ_i and the expected strategies of all other experts $\mathbf{m}_{-i}$.

Furthermore, a mechanism satisfies *Strict Interim Individual Rationality* if this inequality is strict. This strict version of Interim IR serves as the formal definition for Sustainable Participation. It guarantees that experts derive a strictly positive expected net benefit from participating in the decision-making mechanism. Note that not participating in the mechanism does not yield 0 utility to the expert since they will be affected by the outcomes of the decision. This includes the reward rules, that are transferred independently of the participation of the expert in the mechanism, i.e., the transfers are applied just because the expert is a member of the council.

Having established the conditions for expert participation, we must also consider the financial viability of the mechanism from the perspective of the DAO. A mechanism satisfies the Budget Constraint if the total net compensation issued by the DAO to the experts is bounded by a maximum subsidy. Formally,

▶ **Definition 5** (Budget Constraint). *A decision-making mechanism satisfies the Budget Constraint if for $c \geq 0$:*

$$\forall \mathbf{m} \in \mathcal{R}^n, \forall \Delta \in \{-1, 1\} : \quad \sum_{i \in N} \Big(t_i(\mathbf{m}) + r_i(\mathbf{m}, \Delta) \Big) \leq c$$

To guarantee that the decision-making mechanism remains impartial, we must ensure it does not possess a bias toward any specific alternative. We formalize this requirement through two related properties: Symmetry and Belief Neutrality.

▶ **Definition 6** (Symmetry). *A decision-making mechanism satisfies Symmetry if*

$$\forall i \in N, \forall \tau_i \in \mathcal{T}_i \;\; s.t. \;\; \theta_i^A = \theta_i^B \wedge p_i^A = p_i^B : \quad \forall \mathbf{m}_{-i} \in \mathcal{R}^{n-1}, \; x(m_i^\star, \mathbf{m}_{-i}) = x(\mathbf{m}_{-i})$$

where $m_i^\star$ is an optimal report.

▶ **Definition 7** (Belief Neutrality). *A decision-making mechanism satisfies Belief Neutrality if*

$$\forall i \in N, \forall \tau_i \in \mathcal{T}_i \;\; s.t. \;\; p_i^A = p_i^B = 0.5 : \quad \mathbb{E}_\Delta[r_i(\mathbf{m}, \Delta) \mid x(\mathbf{m}) = A] = \mathbb{E}_\Delta[r_i(\mathbf{m}, \Delta) \mid x(\mathbf{m}) = B]$$

where $m_i^\star$ is an optimal report.

Informally, a decision-making mechanism satisfies Symmetry if when an expert is indifferent between the two alternatives and has identical subjective beliefs, their optimal message must not unilaterally alter the allocation rule, i.e., their participation must be mathematically equivalent to abstaining. Similarly, a mechanism satisfies Belief Neutrality if an expert who assigns equal probability to the success of either option does not contribute with directional weight to the belief-aggregation components of the mechanism. Formally, their submitted message $m_i^\star$ must yield an expected ex-post reward that is perfectly symmetric across both alternatives.

Incentive compatibility is perhaps the most desired property in mechanism design. A mechanism is incentive compatible if experts maximize their expected utility by honestly revealing this information rather than behaving strategically. A DSIC mechanism guarantees that reporting honestly is an expert's best strategy regardless of the reports of other participants. The formal definition of DSIC is

▶ **Definition 8** (Dominant Strategy Incentive Compatibility). *A decision-making mechanism is Dominant Strategy Incentive Compatible (DSIC) if,*

$$\forall i \in N, \forall \tau_i \in \mathcal{T}_i, \forall m_i \in \mathcal{R}, \forall \mathbf{m}_{-i} \in \mathcal{R}^{n-1} :$$

$$\mathbb{E}[u_i(m_i^\star, \mathbf{m}_{-i} \mid \tau_i)] \geq \mathbb{E}[u_i(m_i, \mathbf{m}_{-i} \mid \tau_i)]$$

where $m_i^\star$ denotes the honest reporting strategy for type τ_i.

This property is very important in the maximization problems because it ensures the allocation rule sums the true valuations. In this context, a strategic report that deviates from the truth is only problematic if it shifts the aggregate message across the decision boundary toward a suboptimal alternative. Unlike classical settings, full DSIC may not hold for all expert types in our environment.

4 Decision-Making Mechanism

This section presents the mechanism construction. We first characterize the space of admissible reward rules (Section 4.1), then describe the complete mechanism (Section 4.2).

4.1 Characterizing the Reward Rule

Before constructing the mechanism, we give an observation that serves as its building block.

▶ **Observation 9** (Additive Adjustments). *Consider a VCG mechanism as in Section 3.3. Suppose that, in addition to the VCG transfer, each expert i receives a fixed reward $r \in \mathbb{R}$ whenever a specific alternative is selected, where r does not depend on the submitted messages. Then a risk-neutral expert's dominant strategy shifts from reporting their true valuation true preference intensity v_i to:*

$$m_i^\star = v_i + r$$

We now show that the properties of Symmetry and Belief Neutrality constrain the reward rule to a single degree of freedom. Suppose the mechanism applies a report-independent reward rule $r_a(\Delta)$ conditional on the chosen decision $a \in \{A, B\}$ and the realized state $\Delta \in \{1, -1\}$:

$$r_A(\Delta) = \begin{cases} r_1^A & \text{if } \Delta = 1 \\ r_2^A & \text{if } \Delta = -1 \end{cases} \qquad r_B(\Delta) = \begin{cases} r_1^B & \text{if } \Delta = 1 \\ r_2^B & \text{if } \Delta = -1 \end{cases}$$

▶ **Lemma 10** (Reward Symmetry and Zero-Sum Constraint). *Any mechanism satisfying Symmetry and Belief Neutrality must restrict its ex-post rewards to the symmetric, zero-sum form: $r_1^A = -r_2^A = r_1^B = -r_2^B$.*

Proof. Let $\tau_i = (\theta_i^A, \theta_i^B, p_i^A, p_i^B)$. By Observation 9, a risk-neutral expert's optimal report incorporates the expected reward difference: $\mathbb{E}[r_A \mid p_i^A] - \mathbb{E}[r_B \mid p_i^B]$.

Symmetry. If an expert is indifferent, $\theta_i^A = \theta_i^B$ and $p_i^A = p_i^B = p$, their report must not bias the allocation rule ($m_i = 0$). This requires:

$$p\, r_1^A + (1-p)\, r_2^A = p\, r_1^B + (1-p)\, r_2^B$$

For this to hold for all $p \in [0,1]$, we need $r_1^A = r_1^B$ and $r_2^A = r_2^B$.

Belief Neutrality. If an expert is completely uncertain ($p_i^A = p_i^B = 0.5$), the expected reward must be zero:

$$0.5\, r_1^A + 0.5\, r_2^A = 0 \implies r_1^A = -r_2^A$$

Combining both conditions yields $r_1^A = -r_2^A = r_1^B = -r_2^B$. ◀

The reward function is thus entirely determined by a single positive scalar magnitude.

4.2 Mechanism Construction

Having characterized the reward rule, we construct the complete mechanism $\mathcal{M} = (x, \mathbf{t}, \mathbf{r})$. The design rationale is to decouple an expert's idiosyncratic preferences from their subjective beliefs about the organization's success. While VCG transfers incentivize truthful preference revelation by charging experts the externality they impose on others, they ignore the realized outcome. Our mechanism augments VCG with an ex-post reward that forces pivotal experts to internalize the risk of their influence, thereby satisfying accountability and aligning incentives with the designer's objective. The reward is set to consume the full available budget, maximizing the weight of outcome information in the expert's report.

The mechanism operates as follows. Each expert $i \in N$ submits a single message $m_i \in \mathbb{R}$, and the allocation rule (2) selects an alternative. The standard VCG transfer $t_i(\mathbf{m})$ is computed as in (3). After the predetermined evaluation period, the governance smart contract observes $\Delta \in \{-1, 1\}$ via the on-chain evaluation tool and distributes an ex-post reward contingent on the realized outcome:

$$r_i(\mathbf{m}, \Delta) = \begin{cases} -t_i(\mathbf{m}) + \dfrac{c}{n} & \text{if } \Delta > 0 \\ t_i(\mathbf{m}) - \dfrac{c}{n} & \text{if } \Delta < 0 \end{cases} \tag{4}$$

When the outcome is positive, the reward refunds the VCG tax and adds a per-expert share of the budget; when the outcome is negative, the reward doubles the tax and deducts the same share. This rule is obtained by maximizing the reward subject to the available budget, and its form follows from Lemma 10. Unlike the fixed additive adjustment in Observation 9, the reward here depends on $t_i(\mathbf{m})$, which itself varies with the submitted messages, altering each expert's report in a way that reflects both their beliefs and their influence on the outcome. The full procedure is given in Algorithm 1.

Algorithm 1 Decision-Making Mechanism $\mathcal{M}$

Input: Report profile $\mathbf{m} = (m_1, \ldots, m_n)$ with $m_i \in \mathbb{R}$; budget $c \geq 0$
$S \leftarrow \sum_{i=1}^{n} m_i$;
if $S \geq 0$ **then**
  $a \leftarrow A$;
else
  $a \leftarrow B$;
end
Observe the realized outcome $\Delta \in \{-1, 1\}$;
for $i = 1, \ldots, n$ **do**
  if $\Delta > 0$ **then**
    $r_i \leftarrow -t_i + \frac{c}{n}$;
  else
    $r_i \leftarrow t_i - \frac{c}{n}$;
  end
  $\pi_i \leftarrow t_i + r_i$;
end
return *Allocation* a, *payoffs* $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_n)$;

5 Analysis

We now analyze the mechanism. We first verify the structural properties that follow directly from the construction (Section 5.1), then analyze experts' strategic behavior and establish dominant strategies, safe deviation, and individual rationality (Section 5.2), and conclude with the main theorem on information aggregation and a worked example (Section 5.4).

5.1 Structural Properties

We verify the properties that follow directly from the mechanism's construction, without requiring the equilibrium characterization. Since the reward rule is determined by a single scalar (Lemma 10), Symmetry and Belief Neutrality hold by construction.

▶ **Lemma 11.** *The mechanism $\mathcal{M}$ satisfies Budget Constraint.*

Proof. For any message profile $\mathbf{m}$ and any realization $\Delta \in \{-1, 1\}$:

Case $\Delta > 0$:

$$\sum_{i \in N} \left(t_i + r_i\right) = \sum_{i \in N} t_i + \sum_{i \in N} \left(-t_i + \frac{c}{n}\right) = \sum_{i \in N} \frac{c}{n} = c \leq c$$

Case $\Delta < 0$:

$$\sum_{i \in N} \left(t_i + r_i\right) = \sum_{i \in N} t_i + \sum_{i \in N} \left(t_i - \frac{c}{n}\right) = 2 \sum_{i \in N} t_i - c \leq -c \leq c$$

where the first inequality follows from $\sum_{i \in N} t_i \leq 0$, since VCG transfers are non-positive. ◀

▶ **Lemma 12.** *The mechanism $\mathcal{M}$ satisfies Weak Accountability.*

Proof. Let expert i be pivotal $\big(x(\mathbf{m}_{-i}) \neq x(\mathbf{m})\big)$ and expert j non-pivotal $\big(x(\mathbf{m}_{-j}) = x(\mathbf{m})\big)$. W.l.o.g. assume $m_i > 0$, so $t_i = \sum_{k \neq i} m_k < 0$ and $t_j = 0$; the case $m_i < 0$ is symmetric, since the VCG transfer satisfies $t_i < 0$ for any pivotal expert (regardless of the sign of m_i). The total monetary transfers $T_i = t_i + r_i$ and $T_j = t_j + r_j$ evaluate to:

$$\begin{cases} T_i = t_i + (-t_i) + \frac{c}{n} = \frac{c}{n} = T_j & \text{if } \Delta > 0 \\ T_i = t_i + t_i - \frac{c}{n} = 2t_i - \frac{c}{n} < -\frac{c}{n} = T_j & \text{if } \Delta < 0 \end{cases}$$

The first case gives $T_i \geq T_j$ and the second gives $T_i < T_j$, satisfying Weak Accountability. ◀

5.2 Incentive Properties

We analyze the strategic behavior of experts under $\mathcal{M}$. First we provide a formal definition of alignment, then we derive the expected utility expressions and establish the dominant strategy for aligned experts, then show that deviations by unaligned experts are safe. We verify individual rationality and conclude with the main theorem on information aggregation.

We say an expert is *aligned* when their preferences and beliefs point in the same direction.

▶ **Definition 13** (Alignment). *An expert i is aligned if the direction of their preference parameter strictly matches the direction of their probabilistic belief:*

$$(\theta_i^a - \theta_i^b)(p_i^a - p_i^b) > 0$$

W.l.o.g. throughout the section, we consider an expert that strictly prefers Option A ($\theta_i^A > \theta_i^B$); the opposite case is symmetric. Given an outcome $a \in \{A, B\}$ and VCG transfer t_i, the expert's expected utility is:

$$\mathbb{E}[u_i \mid a] = \theta_i^a + 2t_i(1 - p_i^a) + (2p_i^a - 1)\frac{c}{n} \tag{5}$$

When the expert is non-pivotal, $t_i = 0$ and (5) reduces to $\theta_i^a + (2p_i^a - 1)\frac{c}{n}$. We partition the analysis based on $\sum_{j \neq i} m_j$, since the expert's report can only change the allocation by crossing the decision boundary. In Case 1 ($\sum_{j \neq i} m_j > 0$), A is the default and the expert can only pivot to B. In Case 2 ($\sum_{j \neq i} m_j < 0$), B is the default and the expert can only pivot to A.

▶ **Lemma 14** (Dominant Strategy for Aligned Agents). *If an expert is aligned, their weakly dominant strategy under $\mathcal{M}$ is to report:*

$$m_i^\star = \begin{cases} \frac{\theta_i^A - \theta_i^B}{2(1 - p_i^A)} + \frac{c}{n}\frac{p_i^A - p_i^B}{1 - p_i^A} & \text{if } \theta_i^A > \theta_i^B \\ \frac{\theta_i^A - \theta_i^B}{2(1 - p_i^B)} + \frac{c}{n}\frac{p_i^A - p_i^B}{1 - p_i^B} & \text{if } \theta_i^B > \theta_i^A \end{cases} \tag{6}$$

Proof. W.l.o.g. assume the expert is aligned with $\theta_i^A > \theta_i^B$ and $p_i^A > p_i^B$.

In Case 1 ($\sum_{j \neq i} m_j > 0$), A is the default. Evaluating (5) for $a = A$ with $t_i = 0$ and for $a = B$ with $t_i < 0$:

$$\mathbb{E}[u_i \mid A] - \mathbb{E}[u_i \mid B] = (\theta_i^A - \theta_i^B) + 2(p_i^A - p_i^B)\frac{c}{n} + (-2t_i)(1 - p_i^B) > 0$$

Each term is positive: the first two by alignment, and the third because $t_i < 0$ and $1 - p_i^B > 0$. Hence pivoting to B is dominated.

In Case 2 ($\sum_{j \neq i} m_j < 0$), B is the default. Evaluating (5) for $a = A$ with $t_i < 0$ and for $a = B$ with $t_i = 0$, the expert prefers pivoting to A when:

$$t_i > -\frac{\theta_i^A - \theta_i^B}{2(1 - p_i^A)} - \frac{c}{n}\frac{p_i^A - p_i^B}{1 - p_i^A} = -m_i^\star \tag{7}$$

It remains to show that deviations from $m_i^\star$ are dominated. Since $t_i = \sum_{j\neq i} m_j$ and $m_i^\star > 0$ by alignment, consider a deviation $m_i' > m_i^\star$ (over-reporting). The outcome changes only if $m_i^\star < -t_i < m_i'$, which implies $t_i < -m_i^\star$. By (7), this yields $\mathbb{E}[u_i \mid a = A] < \mathbb{E}[u_i \mid a = B]$: the expert forced an outcome it prefers less. Conversely, for $m_i' < m_i^\star$ (under-reporting), the outcome changes only if $m_i' < -t_i < m_i^\star$, which implies $t_i > -m_i^\star$. By (7), this yields $\mathbb{E}[u_i \mid a = A] > \mathbb{E}[u_i \mid a = B]$: the expert lost a beneficial pivot. ◀

Although Lemma 14 establishes $m_i^\star$ as a dominant strategy only for aligned experts, Equation (6) defines $m_i^\star$ for every expert as a function of their private type $\tau_i = (\theta_i^A, \theta_i^B, p_i^A, p_i^B)$ and the public parameters c and n. For unaligned experts, $m_i^\star$ is not a dominant strategy — deviations toward the alternative the expert believes is more likely to succeed may be strictly profitable. However, $m_i^\star$ remains a known reference strategy that each expert can compute from their own private information. The following proposition shows that deviations from $m_i^\star$ are *safe*.

▶ **Proposition 15** (Safe Deviation). *Under mechanism $\mathcal{M}$, for any expert i (aligned or unaligned), every deviation from $m_i^\star$ that changes the allocation toward the alternative with lower success probability from the expert's own perspective is weakly dominated. Formally, any $m_i' \neq m_i^\star$ such that the allocation changes from b to a with $p_i^a < p_i^b$ satisfies:*

$$\mathbb{E}[u_i(m_i', \mathbf{m}_{-i} \mid \tau_i)] \leq \mathbb{E}[u_i(m_i^\star, \mathbf{m}_{-i} \mid \tau_i)]$$

Proof. For aligned experts, Lemma 14 establishes that $m_i^\star$ is a weakly dominant strategy, so every deviation—including those that change the allocation toward the alternative with lower success probability—is weakly dominated.

It remains to prove the claim for unaligned experts. Let $m_i' \neq m_i^\star$ be any deviation such that the allocation changes from b to a with $p_i^a < p_i^b$, i.e., under $m_i^\star$ the selected alternative is b and under m_i' it is a. Since the expert is unaligned and $p_i^a < p_i^b$, we have $\theta_i^a > \theta_i^b$: the expert idiosyncratically prefers the lower-probability alternative a. We show that $\mathbb{E}[u_i \mid a] \leq \mathbb{E}[u_i \mid b]$.

W.l.o.g. assume $a = A$ and $b = B$. Then $\theta_i^A > \theta_i^B$, and $p_i^B > p_i^A$. Let $s = \sum_{j\neq i} m_j$. The deviation changes the outcome from b to a only when $m_i^\star + s < 0 \leq m_i' + s$, i.e., $s < -m_i^\star$. We verify $\mathbb{E}[u_i \mid a] \leq \mathbb{E}[u_i \mid b]$ in two sub-ranges.

If $s < 0$: under $m_i^\star$ the expert is non-pivotal (the outcome is B, $t_i = 0$); under m_i' the expert pivots $B \to A$ with $t_i = s$. Evaluating (5):

$$\begin{aligned}\mathbb{E}[u_i \mid a] - \mathbb{E}[u_i \mid b] &= \left[\theta_i^A + 2s(1 - p_i^A) + (2p_i^A - 1)\tfrac{c}{n}\right] - \left[\theta_i^B + (2p_i^B - 1)\tfrac{c}{n}\right] \\ &= (\theta_i^A - \theta_i^B) + 2s(1 - p_i^A) + 2(p_i^A - p_i^B)\tfrac{c}{n} \\ &= 2(1 - p_i^A)\, m_i^\star + 2s(1 - p_i^A) = 2(1 - p_i^A)(m_i^\star + s) \leq 0\end{aligned}$$

where the third equality uses $2(1 - p_i^A)\, m_i^\star = (\theta_i^A - \theta_i^B) + 2(p_i^A - p_i^B)\frac{c}{n}$ from (6), and the inequality holds since $m_i^\star + s < 0$.

If $0 \leq s < -m_i^\star$ (non-empty only when $m_i^\star < 0$): under $m_i^\star$ the expert pivots $A \to B$ with $t_i = -s$; under m_i' the expert is non-pivotal (outcome $a = A$, $t_i = 0$). Evaluating (5):

$$\begin{aligned}\mathbb{E}[u_i \mid a] - \mathbb{E}[u_i \mid b] &= \left[\theta_i^A + (2p_i^A - 1)\tfrac{c}{n}\right] - \left[\theta_i^B + 2(-s)(1 - p_i^B) + (2p_i^B - 1)\tfrac{c}{n}\right] \\ &= (\theta_i^A - \theta_i^B) + 2s(1 - p_i^B) + 2(p_i^A - p_i^B)\tfrac{c}{n} \\ &= 2(1 - p_i^A)\, m_i^\star + 2s(1 - p_i^B)\end{aligned}$$

This is increasing in s (coefficient $2(1 - p_i^B) > 0$) and at $s = -m_i^\star$ equals $2m_i^\star[(1 - p_i^A) - (1 - p_i^B)] = 2m_i^\star(p_i^B - p_i^A) < 0$, since $m_i^\star < 0$ and $p_i^B > p_i^A$. Hence $\mathbb{E}[u_i \mid a] < \mathbb{E}[u_i \mid b]$ throughout. ◀

For unaligned experts, deviations from $m_i^\star$ toward the alternative they believe is *more* likely to succeed may be strictly profitable. Such deviations are beneficial for information aggregation: they trade idiosyncratic noise for belief signal; this is made clear in Theorem 17.

We now verify that participation is individually rational for all expert types. Recall that Interim IR (Definition 3) only requires the existence of some strategy $m_i^\star$ such that participating yields at least as much expected utility as abstaining. The strategy $m_i^\star$ from Equation (6) serves this role for every expert type: the proof below shows that reporting $m_i^\star$ is always weakly better than not participating, even for unaligned experts who may profitably deviate from $m_i^\star$ to another report.

▶ **Lemma 16.** *The mechanism $\mathcal{M}$ satisfies Interim Individual Rationality.*

Proof. We show that for all $i \in N$, $\tau_i \in \mathcal{T}_i$, and $\mathbf{m}_{-i} \in \mathcal{R}_{-i}$:

$$\mathbb{E}[u_i(m_i^\star, \mathbf{m}_{-i} \mid \tau_i)] \geq \mathbb{E}[u_i(\mathbf{m}_{-i} \mid \tau_i)]$$

If expert i's report does not alter the allocation, i.e., $x(m_i^\star, \mathbf{m}_{-i}) = x(\mathbf{m}_{-i})$, then $t_i = 0$ and both sides coincide.

Consider the case where expert i is pivotal, changing the outcome from alternative b (the default without participation) to alternative $a \neq b$. W.l.o.g. assume $\theta_i^A \geq \theta_i^B$. Using the reward rule from Algorithm 1, the expected utility difference between participating and abstaining is:

$$\mathbb{E}[u_i(m_i^\star, \mathbf{m}_{-i} \mid \tau_i)] - \mathbb{E}[u_i(\mathbf{m}_{-i} \mid \tau_i)] = (\theta_i^a - \theta_i^b) + 2(p_i^a - p_i^b)\frac{c}{n} + 2(1 - p_i^a)\, t_i$$

where $t_i \leq 0$ is the VCG transfer.

Case expert pivots toward their preference (Outcome is A). Here $a = A$ and $b = B$ and $m_i^\star \geq 0$ and $t_i = \sum_{j \neq i} m_j \geq -m_i^\star$. The IR condition reduces to $t_i \geq -m_i^\star$, which is exactly the pivotality constraint, and thus it must hold.

Case expert pivots against their preference (Outcome is B). This occurs only for unaligned experts whose beliefs dominate their preferences, so that $m_i^\star < 0$ and necessarily $p_i^A < p_i^B$. Here $a = B$ and $b = A$, so the expected utility difference from above becomes:

$$(\theta_i^B - \theta_i^A) + 2(p_i^B - p_i^A)\frac{c}{n} + 2(1 - p_i^B)\, t_i \geq 0$$

Solving for t_i:

$$t_i \geq \frac{(\theta_i^A - \theta_i^B) + 2(p_i^A - p_i^B)\frac{c}{n}}{2(1 - p_i^B)}$$

Recalling that $m_i^\star = \frac{\theta_i^A - \theta_i^B}{2(1-p_i^A)} + \frac{c}{n} \cdot \frac{p_i^A - p_i^B}{1-p_i^A}$, we can multiply $m_i^\star$ by $\frac{1-p_i^A}{1-p_i^B}$ to obtain:

$$m_i^\star \cdot \frac{1 - p_i^A}{1 - p_i^B} = \frac{\theta_i^A - \theta_i^B}{2(1 - p_i^B)} + \frac{c}{n} \cdot \frac{p_i^A - p_i^B}{1 - p_i^B}$$

which equals the right-hand side above. The IR condition therefore reduces to $t_i \geq m_i^\star \cdot \frac{1-p_i^A}{1-p_i^B}$. It remains to verify this holds. Since $p_i^A < p_i^B$ implies $\frac{1-p_i^A}{1-p_i^B} > 1$, and $m_i^\star < 0$, multiplying by a factor greater than one yields $m_i^\star \cdot \frac{1-p_i^A}{1-p_i^B} < m_i^\star$. Chaining the inequalities: $t_i > m_i^\star > m_i^\star \cdot \frac{1-p_i^A}{1-p_i^B}$, which is exactly the pivotality constraint, and thus it must hold.

◀

Note that the IR inequality compares participating to abstaining, not to a zero-utility baseline. An abstaining expert still receives utility from the decision outcome chosen by the remaining participants, so their outside option is generally nonzero. The proof, therefore, first establishes that participating and reporting $m_i^\star$ yields at least as much expected utility as abstaining (the true outside option).

The inequality is strict whenever $t_i > -|m_i^\star|$. If experts are uncertain about the reports of others and assign positive probability to being pivotal, the inequality holds strictly in expectation. Sustainable Participation follows: experts derive a strictly positive expected net benefit from participating.

We now connect the individual incentive guarantees to the designer's aggregate objective. By substituting the strategy profile $m^\star$ into the allocation rule, we characterize when the mechanism achieves correct classification and how the budget c governs the tradeoff between idiosyncratic preferences and belief aggregation.

▶ **Theorem 17** (Information Aggregation). *Under mechanism $\mathcal{M}$, the strategy profile $m^\star$ yields the aggregate signal:*

$$\sum_{i=1}^{n} m_i^\star = N(\tau) + \frac{c}{n} B(\tau) \tag{8}$$

where $N(\tau)$ is the residual idiosyncratic noise and $B(\tau)$ is the aggregate belief signal, both defined explicitly in the proof below. The belief signal satisfies $\operatorname{sgn}(B(\tau)) = \operatorname{sgn}(x^*(\mathbf{p}))$. The mechanism achieves correct classification i.e., $x(\mathbf{m}^\star) = x^*(\mathbf{p})$ whenever

$$\frac{c}{n} > \bar{c}(\tau) \tag{9}$$

where $\bar{c}(\tau) = \begin{cases} \dfrac{|N(\tau)|}{|B(\tau)|} & \text{if } \operatorname{sgn}(N(\tau)) \neq \operatorname{sgn}(B(\tau)) \\ 0 & \text{otherwise} \end{cases}.$

Proof. Substituting $m_i^\star$ from Equation (6) and grouping terms by preference direction yields:

$$\sum_{i=1}^{n} m_i^\star = \underbrace{\sum_{\theta_i^A \geq \theta_i^B} \frac{\theta_i^A - \theta_i^B}{2(1-p_i^A)} - \sum_{\theta_i^B > \theta_i^A} \frac{\theta_i^B - \theta_i^A}{2(1-p_i^B)}}_{N(\tau)} + \underbrace{\frac{c}{n}\left(\sum_{\theta_i^B > \theta_i^A} \frac{p_i^A - p_i^B}{1-p_i^B} + \sum_{\theta_i^A \geq \theta_i^B} \frac{p_i^A - p_i^B}{1-p_i^A}\right)}_{B(\tau)}$$

The belief signal can be written as $B(\tau) = \sum_{i=1}^{n} w_i(p_i^A - p_i^B)$ with strictly positive weights. Since $x^*(\mathbf{p})$ selects A if and only if $\sum w_i(p_i^A - p_i^B) > 0$, we have $\operatorname{sgn}(B(\tau)) = \operatorname{sgn}(x^*(\mathbf{p}))$.

For correct classification we need $\operatorname{sgn}(N + \frac{c}{n}B) = \operatorname{sgn}(B)$. If $\operatorname{sgn}(N) = \operatorname{sgn}(B)$, this holds for any $c \geq 0$. Otherwise, the condition requires $\frac{c}{n}|B| > |N|$, yielding the threshold (9). ◀

The proof reveals that the equilibrium induces the weights $w_i = \frac{1}{1-p_i^{(\cdot)}}$, where $p_i^{(\cdot)}$ is the success probability of expert i's preferred alternative.

By Proposition 15, the strategy profile $m^\star$ yields the noisiest possible aggregate: any rational deviation from $m_i^\star$ can only shift the aggregate toward the alternative the deviating expert believes is more likely to succeed. These deviations are beneficial for information aggregation when an unaligned expert who deviates from $m_i^\star$ toward their belief effectively trades preference-driven noise for belief-driven signal, reducing $|N(\tau)|$ and lowering the budget threshold $\bar{c}(\tau)$.

The decomposition (8) makes precise the relation between preferences and beliefs in the designer's classification problem. The noise term $N(\tau)$ captures the bias that idiosyncratic preferences inject into the aggregate signal, while $B(\tau)$ aggregates the experts' private beliefs. The budget parameter c/n controls the relative strength of the belief component: larger budgets amplify the informational signal, enabling the mechanism to override stronger idiosyncratic biases.

Note that the threshold $\bar{c}(\tau)$ depends on the realized type profile τ, which the designer does not observe ex ante. In practice, the designer selects the budget c before the mechanism runs, without knowing whether c/n exceeds $\bar{c}(\tau)$. The theorem therefore provides a structural guarantee — for any type profile, there exists a finite per-expert budget that suffices — rather than a recipe for choosing c in advance.

The mechanism performs best when experts' beliefs converge. Each term in $N(\tau)$ carries the sign of the expert's preference direction, but the denominators $2(1 - p_i^{(\cdot)})$ depends on belief strength. As experts' beliefs converge toward the same alternative — say all experts come to share the belief that $p_i^B > p_i^A$ — experts who prefer A see their weight shrink, which reduces their positive contribution to $N(\tau)$. At the same time, experts who prefer B see their weight grow, amplifying their negative contribution. This leads $N(\tau)$ to be driven towards the same sign as $B(\tau)$. When $\mathrm{sgn}(N) = \mathrm{sgn}(B)$, the threshold $\bar{c}(\tau) = 0$ by (9), and the mechanism classifies correctly for *any* budget $c \geq 0$. This convergence property makes the mechanism particularly suited to DAO governance, where proposals are typically preceded by community discussions and temperature checks that foster belief alignment among experts before the formal vote.

Figure 2 illustrates the classification plane. Every type profile τ, regardless of the number of experts n, maps to a single point $(N(\tau), B(\tau))$. The decision boundary $N + \frac{c}{n}B = 0$ is a line through the origin with slope $-n/c$; increasing the per-expert budget $\frac{c}{n}$ rotates this line toward the N-axis, shrinking the misclassification wedge (shaded red for c_2/n). In the green quadrants noise and signal agree in sign, so $\bar{c}(\tau) = 0$ and the mechanism classifies correctly for any budget. A point's distance from the N-axis relative to its distance from the B-axis determines its threshold $\bar{c}(\tau) = |N|/|B|$: points closer to the N axis (strong noise, weak signal) are harder to classify and require a larger budget.

5.3 Complexity

The mechanism is computationally lightweight. The allocation rule computes a single sum $\sum_{i=1}^{n} m_i$ in $O(n)$ arithmetic operations. The VCG transfer for each expert requires evaluating $\sum_{j \neq i} m_j = \sum_{i=1}^{n} m_i - m_i$, which is $O(1)$ per expert given $\sum_{i=1}^{n} m_i$, yielding $O(n)$ total. The reward rule (4) performs a constant number of operations per expert, again $O(n)$ in total. The entire mechanism therefore runs in $O(n)$ time. The remaining cost is the evaluation of the KPI that resolves Δ. This cost is $O(1)$ with respect to n—a single evaluation per decision, regardless of the number of experts.

From a smart contract perspective, the on-chain cost is equally modest. The contract stores n messages and the budget c, computes the sum, determines the allocation, and records the VCG transfers—all in a single transaction with $O(n)$ storage writes and arithmetic. The ex-post reward is computed in a second transaction after Δ is resolved, requiring $O(n)$ operations. No sorting, optimization, or iterative procedures are needed, making the mechanism well-suited for on-chain execution where gas costs scale with computational steps.

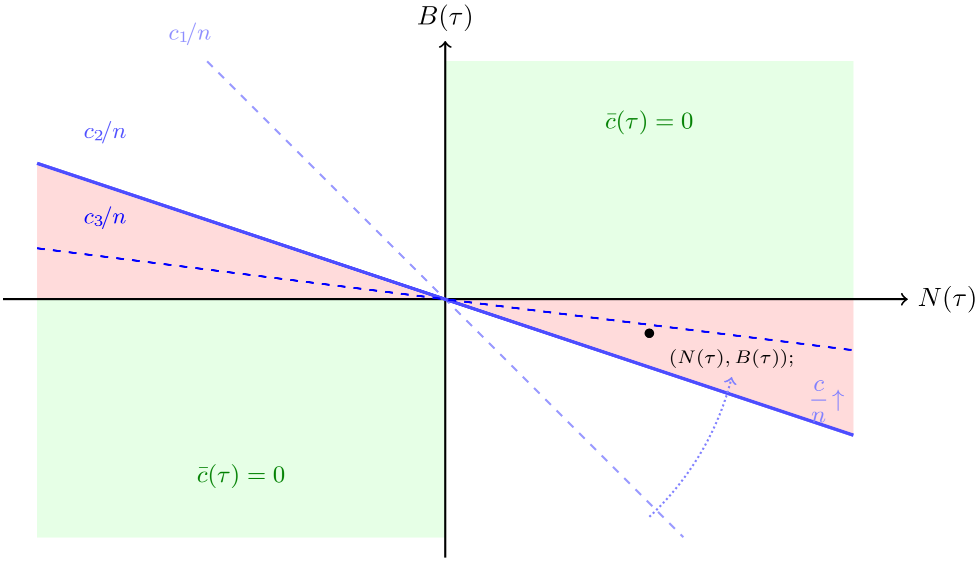

Figure 2 The classification plane. The $N(\tau)$ - axis is the idiosyncratic noise gathered by the allocation rule and the $B(\tau)$ - axis is the belief signal.

5.4 Example

We illustrate the mechanism with a concrete scenario. A DAO governance council of $n = 3$ experts must decide whether to integrate a new DeFi protocol (A = approve) or reject it (B = reject). The DAO allocates a budget of $c = 24$, yielding $c/n = 8$ per expert. Table 2 summarizes the experts' private types.

	θ_i^A	θ_i^B	p_i^A	p_i^B	Alignment
Expert 1	2	0	0.75	0.50	Aligned (favors A)
Expert 2	0	1	0.50	0.75	Aligned (favors B)
Expert 3	1	0	0.50	0.75	Unaligned (prefers A, believes B)

Table 2 Expert type profiles in the worked example.

Expert 1 personally benefits from approval and believes it will succeed. Expert 2 prefers rejection and believes it is the better outcome. Expert 3 personally prefers approval but believes rejection is more likely to benefit the DAO — a conflict between preferences and beliefs.

Step 1: Reports

Each expert computes their dominant strategy report using Equation (6):

$$m_1^\star = \frac{2-0}{2(1-0.75)} + 8 \cdot \frac{0.75-0.50}{1-0.75} = 4 + 8 = 12$$

$$m_2^\star = \frac{0-1}{2(1-0.75)} + 8 \cdot \frac{0.50-0.75}{1-0.75} = -2 - 8 = -10$$

$$m_3^\star = \frac{1-0}{2(1-0.50)} + 8 \cdot \frac{0.50-0.75}{1-0.50} = 1 - 4 = -3$$

Expert 3 submits a negative report ($m_3^\star = -3$) despite personally preferring A. The belief component (-4) outweighs the preference component ($+1$), pulling the report toward B — the alternative Expert 3 believes will benefit the organization.

Step 2: Allocation

The aggregate signal is $\sum_{i=1}^{3} m_i^\star = 12 - 10 - 3 = -1 < 0$, so the mechanism selects B.

Step 3: VCG transfers

We compute $\sum_{j \neq i} m_j$ for each expert and check pivotality:

- Expert 1: $\sum_{j \neq 1} m_j = -13 < 0$. Without Expert 1, B is already selected. *Not pivotal*: $t_1 = 0$.
- Expert 2: $\sum_{j \neq 2} m_j = 9 > 0$. Without Expert 2, A would be selected. Expert 2 flips the outcome from A to B. *Pivotal*: $t_2 = -9$.
- Expert 3: $\sum_{j \neq 3} m_j = 2 > 0$. Without Expert 3, A would be selected. Expert 3 also flips the outcome. *Pivotal*: $t_3 = -2$.

Step 4: Outcome-contingent rewards

After the evaluation period, the on-chain tool observes Δ. Table 3 shows the resulting transfers.

	t_i	$\Delta = 1$ (B succeeds)		$\Delta = -1$ (B fails)	
		r_i	$\pi_i = t_i + r_i$	r_i	$\pi_i = t_i + r_i$
Expert 1 (non-piv.)	0	8	8	-8	-8
Expert 2 (pivotal)	-9	17	8	-17	-26
Expert 3 (pivotal)	-2	10	8	-10	-12

Table 3 Monetary transfers and payoffs in the worked example.

If B succeeds ($\Delta = 1$), the reward cancels each expert's VCG tax and everyone receives $c/n = 8$, satisfying Weak Accountability: pivotal experts are at least as well off as non-pivotal ones ($8 \geq 8$). If B fails ($\Delta = -1$), pivotal experts bear strictly heavier penalties: $\pi_2 = -26 < -8 = \pi_1$ and $\pi_3 = -12 < -8 = \pi_1$, again satisfying Weak Accountability.

Information aggregation

We verify the decomposition from Theorem 17:

$$N(\tau) = \frac{2}{2(0.25)} + \frac{1}{2(0.50)} - \frac{1}{2(0.25)} = 4 + 1 - 2 = 3, \qquad B(\tau) = \frac{0.25}{0.25} + \frac{-0.25}{0.25} + \frac{-0.25}{0.50} = 1 - 1 - 0.5 = -0.5$$

The noise $N(\tau) = 3 > 0$ pushes toward A (driven by idiosyncratic preferences), while the belief signal $B(\tau) = -0.5 < 0$ pushes toward B. Since $\operatorname{sgn}(B) < 0$, the designer's optimal decision is B. The aggregate confirms: $N + \frac{c}{n} B = 3 + 8(-0.5) = -1 < 0$, correctly selecting B. The budget threshold is $\bar{c}(\tau) = |N|/|B| = 6$, and indeed $c/n = 8 > 6$.

Insufficient budget

With a reduced budget $c = 12$ ($c/n = 4 < 6 = \bar{c}$), the reports become $m_1^\star = 8$, $m_2^\star = -6$, $m_3^\star = -1$, yielding $\sum m_i^\star = 1 > 0$. The mechanism selects A — a misclassification, because the budget is insufficient to overcome the preference-driven noise.

Safe deviation

Suppose Expert 3 deviates from $m_3^\star = -3$ to $m_3' = 5$, following their personal preference for A. Since $\sum_{j \neq 3} m_j = 2 > 0$ and $m_3' + 2 = 7 > 0$, Expert 3 is no longer pivotal ($t_3 = 0$) and the mechanism selects A. Expert 3's expected utility drops from 3 (at $m_3^\star$) to 1 (at

$m'_3 = 5$): the deviation toward the alternative Expert 3 believes is less likely to succeed is costly, confirming Proposition 15.

6 Conclusions

We presented a decision-making mechanism for governance councils in DAOs that augments VCG transfers with outcome-contingent rewards. Unlike classical mechanism design, where the objective is to maximize social welfare, we frame decision-making as a classification problem: the designer seeks to align the collective decision with the aggregate expert beliefs. The mechanism satisfies Budget Constraint, Symmetry, Belief Neutrality, Weak Accountability, and Interim Individual Rationality. For aligned experts it is DSIC; for unaligned experts, the Safe Deviation property guarantees that no expert can profitably deviate from $m_i^\star$ toward an alternative they individually believe is less likely to succeed. The main result (Theorem 17) decomposes the aggregate signal into idiosyncratic noise and a belief signal whose sign matches the designer's optimal decision, with correct classification achieved whenever the per-expert budget exceeds a type-dependent threshold $\bar{c}(\tau)$. This threshold decreases as experts' beliefs converge — a condition naturally fostered by deliberative processes that precede governance votes in DAOs.

The mechanism operates under several assumptions. It requires independent experts; coalitions could manipulate transfers, though the pseudonymous, trustless DAO environment partially mitigates this since collusive agreements are unenforceable on-chain. The framework targets small councils, as the budget must be distributed across participants. The ex-post reward rule requires an on-chain evaluation tool that resolves outcomes into a boolean; DAOs whose objectives cannot be captured by measurable KPIs fall outside the mechanism's scope.

Several directions extend this work. The mechanism addresses binary decisions; extending to three or more alternatives faces barriers imposed by Roberts' Theorem. A richer outcome space could enable finer-grained rewards. Relaxing outcome observation via inter-expert agreement proxies [33] could broaden applicability at the cost of weaker incentive guarantees. Finally, formally characterizing the budget threshold under parametric correlation models between preferences and beliefs is left to future work.

References

1. János Aczél and Carl Wagner. A characterization of weighted arithmetic means. *SIAM Journal on Algebraic Discrete Methods*, 1(3):259–260, 1980.
2. David Austen-Smith and Jeffrey S. Banks. Information aggregation, rationality, and the condorcet jury theorem. *American Political Science Review*, 90(1):34–45, 1996.
3. Martin Becze, Hudson Jameson, et al. Eip-1: Eip purpose and guidelines. *Ethereum Improvement Proposals*, 27, 2015.
4. Craig Boutilier. Eliciting forecasts from self-interested experts: scoring rules for decision makers. *arXiv preprint arXiv:1106.2489*, 2011.
5. Vitalik Buterin, Zoë Hitzig, and E. Glen Weyl. A flexible design for funding public goods. *Manage. Sci.*, 65(11):5171–5187, November 2019. doi:10.1287/mnsc.2019.3337.
6. Yiling Chen, Ian Kash, Mike Ruberry, and Victor Shnayder. Decision markets with good incentives. In *International Workshop on Internet and Network Economics*, pages 72–83. Springer, 2011.
7. Yiling Chen, Ian A Kash, Michael Ruberry, and Victor Shnayder. Eliciting predictions and recommendations for decision making. *ACM Transactions on Economics and Computation (TEAC)*, 2(2):1–27, 2014.

8 Lin William Cong, Daniel Rabetti, Charles CY Wang, and Yu Yan. Centralized governance in decentralized organizations. *Available at SSRN 5168660*, 2025.

9 DAOstack. Daostack whitepaper: An operating system for collective intelligence. https://www.allcryptowhitepapers.com/daostack-whitepaper/, 2018. Accessed: 2026-03-14.

10 Luke Dashjr. Bip 2: Bip process, revised. https://github.com/bitcoin/bips/blob/master/bip-0002.mediawiki, 2016. Bitcoin Improvement Proposal.

11 David Davó, Javier Arroyo, Samer Hassan, and Silvia Semenzin. The rise and fall of daostack: lessons for decentralized autonomous organizations. *PeerJ Computer Science*, 11:e3320, 2025.

12 Nicola Dimitri. Quadratic voting in blockchain governance. *Information*, 13(6):305, 2022.

13 Nicola Dimitri. Voting in daos. *Distributed ledger technologies: Research and practice*, 2(4):1–12, 2023.

14 Jeff Emmett. Conviction voting: A novel continuous decision making alternative to governance. https://blog.giveth.io/conviction-voting-a-novel-continuous-decision-making-alternative-to-governance-aa746cfb9475, 2019. Giveth Blog. Accessed: 2026-03-14.

15 Timothy Feddersen and Wolfgang Pesendorfer. Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *American Political Science Review*, 92(1):23–35, 1998.

16 Jerry Green and Jean-Jacques Laffont. *Incentives in public decision-making*. Elsevier North-Holland, 1979.

17 Jiasun Han, Jongsub Lee, and Tao Li. Dao governance. *Journal of Corporate Finance*, 2025.

18 Robin Hanson. Shall we vote on values, but bet on beliefs? *Journal of Political Philosophy*, 21(2):151–178, 2013.

19 John H Kagel and Dan Levin. The winner's curse and public information in common value auctions. *The American economic review*, pages 894–920, 1986.

20 Aggelos Kiayias and Philip Lazos. Sok: blockchain governance. In *Proceedings of the 4th ACM Conference on Advances in Financial Technologies*, pages 61–73, 2022.

21 Stefan Kitzler, Stefano Balietti, Pietro Saggese, Hans-Martin Bernhard, and Markus Strohmaier. The governance of decentralized autonomous organizations: A study of contributors' influence, networks, and shifts in voting power. In *Financial Cryptography and Data Security (FC)*, 2024.

22 Steven P Lalley and E Glen Weyl. Quadratic voting: How mechanism design can radicalize democracy. In *AEA Papers and Proceedings*, volume 108, pages 33–37. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, 2018.

23 Clément Lesaege, Federico Ast, and William George. Kleros: Short paper v1.0.7, 2019. Accessed: 2026-03-25. URL: https://kleros.io/whitepaper.pdf.

24 Hao Li, Sherwin Rosen, and Wing Suen. Conflicts and common interests in committees. *NBER Working Paper*, (7158), 1999.

25 MetaDAO. Metadao, n.d. Accessed: 2026-03-13. URL: https://www.metadao.fi/.

26 Caspar Oesterheld and Vincent Conitzer. Decision scoring rules. In *WINE*, page 468, 2020.

27 Abraham Othman and Tuomas Sandholm. Decision rules and decision markets. In *AAMAS*, pages 625–632, 2010.

28 Polkadot Wiki. Web3 and polkadot, 2025. Accessed: 2026-03-12. URL: https://wiki.polkadot.com/general/web3-and-polkadot/.

29 Project Catalyst. Project catalyst documentation. https://docs.projectcatalyst.io/, n.d. Accessed: 2026-03-12.

30 Roopesh Ranjan and Tilmann Gneiting. Combining probability forecasts. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(1):71–91, 2010.

31 Olivier Rikken, Marijn Janssen, and Zenlin Kwee. Governance challenges of blockchain and decentralized autonomous organizations. *Information Polity*, 24(4):397–417, 2019.

32 Kevin Roberts. The characterization of implementable choice rules. *Aggregation and revelation of preferences*, 12(2):321–348, 1979.

33 Siddarth Srinivasan, David Karger, and Yiling Chen. Self-resolving prediction markets for unverifiable outcomes. *arXiv preprint arXiv:2306.04305*, 2023.
34 Nimrod Talmon. Social choice around decentralized autonomous organizations: On the computational social choice of digital communities. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1768–1773, 2023.
35 John Von Neumann and Oskar Morgenstern. Theory of games and economic behavior: 60th anniversary commemorative edition. In *Theory of games and economic behavior*. Princeton university press, 2007.
36 Qin Wang, Guangsheng Yu, Yilin Sai, Caijun Sun, Lam Nguyen, and Shiping Chen. Understanding daos: An empirical study on governance dynamics. *IEEE Transactions on Computational Social Systems*, PP:1–19, 01 2025. doi:10.1109/TCSS.2025.3539889.
37 Natthawat Wanitchaicharoen and Srinath Perera. Dao voting mechanism resistant to whale and collusion problems. *Frontiers in Blockchain*, 7, 2024.
38 Robert L. Winkler, Yael Grushka-Cockayne, Kenneth C. Lichtendahl, and Victor Richmond R. Jose. Probability forecasts and their combination: A research perspective. *Decision Analysis*, 16(4):239–260, 2019.
39 Bingsheng Zhang, Roman Oliynykov, and Hamed Balogun. A treasury system for cryptocurrencies: Enabling better collaborative intelligence. *Cryptology ePrint Archive*, 2018.

A Proofs of no Weak Accountability

▶ **Proposition 18.** *Majority Voting (MV) does not satisfy Weak Accountability.*

Proof. In Majority Voting, the transfer and reward rules are identically zero for all experts: $\forall n \in N, t_n = 0$ and $r_n = 0$. Consider a state where expert i is strictly pivotal in implementing a decision that yields a negative outcome ($\Delta < 0$), while expert j is non-pivotal. By the strict penalty condition of Weak Accountability, we must have $t_i + r_i < t_j + r_j$. Substituting the mechanism's rules yields $0 < 0$, which is a contradiction. Thus, Majority Voting fails to satisfy Weak Accountability. ◀

▶ **Proposition 19.** *Quadratic Voting (QV) does not satisfy Weak Accountability.*

Proof. In Quadratic Voting, transfers are determined ex-ante by the number of votes purchased v_n, such that $t_n = -v_n^2$, and ex-post rewards are absent ($r_n = 0$). Consider a message profile where expert i purchases 1 vote ($v_i = 1 \implies t_i = -1$) and is pivotal in passing a proposal. Agent j strongly opposes the proposal, purchases 3 votes ($v_j = 3 \implies t_j = -9$), but remains non-pivotal. Assume the enacted proposal yields a negative outcome ($\Delta < 0$). Weak Accountability requires the pivotal expert to incur a strictly greater penalty: $t_i + r_i < t_j + r_j$. Substituting the respective transfers yields $-1 < -9$, which is a contradiction. Thus, QV fails to satisfy Weak Accountability. ◀

▶ **Proposition 20.** *The Pivotal Mechanism (VCG) does not satisfy Weak Accountability.*

Proof. In the VCG mechanism, an expert's transfer equals the externality they impose on others, meaning $t_n \leq 0$, and there are no ex-post rewards ($r_n = 0$). For a non-pivotal expert j, the imposed externality is zero, hence $t_j = 0$. For a strictly pivotal expert i, the transfer is strictly negative, $t_i < 0$. Assume expert i is pivotal in implementing a decision that subsequently yields a positive outcome ($\Delta > 0$). Weak Accountability requires that the pivotal expert is weakly better off: $t_i + r_i \geq t_j + r_j$. Substituting the known transfers yields $t_i \geq 0$, which contradicts the condition that $t_i < 0$. Thus, VCG fails to satisfy Weak Accountability. ◀

▶ **Proposition 21.** *Decision Scoring Rules (DSR) do not satisfy Weak Accountability.*

Proof. Under Decision Scoring Rules, experts are rewarded based on the accuracy of their individual predictions regarding the realized state, meaning $t_n = 0$ and $r_n = S(p_n, \Delta)$ for some proper scoring rule S. Suppose a highly-weighted expert i submits a forecast $p_i = 0.9$ and is pivotal in selecting an alternative. A lower-weighted expert j submits an identical forecast $p_j = 0.9$ but remains non-pivotal. Assume the chosen alternative results in a negative outcome ($\Delta < 0$). Because both experts submitted identical forecasts, the scoring rule yields $r_i = r_j$. Weak Accountability requires a strict penalty for the pivotal expert in a negative state: $t_i + r_i < t_j + r_j$, which reduces to $r_i < r_j$. This contradicts $r_i = r_j$. Thus, DSR fails to satisfy Weak Accountability. ◀

▶ **Proposition 22.** *Decision Markets (DM) do not satisfy Weak Accountability.*

Proof. In a Decision Market, payoffs depend strictly on the capital deployed to acquire shares (t_n) and the final market payout (r_n). Suppose expert i executes a marginal trade that pushes the market price past the execution threshold, making them strictly pivotal, and acquires exactly k shares of Alternative A for a total cost of c. Agent j executed a trade earlier, is non-pivotal to the threshold, but also acquires exactly k shares of Alternative A for a cost of c. Assume the outcome is negative ($\Delta < 0$) and the shares expire worthless ($r_i = r_j = 0$). Both experts incur an identical net loss: $t_i + r_i = -c = t_j + r_j$. Weak Accountability requires $t_i + r_i < t_j + r_j$, which contradicts the equality. Thus, DM fails to satisfy Weak Accountability. ◀