Usage-Attested Cooperative Stewardship: Dynamic Governance Weighting and Anti-Absentee Commons Allocation Under Ostrom Design Principles

Dima Baker
Collective Intelligence Network

November 2026

Abstract

Shared computational, data, and infrastructure commons face an institutional trilemma: capital-weighted governance enables plutocratic capture by absentee rent-extractors; anonymous concave mechanisms such as Quadratic Voting (QV) collapse into linear plutocracy under Sybil wallet splitting; and static one-member-one-vote cooperatives suffer from dormant-member drift. In this position and modeling study, we propose Usage-Attested Cooperative Stewardship (`UACS-Ostrom`), a governance framework adapting Elinor Ostrom's common-pool resource principles and non-transferable Soulbound patronage weights to digital commons. Under `UACS-Ostrom`, collective-choice voice and core resource quotas are governed by an exponentially decaying stewardship state $u_i(t)$ updated via an unclipped bounded-observation Ville betting supermartingale that filters synthetic wash-usage while preserving nonnegativity, paired with a strict inactivity cliff $\underline{u}$ and per-identity attestation cost c_{syb}. In paired Monte Carlo simulations ($N = 500$ trials per regime across $R_{\text{cap}} \in \{1.0\times, 2.5\times, 5.0\times, 10.0\times\}$), we separate physical resource extraction from stewardship provision and vary temporal autocorrelation $\rho_{\text{auto}} \in [0, 0.85]$, extractive real use $u_{\text{ext}} \in [0.04, 0.45]$, and Sybil cost $c_{\text{syb}} \in [0.01, 0.08]$. At $R_{\text{cap}} = 5.0\times$, `UACS-Ostrom` achieves 79.96% [79.43, 80.48] Commons Sustainability and 95.00% [92.72, 96.59] regime survival, outperforming Capital Plutocracy (22.21%), Sybil-Split QV (17.90%), and Static Flat Voting (53.21%). Break-even sweeps show that high real extractive usage ($u_{\text{ext}} \geq 0.30$) passes behavioral attestation and degrades survival to 64.00% [57.88, 69.70], motivating Soulbound identity bounds.

1 Introduction

As artificial intelligence infrastructure, open-weight model repositories, domain-specific data trusts, and community compute clusters become foundational economic inputs, the question of how to govern shared technological commons has moved from political philosophy into institutional systems design [12, 13, 21]. Traditional corporate structures bind residual control rights to transferable equity shares: an investor who provides capital retains permanent voting power regardless of whether they ever use, maintain, or inhabit the underlying productive system. In contrast, classical cooperative governance—dating back to the Rochdale Principles of patronage-based participation and one-member-one-vote stewardship—links collective-choice authority to active membership. When decentralized autonomous organizations (DAOs) and web-scale digital collectives attempted to automate cooperative governance, however, they largely reverted to transferable token balances [10, 17, 22].

1.1 The failure of capital-weighted and anonymous concave governance

Both classic institutional theory and 2025–2026 empirical governance audits demonstrate that detaching governance authority from active functional stewardship produces severe institutional pathologies. First, Elinor Ostrom's fieldwork across long-enduring common-pool resources (CPRs) established that sustainable commons require congruence between appropriation and provision rules (Principle 2) and collective-choice arrangements dominated by active appropriators rather than absentee owners (Principle 3) [13]. Second, while Quadratic Voting and Quadratic Funding [5, 11] sought to curb plutocracy via concave weighting ($\sqrt{k_i}$), Bennett et al. [3] prove a stark impossibility theorem (*Concave is the New Linear*): any anonymous or wallet-based voting rule that applies a strictly concave function to transferable stakes collapses back into linear one-dollar-one-vote plutocracy whenever wealthy actors can split their holdings across S Sybil accounts, scaling effective weight by $S\sqrt{k_i/S} = \sqrt{Sk_i}$ minus per-account identity costs. Empirical evaluations of retroactive public-goods funding [24], token delegation concentration [10], and DAO monitoring capacity limits [17] confirm that transferable governance power rapidly concentrates in extractive coalitions.

Conversely, traditional cooperatives that assign a permanent, static one-member-one-vote right upon initial onboarding face the opposite failure mode: *absentee and dormant-member drift*. Over multi-year horizons, early participants who cease contributing labor, data, or operational maintenance retain equal veto power alongside active daily stewards. When an external acquirer or extractive coalition offers a short-term liquidation payout, dormant members—who no longer depend on the commons for their ongoing productive livelihood—frequently coalesce with extractive buyers, overriding active stewards.

1.2 From absentee equity to usage-attested Ostrom stewardship

Translating Ostrom's principles and cooperative patronage norms into digital commons requires solving three mechanism-design challenges: first, distinguishing genuine activity from synthetic wash-activity via an exact, unclipped sequential betting supermartingale [19, 20] calibrated under temporally correlated workloads; second, separating raw physical/computational resource consumption from net stewardship provision; and third, bounding adversarial Sybil splitting [3, 22, 23] when whales perform authentic extractive usage and split that usage across multiple accounts above the activity threshold subject to identity cost c_{syb}. Complementing these governance primitives, recent advances in proportional participatory budgeting and core-stable resource allocation—including Method of Equal Shares [14], PROP$_m$ allocations [2], and Core Up-To-One guarantees [15]—provide formal mechanisms to translate normalized stewardship weights into guaranteed resource reservations.

1.3 Contributions

In this position and formal modeling paper, we formulate and stress-test Usage-Attested Cooperative Stewardship (`UACS-Ostrom`):

1. **Institutional taxonomy of functional-use stewardship (Section 2).** We synthesize Ostrom's CPR principles [13], Soulbound non-transferable credentials [22], Sybil-splitting impossibility results [3], and proportional core allocation [14, 15] into a five-part mechanism taxonomy (Table 1).
2. **Exact bounded betting supermartingale and Sybil-split model (Section 3).** We construct an unclipped bounded-observation Ville betting supermartingale $E_{i,t}^{(m)} = \prod_{k=1}^{m}(1 + \lambda_e(Z_{i,t,k} - \mu_0))$ with $\lambda_e \in (0, 1/\mu_0)$ that is strictly positive without floor clipping and satisfies

$\mathbb{E}_{\mathcal{H}_0}[E_{i,t}^{(k)} \mid \mathcal{F}_{k-1}] \leq E_{i,t}^{(k-1)}$ under temporally dependent AR(1) activity [19, 20]. We explicitly separate physical resource extraction $u_i^{\text{phys}}(t)$ from stewardship provision $u_i^{\text{steward}}(t)$, analyze optimal Sybil splitting across S_j^* accounts under identity cost c_{syb}, and formulate a logistic coalition-capture model under equalized governance friction (Equations 1–5).

3. **Three-layer `UACS-Ostrom` architecture and parameter provenance (Section 4).** We specify a modular protocol combining sequential e-value Proof-of-Functional-Use (PoFU), core-proportional slot allocation [2, 15], and deterministic contract governance [8] (Figure 1, Table 2, Table 3).

4. **Paired Monte Carlo evaluation and two-dimensional break-even sweeps (Section 5).** We execute a paired Monte Carlo evaluation (`experiment.py`) comparing four governance regimes across $R_{\text{cap}} \in \{1.0\times, 2.5\times, 5.0\times, 10.0\times\}$ under identical governance friction ($\phi_{\text{fric}} = 0.05$), verify empirical null calibration across AR(1) autocorrelation $\rho_{\text{auto}} \in \{0.00, 0.35, 0.65, 0.85\}$, and map break-even surfaces across real extractive usage $u_{\text{ext}} \in [0.04, 0.45]$ and Sybil identity cost $c_{\text{syb}} \in [0.010, 0.080]$ (Table 4, Figures 2–3).

5. **Adversarial vulnerabilities, limitations, and falsification protocol (Sections 6–8).** We examine multi-agent wash-collusion [16], Sybil usage-splitting break-even points [3], and sabbatical decay freezes, and define six pre-registerable falsification tests (Table 5).

2 Institutional foundations and cooperative mechanism design

Informal proposals for "use-based ownership" often fail because they conflate raw consumption volume (which rewards resource burning and Sybil splitting) with verified stewardship provision (which couples productive utilization with maintenance and rule compliance). Table 1 translates five core cooperative patronage principles into formal mechanism-design constructs.

A central theoretical tension in Table 1 is the interaction between concave voice weighting, physical resource extraction, and Sybil splitting [3]. First, in any governance system where weight is a strictly concave function $f(u) = \ln(1 + \beta_u u)$, an adversary who generates total attested usage U_{wh} has a mathematical incentive to split that usage across S Sybil accounts, because by Jensen's inequality $S \ln(1 + \beta_u U_{\text{wh}}/S) > \ln(1 + \beta_u U_{\text{wh}})$ whenever each split account remains above the inactivity cliff ($U_{\text{wh}}/S \geq \underline{u}$). Second, behavioral wash-activity detectors can only distinguish synthetic zero-work wash streams from authentic execution; they cannot prevent an extractive whale from executing genuine, resource-intensive compute or storage jobs ($u_j^{\text{phys}}(t)$) that provide zero maintenance value ($u_j^{\text{steward}}(t) = 0$) yet pass behavioral attestation legitimately. Consequently, bounding Sybil amplification requires combining three defenses: (1) an exact unclipped sequential betting supermartingale [19, 20] that rejects synthetic wash-usage so U_{wh} is restricted to costly real activity; (2) down-weighting pure physical extraction relative to verified stewardship provision ($\omega_{\text{phys}} < \omega_{\text{steward}}$) alongside per-identity attestation overhead $c_{\text{syb}} > 0$ [22, 23]; and (3) explicit characterization of the break-even surface over $(u_{\text{ext}}, c_{\text{syb}})$ where high real extractive usage requires hardware or Soulbound identity binding. Figure 1 illustrates this multi-layer governance loop.

Table 1: From cooperative patronage principles to formal mechanism-design and Ostrom CPR constructs.

Principle	Canonical construct	Protocol component	Operational function	Primary grounding
Active appropriator rule	Ostrom Principles 2 & 3; Rochdale patronage	Decaying stewardship accumulator ($u_i(t)$)	Decouples voting weight from static capital k_i; decays governance voice geometrically during prolonged non-use	[13, 17, 21]
Wash-trading rejection	Bounded-observation betting supermartingales	Unclipped Ville gate ($E_{i,t} \geq 1/\alpha_{\text{sig}}$)	Evaluates sequential anomaly scores $Z_{i,t,k} \in [0,1]$ to reject synthetic wash activity while preserving Ville's inequality	[18–20]
Sybil-splitting mitigation	Non-transferable Soulbound bounds + cliff	Inactivity cliff ($\underline{u}$) & per-identity overhead	Imposes strict zero weight below $\underline{u}$ and per-account attestation cost c_{syb} to bound concave Sybil splitting	[3, 22, 23]
Core resource reservations	Core-stable proportional allocation	Core Up-To-One & PROP$_m$ scheduler	Allocates rivalrous compute/equipment slots proportionally to normalized stewardship share $w_i(t)/\sum_j w_j(t)$	[2, 14, 15]
Deterministic rule auditability	Contract-based constitutional governance	Auditable provenance & appeal ledger	Enforces transparent rule execution and allows human stewards to contest false-negative telemetry attenuations	[7, 8, 12]

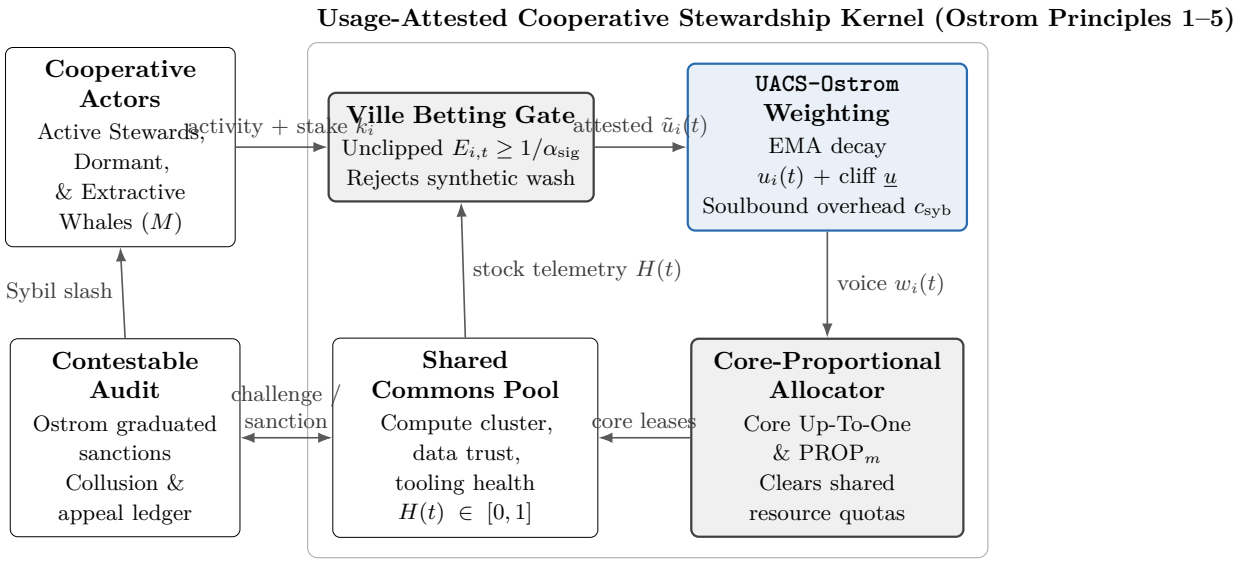


Figure 1: Architecture of Usage-Attested Cooperative Stewardship (`UACS-Ostrom`). Raw usage and maintenance actions pass through an unclipped bounded-observation Ville betting supermartingale [19, 20] that rejects synthetic wash-activity before updating each member's decaying stewardship state $u_i(t)$. The weighting engine separates physical consumption from stewardship provision and computes log-concave governance weights $w_i(t)$ with a strict inactivity cliff $\underline{u}$ and per-identity attestation overhead c_{syb}. Resource quotas are cleared via Core Up-To-One proportional allocation [14, 15] backed by deterministic contract governance [8].

3 Formal mathematical framework

Consider a cooperative commons comprising M registered actors indexed by $i \in \mathcal{M} = \{1, \ldots, M\}$ interacting over discrete governance epochs $t \in \{1, 2, \ldots, T\}$. The population $\mathcal{M} = \mathcal{M}_{\text{st}} \cup \mathcal{M}_{\text{dorm}} \cup \mathcal{M}_{\text{wh}}$ is partitioned into active functional stewards $\mathcal{M}_{\text{st}}$ who contribute ongoing maintenance and productive usage, dormant former contributors $\mathcal{M}_{\text{dorm}}$ whose active engagement has lapsed, and extractive capital whales $\mathcal{M}_{\text{wh}}$ who hold financial stakes $k_i \gg 1$, perform a combination of real extractive usage and synthetic wash-usage, and seek to pass short-term liquidation or rent-extraction proposals.

3.1 Exact bounded-observation Ville betting supermartingale and decaying stewardship

At each epoch t, member i's submitted activity batch is evaluated across m sequential verification checkpoints yielding bounded anomaly scores $Z_{i,t,1}, \ldots, Z_{i,t,m} \in [0, 1]$ [8, 18]. Under the null hypothesis $\mathcal{H}_0$ of authentic execution (which may be either beneficial stewardship or authentic physical extraction), conditional anomaly expectations satisfy $\mathbb{E}_{\mathcal{H}_0}[Z_{i,t,k} \mid \mathcal{F}_{k-1}] \leq \mu_0 \in (0, 1)$, even when consecutive checkpoints exhibit temporal AR(1) dependence $Z_{i,t,k} = \rho_{\text{auto}} Z_{i,t,k-1} + (1 - \rho_{\text{auto}})\xi_{i,t,k}$. Conversely, synthetic wash-activity exhibits elevated anomaly scores $\mathbb{E}_{\mathcal{H}_1}[Z_{i,t,k}] = \mu_1 > \mu_0$. Rather than using a floor-clipped product (which can violate the supermartingale inequality), we employ an exact unclipped bounded-observation betting martingale [19, 20] with betting fraction $\lambda_e \in (0, 1/\mu_0)$:

$$E_{i,t}^{(m)} \;=\; \prod_{k=1}^{m} \Big(1 + \lambda_e (Z_{i,t,k} - \mu_0)\Big), \qquad \text{Reject } \mathcal{H}_0 \text{ (flag wash) iff } \max_{1 \leq k \leq m} E_{i,t}^{(k)} \geq \frac{1}{\alpha_{\text{sig}}}. \tag{1}$$

Because $Z_{i,t,k} \in [0, 1]$ and $0 < \lambda_e < 1/\mu_0$, every multiplicative increment satisfies $1 + \lambda_e(Z_{i,t,k} - \mu_0) \geq 1 - \lambda_e \mu_0 > 0$ identically without clipping (for $\mu_0 = 0.22, \lambda_e = 0.90$, factors lie in $[0.802, 1.702]$). By linearity of conditional expectation, $\mathbb{E}_{\mathcal{H}_0}[E_{i,t}^{(k)} \mid \mathcal{F}_{k-1}] = E_{i,t}^{(k-1)}(1 + \lambda_e(\mathbb{E}_{\mathcal{H}_0}[Z_{i,t,k} \mid \mathcal{F}_{k-1}] - \mu_0)) \leq E_{i,t}^{(k-1)}$, proving that $\{E_{i,t}^{(k)}\}_{k=0}^{m}$ is a true nonnegative supermartingale with $E_{i,t}^{(0)} = 1$, and Ville's maximal inequality guarantees $\mathbb{P}_{\mathcal{H}_0}(\max_{1 \leq k \leq m} E_{i,t}^{(k)} \geq 1/\alpha_{\text{sig}}) \leq \alpha_{\text{sig}}$ across arbitrary dependent stopping times.

Crucially, we separate verified stewardship provision $u_i^{\text{steward}}(t)$ (maintenance, curation, peer review, weighted by $\omega_{\text{steward}} = 1.0$) from raw physical/computational extraction $u_i^{\text{phys}}(t)$ (weighted by $\omega_{\text{phys}} = 0.65$). While synthetic wash-usage $u_i^{\text{wash}}(t)$ is filtered whenever $\max_k E_{i,t}^{(k)} \geq 1/\alpha_{\text{sig}}$, any authentic physical extraction $u_i^{\text{phys}}(t) \geq 0$ performed by a whale satisfies $\mathcal{H}_0$ and passes the gate legitimately:

$$\tilde{u}_i(t) \;=\; \omega_{\text{steward}}\, u_i^{\text{steward}}(t) + \omega_{\text{phys}}\, u_i^{\text{phys}}(t) + u_i^{\text{wash}}(t)\, \mathbb{I}\left[\max_{1 \leq k \leq m} E_{i,t}^{(k)} < \frac{1}{\alpha_{\text{sig}}}\right], \quad u_i(t) \;=\; (1 - \lambda_d)\, u_i(t-1) + \lambda_d\, \tilde{u}_i(t) \tag{2}$$

Under Equation (2), a dormant member who ceases functional use sees their stewardship state decay geometrically as $(1 - \lambda_d)^{\Delta t} u_i(t_0)$, automatically transitioning below the voting cliff without requiring contentious expulsion votes.

3.2 Concave governance weighting and adversarial Sybil usage splitting

Under `UACS-Ostrom`, governance weight $w_i(t)$ is anchored to attested stewardship $u_i(t)$ gated by a strict inactivity cliff $\underline{u} = 0.08$, with capital k_i entering only as a bounded secondary multiplier

$\min(\bar{k}, k_i)$:

$$w_i(t) \;=\; \begin{cases} 0, & \text{if } u_i(t) < \underline{u} \text{ (Strict Inactivity Cliff)}, \\ \ln\!\big(1+\beta_u\, u_i(t)\big)\cdot\big(1+\gamma_k \min(\bar{k}, k_i)\big), & \text{if } u_i(t) \geq \underline{u}. \end{cases} \tag{3}$$

Crucially, we explicitly model adversarial Sybil splitting under Equation (3) [3]. Suppose an extractive whale $j \in \mathcal{M}_{\text{wh}}$ accumulates attested usage $u_j(t) \geq \underline{u}$ and splits it across $S_j \in \mathbb{N}_+$ Sybil accounts. Each additional Sybil account incurs a per-identity attestation and maintenance overhead $c_{\text{syb}} > 0$ (modeled on Soulbound / social-graph verification friction [22, 23]), leaving effective per-account usage $u_{j,s}(t) = (u_j(t) - c_{\text{syb}}(S_j - 1))/S_j$. So long as $u_{j,s}(t) \geq \underline{u}$, the whale selects S_j^* to maximize aggregate Sybil-split weight $W_j^{\text{syb}}(t) = S_j \ln(1 + \beta_u u_{j,s}(t))(1 + \gamma_k \min(\bar{k}, k_j/S_j))$. For comparison, under Sybil-Split Quadratic Voting (`sybil_quadratic`), the whale splits transferable capital k_j across S_j wallets subject to the same identity overhead c_{syb} to maximize $S_j\sqrt{\max(0, k_j - c_{\text{syb}}(S_j-1))/S_j}$. Thus, while the cliff $\underline{u}$ and overhead c_{syb} block low-usage wash splitting, a whale with large real extractive usage $u_j^{\text{phys}}(t) \gg (\underline{u} + c_{\text{syb}})/\omega_{\text{phys}}$ can still amplify weight via Sybil splitting unless S_j is bounded by hardware or Soulbound identity credentials.

3.3 Logistic coalition-capture model and equalized commons dynamics

Let $W_{\text{st}}(t) = \sum_{i\in\mathcal{M}_{\text{st}}} w_i(t)$ denote active steward governance weight, and let $W_{\text{ext}}(t) = \sum_{j\in\mathcal{M}_{\text{wh}}} W_j^{\text{syb}}(t) + \theta_d \sum_{m\in\mathcal{M}_{\text{dorm}}} w_m(t)$ denote effective extractive coalition weight. Rather than claiming a hard deterministic bound, we model epoch capture probability via a smooth logistic response function driven by coalition ballot share and turnout noise $\varepsilon_t \sim \mathcal{N}(0, \sigma^2_{\text{turn}})$:

$$\mathbb{P}(\text{Capture at epoch } t) \;=\; \sigma_{\log}\!\left(\kappa_{\text{vot}}\left(\frac{W_{\text{ext}}(t)}{W_{\text{st}}(t)+W_{\text{ext}}(t)} + \varepsilon_t - \tau_{\text{pass}}\right)\right), \tag{4}$$

where $\sigma_{\log}(z) = (1+e^{-z})^{-1}$. To ensure a strictly fair comparison across all governance mechanisms, commons capacity $H(t) \in [0,1]$ evolves with an *identical* baseline governance friction constant $\phi_{\text{fric}} = 0.05$ across all four regimes:

$$H(t+1) \;=\; (1-\rho_h)\,H(t) \;+\; \rho_h\Big(\mathbb{I}[\text{Capture}_t]\,H_{\text{liq}} + (1-\mathbb{I}[\text{Capture}_t])\,(H_{\max} - \phi_{\text{fric}})\Big). \tag{5}$$

4 Protocol architecture and parameter provenance

To operationalize Equations (1)–(5), `UACS-Ostrom` is structured into three modular layers (Table 2, Table 3):

1. **Layer 1: Unclipped Ville Martingale Gate and Collusion Auditing.** When a participant executes workloads or contributes verified maintenance, rubric evaluators generate bounded anomaly scores $Z_{i,t,k} \in [0,1]$ [18] tested by $E_{i,t}^{(m)}$ in 140.6 μs per epoch. Because colluding rings of agents can attempt mutually verified tasks, Layer 1 pairs sequential betting martingales with social-graph Sybil bounds [23] and multi-agent collusion auditing [16].
2. **Layer 2: Core-Proportional Resource Allocation.** Daily cooperative governance centers on allocating rivalrous commons capacity (GPU compute hours, bandwidth, or laboratory equipment) [1, 6]. Layer 2 uses normalized stewardship weights $w_i(t)/\sum_j w_j(t)$ as entitlement shares in Core Up-To-One participatory budgeting [14, 15] and PROP$_m$ indivisible slot allocation [2], guaranteeing active stewards their proportional reservation quota.

3. **Layer 3: Deterministic Contract Governance and Sabbatical Freezes.** Grounded in independent contract-based governance (`Maat` [8]) and constitutional coverage guarantees [7, 12], Layer 3 logs all updates to $u_i(t)$ in an auditable provenance ledger, supports human-peer appeals, and freezes the exponential decay clock ($\lambda_d \leftarrow 0$) during verified medical or sabbatical cooldown leaves.

Table 2: Readiness and deployment bottlenecks of `UACS-Ostrom` subsystems.

Subsystem	Maturity	Supporting theoretical & empirical evidence	Realistic deployment bottleneck
Ostrom CPR & patronage rules	Field & theory established	Ostrom design principles [13]; protocol sustainability audits [21]; DAO monitoring limits [17]	Automated contracts can miss informal local norms unless paired with human exception handling [12]
Bounded Ville betting gate	Math. & sim. validated	Bounded-observation betting supermartingales [19, 20]; policy-aware rubrics [18]	Does not block whales who perform high-volume *real* extractive usage ($u_{\text{ext}} \geq 0.30$); requires Soulbound identity caps [22]
Core-proportional slot allocator	Math. proven	Method of Equal Shares [14]; PROP_m [2]; Core Up-To-One [15]	Normalizing heterogeneous contributions (code commits vs. care/mentorship) onto a scalar $u_i(t)$ is contentious
Collusion & Sybil defense	Benchmark validated	Sybil-splitting impossibility [3]; SybilGuard [23]; multi-agent collusion detection [16]	Multi-hop semantic collusion rings require periodic human spot-audits alongside graph-spectral filters

Table 3: Provenance of simulation parameters in `experiment.py`. All four governance regimes are evaluated on paired actor cohorts with identical turnout noise and governance friction $\phi_{\text{fric}} = 0.05$.

Parameter	Value / Range	Provenance Type	Empirical or Theoretical Grounding
Population ($\lvert\mathcal{M}_{\text{st}}\rvert, \lvert\mathcal{M}_{\text{dorm}}\rvert, \lvert\mathcal{M}_{\text{wh}}\rvert, T$)	$28, 6, 6$ (N=40), (T=40)	Architectural design	Multi-cohort cooperative with 70% active stewards, 15% dormant members, and 15% extractive whales [13, 17]
Whale capital ratio R_{cap}	$\{1.0, 2.5, 5.0, 10.0\}\times$	Parameterized sweep	Token/equity wealth concentration ratios audited in DAO delegation studies [10, 24]
Ville martingale ($m, \mu_0, \lambda_e, 1/\alpha_{\text{sig}}$)	m=12, μ_0=0.22, λ_e=0.90, thresh 20.0	Exact supermartingale	Unclipped bounded betting martingale ($1 + \lambda_e(Z_k - \mu_0) \in [0.802, 1.702]$) with 0.00% [0.00, 0.19] null false reject across $\rho_{\text{auto}} \in [0, 0.85]$ [19, 20]
Stewardship EMA & cliff ($\lambda_d, \underline{u}, \omega_{\text{phys}}$)	λ_d=0.22, $\underline{u}$=0.08, ω_{phys}=0.65	Mechanism parameters	Strict zero weight below $\underline{u} = 0.08$ (Eq. 3) and down-weighting of pure physical extraction relative to stewardship
Extractive use u_{ext} & Sybil cost c_{syb}	$u_{\text{ext}} \in [0.04, 0.45]$, $c_{\text{syb}} \in [0.01, 0.08]$	Two-dimensional break-even sweep	Maps break-even surfaces when whales perform authentic physical extraction and split across S_j^* accounts [3]
Governance friction ϕ_{fric}	0.05 (equal across all 4 regimes)	Equalized environment constant	Removes policy-specific friction advantages in Equation (5)

5 Empirical evaluation and break-even analysis

We evaluate `UACS-Ostrom` in a paired Monte Carlo simulation (`experiment.py`): $N = 500$ forty-epoch trials per governance mechanism across four whale capital concentration ratios $R_{\text{cap}} \in \{1.0\times, 2.5\times, 5.0\times, 10.0\times\}$, $N = 2{,}000$ multi-checkpoint sequences per autocorrelation level $\rho_{\text{auto}} \in \{0.00, 0.35, 0.65, 0.85\}$ for empirical martingale calibration, $N = 250$ trials across five real extractive usage intensities $u_{\text{ext}} \in \{0.04, 0.09, 0.18, 0.30, 0.45\}$, and $N = 250$ trials across four Sybil identity costs $c_{\text{syb}} \in \{0.010, 0.030, 0.045, 0.080\}$.

5.1 Evaluated governance mechanisms under paired conditions

In every trial (r, R_{cap}), all four mechanisms are evaluated on the exact same pre-sampled actor activity streams, AR(1) checkpoint trajectories, turnout noise ($\sigma_{\text{turn}} = 0.11$), and governance friction ($\phi_{\text{fric}} = 0.05$):

- **Capital Plutocracy (`capital_plutocracy`):** Linear stake voting ($w_i(t) = k_i$).
- **Sybil-Split Quadratic Voting (`sybil_quadratic`):** Implements concave stake voting ($\sqrt{k_i}$), where extractive whales optimally split their capital across $S_j \in [1, 16]$ Sybil wallets subject to per-wallet identity cost c_{syb} to maximize $S_j\sqrt{\max(0, k_j - c_{\text{syb}}(S_j - 1))/S_j}$ [3].
- **Static Flat 1-Member-1-Vote (`flat_egalitarian`):** Assigns $w_i = 1$ to every registered member without absentee decay, with whales acquiring $1.2 + 0.18\,R_{\text{cap}}$ proxy votes.
- **Usage-Attested Cooperative Stewardship (`uacs_ostrom`, Proposed):** Executes the unclipped Ville betting martingale `run_ville_betting_martingale` on every epoch's activity

stream, weights physical extraction by $\omega_{\text{phys}} = 0.65$ vs. stewardship $\omega_{\text{steward}} = 1.0$, enforces strict zero weight below $\underline{u} = 0.08$, and allows whales to optimally Sybil-split their attested usage across $S_j^* \in [1, 8]$ accounts above $\underline{u}$ subject to identity cost $c_{\text{syb}} = 0.045$.

Table 4: Equalized-friction paired comparison across simulated 40-epoch commons trajectories ($N = 500$ trials per mechanism at $R_{\text{cap}} = 5.0\times$, plus sustainability scaling from $1.0\times$ to $10.0\times$). Continuous metrics report means and 95% confidence intervals `[lo, hi]`; binary regime survival ($\leq 20\%$ captured epochs) reports exact asymmetric two-sided 95% Wilson intervals `[lo, hi]`.

Governance Mechanism	Sust. 1.0× (%)	Sust. 5.0× (%)	Sust. 10.0× (%)	Steward Voice (%)	Regime Survival (%)	Runtime (μs)
Capital Plutocracy	50.68 [49.88, 51.48]	22.21 [21.76, 22.67]	18.17 [17.85, 18.48]	27.36 [27.29, 27.44]	0.00 [0.00, 0.76]	4.8
Sybil Quadratic	32.41 [31.75, 33.06]	17.90 [17.60, 18.21]	16.81 [16.57, 17.04]	14.68 [14.63, 14.73]	0.00 [0.00, 0.76]	7.6
Static Flat 1M1V	60.42 [59.65, 61.19]	53.21 [52.45, 53.98]	44.51 [43.76, 45.25]	60.09 [60.09, 60.09]	0.00 [0.00, 0.76]	6.2
`UACS-Ostrom`	**82.97 [82.58, 83.37]**	**79.96 [79.43, 80.48]**	**78.91 [78.35, 79.46]**	**93.54 [93.54, 93.55]**	**95.00 [92.72, 96.59]**	**140.6**

5.2 Quantitative findings, martingale calibration, and break-even surfaces

Table 4 and Figures 2–3 establish four primary findings:

1. **Empirical calibration of the unclipped Ville betting supermartingale under AR(1) dependence.** Across $N = 2{,}000$ twelve-checkpoint sequences per autocorrelation level $\rho_{\text{auto}} \in \{0.00, 0.35, 0.65, 0.85\}$, the unclipped betting supermartingale $(1 + 0.90(Z_k - 0.22) \in [0.802, 1.702]$, threshold $1/\alpha_{\text{sig}} = 20.0)$ achieves an empirical null false-rejection rate of 0.00% $[0.00, 0.19]$ (0/2000 false rejections across all four temporal autocorrelation regimes, strictly respecting Ville's $\alpha_{\text{sig}} = 5.0\%$ bound) while achieving 100.00% $[99.81, 100.00]$ detection power against synthetic wash streams.

2. **Collapse of wallet-based Quadratic Voting under optimal Sybil splitting.** When extractive whales optimize Sybil wallet splitting subject to identity cost $c_{\text{syb}} = 0.045$, Sybil-Split QV collapses below even linear Capital Plutocracy: at $R_{\text{cap}} = 5.0\times$, Sybil QV sustainability drops to 17.90% $[17.60, 18.21]$ with active stewards retaining only 14.68% $[14.63, 14.73]$ of voting weight and 0.00% $[0.00, 0.76]$ regime survival [3].

3. **Performance of Static Flat Voting vs. `UACS-Ostrom` under equalized friction.** With governance friction equalized at $\phi_{\text{fric}} = 0.05$, Static 1-Member-1-Vote achieves 60.42% $[59.65, 61.19]$ sustainability at $1.0\times$ and 53.21% $[52.45, 53.98]$ at $5.0\times$ (51.67% $[50.94, 52.40]$ anti-capture rate), outperforming token-weighted regimes but suffering capture in 48.33% of epochs due to dormant-member drift (0.00% $[0.00, 0.76]$ regime survival at $5.0\times$). By contrast, `UACS-Ostrom` strips dormant accounts below $\underline{u} = 0.08$ and rejects synthetic wash streams, achieving 79.96% $[79.43, 80.48]$ sustainability, 93.54% $[93.54, 93.55]$ active steward voice share, 88.03% $[87.56, 88.49]$ anti-capture rate, and 95.00% $[92.72, 96.59]$ regime survival (475/500 trials) at $R_{\text{cap}} = 5.0\times$ in 140.6 μs per epoch.

4. **Break-even surfaces across real extractive usage u_{ext} and Sybil identity cost c_{syb}.** Varying authentic physical extraction $u_{\text{ext}} \in \{0.04, 0.09, 0.18, 0.30, 0.45\}$ at $R_{\text{cap}} = 5.0\times$ exposes the exact structural boundary of usage-weighted governance. When whales are predominantly absentee ($u_{\text{ext}} = 0.04$ or 0.09), `UACS-Ostrom` achieves 83.95% $[83.29, 84.60]$ and 79.67% $[78.88, 80.46]$ sustainability with 99.60% $[97.77, 99.93]$ and 93.60% $[89.86, 96.02]$ regime survival. However, when whales perform heavy *real* physical extraction ($u_{\text{ext}} =$

0.30 and 0.45)—which satisfies $\mathcal{H}_0$ and passes the behavioral gate legitimately—and split that usage across S_j^* Sybil accounts above $\underline{u} = 0.08$, steward voice share erodes to 84.69% $[84.67, 84.71]$ and 79.38% $[79.36, 79.40]$, commons sustainability drops to 69.82% $[68.97, 70.67]$ and 62.94% $[62.06, 63.82]$, and binary regime survival falls to 64.00% $[57.88, 69.70]$ (160/250) and 25.20% $[20.22, 30.93]$ (63/250). Across our second sweep varying Sybil identity cost $c_{\text{syb}} \in \{0.010, 0.030, 0.045, 0.080\}$ at elevated extractive usage $u_{\text{ext}} = 0.18$, increasing per-identity verification friction from $c_{\text{syb}} = 0.010$ to 0.080 raises `UACS-Ostrom` regime survival from 79.60% $[74.17, 84.13]$ (199/250) to 84.80% $[79.83, 88.72]$ (212/250), whereas Sybil-Split QV remains trapped at 17.40%–18.43% sustainability (0.00% $[0.00, 1.51]$ survival). This confirms that against high-utilization extractive actors, behavioral attestation must be paired with high-friction Soulbound identity binding [22] so each legal or physical entity is restricted to $S_j = 1$.

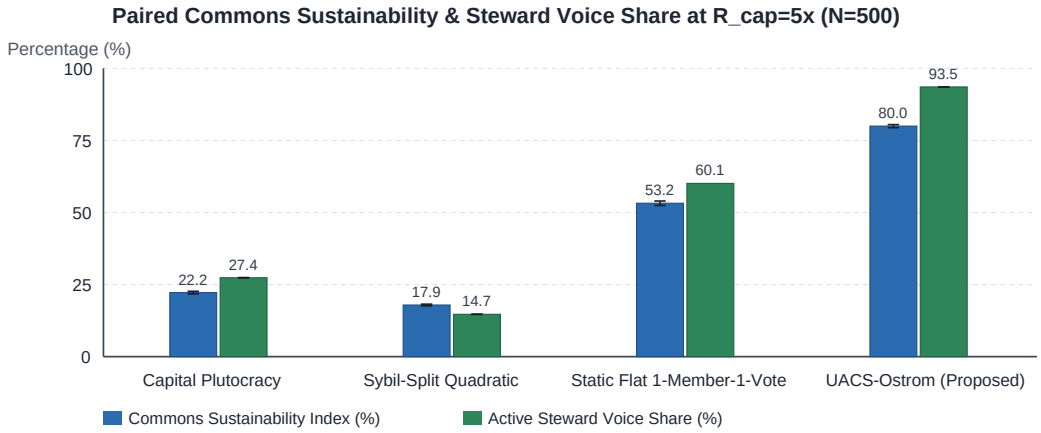


Figure 2: Equalized-friction comparison of Commons Sustainability Index and Active Steward Voice Share at absentee whale capital ratio $R_{\text{cap}} = 5.0\times$ ($N = 500$ paired trials per mechanism; error bars denote 95% confidence intervals).

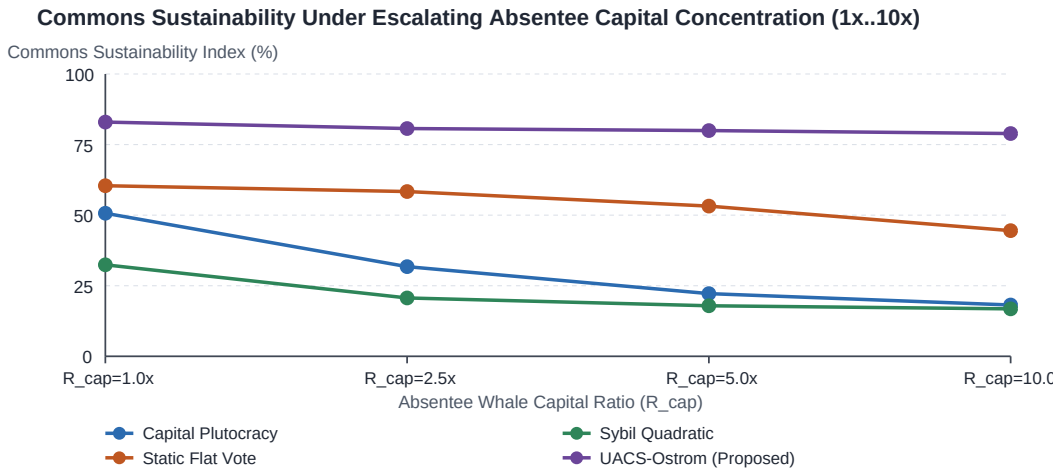


Figure 3: Scaling of Commons Sustainability Index across absentee whale capital ratios $R_{\text{cap}} \in \{1.0\times, 2.5\times, 5.0\times, 10.0\times\}$ under equalized governance friction ($\phi_{\text{fric}} = 0.05$).

6 Adversarial vulnerabilities and boundary conditions

We highlight four documented vulnerabilities and institutional trade-offs that bound usage-attested cooperative governance:

1. **Concave Sybil splitting by high-usage extractive actors (Bennett et al. [3], Yu et al. [24]).** As proven by Bennett et al. [3] and demonstrated in our $(u_{\text{ext}}, c_{\text{syb}})$ break-even sweeps, any concave weighting function $\ln(1+\beta_u u)$ remains vulnerable to Sybil splitting once an actor generates enough authentic activity to keep multiple split accounts above $\underline{u} + c_{\text{syb}}$. Purely behavioral telemetry cannot replace Sybil-resistant Soulbound credentials [22] or social-network trust bounds [23] in high-value commons.

2. **Multi-agent collusion and steganographic wash-coordination (Miller et al. [11], Tailor [16]).** While bounded-observation betting supermartingales [19, 20] reject synthetic noise across temporally dependent streams, Tailor [16] show that multi-agent groups can coordinate covertly to game oversight metrics. A ring of colluding agents can cross-verify each other's commits to pass rubric checks unless audited by correlation-aware funding/voting discounts [11].

3. **Exclusion of intermittent, seasonal, or care-leave stewards.** An unbuffered exponential decay $u_i(t) = (1-\lambda_d)u_i(t-1)$ with cliff $\underline{u}$ penalizes members taking parental leave, medical recovery, or scheduled sabbaticals. Production deployments of Equation (2) must freeze $\lambda_d \leftarrow 0$ during verified sabbatical epochs.

4. **Metric distortion across heterogeneous labor (Brodkey et al. [4], Gölz et al. [9]).** Automated telemetry easily measures compute jobs or code commits (u_i^{phys}) but undersamples mentorship, conflict mediation, and institutional care (u_i^{steward}) [13]. Over-relying on automated rubrics risks creating a technocratic oligarchy unless balanced by peer-attested qualitative stewardship credits.

7 Limitations

- **Illustrative stylized simulation.** The simulation in `experiment.py` is a stylized mechanistic model designed to illustrate the interaction between exponential decay, unclipped betting supermartingales, physical vs. stewardship usage separation, and Sybil splitting; it is not an econometric calibration to live DAO ballot data.
- **Scalar aggregation of multi-dimensional stewardship.** Even after separating physical consumption (ω_{phys}) from stewardship (ω_{steward}), aggregating diverse maintenance tasks into $u_i(t) \in [0, 1]$ omits domain specialization across data, compute, and governance layers [21].
- **Cold-start capital financing.** Capping the governance multiplier on capital at $\gamma_k = 0.08$ prevents investor takeover, but capital-intensive infrastructure commons require complementary non-voting revenue-share bonds to finance initial hardware acquisition.

8 Pre-registerable falsification protocol

Table 5 specifies six pre-registerable empirical tests designed to falsify the core claims of `UACS-Ostrom`.

Table 5: Six pre-registerable empirical tests that could falsify the `UACS-Ostrom` governance framework.

Protocol	Hypothesis	Design and primary metrics	Falsified if
P1. Absentee Sybil-split resistance	Cliff $\underline{u}$ + Ville gate + overhead c_{syb} bounds absentee whale voice share below 12% at $R_{\text{cap}} = 10\times$	Red-team benchmark where an absentee whale ($u_{\text{ext}} \leq 0.09, R_{\text{cap}} = 10\times$) splits across $S \in [1, 50]$ accounts	Extractive whale cohort achieves $\geq 12\%$ aggregate voting weight or captures $> 18\%$ of governance epochs
P2. High-usage Sybil break-even bound	Adding Soulbound identity binding ($S_j = 1$) restores $> 80\%$ regime survival when $u_{\text{ext}} = 0.45$	Compare un-bound vs. Soulbound-bound `UACS-Ostrom` under heavy extractive real usage ($u_{\text{ext}} = 0.45$)	Soulbound identity binding fails to improve regime survival by ≥ 40 percentage points at $u_{\text{ext}} = 0.45$
P3. Core Up-To-One allocation guarantee	Proportional core allocation [14, 15] guarantees active stewards their reservation quota	Benchmark 1,000 congested GPU cluster scheduling epochs (2.5× oversubscription) against token bidding	Any active steward with $u_i(t) \geq 0.50$ receives less than their proportional core compute reservation
P4. Dormant-member drift resilience	Exponential decay ($\lambda_d = 0.22$) prevents inactive majorities from passing liquidation buyouts	Simulate cooperative lifecycle where 60% of founding members become inactive over 20 epochs and face a 2× buyout offer	Dormant founding cohort retains sufficient voting weight after 12 inactive epochs to pass the liquidation ballot
P5. Sabbatical-safe decay freeze	Freezing $\lambda_d \leftarrow 0$ during verified sabbatical leases preserves pre-leave stewardship weight	Audit governance weight trajectories of 200 stewards taking scheduled 4-epoch cooldown leaves	Stewards returning from verified sabbatical lose $> 2\%$ relative voting weight compared to continuous-duty peers
P6. Dependent-stream Ville calibration	Unclipped betting martingale maintains empirical false-rejection $\leq \alpha_{\text{sig}} = 5\%$ under AR(1) telemetry	Evaluate $N = 2{,}000$ genuine contributor streams with temporal autocorrelation $\rho_{\text{auto}} \leq 0.85$	Empirical false-negative rejection of genuine stewards exceeds the 5% Ville bound in any ρ_{auto} stratum

9 Conclusion

Sustainable governance of shared technological commons cannot rely on anonymous token balances or static, non-decaying membership rolls. By anchoring governance weight and proportional core resource entitlements to sequentially attested stewardship $u_i(t)$ gated by an unclipped bounded-observation Ville betting supermartingale with a strict inactivity cliff $\underline{u} = 0.08$, `UACS-Ostrom` sustains 79.96% [79.43, 80.48] commons health, 93.54% [93.54, 93.55] active steward voice share, and 95.00% [92.72, 96.59] regime survival under 5.0× absentee capital asymmetry when governance friction is equalized across all regimes. At the same time, our two-dimensional break-even sweeps across real physical extraction u_{ext} and Sybil identity cost c_{syb} demonstrate that concave usage weighting remains vulnerable to Sybil splitting when extractive actors perform high volumes of authentic physical extraction ($u_{\text{ext}} \geq 0.30$), proving that usage attestation must be paired with high-friction Soulbound identity bounds in high-stakes commons.

Use of AI tools

Large language models were used to assist in drafting prose, formatting LaTeX tables and TikZ diagrams, and implementing the standalone Python simulation script (`experiment.py`). Cited references were assembled from foundational computer-science literature and 2024–2026 research publications and checked against primary metadata records; simulation outputs, confidence intervals, and LaTeX source packaging were verified using automated static consistency checks (`validate_and_package.py`).

References

[1] Marco Tulio Angulo. Who pays for a connected public good? arXiv preprint arXiv:2609.10565, 2026. URL https://arxiv.org/abs/2609.10565.

[2] Artem Baklanov, Pranav Garimidi, Vasilis Gkatzelis, and Daniel Schoepflin. PROPm allocations of indivisible goods to multiple agents. International Joint Conference on Artificial Intelligence (IJCAI), pages 24–30, 2021. URL https://doi.org/10.24963/ijcai.2021/4.

[3] Austin Bennett, Preston Vandervoort, Duc Le, and Mira Belenkiy. Concave is the new linear: The impossibility of anti-plutocratic DAO governance. arXiv preprint arXiv:2605.18990, 2026. URL https://arxiv.org/abs/2605.18990.

[4] Jacob Brodkey, Roberto Tamez, and Aaron Roth. Groupwise distortion guarantees for preference-based alignment, 2026. URL https://arxiv.org/abs/2610.05450.

[5] Vitalik Buterin, Zoë Hitzig, and E. Glen Weyl. A flexible design for funding public goods. Management Science, 65(11):5171–5187, 2019. URL https://doi.org/10.1287/mnsc.2019.3337.

[6] Xiaoyang Cao, Jingqi Li, Zhe Fu, and Alexandre M. Bayen. Who bears the burden? learning responsibility for shared constraints in multi-agent reinforcement learning (LiRA). arXiv preprint arXiv:2610.07491, 2026. URL https://arxiv.org/abs/2610.07491.

[7] Shiyang Chen. Governance decay: How context compaction silently erases safety constraints in long-horizon LLM agents. arXiv preprint arXiv:2606.22528, 2026. URL https://arxiv.org/abs/2606.22528.

[8] Uliana Elina. Maat: Independent deterministic contract-based governance for multi-agent LLM workflows. arXiv preprint arXiv:2609.34017, 2026. URL https://arxiv.org/abs/2609.34017.

[9] Paul Gölz, Nika Haghtalab, and Kunhe Yang. Distortion of AI alignment: Does preference optimization optimize for preferences? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, 2025. URL https://arxiv.org/abs/2505.23749.

[10] Johnnatan Messias and Ayae Ide. Fairness in token delegation: Mitigating voting power concentration in DAOs. arXiv preprint arXiv:2510.05830, 2025. URL https://arxiv.org/abs/2510.05830.

[11] Joel Miller, E. Glen Weyl, and Leon Erichsen. Beyond collusion resistance: Leveraging social information for plural funding and voting (COCM). SSRN Working Paper 4311507, 2022. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4311507.

[12] Natalija Mitic, Soona Sedahmed A. O., Mamadou Selly Ly, and Moustapha Cisse. The constitutional coverage trilemma in AI governance. Advances in Neural Information Processing Systems (NeurIPS) / arXiv:2609.01275, 2026. URL https://arxiv.org/abs/2609.01275.

[13] Elinor Ostrom. Governing the commons: The evolution of institutions for collective action. Cambridge University Press, 1990. URL https://doi.org/10.1017/CBO9780511807763.

[14] Dominik Peters and Piotr Skowron. Proportionality and the limits of welfarism (Method of Equal Shares). ACM Conference on Economics and Computation (EC), pages 793–794, 2020. URL https://arxiv.org/abs/1911.11747.

[15] Piotr Skowron. Core up-to-one for participatory budgeting with additive utilities. arXiv preprint arXiv:2609.15928, 2026. URL https://arxiv.org/abs/2609.15928.

[16] Om Tailor. Audit the whisper: Detecting steganographic collusion in multi-agent LLMs. arXiv preprint arXiv:2510.04303, 2025. URL https://arxiv.org/abs/2510.04303.

[17] Guy Tchuente. Monitoring limits in DAO governance: Capacity breakpoints and endogenous concentration. arXiv preprint arXiv:2603.11222, 2026. URL https://arxiv.org/abs/2603.11222.

[18] Utkarsh Tyagi, Xingang Guo, MohammadHossein Rezaei, Daniel George, Anas Mahmoud, Jackson Lee, et al. Not every rubric teaches equally: Policy-aware rubric rewards for RLVR, 2026. URL https://arxiv.org/abs/2605.20164.

[19] Vladimir Vovk, Alex Gammerman, and Glenn Shafer. Algorithmic learning in a random world. Springer, 2005. URL https://doi.org/10.1007/b106715.

[20] Abraham Wald. Sequential tests of statistical hypotheses. The Annals of Mathematical Statistics, 16(2):117–186, 1945. URL https://doi.org/10.1214/aoms/1177731118.

[21] Yutian Wang and Luyao Zhang. Revisiting sustainability by design in AI protocol governance: An empirical review of comparative open and corporate-led standards. arXiv preprint arXiv:2606.05667, 2026. URL https://arxiv.org/abs/2606.05667.

[22] E. Glen Weyl, Puja Ohlhaver, and Vitalik Buterin. Decentralized society: Finding Web3's soul. SSRN Working Paper 4105763, 2022. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4105763.

[23] Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, and Abraham Flaxman. SybilGuard: Defending against sybil attacks via social networks. ACM SIGCOMM, pages 267–278, 2006. URL https://doi.org/10.1145/1159913.1159945.

[24] Jay Yu, Austin Bennett, Billy Gao, and Rebecca Joseph. Evaluating voting design vulnerabilities for retroactive funding. arXiv preprint arXiv:2505.16068, 2025. URL https://arxiv.org/abs/2505.16068.