# Woven Trust: Consensus and Governance Anchored in Verifiable Reputation

Technical whitepaper, v3 — English edition

## Contents




### Abstract

We present Woven Trust (*Confianza Tejida*), the architecture of a decentralized network in which authority — eligibility for the finality committee, dispute juries, and governance — is derived from a reputation metric computed deterministically over evidence recorded on chain, rather than from computational work or capital at stake. Reputation is computed per domain from two objective sources: bilateral attestations of settled transactions and a graph of vouch bonds, propagated by an algorithm of the EigenTrust family in fixed-point (i128) arithmetic, so that every node arrives at bit-identical values. Powers that require cross-domain integrity use a concave aggregate of the four domains that vetoes profiles lacking minimum coverage. Sybil and collusion resistance is achieved through defense in depth: graph-proximity discounting, dense-community damping, vouch bonds with probation windows, and a weighted-sortition judicial layer. Finality is granted by a committee rotated per epoch, drawn by sortition seeded from a grinding-hardened commit-reveal beacon, and protected by an entrenchment guard that freezes finality, reversibly,

if any single entity's weight crosses a ceiling. We document the dual-asset economic model with issuance tied to measured service, the bootstrap procedure with protocol-forced founder dilution, and an adversarial evaluation backed by a public validation ledger: 17/17 attack scenarios resisted by the reputation engine, two-hop collusion jury capture quantified and closed, and consensus survival across validator dropout and rejoin at epoch boundaries in two independent test environments. We close with the open limitations: proof of unique personhood, transport metadata, and parameters still under calibration.

## 1. Introduction

Open consensus systems anchor their security in a resource the adversary must acquire. Proof of work anchors it in energy [8]; proof of stake, in capital. Both resources are market-purchasable, so network security is, in the limit, a function of the attacker's budget. This work explores a different anchor: **reputation-time** — verifiable behavioral history within the system itself, which accrues slowly, decays with disuse, and cannot be transferred. An adversary cannot rent years of history attested by independent counterparties; it must produce that history under the same rules that watch every other participant.

The classical difficulty with reputation metrics is twofold. First, the Sybil attack [2]: if identity creation is free, any aggregable metric can be inflated with fabricated identities. Second, collusion: coordinated subsets that attest to one another can manufacture signal without any real economic activity. A third problem, less treated in the literature, is institutional: in platform reputation systems [7], the operator custodies the database and can edit it, so the metric inherits the trust — and the corruptibility — of its custodian.

**Contributions.** This paper describes a complete system, implemented and adversarially validated, with the following contributions:

1. A **multi-domain reputation scheme** computed exclusively from objective on-chain evidence (bilateral attestations of settled transactions; vouch bonds), with no subjective scoring, per-domain temporal decay, and anti-specialist concave aggregation (§4).
2. An **anti-collusion battery in depth** — proximity discounting, community damping, probation-windowed bonds, and a capped liability cascade — evaluated against a catalogue of attacks (§5, §6).
3. A **reputational-committee finality mechanism**: weighted sortition over a commit-reveal beacon with a reputable-weight guard (anti-grinding and anti-freeze), domain-separated signed votes, and an entrenchment guard with reversible halt (§7).
4. A **dual-asset economic model** in which issuance pays for algorithmically measured service and no discretionary authority mints or allocates (§9).
5. A **validation methodology** of evidence-over-claim, with a public ledger tying every declared property to a reproducible run and the hash of the binary that produced it (§11).

**Implementation.** The system is written in **Rust** on top of a general-purpose blockchain runtime framework (Substrate family): the mechanisms in this paper (reputation, justice, beacon, tokens) are independent runtime modules, and the computation engines (reputation propagation, committee election, beacon) live in dependency-free core crates that compile to `no_std`/wasm. Every computation path relevant to consensus runs in fixed-point integer arithmetic (i128, fixed scale) to guarantee determinism across architectures. The adversarial analysis that preceded the wiring was performed on standard-library Python prototypes with seeded randomness, later

cross-checked against the Rust cores through exact-parity oracle tests (Appendix A).

## 2. Related Work

**Trust propagation.** EigenTrust [1] computes a global trust value as the fixed point of a Markov chain over the interaction graph, anchored by a pre-trust vector. Our engine belongs to this family (§4.3), with two differences: pre-trust is derived from objective evidence rather than from an a priori trusted set, and the computation runs per domain in deterministic fixed-point arithmetic. The teleportation structure is analogous to PageRank's [6].

**Structural Sybil defenses.** SybilGuard [3] and its successors exploit the fact that fabricated identities connect to the honest graph through few cuts. Our proximity discount and dense-community damping (§5) follow the same structural principle, applied to the weight of evidence rather than to identity admission.

**Committee finality.** The separation between block production and committee-voted finality parallels BFT-style finality gadgets (e.g., GRANDPA [9]). The central difference is the source of voting weight: non-transferable reputation rather than economic stake on deposit.

**Verifiable delay functions.** The proposed hardening of the beacon against last-revealer bias uses VDFs [4].

**Platform reputation.** Centralized review systems [7] aggregate subjective opinions under an editable custodian; the present design inverts all three decisions (facts instead of opinions, multi-domain instead of scalar, universal recomputation instead of custody).

**Institutional precedents.** The viability of exchange orders sustained by reputation without state enforcement is historically documented; see Milgrom, North, and Weingast's analysis of medieval mercantile law [5] and, on self-governance of common-pool resources, Ostrom [10].

## 3. System Model and Threat Model

### 3.1 Actors and Assumptions

Participants are identified by cryptographic pseudonyms: sr25519 keypairs; the address derived from the public key is the member's complete identity within the system. There is no link to any legal identity, by design (the network layer is discussed in §12). Nodes replicate the chain and deterministically recompute the entire reputation state; no node holds privileged state.

We assume the usual partially synchronous network model, standard cryptography (sr25519 signatures, collision-resistant hashes), and an adaptive adversary who can: (a) create unlimited identities at negligible marginal cost, (b) coordinate subsets of owned or bribed identities, (c) withhold or reorder its own protocol messages, and (d) participate honestly for extended periods before deviating.

### 3.2 Security Objectives

- **S1 (anti-Sybil):** identities without independent evidence accrue no consensus weight, regardless of their number.
- **S2 (anti-collusion):** the weight obtainable by a coordinated subset grows with its evidence external to the subset, not with its internal activity.

- **S3 (non-transferability):** weight cannot be bought, sold, or inherited; it decays without activity.
- **S4 (determinism):** every node computes bit-identical weights from the same chain prefix.
- **S5 (finality safety):** no block is finalized without a supermajority of the reputational weight of the sitting committee; under excessive weight concentration, the system prefers to halt finality (reversibly) over finalizing under capture.
- **S6 (non-discretionarity):** no account holds discretionary interpretation, editing, or allocation power; wherever human judgment is required, it is resolved by sortition (§8).

### 3.3 Considered Attack Catalogue

Sybil rings; collusion clusters (mutual admiration); whitewashing (exit and re-entry); slashing targeted at one of the victim's domains; vouch-kamikaze and vouch-bomb attacks (§6); beacon grinding and reveal withholding (§7.3); finality freeze by a zero-weight participant (§7.3); two-hop collusion jury capture (§8); committee capture during the bootstrap window (§10).

## 4. The Reputation System

### 4.1 Admissible Evidence

Only two types of events generate reputation signal, both objective and verifiable on chain:

1. **Bilateral attestation.** One counterparty records the fact and the beneficiary claims it — two signatures, two wills. There are no graded scores or third-party reviews: only the testimony of a counterparty staking its own name counts, and four deterministic brakes put a price on signal fabrication: a per-epoch attestation budget, a reputation floor required of the attester, a magnitude cap (nobody attests to more weight than they themselves hold), and per-pair geometric wear that saturates repeated signal. The hard link between attestation and an on-chain settled transaction — so that only a settled deal can be attested — is a declared hardening, pending the settlement layer (§12).
2. **Vouch bond** (§4.2).

Which event types are admissible is fixed by public, recomputable rules, with no discretionary admission authority (objective S6). The bootstrap key that seeds the initial event types expires automatically at a fixed block height (§10).

### 4.2 The Vouch as a Bond

A member may vouch for another's entry. The vouch: (i) locks a fraction of the sponsor's reputation, in the same domain, as a pledge; (ii) the pledge is released gradually as the sponsored member accrues **independent** reputation, defined as evidence from counterparties whose graph proximity to the sponsor falls below a threshold; (iii) the vouch transfers no weight: a member whose only edge is its sponsor has zero structural weight; (iv) fraud by the sponsored member propagates loss up the vouch chain, capped per incident (§6). The vouch structure forms a directed graph rooted at genesis, public and analyzable: Sybil rings present detectable topological signatures.

### 4.3 Propagation Engine

Let $G_d$ be the evidence graph of domain $d$ (edges weighted by attestations and vouches, with the discounts of §5). The weight $s_d$ of each identity in domain $d$ is computed as the fixed point of a propagation iteration of the EigenTrust family [1]:

- **Pre-trust:** the anchor vector is not an a priori trusted set but the distribution of objective evidence (settled transactions, verified work) held by each identity.
- **Teleportation:** with fixed probability, the iteration jumps to the pre-trust distribution, which bounds the influence obtainable by dense subgraphs lacking external evidence (analogous to PageRank's damping factor [6]).
- **Deterministic fixed point:** all arithmetic is integer (i128, fixed scale). No floating point appears on any consensus path; the result is bit-identical across every architecture (objective S4).

### 4.4 Four Domains and Concave Aggregation

Reputation is computed in four independent domains: exchange, work, judgment, and governance (the four suits of the deck, in the project's nomenclature). Each domain admits only evidence of its own kind. For powers that require cross-domain integrity (finality committee, juries, vouching authority), the four values are aggregated as:

$$\mathrm{agg}(s) = (1-\lambda)\cdot \min_d(s_d) + \lambda \cdot \mathrm{med}_d(s_d)$$

with $\lambda \in [0,1]$ a blending dial (provisionally $\lambda = 0.5$; an open parameter in final calibration, §12). The concave form preserves the anti-specialist property — covering fewer than half of the domains yields a zero median and a near-zero aggregate for every $\lambda$, so $\lambda$ **cannot** rescue an under-covered profile — while it softens the pure minimum, which adversarial analysis showed to be weaponizable: with agg = min, a strike targeted at a single domain zeroes out the victim ($-100\,\%$); with $\lambda = 0.5$ the wound is bounded to $\approx -50\,\%$ in the ideal symmetric case ($-47\,\%$ measured on real profiles on hardware; ledger, row 16). The value $\lambda = 0.5$ is provisional and pending re-validation against the full attack battery (§12).

### 4.5 Decay

Weight decays multiplicatively per epoch in the absence of new evidence, per domain, with a half-life of $\approx 24$ months (per-epoch retention $\rho = 0.999052$). Decay implements temporal non-transferability (S3): there are no aged accounts with resale value, and the advantage of early participants dissolves without intervention (§10).

## 5. Anti-Collusion Defenses

No single defense satisfies S2 in isolation; the system stacks four, all deterministic:

1. **Proximity discounting.** The weight of every attestation is discounted as a function of the graph distance between attester and attested. Signal fabricated within a cluster is structurally cheap.

2. **Community damping.** Dense communities whose internal edge mass exceeds their external evidence by a configured proportion are damped as a group (detection and enforcement in fixed-point arithmetic, deterministic).
3. **Vouch bonds and quotas** (§4.2) bound the fan-out of fabricated identities financeable by an attacker with a given reputation budget.
4. **Judicial layer.** Collusion proven before a jury (§8) is sanctioned across all four domains simultaneously, the sole case of cross-domain sanction (§6).

## 6. Sanctions and the Liability Cascade

Trust failure propagates liability with three properties, each validated against the specific attack it targets:

- **Capped proportionality.** The sponsor's loss from a sponsored member's fraud is bounded per incident. This closes the **vouch-kamikaze** attack: using a massive fraud from a disposable identity to annihilate an honest sponsor.
- **Probation window.** Fraud committed immediately after receiving a vouch does not propagate a cascade. This closes the **vouch-bomb** attack: soliciting a vouch and detonating it within the same window.
- **Domain locality.** The sanction lands in the domain of the offense; the sole cross-domain exception is structural betrayal (proven Sybil operation, proven collusion, attack on the network).

Sanctions are graduated (public mark, suspension, expulsion), prioritize restitution, and are reversible through the same process that imposes them. The full exclusion-without-confiscation mechanism — how expulsion carries real consequences while coin balances stay structurally untouchable — is developed in Appendix B.

## 7. Consensus

### 7.1 Architecture

Block production is decoupled from finality. Production uses a conventional round-based mechanism; **finality** is granted by a committee drawn by sortition weighted by $\mathrm{agg}(s)$ (§4.4), rotated every epoch. A block is final once it accumulates signed votes representing a supermajority of the sitting committee's reputational weight.

### 7.2 Votes

Votes are domain-separated sr25519 signatures over (height, hash). At the gossip layer, each vote's signature is verified before it is relayed, so a peer flooding the network with forged votes is not amplified through honest nodes. At the tally, each vote passes the committee-membership check and signature verification, and is deduplicated by (height, voter). Epoch rotation prunes and refreshes committee state at the transition; a boundary race condition (outgoing committee members voting across the edge), discovered in multi-node dropout-and-rejoin testing, was reproduced and closed (validation ledger, rows 4 and 33).

### 7.3 The Randomness Beacon

Committee and jury sortition is seeded by a **commit-reveal** beacon: participants commit a value, reveal it in the following window, and the combined entropy of the reveals seeds the selection. Two hardenings:

1. **Last-revealer bias.** We propose layering a verifiable delay function [4] over the beacon's output, removing the advantage of deciding whether or not to reveal after having seen the other reveals.
2. **Reputable-weight guard.** The original design froze selection whenever the set of reveals was weak, which allowed a single reputation-less participant — committing and never revealing — to halt the entire network's finality indefinitely. The corrected version defers to the beacon only when the committed set accumulates sufficient reputable weight; otherwise selection falls back to a deterministic backup mechanism. The attack was reproduced in test before the fix and fails after it.

### 7.4 The Entrenchment Guard

An on-chain guard monitors the concentration of finality power. If any single entity's weight exceeds a configured ceiling, finality is **reversibly frozen** and is designed to recover on its own once concentration re-disperses (the current status of the automatic-recovery path is stated in §12). The justification is security-theoretic: halting finality is a recoverable liveness failure; finalizing history under a captured committee is an irreversible safety failure. The guard prefers the former.

## 8. Dispute Resolution

Disputes between members are resolved by sortitioned juries, with the sortition weighted by the judgment domain and three structural constraints:

1. **Interested-party exclusion:** the parties and any identity with measurable graph proximity to the dispute are excluded from the pool.
2. **Pairwise independence:** the jury must satisfy a minimum graph distance between every pair of jurors — a dispute cannot be adjudicated by a cluster.
3. **Supermajority closure:** the verdict is decided by a supermajority of the sortitioned jury; the two preceding protections ensure that this majority is not that of an interested clan. (An additional layer scoring each juror for coherence with the evidence — rewarding or penalizing the vote according to its alignment with the case file, along the lines of a Schelling point [11] — is future work, §12; it is not implemented today.)

Two-hop collusion jury capture was quantified in simulation ($8$–$10 \times 10^3$ trials per configuration) and closed with depth-2 weight penalties, verified in unit tests and on devnet. The verdict acts solely on the pseudonym and the proven fact; the system holds no information with which to de-pseudonymize, and no procedure generates such information (§12). The jury *certifies* the act rather than decreeing a punishment; the status-only consequences that follow, and their doctrinal and empirical grounding, are set out in Appendix B.

## 9. Economic Model

**Money is not power.** Before any mechanism, one rule governs the economy and ties it to the constitution (§6, "the origin of power"): coin buys no authority. Neither HLQ nor SOV can purchase reputation, consensus weight, a jury seat, or any say in common decisions; there is no staking-for-power anywhere in the protocol. Wealth and influence are *orthogonal axes* — one may hold much of either and none of the other. This is what makes the model anarcho-capitalist rather than plutocratic, and it is the property every design choice below is arranged to protect: the reward layer must never become a back door from balance to authority.

**Dual asset.** Two coins, each with a distinct role and a hard cap that can never be exceeded — no artificial inflation, no central bank, no discretionary mint. **HLQ** (hard cap $540 \times 10^6$, finely divisible) is the medium of exchange — the coin of daily dealing, fees and payments. **SOV** (hard cap $54 \times 10^6$, indivisible) is the reserve — scarce, slow, held rather than spent. Transaction fees are paid in HLQ and validated against the same ledger that collects them — a mismatch between the fee-validation ledger and the collection ledger, found under adversarial stress testing, would have prevented every new account from paying its first fee; it is fixed and covered by regression.

**Issuance by measured service — the mechanism.** There is no calendar-based issuance. At each era boundary the runtime reads, from a non-forgeable participation feed, the verified service each node rendered that era — blocks authored in its elected slots plus finality votes cast as a committee member, both re-checked against sr25519 signatures and committee membership before they count. This yields a weight vector over the nodes that served. Before any coin is split, the vector passes an **anti-Sybil per-node cap**, $w_i \leftarrow \min(w_i,\ \lceil \mathrm{median}(w) \cdot k \rceil)$ with $k \approx 2$: the cap flattens the high tail so that no single operator — or a farm of masks feigning one — can draw a disproportionate slice by out-weighing the median. The era's scheduled pool is then divided **in proportion to the capped weights** (`split_service_reward`), and credited. The order of these steps matters: recipients are verified *before* minting, so an era with no verified service mints nothing and consumes nothing against the cap — closing the phantom-mint defect (an earlier version reserved the cap first and vaporized unclaimed issuance into permanent phantom supply).

**Why service and not stake.** Tying issuance to measured, signature-verified service — rather than to a calendar or to holdings — is what keeps the coin's entry aligned with the network's security: the only way to receive newly minted coin is to have done consensus work the chain could verify and that the anti-Sybil cap could not inflate. Purchasable weight (the flash-loan and paid-delegation attacks that recur across on-chain governance) has no purchase here, because minting reads *service*, not balance.

**SOV by endurance.** The two assets are earned differently on purpose. HLQ, the medium of exchange, is paid for spot service each era. SOV, the reserve, is paid for **constancy**: a node's SOV weight in an era is its service multiplied by its consecutive-era streak (capped at a ceiling $k$; the streak resets to zero the first era it fails to serve). Two nodes doing identical work in one era earn the same HLQ, but the one that has served without interruption earns far more SOV. This makes the reserve accrue to endurance — the one resource a Sybil cannot fabricate, since it is time actually spent, not capital or reputation that can be multiplied across masks — and directly defeats the in-and-out cartel that harvests spot rewards without ever sustaining the network.

**Disinflation — the curve.** Issuance follows a geometric curve: each epoch the pool retains

a fixed fraction $\rho$ of the previous epoch's, so the scheduled amount decays with a half-life of $\approx 24$ months ($\rho = 0.999052$ per epoch), always bounded by the hard cap it can never exceed. The same $\rho$ governs reputational decay (§4.5): coin issuance and reputation weight implement one property in two layers — early advantage dilutes on its own, by arithmetic, without anyone deciding it should.

**Fees and the burn — a two-way supply.** Service-earned issuance is one-directional: it can only add coin. To give supply a way to *contract* with demand — the discipline that redemption provides in free-banking systems, which an unbacked coin otherwise lacks — a fixed fraction of every HLQ transaction fee is **burned** (destroyed, not recycled to a treasury), the remainder paying the processing node. Coin thus enters only up the capped curve and leaves only through the fire, so the float responds in both directions to how much the network is actually used.

**Separation of rewards.** The principle is twofold: being paid gives no voice (payment confers no governance weight), and the seat is not paid for being a seat (issuance pays measured service, not rank). The service the chain measures and pays today is committee service (block authorship and finality votes, with an anti-hoarding cap); paying pure infrastructure service (hosting, custody, availability) through self-verifying proofs is designed and not yet wired (§12). Founding identities start with no balance at all — genesis allocates coin to no one — and the design commitment is that they receive no issuance for their bootstrap service either; the mechanism that enforces it (in-protocol exclusion or verifiable operational separation) is under final decision and will be documented with its validation.

**No premine, no sale, no discretionary mint.** There is no ICO, no venture allocation, and no premine: no coin is created for insiders before the network runs, and none is handed out per person or per account. Coin enters *only* by paying for measured service, on the decreasing published schedule, and circulates afterwards through voluntary exchange — never by decree, never by a mint anyone controls. Founding identities are seeded with zero balance and receive no issuance for their bootstrap service (§10). Collective instruments (welcome funds, mutual aid, exchanges) belong to the free-association layer: whoever builds them owns them, outside the protocol, which neither forbids nor operates them.

*Open parameters:* the exact shape of the issuance curve; $\lambda$; the proof-of-service schedule; the internal split of rewards. These remain under calibration; the values the network runs today are provisional, adjustable through the upgrade process, and each will be published with its evidence as it is fixed (§12).

## 10. Bootstrap Procedure

Bootstrapping a reputation economy is its most delicate security problem: every initial asymmetry is a seed for capture. Genesis rules:

1. **Minimal, public, and expiring seeding.** Genesis seeds a founding cohort labeled as scaffolding, with decay active from block 0 and zero issuance. Verifiable ceremony rule: every seeded identity corresponds to an operating node at genesis.
2. **Protocol-enforced founder dilution.** The founding nodes' advantage dissolves through reputation decay (§4.5): without attested activity, their weight falls below that of the average citizen in little more than one half-life, with no human intervention. The bootstrap key that admits the initial evidence types expires automatically at a fixed block height —

by protocol, not by goodwill — a condition implemented and tested on hardware (ledger, row 14). The founding upgrade authority (a multisig, valid only during construction) additionally carries an expiration in the upgrade pallet itself, with a convergence margin still under review (§12). An explicit sunset of the founding *validator role* tied to an on-chain dispersion metric is future work (§12), not a mechanism wired today: effective dispersion is produced, for now, by decay.

3. **Concentration ceiling** (§7.4), active from block 0, when the network is smallest and most vulnerable.
4. **External entropy.** The genesis protocol anchors the seed to a public Bitcoin block fixed at the ceremony, removing the founders' own ability to grind their genesis. The launch ceremony was executed on 18 July 2026: the seed is anchored to Bitcoin block 958536, and the network runs from genesis hash `0xa1ab5e0b...da1fed6`.
5. Bootstrap-window capture and expiration were adversarially swept in simulation before any value existed (ledger, row 10), and capture-free bootstrap was additionally validated on hardware (row 13).

## 11. Evaluation

**Methodology.** The project operates under an evidence-over-claim protocol: no property is declared without a run that exhibits it, and audits precede the existence of value. Multi-node behavior is validated on devnets in two independently operated environments; findings are cross-checked against independent adversarial review panels under an anti-hallucination protocol (every claim is verified against a run, never accepted from prose). A **public validation ledger** (`VALIDATION-LEDGER.md` in the project's public repository) ties every claim to the run that supports it, the hash of the binary that produced it, and the raw outputs (33 rows as of this version's close). **Appendix A** enumerates the full adversarial battery and traces every result in the table below to its ledger row, its reproduction command, and the environment it ran in.

**Selected results.**

| Property | Method | Result |
|---|---|---|
| Reputation engine resistance (Sybil rings, collusion clusters, targeted slashing, whitewashing) | adversarial battery | 17/17 scenarios resisted |
| Consensus safety and liveness under validator dropout and rejoin across epoch boundaries | multi-node devnet, two environments | survival verified; boundary race closed |
| Two-hop collusion jury capture | simulation, 8–10 $\times 10^3$ trials/configuration | quantified; closed with depth-2 penalty |
| Finality freeze via beacon (reveal withholding) | test reproduction | reproduced; closed with reputable-weight guard |
| Vouch-kamikaze and vouch-bomb | adversarial battery | closed (per-incident caps; probation window) |

| Property | Method | Result |
|---|---|---|
| Determinism of reputation computation | cross-recomputation | bit-identical across nodes and architectures |
| Fee validation/collection mismatch (unpayable new account) | adversarial stress test | found; fixed; regression active |

**Normative layer.** The system's constitutional document was subjected to the same adversarial process (12 flaws found and resolved before sealing), and every mechanism in this paper is traced to the normative article it serves.

## 12. Limitations

A study earns trust by naming what it has *not* settled. The following are the honest residuals of the system as evaluated in this paper — stated as current bounds, not as a roadmap:

- **Proof of unique personhood.** The unique-personhood gate has no on-chain implementation yet. The reputation wall (§4) makes it non-critical to consensus security; it remains relevant as a bound on cold-start capture (§10) and for the base-citizenship unit. An objective, substitutable mechanism, under evaluation.
- **Settlement and attestation.** There is no on-chain transaction-settlement layer yet; attestation rests today on the four deterministic brakes of §4.1. Tying every attestation to a settled transaction is the planned hardening.
- **Infrastructure service.** The measured service issuance pays today is committee service (authorship and finality); paying pure hosting through self-verifying proofs is designed and not wired.
- **Transport metadata.** Content is end-to-end blind, but network-layer correlation (timing, sizes, addresses) is the real privacy frontier. Under construction: an onion-routed transport layer. Until then, the honest claim is content privacy, not traffic privacy.
- **Key separation.** Authentication and spending currently share a single key with cryptographic domain separation (v1); separation into subkeys is designed and pending.
- **Parameters under calibration.** $\lambda$ (a sweep is underway with the full adversarial battery), the shape of the issuance curve, and the proof-of-service schedule: the launch values are provisional, and each will be fixed with published evidence through the upgrade process, never by decree.
- **Beacon VDF.** Proposed, not implemented; the beacon currently operates with the reputable-weight guard and deterministic fallback.
- **Finality vote diffusion.** An internal audit found that the diffusion (gossip) layer relayed finality votes without verifying their sr25519 signature before propagating them, which allowed an external peer to amplify forged votes across the network (validity was still checked at the tally, so finality safety was unaffected; it only opened an amplification vector). Fixed: the signature is now verified before relaying (§7.2); pending hardware validation on the relaunch binary.
- **Committee entrenchment guard.** The guard's threshold uses an absolute floor of reputable accounts, not a fraction of the reputable set; this leaves a narrow, self-healing

window during bootstrap, when the set is small. An acknowledged, low-severity residual.

- **Automatic recovery from a concentration halt.** The reversible halt of §7.4 has two halves: the *freeze* under excessive concentration, and the *automatic thaw* once concentration re-disperses. The freeze — the safety-critical half — holds: the chain refuses to finalize under capture and prefers to stall, verified on hardware. The thaw is where the residual sits. Automatic recovery passed adversarial iron on an earlier binary; a later hardening of the committee re-anchor, made deterministic to close an unrelated split vector, regressed the node-side recovery path on the launch binary. It was found in adversarial testing and is under fix. The distinction is the load-bearing one: a frozen chain is safe but not live, so this is a **liveness** residual, not a safety one — the guard never finalizes captured history, it only fails, for now, to resume on its own without operator action.
- **Dependencies.** The underlying chain framework carries known advisories in transitive dependencies, tracked; the analysis of which ones reach the live node's surface is ongoing.

## 13. Conclusion

Every open consensus system answers one question: what must an attacker acquire to seize it? Proof of work answers *energy*; proof of stake answers *capital*. Both answers are prices, and a price can be paid — so the security of these networks is, at the limit, the size of an adversary's budget. This work set out to give a different answer, one that cannot be bought outright: **time spent behaving well, under public rules, inside the system itself**.

Woven Trust makes that answer concrete. Authority — eligibility for the finality committee, for dispute juries, for governance — is derived from a reputation metric computed deterministically from objective, on-chain evidence: bilateral attestations of settled dealings and a graph of vouch bonds, never subjective scoring, never a custodian who can edit the ledger. That metric is non-transferable and it decays, so an advantage neither sells nor keeps. Around it the paper composes a defense in depth against the two classical failures of reputation systems — the Sybil attack and collusion: multi-domain concave aggregation that vetoes narrow profiles, proximity-discounted and community-damped propagation, vouch bonds with capped liability cascades and probation windows, sortition seeded from a grinding-hardened commit-reveal beacon, and an entrenchment guard that freezes finality *reversibly* the instant any single entity's weight crosses a ceiling. The net effect is an inversion of the attacker's problem: to capture finality one must first spend real time earning verified reputation within the very order one intends to subvert — and reputation earned that way is, by construction, the behavior the order wants.

None of this is asserted from prose. Every property in the paper is tied, through a public validation ledger (Appendix A), to a run that exhibits it, the hash of the binary that produced it, and its raw output — reproducible by anyone: seventeen of seventeen adversarial scenarios resisted by the reputation engine, two-hop jury capture quantified in simulation and closed, finality surviving validator dropout and rejoin across epoch boundaries in two independently operated environments, and bit-identical reputation across nodes and architectures. The constitutional text that governs the system was put through the same adversarial process before it was sealed. We also state, in the same register, what is *not* yet done (§12) — because a study that hides its residuals earns less trust, not more.

The claim we believe the evidence supports is narrow and testable: **a network in which authority is not for sale is buildable with today's tools, and its security is not a**

**slogan but a measured quantity.** The larger reason to build it is older than any of these mechanisms — that people can cooperate, keep their word, and settle their disputes by a name earned through deeds rather than by force or by wealth. Woven Trust is one attempt to give that ancient possibility a machine that cannot be quietly captured. What it becomes is now a matter for the people who use it.

## References


[1] S. D. Kamvar, M. T. Schlosser, H. Garcia-Molina. *The EigenTrust Algorithm for Reputation Management in P2P Networks.* Proc. WWW 2003.

[2] J. R. Douceur. *The Sybil Attack.* Proc. IPTPS 2002.

[3] H. Yu, M. Kaminsky, P. B. Gibbons, A. Flaxman. *SybilGuard: Defending Against Sybil Attacks via Social Networks.* Proc. SIGCOMM 2006.

[4] D. Boneh, J. Bonneau, B. Bünz, B. Fisch. *Verifiable Delay Functions.* Proc. CRYPTO 2018.

[5] P. R. Milgrom, D. C. North, B. R. Weingast. *The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs.* Economics & Politics 2(1), 1990.

[6] L. Page, S. Brin, R. Motwani, T. Winograd. *The PageRank Citation Ranking: Bringing Order to the Web.* Stanford InfoLab, 1999.

[7] P. Resnick, K. Kuwabara, R. Zeckhauser, E. Friedman. *Reputation Systems.* Communications of the ACM 43(12), 2000.

[8] S. Nakamoto. *Bitcoin: A Peer-to-Peer Electronic Cash System.* 2008.

[9] A. Stewart, E. Kokoris-Kogias. *GRANDPA: A Byzantine Finality Gadget.* arXiv:2007.01560, 2020.

[10] E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, 1990.

[11] T. C. Schelling. *The Strategy of Conflict.* Harvard University Press, 1960.


---

## Appendix A: Evidence Map

This appendix replaces the number with the list: what the tests are, where they live, and how to re-run them. The environment convention follows the validation ledger: **local** = the build station; **hardware** = real nodes on separate infrastructure (different machine and network), the stronger evidence.

### A.1 The reputation engine's adversarial battery (the 17/17 in the Abstract)

Suite `tests/test_engine.py` of the reputation prototype (`reputation-engine/` in the public repository, standard-library Python, seeded PRNG — reproducible by anyone). The 17 tests, by exact name and the property each one pins down:

1. `test_sybil_has_no_power` — 200 fabricated identities (40% of the network) capture 0.00% of the committee (S1).

2. `test_independence_penalises_inbreeding` — the proximity discount penalizes inbred signal (§5, defense 1).
3. `test_damping_reduces_laundering` — damping cuts reputation laundering toward a dense ring 8.7× (§5, defense 2).
4. `test_community_reduces_scattered_ring` — community detection cuts the scattered ring that evades the local defense.
5. `test_adaptive_community_does_not_worsen` — fragmenting the ring to evade the label does not increase what is laundered under defense.
6. `test_adaptive_has_no_consensus_power` — at any fragmentation, the puppets' consensus power is $\approx 0$ (§4.4).
7. `test_asymmetric_evades_damping_but_no_power` — the directed funnel (many vouch for one) evades local damping but buys no structural power.
8. `test_in_concentration_closes_funnel` — the in-degree-concentration signal cuts the funnel ~70–80% at every diversification.
9. `test_in_concentration_closes_cross_dim_funnel` — the cross-domain funnel falls too: the evidence deficit is per domain (§4.4).
10. `test_whitewashing_loses_everything` — abandoning the pseudonym and re-entering clean costs all earned reputation (S3).
11. `test_cascade_slashing` — fraud propagates loss up the vouch chain with per-hop attenuation; the non-voucher is untouched (§6).
12. `test_sublinear_quota` — the live-vouch quota grows sublinearly with reputation (§5, defense 3).
13. `test_graduation_frees_mentor_quota` — a vouchee who accumulates independent evidence graduates and frees the voucher's quota (§4.2).
14. `test_temporal_decays_inactive` — an inactive account's reputation decays epoch by epoch; an active one's holds (§4.5).
15. `test_edge_aging_freshness_premium` — vouches themselves age: renewing trust earns a premium over letting it sit.
16. `test_conservative_aggregate` — the anti-specialist aggregate nullifies a profile without domain coverage (§4.4).
17. `test_mass_conservation` — internal engine invariant: propagation conserves reputation mass.

Reproduction: `cd reputation-engine && python3 tests/test_engine.py` (suite) and `python3 run_all.py` (regenerates the `RESULTS.md` report with the cited figures). Last verified re-run: 2026-07-14, 17/17.

### A.2 Traceability of the remaining results (§11)

The *ledger row* numbers below are the entries in the public validation ledger (`VALIDATION-LEDGER.md`), which records more runs than are summarized here; each selected result is mapped to its own row, so the numbering is the ledger's and is not consecutive.

| Result (§11) | Ledger row | Tests / method | Environment |
|---|---|---|---|
| Reputation engine battery | 1 | 17/17 (A.1) | local |

| Result (§11) | Ledger row | Tests / method | Environment |
|---|---|---|---|
| Consensus simulator (safety/liveness under attack, partition, loss) | 2 | 11/11 testrig + 13/13 property | local |
| Rust ↔ prototype parity (deterministic fixed point, `no_std`) | 3 | 35/35 oracle tests | local |
| Multi-node finality gadget (signed votes, proof on import, rotation) | 4 | 6/6 + 6/6 + 8/8; 3/6/7-node devnets | local + hardware |
| Decay / anti-entrenchment | 5 | sim sweep + real account at exactly 0.90/epoch | sim + hardware |
| Justice: sortition jury with interest exclusion | 6 | 8/8 + 14/14 unit; 9/9 on chain | local + hardware |
| Recount with no privileged trigger (automatic epoch) | 7 | 15/15; extrinsic absent from metadata | local + hardware |
| Production network parameters (cadences, $\alpha$, $\tau$) | 8 | simulator re-run + verified metadata | local + hardware |
| Cold-start capture sweep | 10 | 3 adversarial tranches | sim |
| Two-hop jury capture + depth-2 penalty | 11, 11b | 8–10 $\times 10^3$ trials; 44 unit + 8/8 on chain | sim + hardware |
| Commit-reveal anti-grinding beacon (jury + committee) | 12 | 70 unit; 7/7 on chain + verified switch-over | local + hardware |
| Bootstrap with declared seed (no halt, guards sane) | 13 | sustained finalization; founders at exactly 20% | hardware |
| Concave aggregate (targeted blow −47 % vs −100 % under MIN) | 16 | 20/20 unit + on-chain verification | local + hardware |

| Result (§11) | Ledger row | Tests / method | Environment |
|---|---|---|---|
| Cascade caps and probation window (anti-kamikaze, anti-vouch-bomb) | 17, 18 | 21/21, 22/22 unit + on-chain verification | local + hardware |
| Fee validation/charge mismatch (new account unable to pay) | 32 | 21/21 unit + real spend from a fresh account | local + hardware |
| Epoch-boundary race (stale committee voting across the border) | 33 | captured reproduction + KILL/REJOIN after the fix | hardware |

The raw evidence bundles (run outputs, validation scripts) are inventoried in the validation ledger itself with their sha256 hashes, so any published copy is verifiable against what the ledger commits to.

### A.3 Reproducible iron runs

The following runs exercise the *launch binary* (sha256 `fa79dda9...6666a6`, the build now running in production) on real nodes, and each can be reproduced byte-for-byte on any machine: the node keys used are the well-known deterministic development keys (`0x00...01`...`0x00...06`), so the resulting PeerIDs and genesis are identical everywhere, and no production node address or key appears in the procedure. Each entry states what it proves, how to run it, and what was observed.

**E1 — The running binary is the audited binary.** *Proves:* no substitution between audit and run. *Run:* `sha256sum harlequin-node` and compare to the published `fa79dda9...6666a6`, then `harlequin-node --version`. *Observed:* the hash matches; version `0.1.0`, role `AUTHORITY`.

**E2 — The pallet surface matches the ratified constitution.** *Proves:* every constitutional decision is wired into the runtime, and, above all, no call path lets justice touch a balance (Article II). *Run:* a single node with `--consensus manual-seal-1000` (deterministic block production without a committee); read the runtime metadata with a SCALE client. *Observed:* the upgrade key exposes exactly the incorruptible design (auto-expiry `LifeEnd`, network-cosigned `PendingRenewal/apply_renewal` with geometric decay removed, irreversible `Burn`, costly veto, disaster mode, the five seats, and a keyhole-only `Custody` seat with zero calls); justice exposes only `open_case`, `open_incapacity_case`, `cast_vote`, `close_case`, and not one of them takes or moves a balance — a guilty outcome writes a status mark (`GuiltyMarks`), never coin; participation exposes the endurance accumulators (`ConsecutiveEras`).

**E3 — Finality is live and refuses to finalise below quorum.** *Proves:* the gadget finalises under a healthy committee and *halts rather than forks* when votes fall below quorum — safety over liveness. *Run:* six nodes, `--consensus woven-trust-<ms> --validator --vote-as //Alice ... //Ferdie`, deterministic node keys, node 1 as bootnode. *Observed:* with five live voters

(committee 6, quorum $\alpha = 5$) the chain reported *finality below quorum — waiting, not finalising* and did **not** finalise; bringing up the sixth node produced *finalised #15 (committee 6, α 5/6)*, after which finality advanced and sustained. Running the same scenario at `woven-trust-2000` and at the ceremony cadence `woven-trust-12000` produced the **identical finalised hash** — the gadget's behaviour is deterministic and independent of block cadence.

**E4 — Justice guards reject unsafe cases at the door.** *Proves:* the justice pallet refuses to open a case on foreign evidence or without a valid jury — it fails loudly at opening, never zeroing anything silently at verdict. *Run:* on a `--dev` chain with seeded evidence, submit `open_case` as a signed extrinsic. *Observed:* evidence not belonging to the defendant is rejected at open with `ForeignEvidence`; valid evidence with genesis reputation zero is rejected with `EmptyJury` — the correct cold-start behaviour, since founders hold zero reputation at genesis (it is earned by service) and no jury is eligible until reputation accrues.

**E5 — "Justice touches status, never balance," three independent ways.** *Proves:* Article II is a structural limit, not a policy a jury could bend. *The three lines:* (i) *code* — the guilty hook routes only to the reputation pallet, and a scan of the justice pallet finds zero references to token, coin, or balance; (ii) *surface* (E2) — no justice call moves a balance; (iii) *live* (E4) — the guards make an unsafe verdict path unreachable. The guilty-verdict-to-`GuiltyMarks` transition is additionally covered by a deterministic unit test in which the loss lands in the reputation dimension, never the wallet.

*Honest scope.* The full guilty-verdict-through-to-readmission cycle on a live devnet first needs reputation to accrue over several epochs (the E4 cold-start); it is covered deterministically at the unit-test layer today, and the live multi-epoch run is a declared, reproducible next step — not a gap in the mechanism.

---

*The source code, the validation ledger, and the evidence bundles are public and independently verifiable.*

---

## Appendix B: Exclusion without confiscation: a status-only sanction mechanism

*This appendix is self-contained: its references [1]–[34] are numbered independently of the paper's main bibliography.*

### B.1 Problem statement

A society built on voluntary, pseudonymous, non-coercive association faces an apparent contradiction. If membership is genuinely voluntary and property is genuinely inviolable, how can the society respond to an aggressor within it — a fraud, a thief, a seller of contraband that harms non-consenting third parties — without either (a) tolerating the aggression indefinitely in the name of non-coercion, or (b) reaching for confiscation, which requires exactly the coercive, discretionary power over other people's balances that the society exists to foreclose? A mechanism that resolves this must clear two independent bars. First, it must actually work: expulsion has to carry real consequences, not merely a symbolic gesture that leaves an aggressor free to continue operating under the same identity. Second, the mechanism itself — whoever holds

the power to expel — must resist capture; a tool built to punish agreed-upon aggression is, by construction, also a tool that a faction can point at an unpopular minority. Harlequin's Article II commitment to unconditionally inviolable coin balances makes the first horn of the dilemma structural rather than negotiable, which forces the design question this section addresses: can status and property be separated as sanction levers, and can the lever that remains — status — be kept from collapsing back into a capturable, quasi-property mechanism of its own?

### B.2 Doctrinal and historical grounding

The claim that expulsion and confiscation are separable levers is not a single argument but a convergence independently reached by at least nine traditions that do not cite one another: Icelandic medieval customary law, Athenian democracy, the private commercial law of the medieval fairs (lex mercatoria), Quaker church discipline, Catholic canon law, Rothbardian/Hoppean anarcho-capitalist theory, Stringham and Benson's private-governance economics, analytic jurisprudence (the Hohfeldian distinction between a power of status and a claim-right over property), and Nozick's framework of meta-utopia [1–11]. Athenian ostracism banished a citizen for up to ten years by a quorum vote, yet left estates, rents, and family in place [1]. Icelandic *fjörbaugsgarðr*, the "lesser outlawry," imposed a three-year exile and withdrawal of legal protection without touching property — reserving actual confiscation for *skóggangr*, full outlawry, the harshest and rarest penalty, precisely the contrary model [2]. Quaker disownment and Catholic excommunication both withdraw only what the institution itself conferred — voice in a meeting, sacramental office — never a member's independently held secular property [4, 5]. Amish *Meidung* applies only to those who freely accepted adult baptism, an explicit ex-ante covenant [3]. Contemporary crypto-governance design reproduces the same pattern by convergent, uncoordinated engineering: MolochDAO's guildKick mechanism does not burn a jailed member's shares — it converts them into non-voting LOOT tokens that remain redeemable pro-rata against the treasury via ragequit, so the economic claim survives even as voice is revoked [12].

Hoppe's covenant-community theory gives the clearest doctrinal anchor: expulsion operates within a contractual property framework and is "not violent, but consistent with property rights," a reading confirmed independently across sources and clarified by Hoppe himself in a later interview as meaning physical distance achieved through non-violent means — ostracism, ridicule, collective complaint, never force [6]. Rothbard's derivation of punishment from victim self-defense does *not*, on closer reading, rule out non-agonistic collective boycott or ostracism, which he treats as non-invasive [7]. The negative case is The DAO (2016): when Ethereum's community used collective power to reverse a theft by moving balances, the result was a permanent chain split, with critics arguing the "correction" was itself the gravest property violation of the episode [13]. The lesson is architectural, not moral: once a jury can move a balance for a legitimate reason, there is no bright line separating that first, justified use from a later, corrupted one.

This grounding does not eliminate one genuine doctrinal tension. Kinsella's estoppel theory holds that restricting a victim's remedy to expulsion alone re-victimizes them — they might legitimately choose proportional confiscation instead — which directly contradicts a "balance is untouchable" ceiling [6, 8]. Harlequin resolves this not by claiming a false libertarian consensus, but by making the ceiling itself a matter of ex-ante consent: the remedy cap is accepted when a mask is created, so the future victim has also already consented to it, and nothing is being taken from them that they held unconditionally. The DAO precedent (§above) supplies the structural argument for why this ceiling must be a hard constitutional limit rather than a jury's

case-by-case discretion [13].

### B.3 Empirical grounding

Three independently produced results close the plutocracy question with unusual convergence: measurement across 21 on-chain governance systems shows Gini coefficients of voting power "close to 1.0" (e.g., 0.992 for Uniswap and 0.996 for Compound token holders, Table 2), a majority of voting power controlled by fewer than 10 participants in 17 of the 21 systems, and Nakamoto coefficients of three or fewer in half the sample [14]; a survey of blockchain-governance designs finds none satisfying a majority of desiderata simultaneously [15]; and a 2026 formal impossibility result proves that *every* concave voting scheme (quadratic, logarithmic, or otherwise) collapses to linear, plutocratic outcomes under unrestricted sybil splitting, with measured amplification of 1.172×–4.039× across five real DAOs [16]. This last result is the structural caveat that holds up the entire mechanism below: reputation is exactly as capturable as capital once identities can be freely multiplied. Beanstalk's 2022 flash-loan governance attack ($182M) demonstrated instantaneous vote-buying working exactly as designed [17]. Compound's Proposal 289 shows the same dynamic through patient accumulation rather than a flash loan: the proposal, passed 682,191 to 633,636 votes in July 2024, directed roughly $24M (499,000 COMP) from the treasury to a vault linked to a large holder, using voting power built from roughly $12M in delegated tokens rather than the full $24M figure — a real but more modest capture than headlines suggested [18]. A first-hand forum account from Polkadot describes a single 9M-DOT wallet flipping a referendum from majority-approval to 96% rejection with no discussion; this is reported here as an unverified community anecdote, not an audited on-chain fact [19]. Ferreira's 2026 academic analysis, published in Wiley's *Corporate Governance: An International Review* by an LSE finance professor (not, as sometimes cited, an Oxford paper — the Oxford Business Law Blog merely summarized it), reaches the same conclusion from the literature side: procedural formality does not confer capture resistance [20].

Darknet-market data supply a different empirical strand: vouching-as-entry-gate, tested in the best-documented case (Darkode), failed — 94.5% of applications were accepted, and the platform's own administrators generated 38% of recruitment by soliciting novices, collapsing the selectivity the mechanism was meant to provide [31]. By contrast, community self-regulation around a near-universal moral consensus — CSAM prohibition — held consistently across markets without central coordination, while suppression of a commodity good (opioids) merely displaced listings rather than reducing them, and fentanyl-specific suppression combined with external pressure showed sustained effect [33, 34]. The asymmetry matters for the market layer: entry filters do not reliably suppress bad actors, but a hard, narrowly scoped prohibition on aggression against non-consenting third parties can.

### B.4 The mechanism

Harlequin composes these findings into a layered, status-only sanction architecture. **M2, individual dissociation**, is the inalienable floor: any mask may always stop dealing with anyone. It has no teeth alone but cannot be captured, since no central lever exists to seize. **M5, sortition-selected juries with status-only sanctions**, forms the process layer: jurors are randomly selected and rotated — the anti-bribery defense independently rediscovered by Athens and reimplemented by Kleros [1, 30] — verdicts are public, and the only sanction available is reputation, vouch-edge, or eligibility loss, never a balance; this ceiling is written into Article II

as a structural limit, not jury discretion, precisely because of the Kinsella tension and The DAO precedent (§2) [8, 13]. **M3, a graduated ladder** (warning → reputational penalty → privilege suspension → temporary expulsion with a re-admission path → full expulsion) follows Ostrom's design principle #5, graduated sanctions — distinct from her principle #3, collective-choice arrangements, which the jury's compositional legitimacy separately satisfies [21]. Re-admission is priced in endurance (a streak requirement), making the cost of re-entry time rather than capital or reputation, the one currency the sybil-collapse result cannot multiply cheaply [16]. **M4, vouch revocation**, operates as a verdict *consequence*, never an entry gate — the Darkode result rules out vouching-as-gate specifically, not vouching-as-consequence [31]. **M6, privilege-specific bonds** (for jurors, market-makers, node operators) can be forfeited, but only that deposit, architecturally separated from the base wallet, avoiding the HOA-style drift by which a status stake gradually becomes collateral against the balance itself. Finally, Hoppe's unresolved gap — a boycotted party left formally propertied but practically undefended — is closed explicitly and in writing: property protection is universal and status-independent, so a jury must adjudicate a theft complaint from an expelled mask exactly as it would from anyone else; and the sanction clause is acts-only, never belief or affiliation, foreclosing the ideological-capture risk visible even within Hoppe's own later writing.

A final framing constraint comes from the constitution itself: Harlequin's justice imposes no punishment; it makes truth known. The jury therefore does not decree expulsion — it certifies a proven act and publishes the verdict. Every consequence that follows (the reputational record, loss of eligibility for common functions, forfeiture of a privilege-specific bond, withdrawal of common shelter and tools) is either an effect the member consented to ex ante upon creating the mask, or the aggregate of other members' free decisions to dissociate. Sanction is emergent, not imposed: the jury is the epistemic organ; individual dissociation is the enforcement arm. Two corollaries follow. First, expulsion never blocks peer-to-peer exchange at the protocol level — the order neither privileges nor excludes any merchant; it withdraws common privileges and marks the public record, and each counterparty decides for themselves. Second, no constitutional amendment is required: the mechanism lives entirely in the mutable specification, subordinate to the sealed principles from which it derives.

### B.5 Honest limitations

Three limitations should be stated rather than papered over. First, the entire architecture rests on anti-sybil cost: if masks become cheap to multiply, the formal impossibility result applies directly and the system collapses into plutocracy of sybils regardless of how carefully status and property are separated [16]. Second, sortition has a real deployment record in bounded disputes (Kleros) but none at constitutional-upgrade scale; extrapolating its anti-capture properties to Harlequin's jury system is a declared assumption, not a demonstrated one [30]. Third, the thin-versus-thick libertarianism debate over how far a voluntary order may go in policing consensual "vice" versus non-aggression narrowly defined is untouched by this research pass and remains an open question for future work.

---

**B-References**


1. Wikipedia, "Ostracism"; World History Encyclopedia, "Ostracism"; "Pruning of the People," *Philosophies* 8(5):81 (MDPI, peer-reviewed).
2. Hurstwic, "Viking-age Laws and Legal Procedures" (hurstwic.org); *Norse Mythology for Smart People*, "Outlawry in the Viking Age."
3. Amish Studies, Elizabethtown College, "Church Discipline" (groups.etown.edu/amishstudies).
4. Quaker.org, "Our Understanding of Disownment."
5. Wikipedia, "Excommunication in the Catholic Church."
6. Stephan Kinsella, "Hoppe on Covenant Communities and Advocates of Alternative Lifestyles" (stephankinsella.com, 2010); Hoppe's clarification in his interview with Michael Malice ("YOUR WELCOME," Ep. 018, September 2018) that "physically removed" means physical *distance* achieved by non-violent means — ostracism, ridicule, collective complaint — never force.
7. Murray Rothbard, *The Ethics of Liberty*, Ch. "The Ethics of Boycotts"; Hoppe, introduction to *The Ethics of Liberty* (mises.org).
8. Stephan Kinsella, "A Libertarian Theory of Punishment and Rights" (1997/2021 repost, stephankinsella.com); *Libertarian Papers* 1:17.
9. Murray Rothbard, *The Ethics of Liberty*, Ch. 13 "Punishment and Proportionality" ("two teeth for a tooth"); further developed by Walter Block.
10. Robert Nozick, *Anarchy, State, and Utopia* (1974), meta-utopia framework.
11. Stanford Encyclopedia of Philosophy, "Freedom of Association" (Hohfeldian analysis).
12. MolochDAO documentation, guildKick/ragequit mechanism (LOOT conversion, not burn).
13. Ethereum Classic Cooperative blog, "Code is Law and the Quest for Justice"; CoinDesk, "Ethereum Executes Blockchain Hard Fork to Return DAO Funds" (2016).
14. arXiv:2302.12125, governance concentration across 21 on-chain systems.
15. Kiayias & Lazos, "SoK: Blockchain Governance," arXiv:2201.07188 — survey finding no analyzed system satisfies a majority of desirable governance properties.
16. Bennett, Vander Vos, Le, Belenkiy, "Concave is the New Linear: The Impossibility of Anti-Plutocratic DAO Governance," arXiv:2605.18990 (AFT 2026).
17. Bloomberg; CoinDesk; Cointelegraph; The Register (April 17–18, 2022), "DeFi Project Beanstalk Loses $182 Million in Flash Loan Attack."
18. Compound governance forum, Proposal 289 (passed 682,191–633,636, July 28, 2024).
19. Polkadot governance forum, community report on referendum swing by a single 9M-DOT wallet (unverified, first-hand anecdote); Messari, "Polkadot OpenGov Report."
20. Daniel Ferreira, "The Myths of Blockchain Governance," *Corporate Governance: An International Review* (Wiley, 2026); London School of Economics.
21. Elinor Ostrom, *Governing the Commons* (1990), design principles #3 (collective-choice arrangements) and #5 (graduated sanctions).
22. OpenZeppelin, Governor/Timelock documentation (GovernorTimelockCompound.sol).
23. Tezos documentation, self-amendment process (docs.tezos.com).
24. MakerDAO documentation, Governance Security Module (GSM) pause delay (docs.makerdao.com).
25. Vitalik Buterin, "Credible Neutrality as a Guiding Principle" (2020); governance-minimization writings (vitalik.eth.limo).
26. Nick Szabo, "Trusted Third Parties Are Security Holes" (2001), nakamotoinstitute.org.
27. Arbitrum governance documentation, Security Council (2 staggered cohorts of 6 members,

9-of-12 approval threshold).
28. Wikipedia, “Liberum veto”; Encyclopaedia Britannica, “Polish–Lithuanian Commonwealth.”
29. BIP-148 text; uasf.org (Bitcoin UASF, 2017).
30. Kleros.io; Stanford Law School, “Kleros: A Socio-Legal Case Study of Decentralized Justice & Blockchain Arbitration.”
31. ResearchGate, “Darkode Recruitment Patterns.”
32. Morselli, Décary-Hétu, Paquet-Clouston & Aldridge, “Conflict Management in Illicit Drug Cryptomarkets,” *International Criminal Justice Review* 27(4), 237–254 (SAGE, 2017). DOI 10.1177/1057567717709498.
33. Australian Institute of Criminology, “Impact of Darknet Market Seizures on Opioid Availability” (rr18).
34. ResearchGate, “Crime Displacement in Digital Drug Markets.”