

Clockless Consensus

Truth-Selection via an Unknown-Future Schelling Point

Tanguy Prudat*

August 13, 2026 — working draft

Abstract

Many of the questions that matter are empirical yet **unanchored**: there is a fact of the matter, but no oracle will rule on it and no one can name the day the evidence lands — where a disease outbreak began, what a dietary pattern does over decades. Prediction markets cannot reach them, because they need a settlement event to import. What remains measures what a crowd believes *now*, and nothing in it pulls that belief toward evidence. Nor does adding a stake supply the pull: where belief is traded with no resolution to anchor payoff, momentum dominates and the profitable move is to front-run the crowd rather than to be right.

We introduce *clockless consensus*, which supplies the missing pull without importing an anchor. A participant stakes real value on a True/False position; the position **locks** and settles neither on a date the holder picks nor at an oracle’s ruling, but at a moment set by the **activity of the question itself** — more contested questions settle later, and no one can time their own settlement. Each position is thus its own synthetic resolution event, never a shared date. Reward rises with how lopsided the crowd was against a position at entry, and with whether the crowd has come around by the time it is measured: *bold at entry, correct at settlement*.

The conceptual core is an **unknown-future Schelling point**: a Keynesian beauty contest converges only because every player is scored at one shared instant, and removing that instant leaves nothing to coordinate on except an answer that does not move — for an evidence-bearing question, the truth. We propose no theory of truth and do not claim that perfect epistemic consensus is attainable; this is an incentive mechanism meant to improve the long-run aggregation of dispersed evidence under uncertainty. We give the design rationale, a game-resistance analysis, and — centrally — a precise account of why simulation can validate the mechanism’s **engineering properties but not its epistemic claim**.

Contents

1 Introduction	3
2 Related work	6
2.1 Prediction markets and proper scoring rules	7
2.2 Decentralized and optimistic oracles	7
2.3 Beauty contests and focal points	8
2.4 Opinion dynamics in statistical physics	8

*Independent researcher — sardariuss@gmail.com

2.5	The logistic link and the weight of evidence	8
2.6	Forecasting tournaments (the foil)	9
2.7	Reflexivity, performativity, and the formal bounds	9
3	The mechanism	9
3.1	Belief is a drift-free cursor	10
3.2	Decay and crystallization	11
3.3	Inertia and resistances	11
3.4	Activity-timed settlement	12
3.5	The reward	13
3.6	Yield source	14
3.7	The three tunable parameters	14
4	Design rationale	14
4.1	Why decay: arbitration windows	14
4.2	The pathology raw ratios have	15
4.3	Why a logit for dissent	15
4.4	Why a sigmoid for consent	16
4.5	The asymmetry, and the annealing picture	16
4.6	Why one pot	16
5	The core contribution: the unknown-future Schelling point	18
5.1	Unpredictable is not arbitrary (the fair-game constraint)	18
5.2	The reflexive fixed point (the learning-curve argument)	19
5.3	What learning cannot reach (the horizon)	19
5.4	The beauty contest is not forbidden — it is farmed	20
5.5	Honest bound	20
6	Game resistance	21
6.1	Sybil fragmentation is neutral by construction	21
6.2	Capital wars converge to a finite horizon	22
6.3	Consent-sniping is cheap at small pools, and bounded there	23
6.4	Buying the display: exogenous-payoff manipulation	23
7	The validation boundary	25
7.1	Two different claims, only one is simulable	25
7.2	Why the epistemic claim resists simulation	25
7.3	The boundary reaches further down than the claim	26
7.4	What simulation can honestly do: falsify, never confirm	26
7.5	The only real validation is post-deployment and partial	27
8	Limitations & honest claims	27
8.1	The epistemic claim is a wager (in principle)	28
8.2	The right-too-soon trap	28
8.3	Payoff magnitude is globally coupled	29
8.4	Crystallization sets consensus — including, sometimes, a wrong one	30
8.5	Impartial by construction, but only at the consensus level	30
8.6	Which instrument for an anchored question?	31

1 Introduction

Many of the questions that matter most are *empirical* yet *unresolved*: there is a fact of the matter, but we do not know it and cannot say when we will. The instruments we have address the endpoints of this spectrum — from questions where present opinion is all there is to measure, to questions an arbiter will settle on a known date — but not its middle. Opinion polls and sentiment gauges measure what a crowd believes *now*, but nothing in them pulls that belief toward the truth; carrying no stake, they offer no move to profit from at all. Nor does adding one help: where belief is traded with no resolution to anchor payoff, momentum dominates and the profitable move is to front-run the crowd rather than to be right. Prediction markets *do* tie payoff to truth, but only by importing an **oracle** and a **settlement date** — and most contested empirical claims offer neither. We call such questions **unanchored**: there is a fact of the matter, but nothing exists to anchor its resolution.

The stakes. Shared belief is what lets people coordinate: reciprocity in repeated games runs on accurate mutual expectations of past play (Axelrod and Hamilton 1981). And *any* shared belief can serve locally — bubbles and propaganda coordinate too — which is why coordination fragments into clusters, each cohering around its own orthodoxy. Truth is different in two ways: it is the one focal point available to every cluster without prior agreement, and the one that stays serviceable as the world pushes back and evidence accumulates. Each mind it wins therefore merges coordination spaces rather than multiplying them. A mechanism that pulls shared belief toward truth is not merely an epistemic instrument; it is coordination infrastructure.

That definition of *unanchored* carries a restriction worth making explicit: the fact itself must be **fixed** — a past event, or a time-invariant regularity — so that time can change our *knowledge* of the answer, never the answer. Fixing the referent takes more work for some claims than for others, and a claim whose subject drifts too much cannot be admitted at all; both are admissibility questions, and we settle them with the mechanism in §3.

The resulting taxonomy, with the instrument that serves each type:

The question asks about...	Fact of the matter	When the answer becomes obvious	Instrument
present opinion — “ <i>which candidate do voters prefer?</i> ”	none: opinion is the object	immediately — the measurement <i>is</i> the answer	polls, sentiment gauges
a fact an arbiter can rule on by a set date — “ <i>will inflation run under 2% next year?</i> ”	fixed once realized	on a known date	prediction markets, including those arbitrated by AI or jury

The question asks about...	Fact of the matter	When the answer becomes obvious	Instrument
a fixed fact with no resolution anchor — “ <i>did the outbreak originate in a laboratory?</i> ”	fixed, unknown	unknown, perhaps never	clockless consensus
a drifting generality — “ <i>does raising the minimum wage reduce employment?</i> ”	moves with conditions	never: the answer itself moves	none: ill-posed until its subject is pinned (→ row above)

Put concretely. A prediction market asks *will this be written into the record?* — the answer is not there yet, the world will put it there, and the date it does so **is** the settlement date. Clockless consensus asks *is what the record already contains right?* — the fact is already fixed, and what runs on no schedule is when the evidence surfaces and when interpretation converges on it. This is why unanchored questions so often concern a past event or a standing regularity rather than a scheduled one, a pattern that would otherwise look like coincidence: event-occurrence has a date, and interpretation-convergence does not.

Two horizons. The organizing axis is deliberately *when the answer becomes obvious*, not whether the question is dated — a distinction the word “unanchored” tends to swallow. Two unknowns are in play and the paper keeps them apart. The **epistemic horizon** is when the answer becomes obvious: the column the table is ordered by, and the property that defines the domain. The **settlement horizon** is when a given position is measured and paid. A prediction market requires the epistemic horizon to be known, and then simply sets the settlement horizon equal to it. Clockless consensus is built for questions whose epistemic horizon is unknown, and its central move is to make the settlement horizon unknown as well (§3.4).

The boundary this draws is one of *fit*, not of capability. Nothing in the mechanism requires the anchor’s absence, so the two instruments do not partition the space: prediction markets *cannot* operate without an anchor, while clockless consensus *can* operate with one. That it usually should not is a trade-off with arguments on both sides rather than a boundary, and it turns on machinery introduced later; we set it out in §8.6.

The move. Clockless consensus targets that unanchored middle.¹ A participant stakes real value on a True/False position about such a claim; the position **locks** and is settled not on a date the holder picks, nor at an oracle’s ruling, but at a moment set by the *activity* of the pool itself — the more contested the pool, the later and less predictable each position’s settlement. Removing this common clock converts the dominant strategy from **front-running** into **foresight**: because you cannot time your exit, the only way to be reliably paid is to hold a position that is right *whenever* it happens to settle.

The reward. A position *scores* stake × lock-time × dissent × consent — a claim on a redistributed yield pot, not a return on the stake (§3.5). The two performance factors are a **logit** of how

¹Implemented in the Towerview protocol; yield is redistributed from an external lending venue, not generated by the mechanism (§3.6). “No loss” is precise, not absolute — principal minus a fixed withdrawal fee.

lopsided the crowd was against you at entry (boldness — unbounded, and additive in evidence, since each independent datum shifts log-odds by a fixed amount) times a **sigmoid** of whether the crowd agrees with you at settlement (correctness — saturating and near-binary). The asymmetry is deliberate: rare boldness is worth exponentially more, while being right is an almost-binary fact that should not pay extra for a blowout. The two transforms are the two halves of the logistic link, applied at two different moments to two different quantities (§4). While a position is locked its weight decays, so belief stays movable; at settlement its weight **crystallizes** — freezes — so belief *sets* over time rather than drifting.

Why this can select for truth. The design question underneath the paper: can an institution make the most stable equilibrium over an unknown future *convergence toward reality* rather than *convergence toward one another*? A Keynesian beauty contest — “guess what the average expects the average to guess” — is coordination of the second kind, and it converges only because everyone evaluates at one shared instant. Remove that instant and there is nothing left to coordinate on *except* an answer that does not move: across an uncertain horizon, transient crowd moods wash out and only the time-invariant attractor survives, which for an evidence-bearing question is the truth. We call this the **unknown-future Schelling point**, and it is the paper’s conceptual contribution (§5). We are deliberately honest about its reach: this does not *escape* the beauty contest, it *slows* it — a wager, developed and bounded in §5, not a proof.

What we do not claim. We are as precise about the boundaries as about the claims. We do *not* claim the mechanism is ungameable — no mechanism of this kind is strategy-proof (Gibbard 1973; Satterthwaite 1975) — but show instead that manipulation is *economically irrational at the scale that matters* (§6). We do *not* claim it provably converges to truth: the consensus it produces is reflexive, so what the design guarantees is a *consistent scoring of contrarian-correctness against the eventual consensus*, not that the consensus equals reality (§8.1). And we do *not* claim “no loss” — principal is protected only up to a fixed withdrawal fee. Most consequentially, we argue that the truth-tracking claim **cannot be settled by simulation** (§7), and that drawing that boundary precisely is itself one of our contributions.

Contributions.

1. **The clockless mechanism.** A consensus layer for unanchored, contested empirical claims in which positions settle asynchronously, each at an **activity-timed, individually-unpredictable** moment — synthetic resolution events, one per position, for a question that has no natural one.
2. **The unknown-future Schelling point.** Removing the *common clock* from a Keynesian beauty contest turns rational coordination into a **truth-selecting** force: across an uncertain horizon, the only durable focal point is the time-invariant (= true, where knowable) answer.
3. **The logit×sigmoid reward decomposition.** Score = stake × time × dissent × consent, where dissent is a **logit at entry** (boldness, unbounded, additive in evidence) and consent is a **sigmoid at exit** (correctness, saturating, decisive) — the two halves of the logistic link applied at two moments to two quantities.
4. **Decay → crystallization as annealing.** While locked, mass decays (belief stays “hot” and movable); at settlement it crystallizes/freezes (the unit “cools” and commits to a basin) — simulated annealing into a ground state; **consent_steepness** is the temperature.
5. **A precise account of the limits of validation.** An argument that the *epistemic* claim is

non-simulable by construction (reflexivity, performativity/Lucas, Wittgenstein’s Ruler), separating what the design *guarantees* (engineering) from what it *wagers* (epistemics) — and why that boundary is a feature.

Roadmap. §2 situates the design among prediction markets and decentralized oracles, proper scoring rules, the beauty-contest tradition, and the statistical physics of opinion dynamics; §3 specifies the mechanism; §4 gives the reward-design rationale; §5 develops the unknown-future Schelling point; §6 analyzes game resistance; §7 draws the validation boundary; §8 catalogs limitations and honest claims; §9 concludes.

2 Related work

Clockless consensus sits at the intersection of prediction markets, proper scoring rules, the beauty-contest tradition, and the statistical physics of opinion dynamics. The correspondences below are positioning, not lineage: the design was reached independently, so its alignment with these established results is best read as evidence that its form is principled rather than tuned. For each we mark where the design *coincides* and where it *departs*; the through-line is that every departure follows from operating on questions with **no oracle and no settlement date**.

Before the individual correspondences, it helps to locate the design by what a mechanism *does* when no anchor is available. Three strategies are in play, and they concede different things:

Response to anchor-scarcity	Representative work	What it concedes
Scale the arbitration — make ruling cheap enough to cover more questions	optimistic and jury oracles (UMA Project 2020; Lesaege et al. 2019; Peterson et al. 2018); recent AI-plus-jury resolvers	a resolution moment is still required, so questions no panel can rule on stay out of reach
Self-resolve against the crowd — score reports against the eventual consensus rather than against an outcome	peer prediction (Miller et al. 2005; Prelec 2004; Shnayder et al. 2016), crowdsourced outcome determination (Freeman et al. 2017), self-resolving prediction markets (Srinivasan et al. 2025)	ground truth as the referent — and each design must then supply a stopping rule fixing when the consensus is read
Drop the anchor entirely — measure present belief and stop there	polls, sentiment gauges	any force pulling belief toward evidence

Clockless consensus belongs to the second family, and its departure from the rest of it is the **shape of the stopping rule**. The nearest prior art, self-resolving prediction markets (Srinivasan et al. 2025), stops the market with a fixed probability after each report — exogenous, independent of what anyone reported — and scores all but the last few agents against the terminal agent’s forecast. That clock is *global* (one draw ends the market for everyone) and *lawless* in

the sense of §5: identical actions can be measured at wildly different times for no reason traceable to the participants. Ours is *per-position* and *activity-timed* — a common, public function of observable pool state, unpredictable in instance because it is generated by the crowd’s own future behavior. §5 develops why that distinction is not cosmetic: an exogenous random clock supplies the horizon-uncertainty the argument needs but destroys the fair-game property that makes foresight priceable rather than lucky.

The deeper difference is in what each design *assumes* versus what it *pays for*. The self-resolving construction earns a strict perfect-Bayesian-equilibrium guarantee by assuming a common prior, rational Bayesian agents, and — centrally — that agents’ signals are **conditionally independent given the ground truth**. That last assumption is the Condorcet premise this paper openly rests on (§8.1). We decline to assume it, and instead try to *price* it: *dissent* pays for the log-odds distance between a position and the pool it entered, so a belief merely copied from the crowd books almost nothing while a divergent one books a great deal (§4.3). Where their mechanism has a theorem conditional on independence, ours has an incentive aimed at approximating it — a weaker guarantee, on a domain where the assumption is least tenable, since correlated belief is precisely the phenomenon these questions exhibit.

The substitution is a proxy and not a payment for the thing itself, and we record the shortfall with the other limitations rather than here: dissent prices decorrelation from the *consensus*, which is not pairwise independence of signals, so correlation among the correctors goes unpriced (§8.1). The contrast with the self-resolving construction survives it — an imperfect proxy one can inspect is a different epistemic object from an assumption one must grant — but it is a contrast between an assumption and a proxy, not between an assumption and a guarantee.

2.1 Prediction markets and proper scoring rules

Prediction markets — canonically Hanson’s logarithmic market scoring rule (LMSR) (Hanson 2003, 2007) — and proper scoring rules (Good 1952; Gneiting and Raftery 2007) tie a forecaster’s payoff to a *resolved outcome*, rewarding truthful probabilities against realized ground truth. Our reward is in this family: *dissent* is a log-odds score in the family of *truthful* (proper) scoring rules, and *consent* is a market-scoring-style read-out of the crowd. The departure is the resolution model. Classical scoring needs an **exogenous oracle** to settle the outcome at a **known date**; LMSR in particular scores against a resolved event. Clockless consensus has neither: *consent* is scored against the *endogenous* consensus, at an *activity-timed*, *individually unpredictable* settlement (§3.4). LMSR is the closest prior art to *consent*, but its counterparty is a future fact an oracle will reveal; ours is the crowd itself. Its *sponsored subsidy* is likewise the closest prior art to our global yield pot, which funds thin pools by the same logic without a sponsor (§4.6). This is what lets the mechanism run on unanchored questions — and it is exactly what turns a proof into the epistemic wager of §5 and §7.

2.2 Decentralized and optimistic oracles

Closer to us than a plain prediction market is the *decentralized oracle*: mechanisms that settle subjective or hard-to-verify questions without a trusted arbiter, through Schelling-point coordination among token-holders — UMA’s optimistic oracle (UMA Project 2020), Kleros (Lesaege et al. 2019), Augur’s reporting (Peterson et al. 2018). These share our refusal of a privileged truth-teller and our reliance on a focal point. The difference is the *shape of resolution*. A de-

centralized oracle settles a *posed, dated* question *at a resolution moment*: a claim is submitted, a window opens, reporters vote or a dispute escalates, and the question is decided and closed. Clockless consensus has no such moment — settlement is activity-timed and per-position (§3.4), the question never closes, and the reward is for *foresight* (being contrarian-early and later vindicated, §5), not for reporting the majority at a known deadline. A decentralized oracle answers “how does the panel rule, now that we are resolving?”; clockless consensus answers “which way will belief settle, over a horizon no one controls?” The former still needs an agreed moment of truth; the latter is built precisely for the questions that never get one.

2.3 Beauty contests and focal points

Keynes’s beauty contest (Keynes 1936) — reward for guessing what the average guesses the average will guess — is the canonical model of a market that coordinates on *expectations about others* rather than fundamentals; Schelling’s focal points (Schelling 1960) explain how such coordination settles without communication. This tradition established that a consensus market is at root a coordination game, organized by a focal point; clockless consensus turns on the same recognition. Our departure is the **unknown-future Schelling point** (§5): a beauty contest converges only because all players evaluate at one shared instant, and removing that instant leaves nothing to coordinate on *except* the one focal point invariant across an uncertain horizon — for an evidence-bearing question, the truth. To our knowledge, using horizon-uncertainty itself as the selection filter — turning “predict present sentiment” into “predict the time-invariant answer” — is novel to this work.

2.4 Opinion dynamics in statistical physics

A large literature models how opinions spread and settle using spin systems: Ising/Glauber dynamics (Glauber 1963) and their sociophysics descendants (Castellano et al. 2009), and Hopfield’s attractor networks (Hopfield 1982). The *mathematics* coincides exactly — the Glauber update is literally a sigmoid with temperature, which is *consent* with `consent_steepness` as that temperature, and the decay/crystallization dynamics read as temperature and simulated annealing into a basin (§4.5). The departure is one of purpose: this literature is **descriptive** (it models how opinions in fact evolve), whereas the same functional forms are used here **prescriptively**, as an incentive design. The physics says how a field relaxes; we pay people to relax it toward evidence.

2.5 The logistic link and the weight of evidence

That the reward pairs a logit with a sigmoid is not an aesthetic choice but the two ends of the canonical generalized-linear-model link (McCullagh and Nelder 1989): logit as natural parameter, sigmoid as mean (§4.5). The choice of the log-odds scale for *dissent* rests on the weight-of-evidence tradition (Good 1950; the deciban unit is Turing’s, see Good 1979): independent evidence is *additive* in log-odds, which is precisely why boldness against a lopsided crowd should be scored there (§4.3). These established results are why the reward’s functional form is the principled one and not a tuned one.

2.6 Forecasting tournaments (the foil)

Tetlock’s Good Judgment Project and the *Superforecasting* program (Tetlock and Gardner 2015) elicit calibrated human forecasts and are the nearest existing body of real forecast data. They are best read here as a **clarifying foil**: they operate on *dated, resolvable* questions (“will X happen by date D”), exactly the regime clockless consensus is built to leave behind. They differ from our mechanism on precisely the axis of §1 — whether the date at which the answer becomes obvious is known — which is its whole novelty, and also why they cannot serve as a validation set for it (§7).

2.7 Reflexivity, performativity, and the formal bounds

The *epistemic* status of the truth claim connects to a second lineage, developed in §7: reflexivity and self-fulfilling prophecy (Soros 1987; Merton 1948), the performativity of economics (MacKenzie 2006; Callon 1998), the Lucas critique (Lucas 1976), Wittgenstein’s Ruler (Taleb 2001), and Goodhart’s Law (Goodhart 1975). These explain why the truth-tracking claim *resists validation*, not merely why it might fail. Two formal results bound the claims from the start: **Gibbard–Satterthwaite** (Gibbard 1973; Satterthwaite 1975), which is why we claim economic irrationality of manipulation rather than strategyproofness (§6), and the **Condorcet Jury Theorem** (Condorcet 1785) with its known failure under correlated or insincere voters (Ladha 1992), the premise the truth-aggregation claim openly rests on (§8.1).

3 The mechanism

Scope and objects. The mechanism operates on *pools*, one per question. A pool has two sides, Y (True) and N (False), and aggregates timed, staked contributions into a single scalar — the crowd’s **aggregate belief** that Y holds. A pool is the analogue of a *market* in a prediction market — one question, every position on it — but the word is spatial, not monetary. A pool **locks** money — each stake is committed until its settlement — yet no pool **owns** money: all principal, locked and idle alike, sits in a single platform-wide supply (§3.6), and rewards redeem at a global share price (§3.5, §8.3). What is local to a pool is belief: which stakes, on which side, entered when. Pools are not created freely — each is instantiated by a governance proposal (§8), the only privileged action in the system. That gate also enforces well-posedness (§1): an admissible statement is a single, declarative, empirical claim with a **time-invariant referent** — a claim over an evolving set (“the studies to date”) enters only with its subject pinned by an explicit cutoff.² Once a pool is open, no actor can set or override that belief: it is a pure function of staked mass.

Well-posedness. That rule inherits one requirement from the anchored case, and what it does and does not demand is worth spelling out. A settlement date, where one exists, quietly does two jobs: it *pins the referent* (“unemployment below 5% *on this date*”) and it *schedules the payoff*. Clockless consensus dispenses only with the second; the first remains mandatory and is done by other means. How much work that takes varies with the claim. A time-invariant regularity pins itself — “the earth is round” needs no date at all — while a historical claim carries one explicitly, as in “the outbreak that began in Wuhan in December 2019”. Either way the *referent*

²Full statement-admissibility criteria are specified in the Towerview design documentation; design documents are available from the author on request.

is fixed and the *resolution* is not, which is exactly what the word “undated” obscures, and why we avoid it.

The dependence runs the other way too, and §1’s taxonomy shows it: a prediction market can pose a *relative* referent — “under 2% next year” — because closing on resolution fixes what “next year” meant. A clockless pool never closes, so every date it relies on must be absolute; relative dating is harmless under an anchor and drifts silently without one. A generic causal claim whose truth drifts with conditions (“this policy reduces that harm”) is admissible only once its subject is pinned to a named case, and the pin anchors the *subject*, not a resolution date, which the question still lacks. The mechanism thus targets not the whole unanchored space but its **well-posed** part; the cut is structural, not cosmetic, since the argument of §5 is convergence to a time-invariant attractor, and a drifting fact offers none.

Notation.

Symbol	Meaning
$h, \lambda = \ln 2/h$	pool half-life and the implied decay rate
t	current (running) time — a position enters at s , settles at σ
$c(t) \in (0, 1)$	the belief cursor — the crowd’s aggregate belief (its YES-share of staked mass)
$M(t) = L(t) + X(t)$	mass : the pool’s total weight — a decaying part L plus a crystallized (frozen) part X
(a, y, s)	a contribution of amount a on side $y \in \{Y, N\}$ entered at time s
$D_i, \sigma_i = s_i + D_i$	the lock duration and settlement date of position i
θ	consent_steepness — a temperature
$\alpha, \gamma = \log_{10} \beta$	duration-scaling parameters

3.1 Belief is a drift-free cursor

The aggregate belief is stored directly as a **belief cursor** $c \in (0, 1)$, not read off a ratio of side balances. (A note on words: c is a *level of belief*, not a state of agreement — we reserve **consensus** for belief concentrated near an extreme, the agreed state the mechanism drives toward; at $c = \frac{1}{2}$ there is no consensus, only maximal division.) When an amount a is added on side y at time t , the cursor moves as fresh mass competing against current mass, and the mass grows:

$$c \leftarrow \frac{c \cdot M + a \cdot \mathbf{1}[y = Y]}{M + a}, \quad M \leftarrow M + a.$$

This reproduces the side-mass ratio *at the instant of each contribution* and then holds it fixed until the next one. The reason to store c rather than recompute $\text{total}_Y / (\text{total}_Y + \text{total}_N)$ is **drift**: locked balances decay while crystallized mass does not, so a recomputed ratio would move on its own with the mere passage of time, spuriously shifting who is “right.” Storing the cursor guarantees the first system invariant — **belief changes only when mass is added, never with time** (only the *mass* M decays, which is what makes the cursor easier to move).

3.2 Decay and crystallization

A contribution is *locked* from entry s until its settlement σ . While locked, its weight decays, $a e^{-\lambda(t-s)}$, contributing to $L(t)$. At settlement it **crystallizes**: its weight stops decaying and is moved into X , the permanent (frozen) mass, where it influences c forever after. The macro-dynamics follow: while a pool’s mass is mostly locked, it decays with time, so M shrinks and belief grows *cheaper* to move; as positions settle and crystallize, X accumulates and raises a permanent floor under M , so belief progressively **sets** — it stops getting cheaper to contest. *Time weakens the ground, not the result.* Concretely, since the downstream side-magnitudes are reconstructed as same $= c M$ and opposite $= (1 - c) M$, the two *scoring* signals reduce to functions of c alone — consent $= \sigma((c - \frac{1}{2})/\theta)$ for the Y side (mirrored for N), and dissent as the log-odds of the crowd split — so **they do not drift as M decays** (and a position’s dissent is fixed at entry in any case). Only *magnitude*-scaled quantities move with time: **inertia** $= M (C - c)/(1 - C)$ falls as locked mass decays, and the dissent a *new* fixed-size stake would earn also moves — its self-impact integral depends on a/M , so as M decays that stake is a larger share of the pool, self-dilutes more of the opposition it is scored against, and books *less* dissent.³

Arbitration windows. While a position is locked, its conviction remains contestable by later entrants; crystallized mass, by contrast, has *survived* its window. We call the interval in which a still-locked cohort can be overruled its open **arbitration window** — each cohort is one on the cohorts before it, and is itself at the mercy of the next, until it crystallizes. *Why* the dynamics take this two-tier form — and what the half-life h assumes about evidence tempo — is taken up in §4.1.

3.3 Inertia and resistances

Inertia is the signed mass needed *right now* to move the cursor to a target C . Reconstructing the side totals from the cursor ($\text{total}_Y = c M$, $\text{total}_N = (1 - c) M$) and solving the cursor update for the required push gives

$$\Delta_Y = M \frac{C - c}{1 - C} \quad (\text{push up, } C > c), \quad \Delta_N = M \frac{c - C}{C} \quad (\text{push down, } C < c).$$

Inertia is proportional to M , so it **falls as locked mass decays**, floored by the crystallized mass, which never decays. Crystallization does not push inertia back up — it is mass-conserving at the settlement instant (a decaying contribution is swapped for an equal frozen one), so that portion simply **stops falling**. What “sets” belief over time is thus the *accumulating frozen floor*, not any upward force: inertia stops getting smaller, it does not grow.

The protocol exposes this quantity directly, through limit-order-style conditional entries (“**resistances**”). A resistance pre-commits a position to open when belief crosses a chosen threshold: the mass standing between the current cursor and that threshold is exactly the inertia above, so a pool’s resistances partition the cursor axis into intervals each priced by the push required to cross it. A resistance does not itself affect inertia until it activates. Resistances play no role in the truth-selection argument (§5), though they reappear as a defensive primitive

³Exact freezing accounting is specified in the Towerview design documentation.

against settlement-sniping (§6.3), and we omit their execution details.⁴

3.4 Activity-timed settlement

When position i is created at s_i , its lock duration is a function of pool **hotness** — the standing mass it lands on, plus a decay-weighted sum of the activity that follows it:

$$H_i = a_i + M(s_i) + \sum_{s_j > s_i} e^{-\lambda(s_j - s_i)} a_j, \quad D_i = \alpha \cdot H_i^\gamma,$$

the sum running over contributions that arrive while i is still locked. The two halves are deliberately asymmetric. Backward-looking influence enters at *standing weight* — the pool mass $M(s_i)$ of §3.2, in which locked contributions have decayed but crystallized ones sit at their frozen weight — so a late entrant challenging a settled pool locks *long*, not short: the same frozen floor that makes belief expensive to move (§3.3) buys the crowd a long counter-window against whoever tries. Forward-looking influence decays with time-distance, so a position also locks longer when it opens a busy, contested window. Crucially, because H_i grows with contributions that arrive *after* s_i , the duration D_i — and the settlement date $\sigma_i = s_i + D_i$ — is **recomputed as activity arrives**: at entry the holder sees only a current lower bound, never the final settlement instant. This is the formal content of “removing the common clock” (§1, §5): settlement is individuated and unpredictable. It is also bounded — successive late contributions extend σ_i by an exponentially decaying amount, a geometric series that converges to a finite settlement date even under adversarial capital injection (§6).

Making the settlement horizon unknown is not uncertainty for its own sake: a measurement date no participant can choose is what forces coordination onto an answer that does not move (§5). The design’s sharpest cost lives in the gap between the two horizons of §1: a position whose settlement horizon arrives before its epistemic one is right too soon, and paid as though it were wrong (§8.2).

The lock-duration curve. The curve from hotness to horizon is a power law with $\gamma = \log_{10} \beta$: α anchors the time scale (the lock of a unit-hotness position) and β is the *multiplier per decade* — a pool ten times hotter locks its positions β times longer, whatever the units. Two properties recommend the shape. Practically, a sublinear exponent ($\beta < 10$) compresses the enormous dynamic range of pool sizes into a usable band of horizons, with no absurd durations at either extreme. Structurally, it keeps the design’s log-scale discipline: dissent prices boldness per decade of odds (§4.3); the clock prices contention per decade of hotness. Nothing in the truth-selection argument depends on this particular shape — any monotone, activity-driven map preserves the clockless property (§5) — but the parametrization is where §8.2’s open question lives: adapting the horizon to a claim’s expected evidence-speed means varying α, β per pool rather than globally.

A delayed counterweight. Three couplings compose into a negative feedback with a built-in delay: contested pools draw stake from both sides (§4.5); stake raises hotness, lengthening every lock in the wave; and a position crystallizes at its *decayed* weight, $a_i e^{-\lambda D_i}$ — so the longer the lock, the lighter the permanent trace. A wave of capital that arrives in a contested burst therefore pre-shrinks its own lasting influence: the pool’s heat postpones the wave’s freezing

⁴Resistance lifecycle and execution rules are specified in the Towerview design documentation.

and melts its mass while it waits, whereas mass that arrives in the settled aftermath crystallizes at nearly full weight. **The permanent record is written by the calm, not the battle** — proportionally: a violent pool can still leave a heavier floor than a quiet one, but each unit staked in the heat contributes less of itself than a unit staked in the calm.

One corollary: current mass does not predict final mass. Absent new stakes, a pool’s mass decreases monotonically to its frozen floor $\sum_i a_i e^{-\lambda D_i}$, so a pool that out-weighs another today by virtue of a hot, locked wave can settle to a lighter floor than the smaller, calmer pool it dwarfs — where the curves cross is set by the half-life and the realized lock durations, not by today’s ranking. The shape is Le Chatelier’s principle — the system responds to a forcing in the direction that attenuates it; its economic form is Samuelson’s (Samuelson 1947) — and in the annealing picture (§4.5) it is one sentence: adding mass heats the pool, heat both postpones freezing and evaporates mass, and what crystallizes is what survives the cooling. The adversarial corollary is developed in §6.2: a capital war not only converges in horizon, it self-erases in mass.

3.5 The reward

At settlement, position i is credited with **shares**

$$S_i = a_i \cdot D_i \cdot \text{dissent}_i \cdot \text{consent}_i \quad (D_i \text{ in years}).$$

These shares are a claim on a **single, platform-wide pot** of exogenous yield — accrued on *all* principal the protocol holds: locked positions, resistances, and idle supply alike (idle capital earns nothing on its own, so all yield flows through shares). A position’s payout is its shares valued at the current **share price** — the yield pot divided by all outstanding shares, across every pool. Thus the *number* of shares a position earns is **local** — set by its own pool’s foresight — but the *price* they redeem at is **global**: what a winning position is worth floats with the platform’s total accrued yield and total outstanding shares, coupling otherwise-unrelated pools — a sharp departure from the per-market isolation of a prediction market, and a structural consequence of clocklessness rather than an implementation convenience (§4.6, §8.3). The two performance factors are the two halves of the logistic link, applied at two moments (§4):

- **dissent (logit, fixed at entry)** — how lopsided the crowd was against you. In point form $\text{dissent} = \log_3((1 + \bar{c})/(1 - \bar{c}))$ with $\bar{c}$ the opposite-side share at entry: 0 for a pure conformist, 1 at a 50/50 split, $\rightarrow \infty$ against a near-unanimous wall. The implementation integrates this logit over the position’s *own* contributed mass (§4.3), which is what makes dissent Sybil-resistant (§6).
- **consent (sigmoid, read at crystallization)** — whether the crowd ended up agreeing with you. With q your side’s share at settlement and θ the steepness, $\text{consent} = \sigma((q - \frac{1}{2})/\theta) \in (0, 1)$: $\frac{1}{2}$ while the question is undecided, saturating toward 1 once your side has clearly won and toward 0 once it has clearly lost.

Because payout is pro-rata over shares, the mechanism redistributes the **yield**, not principal: a bold, correct, long, large position captures a large slice; a meek or wrong position captures little. Principal is returned regardless (the “no-loss” property, precise rather than absolute — §1).

3.6 Yield source

The redistributed yield is *exogenous*: principal is supplied to an external lending venue that pays a base rate, and the mechanism reallocates that yield by foresight (§3.5) rather than issuing any itself — the venue is a swappable dependency.

3.7 The three tunable parameters

Everything above is derived from just three knobs. Naming what each one *controls* — as distinct from how it is *shaped* (§4) — is the compact way to hold the whole design in mind:

Parameter	Controls
lock-duration curve (α, β)	the speed of the beauty contest : longer horizons track truth better (evidence is given time to become focal, §8.1–§8.2), shorter ones serve participants — the truth $\leftrightarrow$ usability trade-off
decay half-life (h)	the assumed knowledge gradient between newcomers and incumbents — how fast each new cohort is presumed to out-know the last (§4.1)
consent steepness (θ)	the decisiveness of the reward — how sharply consent separates a settled winner from an undecided split (§4.4); the same dial sets reward volatility near 50/50 (§6.3), and is bounded above rather than merely traded off — past a hard ceiling the mechanism rewards opposition at every cursor and no consensus can form (§8.4)

4 Design rationale

Every part of the mechanism is shaped for a reason; two choices carry the most weight and are the least self-evident. The first is the **two-tier dynamics** — why locked mass decays while crystallized mass does not (§4.1). The second is the **reward’s two transforms** — not arbitrary, but the two halves of the logistic link, each the natural choice for the quantity it scores (§4.2–§4.5).

4.1 Why decay: arbitration windows

The rationale for the two-tier dynamics — locked mass decays, crystallized mass does not (§3.2) — is that decay tracks *outstanding judgment*. A later entrant holds an information set at least as large as any earlier one: whatever evidence has surfaced since a position opened is theirs, and was not the holder’s. So while positions remain locked — their consent unread, their reward still to be decided — the field deliberately weights conviction by recency and keeps itself cheap to move (§3.3): each cohort of entrants is an open **arbitration window** on the cohorts before

it, and is itself at the mercy of the next, for as long as its own positions remain uncrystallized. What has fully settled, by contrast, no longer decays, because nothing hangs on it — no one is left to arbitrate — so the record may legitimately be hard to move: from then on, new evidence must out-stake the frozen floor rather than out-wait it (§8.4).

The half-life h sets the *rate* of this weighting: it is the design’s assumption about **how fast each new cohort out-knows the last** — a short half-life presumes newcomers hold decisively more evidence than those already in (so old conviction goes stale fast), a long one that they barely do (old conviction stays valid longer). It is a *rate*, not a shelf life: a position’s weight fades continuously from the moment it opens until it crystallizes. And h is only *half* of the design’s evidence-tempo assumption: it prices how fast conviction goes stale, while the **lock-duration curve** (§3.4) sets *how long* the arbitration window stays open before it freezes — and it is there, not in h , that the settlement horizon is matched, or mismatched, to a claim’s evidence-speed (§8.2). Both are assumptions about *access* to evidence, not its use — a later crowd can read a larger record worse; that is the §8.1 bet in another form.

4.2 The pathology raw ratios have

Price the four outcomes a position can have — entered in the **minority** or **majority**, resolved into the **majority** (won) or **minority** (lost) — and compare the transformed reward against naive raw ratios ($\text{dissent}_{\text{raw}} = \text{opposite-share}$, $\text{consent}_{\text{raw}} = \text{same/total}$). Normalizing a pool to total = 1, entry split 20/80, $\theta = 0.1$, and holding amount $\times$ duration fixed so only dissent $\times$ consent varies, the meaningful column is each case’s **share of the pot** (payout is pro-rata by shares):

Case (entry $\rightarrow$ resolution)	raw $d \times c$	raw share	logit/sigmoid $d \times c$	combo share
Minority $\rightarrow$ wins (bold, right)	0.64	64%	1.90	80%
Majority $\rightarrow$ wins (meek, right)	0.16	16%	0.35	15%
Minority $\rightarrow$ loses (bold, wrong)	0.16	16%	0.095	4%
Majority $\rightarrow$ loses (meek, wrong)	0.04	4%	0.018	0.7%

The raw column has a fatal degeneracy: the **wrong-but-bold** loser (16%) earns *exactly as much as* the **right-but-meek** winner (16%) — raw ratios pay for boldness even when it was wrong. The logit/sigmoid combo splits them (4% vs 15%), concentrates the pot on the fore-sighted minority winner (64% $\rightarrow$ 80%), and starves the losers (under 5% combined). Sharpening that winner/loser gap — especially lifting the minority winner — is the entire job of the two transforms.

4.3 Why a logit for dissent

Dissent asks: as the crowd gets more lopsided, how fast should the reward for opposing it grow? The answer must be **unbounded** and **additive in evidence**, and the log-odds is the unique natural function with both. Evidence adds on the log-odds scale — each independent datum shifts it by a fixed amount, its log-likelihood ratio — Good’s “weight of evidence” (Good 1950), with Turing’s deciban as its unit (Good 1979). A 99:1 split embodies *exponentially* more apparent evidence than 9:1, so standing against it must pay commensurately more — $\rightarrow \infty$ as the wall nears unanimity; a linear measure badly under-pays the rare against-the-wall

contrarian, exactly the signal the system most needs. Base 3 fixes the neutral baseline: at 50/50, $\text{dissent} = \log_3(1.5/0.5) = 1$. And rewarding in log-odds belongs to the family of *truthful* (proper) forecast rewards, alongside the logarithmic scoring rule (Good 1952; Gneiting and Raftery 2007). The implementation integrates this logit over a position’s *own* mass, scoring boldness net of its own price impact — which is what makes dissent Sybil-resistant (§6).

4.4 Why a sigmoid for consent

Correctness wants the **opposite** shape: it must **saturate**, not explode, and must **not pay out while the crowd is undecided**. A sigmoid does exactly three wanted things: (i) it compresses the ambiguous middle — $\approx \frac{1}{2}$ for everyone near 50/50, so the system refuses to strongly reward or punish while the question is genuinely open, where a step function would pay maximally; (ii) it saturates at the extremes — once you are clearly on the winning side, a bigger blowout carries no extra information about *your* judgment, so it pays no extra; and (iii) it exposes **consent_steepness** as a **temperature** trading decisiveness against reward volatility near the boundary. This is the *output/gate* use of a sigmoid — a soft classifier read-out, exactly what consent is.

4.5 The asymmetry, and the annealing picture

The elegant core is the asymmetry: **unbounded on the input you dared (logit), bounded on the output of whether you were right (sigmoid)**. Rare boldness is worth exponentially more; being right is a saturating, near-binary fact. Swap the two and the mechanism is incoherent. This makes a pool a **logistic unit**: the belief cursor $c \in (0, 1)$ is its activation (a probability), *dissent* reads the pre-activation (the log-odds behind it) at entry, and *consent* re-applies the activation at exit — logit = natural parameter, sigmoid = mean, the two ends of a generalized-linear-model link (McCullagh and Nelder 1989).

The dynamics complete the picture: **decay is temperature, crystallization is annealing**. While locked, mass decays and belief stays hot and movable; as mass crystallizes the unit cools and commits to a basin — simulated annealing into a ground state. The Glauber update of the Ising model is *literally* a sigmoid with temperature T , and **consent_steepness** plays that T (Glauber 1963; Castellano et al. 2009). Finally, the reward is steepest near 50/50, where consent’s gradient is largest, so small moves in belief swing everyone’s payout. That concentration is deliberate: the marginal reward for tipping belief across the $\frac{1}{2}$ line is largest exactly where the question is most contested, which draws capital to resolve it. It is also self-balancing — the consent bandwagon on the fat side is met by higher dissent offered to contrarians on the thin side — with one honest cost: that same near-50/50 region — where a small nudge swings the most consent — is also the cheapest place to manipulate (§6.3).

4.6 Why one pot

The reward’s *shape* is settled above; its *funding* is a separate question with a separate answer. A prediction market partitions the **money** — each market its own matched book — and need not partition belief, because an oracle settles it. Clockless consensus has no oracle, so belief must be scored per question; but nothing then requires the *capital* to be partitioned in order to do it. One fund earning one base rate, many scoring lenses deciding how the return splits by demonstrated foresight. **Money is fungible, belief is not — so pool the money and partition the**

belief. Two steps get us there, and they are worth keeping apart because different things force them.

First, **clocklessness forces the pot to be exogenous.** With no common settlement event there is no moment at which counterparties could be paired and settled against one another, so a payout cannot come from a matched loser — it must come from outside the game. Non-zero-sum is *entailed* by the design rather than chosen alongside it.

That also answers the natural suspicion that the mechanism is a subsidy scheme in disguise. A prediction-market position is a claim on another participant: every YES share was minted as half of a YES+NO pair against collateral, so total out equals total in identically, and fees make it negative-sum. The conservation is a balance-sheet fact, not a pricing one — LMSR has no supply-and-demand price formation at all and is still not zero-sum, but *negative*-sum for the sponsor, who funds a bounded subsidy precisely so participants can collectively win. A position here is a claim on no one: minted against its holder’s own principal, burned against an external pot. Prediction markets are a **trading venue**, where positions transfer between holders; this is a **minting venue**, where they are created and destroyed. No transfer, no conservation law. Conservation has not disappeared so much as moved up one level — total payout stays hard-capped by yield actually accrued, so *lost = won* survives platform-wide against the exogenous inflow instead of per-market against principal.

Second, **exogeneity plus per-pool isolation forces globality.** Yield accrues continuously and steadily while claims are lumpy and reversal-driven, and an isolated pot has nothing keeping the two in balance: it drains or it jackpots (§8.3). Pooling averages that lumpiness away — a plain risk-pooling argument, with the share price as the net asset value that turns lumpy claims into a continuous one. The separation matters: clocklessness forces exogeneity, and it is *isolation*, not clocklessness, that forces globality.

Three properties follow that were not designed for. The first is that **your benchmark is the platform, not your counterparty.** Shares are $a \cdot D \cdot \text{dissent} \cdot \text{consent}$, and $a \cdot D$ is capital $\times$ time — exactly what yield is proportional to — so the share price sits near $r / \langle \text{dissent} \cdot \text{consent} \rangle$ platform-wide, and a payout approximates the base yield one’s own capital-time generated, multiplied by one’s foresight *relative to the average participant anywhere on the platform*. (Approximate rather than an identity: idle capital mints no shares, so it only pushes the price up.) Being right in your own pool is not sufficient — you must beat the index. That is the fair-game structure of §5 obtained a second time, in the reward layer rather than the timing layer.

The second is that **idle and losing capital fund the epistemic work.** Yield accrues on all principal while only settled positions mint shares, so parked and wrong capital pays the participants doing the work. This is what makes the no-loss property structurally honest rather than a marketing claim: under a per-pool pot the natural source of funding for winners is the losers’ principal — which is a prediction market again. Here nobody’s gain is any specific person’s loss.

The third is an **automatic thin-market subsidy.** The classic prediction-market failure is the neglected question: no liquidity, so no incentive to inform, so it pays nothing — for which Hanson’s remedy is a *sponsored* market maker (Hanson 2003, 2007). Global pooling is a self-funded version of the same idea, since a small, neglected pool redeems at the **global** share price, leaving a real return available for being right there. The bounty on error is platform-wide for the

same reason: capital flows wherever dissent $\times$ consent is highest, so a lopsided false consensus anywhere competes for capital against every other pool, and §8.4’s low-hanging fruit is picked from a shared orchard.

The obvious objection to coupling — that connecting pools contaminates their readings — is answered by confining the coupling strictly to *magnitude*. The cursor is a pure function of that pool’s own stake (§3); only the price at which shares redeem is shared. **Money global, belief strictly local.** One consequence looks at first like a degeneracy and is not: two *opposite* positions can both be paid. They are not opposite bets on one proposition at one price. One entered against a 20/80 crowd; the other entered later, after that side had won the argument, against 70/30 the other way. Each booked genuine dissent against the crowd at its own moment, and each had consent read at its own settlement — scored over different intervals, against different crowd states, never against each other. The mechanism does not ask *is it true*; it asks *were you ahead of the crowd*, and two participants can be ahead in opposite directions at different times if the crowd genuinely moved.

5 The core contribution: the unknown-future Schelling point

The gap of §1 in one line: measuring belief without a resolution anchor supplies no force pulling it toward truth, and a prediction market restores one only by importing exactly the anchor — oracle plus settlement date — that unanchored questions lack. Clockless consensus targets that gap. Its move is not “ban selling” but **removing the common clock**: each position settles at a time set by pool *activity*, not by the holder. You cannot time your exit, so you cannot ride a fad you know is a fad — the only prediction with high expected payoff *across all the times you might be measured* is one aligned with a **stable, time-invariant attractor**. For a well-posed, evidence-bearing question (§1), that attractor is the truth.

Keynes’s beauty-contest recursion only converges because everyone evaluates at one shared instant; take the shared instant away and there is nothing to coordinate around *except* the answer that stays put. The unknown horizon is therefore a filter that selects for time-invariant predictions — and *dissent* pays *more* for reaching that attractor before the crowd does. That is “foresight” made precise: identify the eventual stable consensus before it forms. Because dissent is scored net of a position’s own price impact (§4.3), a high-dissent position is by construction too small to move the tide on its own — so booking high dissent is itself a bet that others will come around before it settles, which makes foresight a *priced risk* rather than a free contrarian bonus.

5.1 Unpredictable is not arbitrary (the fair-game constraint)

The argument above needs only an *unknown* horizon — and pure randomness would supply one for free: settle each position at an independent random draw, and no one can time an exit either. The design rejects it, because unpredictability is only half of what settlement must satisfy; the other half is that the timing law must be **common, public, and stable**. Every position faces the same lock-duration curve (§3.4), a deterministic function of observable pool state, so differences in horizon trace to differences in the activity around a position — never to a draw.

A random clock fails on both counts at once. It is unfair in the plain sense: identical actions

receive arbitrarily different treatment, one position measured in days and its twin in years. And it stops selecting for anything: when the mapping from action to measurement date is noise, foresight cannot be priced, and the reward pays luck. A fixed date is lawful but predictable — the beauty contest returns; a random draw is unpredictable but lawless — the game degenerates into a lottery. Activity-timing is the narrow middle: the *rule* is learnable, the *instance* is not, because the rule’s output depends on the crowd’s future activity, which no one — the holder included — controls. Settlement is unpredictable for the right reason, a genuinely unknown future, not by injected noise.⁵

The same constraint is why *stability* of the timing parameters matters as much as their optimality: a law retuned too often is, from the participant’s seat, indistinguishable from a random one, and the same degeneration toward lottery follows (§8.2, §8.5).

5.2 The reflexive fixed point (the learning-curve argument)

The engine runs on the *gap* between present belief and the eventual settling point — that gap is what *dissent* (at entry) $\times$ *consent* (at exit) monetizes. As participants learn that the game is “predict where belief will settle,” they get better at it, the gap narrows, and belief jumps faster to its settling point; the foresight premium compresses toward zero. This splits into two regimes:

- If the settling point is **truth**, the system becomes a near-instant truth aggregator.
- If the settling point is an **arbitrary focal / self-fulfilling point**, the same learning curve merely makes it a *more efficient beauty contest* — faster bubbles.

What decides which regime you are in is **the clockless property itself**. Even a sophisticated player who can predict where the crowd will settle cannot *exploit* a fad settling-point, because they cannot time their exit to it — the settlement is activity-timed and unknown. So the only settling point worth coordinating on, across all possible exit times, is the time-invariant one. **The learning curve amplifies whatever the mechanism selects for; the unknown horizon is what makes “what it selects for” equal truth rather than a focal fad.** Rising participant sophistication therefore *accelerates* convergence and channels that acceleration toward the evidential answer — benign precisely because the horizon is unknown.

5.3 What learning cannot reach (the horizon)

Sophistication has a second target. Besides *where* belief will settle, a participant can try to learn *when* they will be measured: the lock-duration curve is public (§3.4), so the curve plus a forecast of pool activity yields an estimate of one’s own settlement date — and the regime argument above leaned on that date being unknowable. Three structural facts keep the estimate from maturing into an exit clock.

First, the horizon is not a hidden constant that better inference uncovers: it is *generated* by the crowd’s future behavior, including the response to one’s own entry, which itself adds hotness and moves every date in the pool — the estimator’s included. Forecasting one’s settlement is not an exit from the game but another bet inside it, on the same crowd. Second, the error is

⁵The constraint rules out a *lawless* clock, not stochasticity as such: a bounded, identically distributed jitter on top of the lock-duration curve would keep the timing law common — and learnable in distribution — while further blurring instance prediction. A design dial, not a violation.

one-sided: activity only *extends* settlement — at entry a holder sees a lower bound, never the instant (§3.4) — and a fad attracts exactly the attention that pushes its riders’ measurements later, into the future where the fad dies. A fad digs its own arbitration window; riding one safely would require it to stay quiet, a contradiction in terms. Third, horizon predictability anti-correlates with exploitability: dates firm up where activity is stable — quiet, settled pools with no fad to ride — while in the hot, contested windows where timing would pay, the date is at its most volatile. The one place fine-grained date knowledge does emerge — late in a lock, once the geometric extension has converged (§6.2) — arrives long after the holder chose a side, and degenerates into the already-bounded sniping problem of §6.3, not into fad-riding.

The residual is named rather than argued away: a pool with genuinely forecastable activity *and* a live fad would thin this defense; the three facts above say why the combination should be rare, not that it is impossible.

5.4 The beauty contest is not forbidden — it is farmed

Suppose collective learning drifts toward “guess what the crowd will guess” anyway. The drift is not detected and punished; it is *priced*, by the standing asymmetry of the reward (§4.5). A crowd-follower enters on the majority side, which books little dissent, and consent saturates — the upside of mimicry is capped by construction. Meanwhile every position the herd adds makes the split more lopsided, raising the logit bounty offered to whoever stands against it: a beauty-contest excursion posts a growing reward for its own refutation and — the pot being the redistributed yield on all locked capital (§3.5), allocated pro-rata by shares (§4.2) — sets its own capital to work financing whoever collects. And the herd cannot ride and exit, because its horizon is the unknowable one above. Herding is therefore **negative-frequency-dependent**: the more common it becomes, the larger the excursions and the richer the dissent offered to defectors, until defecting to evidence dominates and the herd thins.

It does not thin to zero — that is the Grossman–Stiglitz equilibrium (Grossman and Stiglitz 1980): a mix survives, with just enough mispricing left to finance the participants who correct it. The same subsidy logic the mechanism turns against a deliberate manipulator (§6.4) (Hanson and Oprea 2009) applies to the honest herd — an uninformed force whose pressure raises the return to being informed. One consequence deserves stating plainly: the system has no initialization from which mimicry is safe — the deeper a beauty-contest excursion, the stronger the restoring gradient — so it cannot be path-locked into farming fads alone. The bounty is collected, however, only where evidence eventually arrives — the honest bound below.

5.5 Honest bound

This benign reading — the learning curve amplifying truth-selection, the horizon out of learning’s reach, the farmed contest — is conditional on the same premises as the core claim: that evidence eventually becomes focal *within realized horizons*, and that the unknown horizon genuinely prevents fad-timing. For a *persistent* shared delusion (evidence never becomes focal) or for truths slower than the lock distribution (the **right-too-soon** trap, §8.2), the learning curve could instead accelerate convergence to a wrong-but-stable point. So clockless consensus converts a *fast* beauty contest into a *slow* one and bets that, given enough uncertain time, belief drifts toward evidence — a better wager on truth than either an unanchored measure of belief or an oracle-dependent prediction market, but a wager, not a proof.

6 Game resistance

What “game resistance” can and cannot mean. We do not claim the mechanism is ungameable: by Gibbard–Satterthwaite (Gibbard 1973; Satterthwaite 1975), no non-dictatorial mechanism with three or more outcomes is dominant-strategy-proof — and a continuum of belief levels far exceeds three, so ours is no exception. The defensible claim is narrower and honest — the two capital-based attacks the design is specifically shaped against are *self-defeating*, and the places where it stays soft are bounded and named. “Economically irrational” is, moreover, an *on-platform* accounting: an attacker paid outside the game is bounded differently, and treated separately (§6.4). Two structural properties carry the argument for the capital attacks: the **self-impact integral** in dissent neutralizes identity-splitting, and **hotness decay** bounds any capital war to a finite horizon. §6.1–§6.3 are claims about the *formulas*, and so — unlike the epistemic claim of §7 — both provable and fully simulable; the first two are stated and proved below as Propositions 1 and 2. §6.4 ends at the one point where the analysis must touch audience behavior, and marks it.

6.1 Sybil fragmentation is neutral by construction

The obvious first attack is to split one stake across many identities to book more reward. Dissent forecloses it directly, because it is not the pointwise log-odds at entry but the **integral of the log-odds over a position’s own mass**, credited net of the position’s own price impact (§4.3): each successive unit of a stake faces a pool its earlier units already pushed, so it earns less. That integral is *invariant* under how the stake is partitioned, which we state precisely.

Proposition 1 (partition-invariance of dissent). Let a position entering on one side with cumulative own-side mass x_0 and size a , against opposing mass k , be credited

$$D(x_0, a; k) = \int_{x_0}^{x_0+a} \phi(u; k) du,$$

where u is cumulative own-side mass and ϕ is the pointwise log-odds of §4.3. Let $a = a_1 + \dots + a_n$ be any partition into consecutive same-side sub-entries placed against unchanged k . Then $\sum_{m=1}^n D(x_0 + A_{m-1}, a_m; k) = D(x_0, a; k)$, where $A_m = \sum_{j \leq m} a_j$ and $A_0 = 0$.

Proof. The m -th sub-entry is credited $\int_{x_0 + A_{m-1}}^{x_0 + A_m} \phi(u; k) du$, since its own-side mass runs over exactly that interval. The intervals are consecutive and cover $[x_0, x_0 + a]$ without overlap, so by additivity of the integral the sum telescopes to $\int_{x_0}^{x_0+a} \phi(u; k) du$. ■

Two features of the statement carry the security argument. First, ϕ is a function of *pool state* alone and carries no identity argument, so the proposition holds verbatim when the n sub-entries are held by n distinct identities: Sybil fragmentation is not merely unprofitable but exactly neutral. Second, the result does not depend on the particular form of ϕ — any pointwise score integrated over own contributed mass is partition-invariant — so the property survives retuning of the dissent curve. The hypotheses mark the boundary honestly, and “unchanged k ” is a stronger condition than it looks: it means *the same instant*. Exactness fails not only when opposing mass arrives between sub-entries but under any time gap at all, since decay shrinks the standing mass M (§3.2), so a later sub-entry of the same size faces a smaller pool, self-dilutes

more, and books *less*. The direction favors the mechanism — splitting across time earns strictly less than splitting at once, so the gap opens no attack — and an attacker who splits and waits is additionally exposed to the pool moving against him in the interval, which is a risk, not a discount. The self-impact integral is precisely the reason for all of this — a genuine strength, which we state as one.

6.2 Capital wars converge to a finite horizon

The second attack is the whale *doubling-down*: two opposed large holders repeatedly inject capital near their positions' settlements, each trying to own the agreed state at the instant it pays and to drag opponents into ever-longer locks. It self-limits, because a position's settlement time is a function of **hotness** — a decay-weighted sum of nearby stake — and a later injection raises an earlier position's hotness only by $a e^{-\lambda \Delta t}$ (the reverse-direction pairwise weight $\text{decay}_{\text{prev}}/\text{decay}_{\text{new}}$, with $\text{decay}(t) \propto e^{\lambda t}$ increasing, so the ratio is < 1). Successive rounds land later, Δt grows, and the contributions form a geometric series with ratio below one, so *repetition* buys a bounded amount of delay however long it is sustained.

Proposition 2 (bounded settlement under unbounded repetition). Fix position i created at s_i , with hotness and duration as in §3.4. Suppose the contributions arriving after s_i have size at most $\bar{a}$ and inter-arrival times at least $\delta > 0$. Then, for any number of injections,

$$H_i \leq a_i + M(s_i) + \bar{a} \frac{e^{-\lambda \delta}}{1 - e^{-\lambda \delta}} < \infty,$$

and hence the settlement date $\sigma_i = s_i + \alpha H_i^\gamma$ is finite.

Proof. Index the arrivals after s_i by $n = 1, 2, \dots$; the separation hypothesis gives $s_n - s_i \geq n\delta$. Since $a_n \leq \bar{a}$ and $e^{-\lambda(\cdot)}$ is decreasing, $\sum_{n \geq 1} e^{-\lambda(s_n - s_i)} a_n \leq \bar{a} \sum_{n \geq 1} e^{-\lambda n \delta} = \bar{a} e^{-\lambda \delta} / (1 - e^{-\lambda \delta})$, a convergent geometric series. Substituting into H_i gives the bound, and $\sigma_i = s_i + \alpha H_i^\gamma$ is finite because H_i is. ■

The bound is worth reading twice, because the quantity it controls is self-referential: a longer lock admits more contributions into the sum, and those contributions lengthen the lock further. What makes the recursion terminate is that the geometric bound is *uniform in the length of the window* — it holds for a sum taken over unbounded time, and therefore over whichever window D_i ends up selecting. So the fixed point exists and is finite, and no amount of *repetition* moves it: an attacker facing the exponential discount can sustain the war indefinitely and still not push σ_i past the bound above.

What the proposition does not say deserves stating, since the bound is linear in the per-injection cap $\bar{a}$. An attacker who escalates the *size* of his injections rather than their number is not bounded by it: a single contribution of size A extends the horizon by roughly αA^γ , which grows without limit in A . The claim is therefore about compounding, not about capital — repetition does not accumulate, because each round is discounted by the delay it must wait through. Escalation still buys delay, but at a punishing rate: with $\gamma < 1$ the delay purchased grows *sublinearly* in stake, so each additional unit of delay costs strictly more than the last, and the attacker's capital is locked and decaying for the whole of it. The separation hypothesis $\delta > 0$ rules out the degenerate evasion of injecting infinitely often at a single instant, which is in any case just one large contribution and is priced as one.

The costs are symmetric — both sides lock more capital and face the same exponential discount, so neither gains an edge — and once a position **crystallizes** its settlement is frozen and can no longer be extended at all, so the extension window closes as crystallizing mass runs out. The war also erodes its own monument: each escalation lengthens the belligerents’ locks, and a position crystallizes at its decayed weight (§3.4), so the fiercer and longer the fight, the smaller the fraction of its capital that survives into the permanent record. Capital thus buys a bounded *delay*, never control of the settlement *instant*; and because consent is read only at that instant, no amount of mid-lock spending fixes the one quantity that pays.⁶

6.3 Consent-sniping is cheap at small pools, and bounded there

At a small pool, inertia scales with pool mass (§3.3), so flipping the split at a victim’s settlement instant costs little — and costs little *whatever the current belief*, since moving a small pool is cheap even against a lopsided majority. Two properties compound at small size: low mass makes the flip cheap, and low hotness makes the lock short (§3.4), so the window in which others can counter is itself compressed. The attack is bounded on both ends. It is economically irrational at large, slow pools — flip cost rises with mass while the reward does not — and it is defended *without* touching the reward semantics by two independent means: a holder can pre-commit an automatic counter through the protocol’s **resistance** (limit-order) primitive (§3.3), and ordering pools by *imminence of settlement* lets counter-parties watch and defend those about to crystallize, restoring the counter-window that smallness erodes. These bound the attack rather than prove it impossible.

The economic version of “right too soon” — a correct position that settles before belief converges and takes loser reward — is a *limitation*, not a manipulation, and we treat it in §8.

6.4 Buying the display: exogenous-payoff manipulation

The accounting above is in the platform’s own currency, and it does not bound an attacker whose payoff lies *outside the game*. A tobacco company that pushes a pool toward “harmless” is not buying reward; it is renting **display-time** — the cursor shown as apparent consensus while its product sells — and its on-platform loss is a marketing expense (the documented archetype of manufactured doubt (Oreskes and Conway 2010)). The threat scales with success: the more the display is trusted, the more it is worth renting — Goodhart’s law aimed at the mechanism’s own read-out (Goodhart 1975). What the mechanism can honestly do is not prevent the rental but *meter* it, and pay the meter to the correctors. Holding a false cursor means absorbing the arbitration flow continuously, and the wall itself is a honeypot: a lopsided false consensus offers maximal dissent to informed contrarians, with vindication odds high *because* it is false — the manipulation budget funds the bounty on its own refutation. (The corresponding prediction-market result is Hanson and Oprea’s: a manipulator is a subsidy, and can *improve* accuracy (Hanson and Oprea 2009).)

The contrast with a tradeable venue is the whole of it: there, renting the price is cheap precisely because the renter can close the position before resolution, so the display is bought and the bill is never presented. The clockless machinery compounds the bill instead: the attacker cannot time his exit off the display window; his mass heats the pool, lengthening his own lock through precisely the period in which the fraud unravels; consent read at settlement then pays him

⁶Full geometric-series accounting is specified in the Towerview design documentation.

nothing; the counterweight melts his frozen trace (§3.4); and his locked principal earns yield for the common pot the whole time it waits.

Detection needs no oracle, because manipulation has a *shape*. A cursor propped by fresh, one-sided, still-locked mass is visibly on trial — none of it has survived an arbitration window (§3.2) — while crystallized mass *is* survived judgment. That puts the attacker in a pincer: mass modest enough to look normal is cheap for the honest side to flip back (inertia scales with mass, §3.3); mass overwhelming enough to hold the line is anomalous.

The only escape is to attack the statistics themselves — inflate enough pools that fabricated mass looks ordinary. This is the **flooding** attack — the “firehose of falsehood” (Paul and Matthews 2016), or censorship by noise (Roberts 2018): drown a true signal under manufactured ones until the anomalous shape *is* the norm and the target pool flies under it, chaff against radar. But that requirement is itself the security bound, in the same sense as a 51% attack (Nakamoto 2008): defeating the baseline requires a platform-scale stake. And where flooding is nearly free offline, here the mechanism *prices* it: the camouflage capital is transferred rather than burned — stronger than proof-of-work’s, since it earns yield for everyone, opens honeypots wherever it sits, and melts in the counterweight. Creating pools to dilute the baseline is governance-gated (§3), but that gate is only as strong as the cost of capturing governance — plausibly *below* platform-scale stake, since a protocol’s token tends to be valued well under the capital it secures (§8.5). So the binding constraint is not the gate but the **mass**: minting empty pools buys nothing without the platform-scale mass to fill them — the container is cheap, the camouflage is not.

Flooding has a **second target**, however, and that bound does not reach it. Beyond the statistics, minted pools can be aimed at the **share price** every other position redeems at (§3.5). The attacker’s capital does accrue yield into the same pot, so this pays only insofar as he mints shares per unit of capital-time above the platform average — but an uncontested pool is exactly where he can arrange that, building a wall with one position and booking dissent against it with a second. Proposition 1 does not forbid this: it makes *splitting* a stake neutral, not *staging* two on opposite sides. Here the cheapness of the container is the whole attack rather than its bound, and the only defence is the governance gate on pool creation — which §8.5 concedes is plausibly the cheapest surface in the system. We name this one unresolved.

The reading of the display follows: the cursor is one axis of three. The **cursor** says what the crowd believes; **mass** says how much scrutiny that belief has survived — a 50/50 split over trivial mass means *no one has looked*, over deep mass it means *genuinely unresolved*; the **settled share** X/M , with the arrival shape, says how much of the belief is judgment rather than fresh push.

The equilibrium is the honest close. Trust in the display and manipulation of it form a negative-feedback loop — the more credible the cursor, the more it is worth renting; the more visibly it is rented, the less credible it is — so *perfect* trust is not an equilibrium state, for the same reason perfectly informative prices are not (Grossman and Stiglitz 1980). The design’s aim is an interior equilibrium that sits high: raise the price of display-time, keep manipulation legible through composition, and let the attacker’s budget flow to his refuters. The residual is stated plainly: the rented window is real while it lasts — cheapest exactly where §6.3 is soft. And the mechanism computes and exposes mass and settled share no less than the cursor, but cannot *compel* an audience to weigh them alongside it: surfacing the three co-equally is a matter of presentation, and attending to them, finally, of audience behavior. That last point is a claim about audiences, not formulas; it sits with the epistemic wager (§7), not the engineering.

Net. The design makes its two headline capital attacks self-defeating (splitting is neutral, capital wars are horizon-bounded), meters the attack it cannot prevent (display rental is priced per unit of time and funds its own correction, §6.4), stays soft only in bounded, named regions (small, fast-settling pools; the rented window while it lasts), and — per Gibbard–Satterthwaite — is never provably ungameable. These are engineering claims about the formulas — demonstrable, adversarially checkable, simulable, with §6.4’s audience caveat marked where it arises — in contrast to the epistemic claim that §7 shows cannot be settled that way.

7 The validation boundary

A reader will ask where the experiments are. The honest and more interesting answer is to draw the exact line between what can be demonstrated and what cannot — and to show that the line is structural, not a matter of effort.

7.1 Two different claims, only one is simulable

- The **engineering claim** — “the mechanism does what the spec says, even adversarially” (Sybil splitting leaves total dissent exactly unchanged, Proposition 1; whale doubling-down converges as a geometric series to a finite settlement date, Proposition 2; consent-sniping costs a quantifiable amount at a given pool size) — is a claim about the *formulas*. It is provable, and fully simulable almost mechanically. (See §6.)
- The **epistemic claim** — “consensus tracks truth, and faster than the world otherwise would” — is **not** simulable, for structural reasons.

7.2 Why the epistemic claim resists simulation

1. **No oracle-independent ground truth on the target domain.** To validate “reveals truth faster,” one needs a truth-signal to compare against. But the target domain is *defined* as unanchored (§1): no clean oracle, no fixed resolution date. The cases that have hard ground truth are exactly the cases that do not need this mechanism — a dated prediction market with an oracle dominates there. So the platform can never be validated against truth *on the domain it was built for*, because the existence of such an oracle would obviate the platform. This is a definitional boundary, not a flaw.
2. **The validation target is itself a consensus product — Wittgenstein’s Ruler.** Taleb’s formulation (Taleb 2001): “unless you have confidence in the ruler’s reliability, when you measure the table you may be using the table to measure the ruler.” Checking whether the platform “got it right” means checking against *what was later judged true* — and on these questions, that later judgment is itself a crowd verdict. The instrument and the object are the same kind of thing; one ends up measuring the validation-ruler with the platform-table.
3. **A simulation of a reflexive mechanism returns its own assumptions.** Running one requires assuming (i) how information diffuses into beliefs and (ii) how agents strategize. The output is then determined by those assumptions on the dimension that matters: “if agents are truth-seekers, truth emerges” is circular; “if momentum-chasers, fads emerge” is also just the input. Worse, the relevant behavioral parameters are **endogenous and performative** — participant sophistication co-evolves with adoption and with beliefs about

others’ sophistication (the learning curve of §5). This is the **Lucas critique** (Lucas 1976) (you cannot predict the effect of a mechanism using behavioral relations the mechanism itself changes) and economic **performativity** (MacKenzie: models can make themselves more true as they are adopted (MacKenzie 2006); Soros’s *reflexivity* is the general loop (Soros 1987)). No static agent model contains the learning curve, so a simulation cannot help but assume the outcome.

7.3 The boundary reaches further down than the claim

The three arguments above concern the *aggregate* claim, but the same structure applies one level lower, to a single position. Scoring a position means reading consent against the crowd at its settlement instant, and on an unanchored question that instant has no privileged status: belief keeps moving afterward, and no later state is final enough to adjudicate the earlier reading (§8.2). So the mechanism cannot in general certify its own outcomes case by case — it cannot show that a position it paid was right, nor that a position it scored as wrong was right too soon. This is worth stating separately because it disposes of the natural repair. Faced with the aggregate boundary, a reader may reasonably propose validating the mechanism *internally* — audit its own history, find the positions it mis-scored, and measure the rate. That audit cannot be run across the domain, because the ledger records what belief did, not what was true, and on an unanchored question nothing external converts the first into the second. The exception is precisely the subset treated below: where a question does eventually resolve, the receipt exists and the audit is available for those pools — which is what makes the partial validation of the next paragraph possible, and equally what makes it survivorship-biased, since resolution is exactly the property selecting the sample. Outside that subset the mechanism’s honesty about its own record is structural rather than admirable: it has no record of the relevant kind to be honest about.

7.4 What simulation can honestly do: falsify, never confirm

Its legitimate role is to *find a plausible behavioral model under which the mechanism breaks* — gets stuck, entrenches an early error, makes near-50/50 sniping cheap. That is valuable, but it only ever delivers bad news; it can never deliver “and therefore it finds truth.” (A Popperian asymmetry: open to refutation, closed to confirmation.)

Those three, however, are §6’s engineering claims in behavioral dress, and the section still owes a falsifier for the claim it is actually about. Simulation supplies one, and the circularity of §7.2 is what makes it available rather than what forecloses it. Populate a pool with agents each holding a noisy private signal of a fixed latent proposition and no motive beyond the reward — the assumption set most favorable to the mechanism, and exactly the one whose *confirming* reading (“if truth-seekers, then truth emerges”) is circular. The contrapositive is not circular at all. If under those conditions the cursor fails to converge on the latent proposition, or converges later than a simple average of the same signals would, then the mechanism is broken where it should be strongest and no argument about real participants repairs it. **That is the epistemic falsifier, and we state it as one.** The asymmetry runs one way only: the mirror experiment — a population of pure momentum-chasers — is uninformative in both directions, since success would be the confirmation just disclaimed and failure is the farmed beauty contest §5.4 already concedes.

7.5 The only real validation is post-deployment and partial

Over the subset of pools that *do* eventually resolve, one can check ex-post whether clockless consensus moved toward the resolved answer *earlier* than baseline sentiment measures (polls, media). This is survivorship-biased, partial, longitudinal field data — not a simulation — and it is future work. The nearest existing dataset — Tetlock’s **Good Judgment Project** (Tetlock and Gardner 2015) — is dated-forecast data, and so a clarifying foil rather than a validation set (§2.6).

Because the paper rests on this as the only real validation available, the protocol deserves specifying rather than gesturing at. **Sample:** pools whose statement acquires a determinate answer after opening, by an external criterion fixed before the resolution occurs — the first of the two question-kinds §8.2 distinguishes. **Baseline:** a forecaster panel pre-registered on the identical statement set, since published polling rarely exists for unanchored questions and media stance-coding is noisy; polls and media serve as secondary comparators wherever they do exist. **Statistic:** for each pool and each series, the first time it crosses and thereafter remains beyond a threshold in the direction of the answer eventually determined — the same construct as the last crossing of §9, read from the other end — with the effect reported as the paired median difference between those crossing times, cursor minus baseline. Thresholds ($p = 0.5$, with 0.7 as a robustness check) and the minimum lead counted as non-null are fixed in advance; a median lead shorter than the baseline’s own measurement interval should be read as no effect.

Two choices are what would make this a *bounded* claim rather than a hopeful one. First, **the direction of the survivorship bias is known, which converts it from a caveat into a bound**. Resolution correlates with evidence eventually becoming focal — precisely the condition the epistemic claim needs — so the resolving subset is the mechanism’s best case, and any measured lead is an *upper* bound on the domain-wide effect rather than an estimate of it. Registering the full pool set at open rather than the resolved subset keeps resolution an observed outcome instead of a selection filter, and lets the unresolved complement be reported alongside: accuracy cannot be scored there, but dynamics can — crystallization rate, dissent profile, mass at settlement — and how far those match is what bounds the extrapolation from one to the other. Second, **pool creation is governance-gated (§3), so the gate must not be free to select favorably**. Statement set, admissibility criteria, and analysis plan are deposited externally before the pools open; and because the panel arm runs on identical statements, residual selection enters both arms and cancels in the paired comparison. Neither move makes the evidence conclusive. Both make its limits computable, which is the most this domain admits.

Net. The design *guarantees* its engineering properties and *wagers* its epistemic one. Articulating precisely why the wager cannot be settled by simulation is itself part of the contribution.

8 Limitations & honest claims

We collect the limitations in one place: the in-principle bet (§8.1), the design-level behaviors it produces — some accepted costs, some deliberate properties (§8.2–§8.4) — and the bound that governance places on its impartiality (§8.5).

8.1 The epistemic claim is a wager (in principle)

Argued in §5 and §7: the mechanism produces an *accountable consensus*, which equals truth only where evidence eventually becomes focal within realized horizons. In the language of the Condorcet Jury Theorem (Condorcet 1785), aggregation of this kind needs participants whose beliefs are *independent and sincere* and a majority that is *more often right*, and the theorem’s known failure modes are exactly correlated and insincere voters (Ladha 1992). No mechanism can *enforce* those conditions. What we claim is that this one **prices** two of them, which is a weaker thing than enforcement and a stronger thing than assumption. Sincerity is priced by the unknown horizon: positioning where you expect the crowd to land, rather than where you believe, books little dissent against a majority and a consent that saturates, and the herd cannot exit before it is measured (§5). Independence is priced — imperfectly — by dissent itself, which pays for the log-odds distance between a position and the pool it entered, so a belief copied from the crowd earns almost nothing while a divergent one earns a great deal (§4.3).

The imperfection is named in §2 and we do not paper over it here: dissent rewards decorrelation from the *consensus*, not pairwise independence of signals, so a correlated bloc of contrarians collects in full while an independent judgment that agrees with the majority collects nothing. Correlation among the correctors goes unpriced, and that is the Condorcet failure mode (Ladha 1992) surviving inside the design rather than outside it. This is precisely why the design declines to *assume* conditionally independent signals where the nearest prior art assumes them (§2): it attempts to pay for the condition instead of stipulating it.

The residual bet is therefore narrower than the Condorcet premises taken raw, and it has two parts. The first is that these prices actually move behavior — a claim about real participants, and one §7 shows cannot be settled by simulation. The second is irreducible: that evidence eventually becomes focal within realized horizons. Where a shared delusion never yields to evidence, no pricing helps, because there is no eventual signal for dissent to be vindicated against; clockless consensus then faithfully aggregates the delusion. That last clause is the central bet, and we state it as one.

8.2 The right-too-soon trap

The third row of §1’s taxonomy bundles two cases that the paper treats differently, and this is where the difference bites. In the first, the answer is certain to become obvious eventually and no one can say when — documents surface, forensic methods improve. In the second it may never become obvious at all, because what is contested is the interpretation of a record already complete. The right-too-soon trap bites hardest in the first; the shared-delusion bet (§8.1) bites in the second; and only the second leaves the mechanism’s errors permanently unverifiable, as the rest of this section shows.

The sharpest mechanism-level cost sits exactly in the target domain. A participant correct about a *slow-to-emerge* truth can have their position **settle before belief converges**, and so take *loser* reward for being right early. Crystallization softens this only epistemically — the position’s frozen weight keeps pulling belief toward the eventual answer after the holder has exited (§3.2) — but the *economic* loss to that holder stands.

The cost has a second property, less comfortable than the first: it is **unfalsifiable at the time**, and permanently so wherever the question never resolves. Speaking of a position that settled

“before belief converged” presumes a converged state to have settled before, and a pool of the second kind distinguished in §1 — where what is contested is the interpretation of a complete record — never reaches one. There is no *eventual* consensus in that case, only the consensus at each settlement instant; the cursor stays movable however much mass has crystallized, since a frozen floor raises the cost of moving belief without ever fixing it (§8.4). So when the tide turns after a holder exits, nothing establishes whether that turn was the evidence arriving or one more intermediate focal point — and the holder cannot demonstrate they were right too soon rather than simply wrong. Where the answer *is* certain to become obvious eventually — the first kind, awaiting documents or better methods — the verdict does arrive, and right-too-soon becomes visible in retrospect. That visibility is what the partial validation of §7 runs on, and it is available only for that subset.

This tightens the definition of foresight given in §5. What the mechanism rewards, strictly, is predicting the crowd at one unknown future moment; “identifying the eventual stable consensus” is the limiting description of that, and the limit is approached rather than reached. The property cuts both ways and we claim no comfort from it: the same unfalsifiability that spares the mechanism a public record of its failures also denies it any demonstration of its successes, position by position (§7).

The lock-duration curve is a single global parametrization (§3.4): it maps a pool’s activity to a settlement horizon but does not adapt to how slowly a given claim’s evidence is expected to arrive. Tuning that curve — or introducing tiers matched to expected evidence-speed — is an open design question, weighed against a real cost: per-pool variation in lock behavior erodes the fair-game constraint of §5 — the more the timing law varies, the harder its *rule* is to learn, and the further the game drifts from priced foresight toward lottery — so uniform lock dynamics may be the better default. We treat right-too-soon as a genuine, unresolved cost regardless.

8.3 Payoff magnitude is globally coupled

Shares are minted *locally* per pool, but the share *price* is *global* — total redistributed yield over all outstanding shares across every pool (§3.5). Being right in your pool fixes your *share* of the reward; the *magnitude* of that share is set globally, so what a correct position ultimately earns depends on platform-wide state — shares minted in unrelated pools dilute the price. We frame this as a property, not a discretionary cost, because the compartmentalized alternative — a self-contained yield pot per pool — appears unviable, and because globality is what buys the platform-wide benchmark and the thin-pool subsidy of §4.6.

Rewards concentrate where belief *reverses* — a contrarian entry’s dissent is realized only once consent arrives at settlement — while yield accrues steadily. An isolated pot has nothing keeping the two in balance, and it fails apart in two directions. A pool can be **drained**: a run of correct settlements empties it, and the pool’s consensus can then be wrong with nothing left to pay whoever would correct it. That is the worse mode, because it fails *silently* — the display shows a consensus, not an exhausted incentive behind it. Or a pool can **jackpot**: a run of wrong settlements leaves yield unclaimed, and the pool starts attracting capital for the lump rather than for the question, a farmable balance with a public address. The cost is therefore epistemic rather than distributive, and it is what makes §8.4 a property instead of a hope: under per-pool pots, “a revealed error is low-hanging fruit, not a fortress” would be conditional on that pool’s balance. A global price makes it unconditional. That price also smooths the lumpiness away —

a position’s payout tracks the global price at burn, independent of its own pool’s history. The honest residual is that a pool’s payoff *magnitude* is not self-contained.

8.4 Crystallization sets consensus — including, sometimes, a wrong one

Crystallization is designed to make belief *set* (§3.2) rather than drift forever; the natural worry is that it freezes a *wrong-but-stable* consensus. Two things make this largely a feature rather than a failure. First, a crystallized consensus *revealed* to be false is a low-hanging fruit, not a fortress: opposing a lopsided crystallized split earns high dissent by construction (§4.3), and as the error becomes obvious the risk of that contrarian position falls, since vindication grows likely — so correcting a revealed error is a well-incentivized, near-free reward that draws capital exactly when it is most deserved. (Precisely: opposing it does not *pay more* as the error becomes obvious — dissent is set by the split — it becomes more *attractive*, because being proven right grows likely.) Second, that a hard-won consensus requires real capital to shift — you must **out-stake the frozen floor, not out-wait it** — is a stability property, not a defect: the alternative, a consensus that decays forever, would force honest participants to perpetually re-stake merely to defend settled questions from sniping (§6.3). The genuine residual is narrow, and is the §8.1 bet in another form: the correction incentive only fires once the error *becomes* obvious; where evidence never becomes focal, crystallization holds the delusion in place.

The first of those two can be made precise, and doing so bounds it from both ends. Take the extreme case the argument above skips — a wall that never moves at all, so a contrarian’s consent is read at the same cursor c they entered against. For a position of given size entered at a given moment, stake and lock duration are common to both sides and cancel, leaving $\log_3 \frac{1+c}{1-c} \cdot (1 - \sigma(x))$ against a conformist’s $\log_3 \frac{2-c}{c} \cdot \sigma(x)$, where $x = (c - \frac{1}{2})/\theta$ and the two consents sum to one because the two shares do. Opposing the wall is the better position exactly when

$$\theta > \theta^\dagger(c) := \frac{c - \frac{1}{2}}{\ln R(c)}, \quad R(c) = \ln \frac{1+c}{1-c} / \ln \frac{2-c}{c},$$

and $\theta^\dagger$ falls monotonically in c from $3 \ln 3 / 16 \approx 0.206$. Two things follow. Any consent temperature at or above that value makes opposition the better position at *every* cursor, and a mechanism that pays more for opposition everywhere cannot let a consensus form at all — so θ carries a **ceiling**, not merely the trade-off named in §3.7. Below the ceiling there is a cursor $c^\ddagger$ — about 0.98 at the $\theta = 0.1$ used above — past which opposing the wall pays *whether or not it ever moves*. The claim of the preceding paragraph is therefore unconditional near unanimity and conditional below it: between $\frac{1}{2}$ and $c^\ddagger$, correcting a frozen error still requires the crowd to turn before the corrector’s own settlement. Sharper consent raises $c^\ddagger$ and narrows that unconditional band — the one place in the design where decisiveness (§4.4) and correctability pull against each other.

8.5 Impartial by construction, but only at the consensus level

Consensus is impartial by construction — a pure function of decayed and crystallized stake, with no privileged actor able to set it. The guarantee is bounded on both sides. Above the pool, *which questions are posed at all* is a governance act — pool creation is the system’s only privileged

action (§3). Below it, the guarantee does not extend to the reward *parameters* (the consent temperature, the decay rates) or to the code that computes them: these are set and maintained by whatever process governs the deployment, and changing a reward parameter can re-score positions already committed under the old one. This layer is also where the §6.4 capital bound does *not* reach: defeating the consensus statistics takes platform-scale stake, but capturing the governance that sets parameters and admits pools may cost far less — a protocol’s token is typically valued well below the capital it secures, so governance is plausibly the *cheapest* attack surface, not the safest.

So impartiality at the question, parameter, and code level is *inherited from the host framework and its governance* — who controls the code and how it can be altered — not guaranteed by the mechanism’s math. Standard mitigations narrow the gap (snapshotting reward-affecting parameters at the moment a position opens, and gating changes behind a timelock), but the specific decentralization guarantees depend on the deployment substrate and are out of scope here. The honest claim is bounded: impartial at the consensus level; at the question, parameter, and code level, only as impartial as its governance.

8.6 Which instrument for an anchored question?

Clockless consensus can run on a question that does have an anchor (§1), so the choice there is a trade-off rather than a rule — and it is worth stating, because the answer is usually not clockless consensus. Three things favor the prediction market wherever the anchor is crisp. It scores against the *actual outcome* rather than against a crowd, so it escapes the reflexivity and Wittgenstein’s-Ruler problems of §7 — to the extent that the ruling is accepted as final, which is a caveat rather than a formality, since oracles mis-resolve and a decentralized one may itself be a crowd (§2.2). Its capital recycles at resolution, where ours stays locked for an activity-determined term that may outlast the answer.

And the third is the decisive one, because it inverts what the question is *for*. A forecast question rewards being early: in a prediction market the holder can keep the position to resolution and collect in full, so a correct call made years ahead of the crowd is paid for being early — early detection is the product. Clockless consensus measures a position at a moment its holder does not choose, which turns the same early-and-correct call into a risk: settle before the crowd comes around and it is paid as though wrong, though the answer was always going to arrive on schedule. That is the right-too-soon trap (§8.2), and where the epistemic horizon is *known* it is not merely a cost but close to a category error — the mechanism penalizes exactly the behavior the question exists to elicit.

The decisive difference, though, is not the size of the cost but its **legibility**, and it is what makes “well-defined” (§1) fall short of “advisable.” On an unanchored question, right-too-soon is unfalsifiable: a position settles, belief shifts afterward, and no fact is ever available about whether that shift was the truth arriving or one more intermediate focal point (§8.2). Nobody can demonstrate that the holder was wronged. Resolve the same question by oracle and the receipt becomes public — this position was correct, the mechanism scored it as incorrect, on this date. The defect is identical; only now it is provable. Since what the mechanism offers is a *displayed* consensus whose value rests on being trusted (§6.4, §7), a display accumulating documented mis-scorings is worth materially less, and the cost is borne by the platform rather than by the unlucky holder. Each cleanly-resolving pool, in other words, volunteers a public

audit the mechanism did not ask for: the pools it scores correctly pass unremarked, and every one it gets wrong is charged against the credibility of the display, which is the product.

This gives the recommendation a useful shape: its force scales with the *quality* of the anchor, monotonically. Where resolution is crisp, the errors become provable and the argument against is at its strongest; where an oracle exists but its ruling is contested, no clean receipt is ever issued and the question was effectively unanchored to begin with — a question posed with a date and a resolver whose ruling then proves disputed is not a hybrid instrument but a misfiled member of the third row of §1’s table, and clockless consensus is the appropriate instrument for it after all. Dressing an unanchored question in an anchor yields resolution disputes rather than resolution. One tension survives the rule, and we name rather than resolve it: the manipulation advantage below does *not* weaken as anchor quality rises, so a cleanly-anchored question on which display integrity matters more than scoring accuracy remains a judgment call.

One consideration cuts the other way: manipulation. A prediction-market position can be closed, so renting the displayed probability is cheap — push the price, exit before resolution, and the bill is never presented. In clockless consensus there is no exit, the attacker’s own mass lengthens his lock through the period his push unravels, and consent is read at an instant he does not choose (§6.4). Where the *displayed belief* is itself the product, rather than a by-product of trading, that difference can outweigh the accuracy advantages above. Our own deployment declines anchored questions as a matter of positioning, so that the difference from a prediction market stays legible to users; that is a product decision, not a claim about what the mechanism can do.

9 Conclusion & future work

We have described **clockless consensus**, a mechanism for aggregating belief about unanchored empirical questions — a fact of the matter, but no oracle and no settlement date (§1). Its two moving parts are an **asynchronous, activity-timed, individually unpredictable settlement** — each position its own synthetic resolution event — and a reward that is a **logit** of the crowd you opposed at entry times a **sigmoid** of the crowd you settle into, the two ends of one logistic link applied at two times to two quantities. The conceptual core is the **unknown-future Schelling point** (§5): removing the common clock a beauty contest needs in order to converge redirects rational coordination from present sentiment toward the one focal point invariant across an uncertain horizon, which for an evidence-bearing question is the truth.

The paper’s discipline has been to separate what the design **guarantees** from what it **wagers**. The engineering — Sybil-neutral dissent, horizon-bounded capital wars (§6), and the reward’s principled functional form (§4) — is demonstrable and simulable. The epistemics — that the consensus it produces tracks truth, and faster than the world otherwise would — is a bet on incentives that, we argued (§7), *cannot* be settled by simulation: the target domain has no oracle-independent ground truth, the validation target is itself a consensus product (Wittgenstein’s Ruler), and participant behavior is performative (the Lucas critique). Drawing that boundary precisely is itself one of our contributions, not an evasion of one.

Future work follows the same split. The engineering claims invite exactly the falsification simulation is good for (§7): find a plausible behavioral model under which splitting pays, a capital war fails to converge, or small-pool sniping goes unpunished. The epistemic claim admits only the partial, post-deployment test we can honestly offer — over the subset of pools that *do*

eventually resolve, check whether clockless consensus moved toward the resolved answer earlier than baseline sentiment measures; survivorship-biased, longitudinal field data, not a proof, with Tetlock’s dated-forecast tournaments as a clarifying foil rather than a validation set (§2.6). One design question remains genuinely open: whether the lock-duration curve should adapt to a claim’s expected evidence-speed, at a cost to the fair-game constraint (§5, §8.2).

One *formal* question remains open too, and it is the more interesting of the two. §6 proves what it claims (Propositions 1 and 2), but the conceptual core of the paper is argued rather than proved. The theorem this design wants would say something like: in a coordination game with a common evaluation instant, any focal point is sustainable in equilibrium — the standard multiplicity — whereas if evaluation times are drawn from a non-degenerate distribution that players cannot condition on, a decaying focal point survives only in proportion to the probability that settlement arrives *before* it has lost its majority — write $t^\dagger$ for the last moment at which it still holds one — so that any settlement law placing enough mass beyond $t^\dagger$ leaves the time-invariant focal point as the only one whose expected payoff does not vanish. That would be the unknown-future Schelling point as a proposition rather than an argument. We do not have it. We note also that obtaining it would require modeling the rate at which non-evidential focal points decay, which trades one assumption for another and so relocates the wager of §7 rather than discharging it — but relocating it into a stated hypothesis, where it can be inspected, would itself be progress.

Clockless consensus does not escape the beauty contest; it *slows* it — converting a fast contest over present sentiment into a slow one over eventual evidence, and betting that the most stable equilibrium across an unknown future is *convergence toward reality* rather than convergence toward one another. That is a better wager than measuring belief with no force toward evidence, or a prediction market that needs an anchor these questions, by their nature, do not have — but a wager, and we have tried to mark exactly where it is one and exactly where it is not.

References

- Axelrod, Robert, and William D. Hamilton. 1981. “The Evolution of Cooperation.” *Science* 211 (4489): 1390–96. <https://doi.org/10.1126/science.7466396>.
- Callon, Michel, ed. 1998. *The Laws of the Markets*. Blackwell.
- Castellano, Claudio, Santo Fortunato, and Vittorio Loreto. 2009. “Statistical Physics of Social Dynamics.” *Reviews of Modern Physics* 81 (2): 591–646. <https://doi.org/10.1103/RevModPhys.81.591>.
- Condorcet, Marie Jean Antoine Nicolas de Caritat. 1785. *Essai sur l’application de l’analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale.
- Freeman, Rupert, Sébastien Lahaie, and David M. Pennock. 2017. “Crowdsourced Outcome Determination in Prediction Markets.” *Proceedings of the AAAI Conference on Artificial Intelligence* 31. <https://doi.org/10.1609/aaai.v31i1.10594>.
- Gibbard, Allan. 1973. “Manipulation of Voting Schemes: A General Result.” *Econometrica* 41

- (4): 587–601. <https://doi.org/10.2307/1914083>.
- Glauber, Roy J. 1963. “Time-Dependent Statistics of the Ising Model.” *Journal of Mathematical Physics* 4 (2): 294–307. <https://doi.org/10.1063/1.1703954>.
- Gneiting, Tilmann, and Adrian E. Raftery. 2007. “Strictly Proper Scoring Rules, Prediction, and Estimation.” *Journal of the American Statistical Association* 102 (477): 359–78. <https://doi.org/10.1198/016214506000001437>.
- Good, Irving J. 1950. *Probability and the Weighing of Evidence*. Charles Griffin.
- Good, Irving J. 1952. “Rational Decisions.” *Journal of the Royal Statistical Society. Series B (Methodological)* 14 (1): 107–14. <https://doi.org/10.1111/j.2517-6161.1952.tb00104.x>.
- Good, Irving J. 1979. “Studies in the History of Probability and Statistics. XXXVII A. M. Turing’s Statistical Work in World War II.” *Biometrika* 66 (2): 393–96. <https://doi.org/10.1093/biomet/66.2.393>.
- Goodhart, Charles A. E. 1975. “Problems of Monetary Management: The U.K. Experience.” In *Papers in Monetary Economics*, vol. 1. Reserve Bank of Australia.
- Grossman, Sanford J., and Joseph E. Stiglitz. 1980. “On the Impossibility of Informationally Efficient Markets.” *The American Economic Review* 70 (3): 393–408.
- Hanson, Robin. 2003. “Combinatorial Information Market Design.” *Information Systems Frontiers* 5 (1): 107–19. <https://doi.org/10.1023/A:1022058209073>.
- Hanson, Robin. 2007. “Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation.” *The Journal of Prediction Markets* 1 (1): 3–15. <https://doi.org/10.5750/jpm.v1i1.417>.
- Hanson, Robin, and Ryan Oprea. 2009. “A Manipulator Can Aid Prediction Market Accuracy.” *Economica* 76 (302): 304–14. <https://doi.org/10.1111/j.1468-0335.2008.00734.x>.
- Hopfield, John J. 1982. “Neural Networks and Physical Systems with Emergent Collective Computational Abilities.” *Proceedings of the National Academy of Sciences* 79 (8): 2554–58. <https://doi.org/10.1073/pnas.79.8.2554>.
- Keynes, John Maynard. 1936. *The General Theory of Employment, Interest and Money*. Macmillan.
- Ladha, Krishna K. 1992. “The Condorcet Jury Theorem, Free Speech, and Correlated Votes.” *American Journal of Political Science* 36 (3): 617–34. <https://doi.org/10.2307/2111584>.
- Lesaege, Clément, Federico Ast, and William George. 2019. *Kleros: Short Paper v1.0.7*. <https://kleros.io/static/whitepaper.pdf>.

- Lucas, Robert E., Jr. 1976. “Econometric Policy Evaluation: A Critique.” *Carnegie-Rochester Conference Series on Public Policy* 1: 19–46. [https://doi.org/10.1016/S0167-2231\(76\)80003-6](https://doi.org/10.1016/S0167-2231(76)80003-6).
- MacKenzie, Donald. 2006. *An Engine, Not a Camera: How Financial Models Shape Markets*. MIT Press.
- McCullagh, Peter, and John A. Nelder. 1989. *Generalized Linear Models*. 2nd ed. Chapman & Hall.
- Merton, Robert K. 1948. “The Self-Fulfilling Prophecy.” *The Antioch Review* 8 (2): 193–210. <https://doi.org/10.2307/4609267>.
- Miller, Nolan, Paul Resnick, and Richard Zeckhauser. 2005. “Eliciting Informative Feedback: The Peer-Prediction Method.” *Management Science* 51 (9): 1359–73. <https://doi.org/10.1287/mnsc.1050.0379>.
- Nakamoto, Satoshi. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. <https://bitcoin.org/bitcoin.pdf>.
- Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
- Paul, Christopher, and Miriam Matthews. 2016. *The Russian “Firehose of Falsehood” Propaganda Model: Why It Might Work and Options to Counter It*. Perspective PE-198-OSD. RAND Corporation. <https://doi.org/10.7249/PE198>.
- Peterson, Jack, Joseph Krug, Micah Zoltu, Austin K. Williams, and Stephanie Alexander. 2018. *Augur: A Decentralized Oracle and Prediction Market Platform*. arXiv:1501.01042v2.
- Prelec, Dražen. 2004. “A Bayesian Truth Serum for Subjective Data.” *Science* 306 (5695): 462–66. <https://doi.org/10.1126/science.1102081>.
- Roberts, Margaret E. 2018. *Censored: Distraction and Diversion Inside China’s Great Firewall*. Princeton University Press.
- Samuelson, Paul A. 1947. *Foundations of Economic Analysis*. Harvard University Press.
- Satterthwaite, Mark Allen. 1975. “Strategy-Proofness and Arrow’s Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions.” *Journal of Economic Theory* 10 (2): 187–217. [https://doi.org/10.1016/0022-0531\(75\)90050-2](https://doi.org/10.1016/0022-0531(75)90050-2).
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard University Press.
- Shnayder, Victor, Arpit Agarwal, Rafael Frongillo, and David C. Parkes. 2016. “Informed Truthfulness in Multi-Task Peer Prediction.” *Proceedings of the 2016 ACM Conference on*

- Economics and Computation (EC '16)*, 179–96. <https://doi.org/10.1145/2940716.2940790>.
- Soros, George. 1987. *The Alchemy of Finance: Reading the Mind of the Market*. Simon & Schuster.
- Srinivasan, Siddarth, Ezra Karger, and Yiling Chen. 2025. *Self-Resolving Prediction Markets for Unverifiable Outcomes*. <https://arxiv.org/abs/2306.04305>.
- Taleb, Nassim Nicholas. 2001. *Fooled by Randomness: The Hidden Role of Chance in the Markets and in Life*. Texere.
- Tetlock, Philip E., and Dan Gardner. 2015. *Superforecasting: The Art and Science of Prediction*. Crown.
- UMA Project. 2020. *UMA Data Verification Mechanism: Adding Economic Guarantees to Blockchain Oracles*. <https://github.com/UMAprotocol/whitepaper>.