

The Cream Always Rises to the Top... Well, Not Always

Time, Tail Events, and Why Most Arenas Promote Their Best While Politics Promotes Something Else

Tim Matula

Introduction

“The cream always rises to the top.” It is one of those proverbs people reach for when they want to say that merit wins in the end — that the best player makes the team, the best manager beats the market, the best leader ends up leading. I have spent thirty-five years in and around arenas where the proverb gets tested daily: trading floors, portfolio management, public-company boardrooms, and the long grind of building businesses. And my conclusion is that the proverb is true far more often than cynics allow, but that it is true only under specific conditions, and that we badly misread the world when we assume those conditions hold everywhere.

The key variable — the one hiding inside the word “always” — is time. Cream does not rise instantly. It rises as data accumulates. In the short run, every competitive field is dominated by noise: hot streaks, lucky bounces, one great quarter, one viral moment. Over time, the noise cancels and the signal remains. The tail events — the flukes, the one-hit wonders — get smaller and smaller as the sample grows, and the true measurement comes through. That is not a slogan; it is arithmetic, and the first part of this paper walks through why the arithmetic works.

The second part of the paper tours the arenas where the arithmetic is allowed to operate. Sports is the cleanest case: objective scoreboards, thousands of repetitions, ruthless selection. Finance is noisier and slower, but given enough years it delivers the same verdict — the managers still standing after decades are, overwhelmingly, the ones who were actually good. Chess, surgery, music, science, and corporate life each sit somewhere along the same spectrum, and comparing them reveals the three conditions a field must satisfy before its cream can rise.

The third part of the paper turns to the arena where, I will argue, the proverb fails — not because the arithmetic stops working, but because the selection system points the other way. Politics does not merely fail to filter for excellence; in important respects it filters against it. The politician’s core trade — surrender your independent judgment to the tribe in exchange for re-election — is a trade that rewards underperformance and punishes exactly the traits that make people excel everywhere else. More time in politics does not converge on truth; it converges on incumbency. I will make that case with the same standard of evidence I apply to markets, present the honest counterarguments, and let the

reader weigh them.

Part One: The Arithmetic of “Over Time”

Start with a coin. Flip it ten times and there is nothing unusual about seeing seven heads; roughly one trial in six will produce seven or more. Flip it a thousand times and the share of heads will almost certainly sit within a few points of fifty percent. Nothing about the coin changed. What changed is the sample size. Small samples are where flukes live; large samples are where truth lives. Statisticians call this the law of large numbers, and it is the quiet engine underneath every claim that cream rises “over time.”

Now replace the coin with a person whose outcomes are a blend of skill and luck — which describes essentially everyone in every competitive field. In any single contest, luck can swamp skill entirely. A mediocre poker player wins a big hand; a mediocre fund manager has a big year; a mediocre band writes one perfect chorus. These are tail events: outcomes far from the performer’s true average, produced by variance rather than ability. The defining property of tail events is that they do not repeat on demand. As trials accumulate, each new data point drags the observed record back toward the performer’s true level — the phenomenon known as regression to the mean. The one-hit wonder is not a moral failure; it is a statistical inevitability. A single hit is exactly the sample size at which luck is most able to impersonate skill.

Human beings, unfortunately, are built to misread this arithmetic. We are storytellers before we are statisticians, and a story requires a hero, not a distribution. When a fund manager triples the market, the coverage explains her insight; it does not compute how many equally confident managers held opposite positions and vanished. When a startup founder succeeds, we study his habits; we do not interview the thousand founders with identical habits who failed. This is the narrative fallacy compounded by survivorship bias: the losers exit the sample silently, the winner’s luck is relabeled as genius, and the audience walks away believing the tail event was destiny. The practical consequence is that public reputation is systematically noisier than actual performance data — fame accrues to spectacular small samples, while genuine skill, which reveals itself as an undramatic consistency over long periods, makes for poor television. Any serious attempt to judge whether cream is rising must therefore begin by distrusting the highlight reel and demanding the full record: every trial, including the ones that ended quietly.

Michael Mauboussin, who has written as clearly on this subject as anyone, suggests a simple diagnostic for how much luck a field contains: ask whether you can lose on purpose. A chess grandmaster can lose on purpose instantly, because chess is nearly all skill. A lottery player cannot lose on purpose at all, because the lottery is all luck. Most arenas sit in between, and the position matters enormously, because the more luck a field contains, the more repetitions are required before its results become believable. This gives us the master rule of the whole subject: the honesty of any ranking is a function of three things — how clearly the outcome is measured, how many times the contest is repeated, and whether

the selection mechanism actually promotes on that measurement. Hold that rule; every arena in this paper will be judged against it.

Part Two: Sports — The Cleanest Filter We Have

If you wanted to design a machine for making cream rise, you would design professional sports. The outcome is measured to the decimal and displayed on a scoreboard that no one can spin. The contest repeats constantly — an NBA regular season is eighty-two games, a baseball season one hundred sixty-two, and within each game are hundreds of discrete, recorded events. Feedback is immediate and public. And selection is tied directly to the measurement: underperform and you are benched, cut, waived, or relegated. All three conditions of the master rule are satisfied at full strength.

The results are what the proverb promises. Short samples still produce famous flukes — Leicester City winning the Premier League at pre-season odds reported around five thousand to one, Buster Douglas knocking out Mike Tyson as a forty-two-to-one underdog, a sixteen-seed occasionally toppling a one-seed in March. We remember these precisely because they are rare, and the aftermath is always the same: regression. Leicester spent the following seasons back in mid-table. The Cinderella teams of a single-elimination tournament rarely survive a seven-game series, which is exactly why championship series exist — a best-of-seven is a deliberate enlargement of the sample to shrink the role of luck.

Over career-length samples, the filter becomes almost perfectly honest. The players who sustain excellence across fifteen or twenty seasons — across thousands of games, multiple coaches, rule changes, and aging bodies — are, with virtually no exceptions, the genuinely great ones. And notice that the filter corrects its own errors of prediction. Tom Brady was the one hundred ninety-ninth pick of his draft; scouts misjudged him badly. It did not matter, because the system kept generating performance data, and the data kept promoting him. Michael Jordan was famously assigned to his high school's junior varsity squad as a sophomore. The initial assessment was wrong; ten thousand subsequent measurements were not. Meanwhile the same machinery works just as relentlessly in the other direction: the average NFL career lasts only about three years, and rosters churn constantly as the measurements arrive. Sports is brutal, but its brutality is honest. The cream rises because nothing can stop it from rising.

Sports also offers the best documented case of a field deliberately upgrading its own filter — and being rewarded for it. Through most of baseball's history, talent was judged partly on measurements that turned out to be noise: batting average over on-base percentage, the look of a swing, the build of a body. The analytics movement chronicled in Michael Lewis's *Moneyball* was, at bottom, a project of replacing weaker measurements with stronger ones, and the small-budget teams that adopted it first bought wins at a fraction of the market price, because the market was still paying for the noise. Within a decade the entire league had copied them. That is what a healthy arena does when it discovers its scoreboard is imperfect: it sharpens the scoreboard, and the competitive advantage flows to whoever sharpens it first. Keep that image in mind, because later in this paper we will meet an arena that

responds to flaws in its scoreboard in precisely the opposite way.

Part Three: Finance — Noisier, Slower, Same Verdict

Markets are where I grew up professionally — I started on the retail side in 1991 and later ran quantitative portfolios — and markets are a far noisier arena than sports. A single year of returns tells you almost nothing about a manager's skill. In any given year, roughly half of active managers will beat their benchmark by simple chance, and a lucky minority will beat it spectacularly. The financial industry manufactures one-hit wonders at industrial scale: the fund that returned eighty percent last year, the newsletter writer who called the crash, the trader who was up huge in a single wild quarter. Every one of these is a small-sample tail event wearing the costume of genius.

Time strips the costume off. The most sobering numbers in the industry are the long-horizon scorecards: over fifteen-year periods, on the order of nine in ten actively managed U.S. large-cap funds fail to beat the S&P 500 after fees. Read correctly, that statistic is not evidence that skill does not exist; it is evidence that the filter works. Extend the horizon and the pretenders are removed — by redemptions, by career change, by fund closure. The market is a machine for firing underperformance, and it runs continuously. What survives two and three decades of that machine is a small residue, and the residue looks nothing like luck.

An honest analyst has to pause here for the standard objection: survivorship bias. If thousands of managers start and only the winners remain visible, might the survivors just be the lucky tail of a giant coin-flipping contest? For short records, absolutely — that is the correct null hypothesis. But the luck explanation has to work harder with every additional year, and at the far end of the distribution it collapses. Warren Buffett has compounded at roughly twice the market's rate for some six decades; the probability of doing that by chance is not small, it is effectively zero. Buffett himself made the deeper point in his famous "Superinvestors of Graham-and-Doddsville" argument: if the winners of a coin-flipping contest turn out to cluster around the same teacher and the same method, luck stops being a serious explanation. Renaissance Technologies' Medallion Fund is reported to have returned on the order of thirty-nine percent annually, net of extraordinary fees, for roughly three decades — a sample of thousands of trading days across every market regime. These are not one-hit wonders. They are what the far right tail of genuine skill looks like once time has removed everyone else.

The clientele, it must be said, conspires against its own interests by worshipping exactly the tail events the arithmetic warns about. Fund flows chase last year's winner with mechanical reliability: money floods in after a spectacular year — which is to say, at the top of a small lucky sample — and floods out after the inevitable regression, locking in the round trip. Studies of dollar-weighted returns find that the average investor earns meaningfully less than the average fund he invests in — on the order of a percentage point or more annually — purely from the timing of his own entries and exits. I watched this behavior from inside the industry for decades, and the lesson generalizes well beyond finance: it is not enough for a field to generate honest long-run data. The people doing the selecting must also have the

patience to wait for it. Wherever the selectors act on one-year samples, they will systematically promote luck and fire skill, and they will pay for the privilege. Time only works as a filter for those willing to let it run.

So finance obeys the proverb, with two amendments. First, the required time horizon is much longer than in sports, because the luck content is higher — believing any track record shorter than a full market cycle is a category error. Second, the cream that rises is a much smaller share of the container. But the direction of the filter is right: measurement is objective (returns are returns), trials accumulate daily, and capital — the selection mechanism — eventually flows toward performance and away from its absence. The managers who have been around the longest are not necessarily the best in any given quarter, but over time the longest-tenured cohort tends strongly toward genuine, above-market skill, because tenure itself is the market's accumulated verdict.

Part Four: A Tour of Other Arenas

Put other fields against the master rule and a clear gradient appears. Chess sits at the honest extreme — arguably purer than sports. The Elo rating system converts every game into an updated measurement, luck is nearly absent, and the rankings are therefore close to incorruptible; no one disputes who the strongest players in the world are, and the same names hold the top spots for a decade at a time. Surgery sits high as well: outcomes are countable, and the volume-outcome relationship — surgeons and hospitals that perform a procedure most often tend to achieve the best results — is one of the most consistently replicated findings in health services research. Repetition builds skill, and measurement reveals it.

Music and film are luck-heavy. William Goldman's famous verdict on Hollywood — that nobody knows anything — is a statement about the impossibility of predicting single outcomes, and the music industry literally maintains a category called the one-hit wonder. But watch what time does even here: catalogs are large samples. A single hit can be luck; twenty years of albums that people still pay to hear cannot be. The artists still filling arenas three decades on are, by the only measurement that aggregates honestly — sustained audience demand — the real thing. Science is slower still, and noisier at the level of a single paper, as the replication troubles of recent years attest. Yet its long-run filter is famously self-correcting: fraudulent and flimsy results die when others cannot reproduce them, and the Nobel committees now routinely honor work performed twenty or thirty years earlier — an institutionalized admission that only time can certify which discoveries were cream.

Entrepreneurship and venture capital deserve their own stop on the tour, because they are the most luck-drenched corner of business and therefore the sternest test of the thesis. Startup outcomes follow a power law: in a typical venture portfolio, the great majority of investments return little or nothing and one or two outliers pay for everything. At the level of a single company, luck — timing, a chance customer, a competitor's stumble — can dominate any amount of skill, which is why a lone unicorn on a résumé proves far less than its owner usually believes. Yet even here, repetition slowly separates the

population: researchers who track founders across ventures find that those who have succeeded before succeed again at modestly higher rates than first-timers or previous failures, and the small set of investors who beat the asset class across many funds and multiple technology cycles is stable enough that the industry's limited partners fight for access to them. The signal is weaker than in chess and the sample sizes are painfully small — a career may contain only a handful of companies — but the direction is unmistakable and it is the familiar one: even in the noisiest honest arena, more trials mean more truth. Luck sets the variance; time still finds the mean.

Corporate leadership is the hardest honest case, and it is worth being candid about it because it is the world I know best. CEO outcomes are contaminated by timing, sector winds, and inheritance; plenty of executives have ridden a cycle up and been called geniuses for it. Attribution is genuinely difficult — a company's results are the work of thousands of hands. But the conditions of the master rule are still present in weakened form: there is a scoreboard (profit, share performance, survival), the trials repeat (every quarter, every product cycle, every downturn), and the selection mechanism, though slow and imperfect, has teeth — boards remove executives, competitors take share, and bankruptcy is the market's final grading pen. Executives who compound value across multiple companies, multiple cycles, and multiple decades are rare, and for the same reason long-tenured outperforming fund managers are rare: by that point, luck has largely been wrung out of the record. The pattern across every arena in this tour is now visible. Wherever outcomes are measured honestly, trials accumulate, and selection is chained to the measurement, cream rises — quickly where luck is low, slowly where luck is high, but reliably. Which brings us to the arena where the chain is cut.

Part Five: Politics — The Filter That Points the Other Way

Run politics through the master rule and all three conditions fail at once. First, there is no scoreboard. What is the objective, agreed measurement of a legislator's performance? Bills passed is a volume statistic, not a quality one. Economic outcomes lag policy by years and are shaped by forces no officeholder controls. Constituent service is invisible at scale. Voters cannot mark a politician's performance to market because there is no market and no mark — there is only narrative. Second, the trials barely repeat. A senator faces the electorate once every six years; a president, twice in a lifetime of ambition. Where a ballplayer generates thousands of measured events per season, a politician generates a handful of elections per career — a sample size at which, as Part One established, luck and noise dominate everything. Third, and most fundamentally: the selection mechanism is not connected to governing performance at all. It is connected to electoral performance. The system does not promote people who govern well; it promotes people who campaign well, and those are different skills practiced by, in many cases, different kinds of people.

Now add the mechanism that I believe does the deepest damage, because it actively inverts the filter: the tribe trade. In nearly every other arena, excellence is exercised through independent judgment — the portfolio manager's willingness to stand apart from the crowd, the scientist's willingness to follow

evidence against the fashion, the athlete's obligation to perform alone under the lights. In politics, independent judgment is precisely what the system prices out. A candidate does not reach office, and above all does not remain in office, on the strength of his own convictions; he remains by subordinating them to a party. The whip system, the primary electorate, the donor network, and the media ecosystem of his side all enforce the same contract: vote with the tribe, speak with the tribe, suppress your private judgment when it conflicts with the tribe's, and in exchange the tribe will deliver the only performance metric that matters — re-election. This is, in the most literal sense, a compensated agreement to underperform: the officeholder trades away the exercise of his own best judgment, which is the very thing excellence consists of everywhere else, and receives job security in return. In finance we would call this a catastrophic principal-agent problem. In politics we call it party discipline and treat it as normal.

The numbers say the trade clears. Public approval of Congress as an institution has spent most of recent decades in the teens and twenties, while incumbent members of the House who seek re-election win at rates that are routinely above ninety percent. Sit with that pairing for a moment, because no honest filter could produce it: the customers overwhelmingly rate the product a failure, and the producers are overwhelmingly retained. In any market I have ever operated in, a ninety percent retention rate alongside a twenty percent satisfaction rate is impossible — capital would flee, customers would defect, the scoreboard would clear the field. In politics it is the steady state, sustained by safe districts in which the general election is a formality and the only contest that matters is a low-turnout primary decided by the most tribal fraction of the electorate. Under those conditions, time does not filter; time entrenches. Every additional term generates not better measurement of governing skill but greater name recognition, deeper donor relationships, and more seniority — incumbency compounding like interest, on exactly the wrong account.

Even if voters wanted to grade governing performance, the structure of the job frustrates the attempt, because political outcomes mature on a completely different clock than political careers. A pension reform, an infrastructure program, an education policy, a decision about energy capacity — these reveal their quality over ten, twenty, thirty years, long after their authors have moved on, claimed credit, or assigned blame. The officeholder in the chair when results finally arrive is rarely the one who caused them, and everyone in the system knows it. This produces a predictable and corrosive incentive: optimize for what is visible inside the electoral window — announcements, ribbon-cuttings, positions taken, enemies named — and discount everything whose payoff lands beyond it, however large. In finance we would recognize this instantly as marking positions to a horizon that expires before the trade does; no serious investor would accept a track record constructed that way. Yet it is the only track record politics knows how to construct. The attribution problem does not merely add noise to the measurement; it decouples the measurement from the work itself, so that even an infinitely patient electorate would struggle to reward the right people.

The modern attention economy then bolts an amplifier onto the broken filter. Elections are decided by attention, attention flows to conflict, and so the system's effective scoreboard — the one that actually

moves careers between elections — is a running tally of viral moments, cable hits, and social-media engagement, every unit of which rewards heat over light. The legislator who spends a year quietly mastering a complex bill generates no clips; the one who produces a combative soundbite in a hearing raises money that afternoon. It would be difficult to design a promotion metric more perfectly orthogonal to governing skill if one tried. And notice the compounding effect on the intake valve discussed below: the personality type that thrives on perpetual performative combat and the personality type that patiently builds durable institutions are rarely the same type. The arena is not neutral about which type it admits. It advertises for one and quietly repels the other.

The filter also fails at the intake valve. Ask what the job actually selects for at entry, and what it costs the entrant. Members of Congress are advised by their own party committees to spend roughly four hours a day on fundraising calls — a schedule first revealed in a leaked party orientation presentation and confirmed by numerous former members since. The compensation is a fraction of what comparable talent earns in the private sector; the scrutiny extends to one's family; the work itself, for a junior member of a disciplined caucus, consists largely of reciting positions someone else drafted. Now perform the thought experiment as a recruiter, which is a hat I have worn many times: describe that role honestly — the hours devoted to soliciting money, the prohibition on independent judgment, the pay cut, the public exposure — and ask which candidates accept the offer. Economists Francesco Caselli and Massimo Morelli formalized the grim answer in a paper bluntly titled “Bad Politicians”: when the material and reputational rewards of office are low relative to private alternatives, the candidate pool skews toward those with the least to give up — adverse selection at the point of entry, before a single vote is cast. The people best equipped to govern — those with demonstrated records of building, managing, and deciding under uncertainty — are, with honorable exceptions, exactly the people for whom the trade makes least sense. They stay in the arenas with scoreboards. The result is a political class that is not a fair sample of a nation's talent, let alone its cream: it is a sample of those willing to make a trade that talent rarely accepts.

Consider, finally, what the profession's own career ladder trains. In every arena where cream rises, the apprenticeship consists of practicing the actual work: the surgeon's residency is surgery, the trader's formation is trading, the point guard's development is ten thousand possessions. The standard political ladder — campaign staffer, communications aide, party operative, candidate — is an apprenticeship in campaigning. Its curriculum is message discipline, opposition research, fundraising, and coalition maintenance; at no rung does it require building anything, managing a payroll, delivering a project on budget, or living with the long-term consequences of one's own decisions. By the time the successful climber reaches office, he may have spent twenty years becoming world-class at the acquisition of office and not one day practicing its use. This is not an accusation of bad faith; it is an observation about training. We would not expect excellence from a surgeon whose entire residency was spent interviewing for surgical jobs, however brilliantly he interviewed. Yet that is, structurally, the résumé the political ladder produces — and then we express surprise that the institution staffed by that ladder underperforms.

Notice that this argument requires no cynicism about any individual and no claim about either party. It is a claim about system design, and it would apply to any tribe and any flag. Sports would produce mediocrity too if athletes were selected by popularity contests held every few years among fans who never saw the games, and if coaches benched any player who improvised. Finance would produce mediocrity if managers were retained by marketing skill rather than returns — indeed, in the corners of finance where exactly that is true, mediocrity is exactly what is produced, which rather proves the point. Politics is not populated by worse human beings than other fields. It is governed by a worse filter — one that measures the wrong thing, measures it rarely, and pays its participants to suppress the one trait that constitutes excellence everywhere else.

Part Six: The Honest Counterarguments

A thesis this pointed deserves its strongest objections, fairly stated. The first is definitional: democracy's filter was never designed to select for technical excellence. It selects for consent, legitimacy, and representation — and on its own metric it succeeds. An election answers the question “whom do the people choose?” with perfect accuracy every time. On this view, complaining that elections fail to identify the most capable governors is like complaining that a thermometer fails to measure weight; the instrument is doing a different job, and a valuable one, since governments that lose legitimacy fail in ways far worse than mere incompetence. There is real force here, and it reframes rather than refutes the thesis: the cream does rise in politics — cream as defined by the filter's own metric. The best campaigners, coalition-builders, and tribal signalers reliably reach the top. The trouble is only that we then hand them a job the filter never tested them for. The proverb fails not because selection is random, but because the selected trait and the required trait are different traits.

The second objection is historical: crises appear to make the cream rise even in politics. Lincoln, Churchill in 1940, and other examples suggest that when stakes become existential, systems somehow reach past the usual machinery and elevate genuine capacity. I think the observation is correct and revealing rather than contrary. What a crisis does is restore the missing conditions of the master rule: suddenly there is a scoreboard (the war is being won or lost, visibly), feedback arrives in weeks rather than decades, and the tribe's tolerance for loyal mediocrity collapses because the tribe itself is at risk. Churchill spent the peacetime 1930s exiled to the back benches — the tribal filter working exactly as described — and was summoned only when the scoreboard lit up. The exception operates by temporarily converting politics into the kind of measured arena where cream can rise, which is less a rebuttal of the argument than a demonstration of it.

A related objection comes from the other direction: if politics merely filtered out excellence, then accomplished outsiders — executives, generals, surgeons — should stride into office and outperform the career politicians, and the record shows that they frequently flounder. Fair enough, and worth taking seriously. Legislating through a body of hundreds of independent actors, none of whom can be fired, is genuinely different work from running a company, and the outsider's private-sector scoreboard tested

none of it. But look closely at what this proves. It proves that political skill is real and distinct — coalition assembly, procedural mastery, the management of allies who are also rivals — not that the filter selects for governing excellence. A system can simultaneously require real skills and select for the wrong ones; indeed that is precisely this paper’s claim. The failure of outsiders and the mediocrity of insiders are not competing observations. They are the two halves of the same diagnosis: the job requires one set of abilities, the filter tests for another, and almost no path exists that certifies both.

The third objection is a caution against my own potential bias, and readers should hold me to it. “Underperformer” must not become a synonym for “person whose policies I dislike.” Capable, principled people do serve in public life in every generation and in every party; some sacrifice considerably to do so, and a fair accounting salutes them. My claim is statistical, not universal: it concerns the central tendency the filter produces, not every individual it passes. And a fourth objection deserves standing room: perhaps the comparison is unfair because governing is simply harder to measure than shooting a basketball, through no fault of the system’s designers. Partly true — but only partly, because measurement in politics is not merely difficult; it is actively resisted. Institutions that would sharpen the scoreboard — independent scoring of legislation, sunset clauses that force retrospective review, nonpartisan redistricting that makes general elections competitive again — are perennially proposed and perennially blocked, most vigorously by incumbents. A field can be forgiven for having a noisy scoreboard. It is harder to forgive a field for dismantling the scoreboard on purpose.

Conclusion: You Get What You Select For

The proverb, corrected, reads something like this: cream rises wherever time is allowed to multiply honest measurements, and only there. In sports, the conditions are nearly perfect and the cream rises fast. In finance, the noise is heavier and the rise takes decades, but the long-tenured survivors of the market’s continuous culling are the real thing — time shrinks the tails, retires the one-hit wonders, and lets the true measurement come through. Chess, surgery, science, music, and business each confirm the same law at their own speed. Across all of them, the pattern is not that virtue triumphs; it is that filters deliver exactly what they select for, no more and no less.

Politics delivers exactly what it selects for too — and that is the indictment. It selects for electability, tribal reliability, and the willingness to trade one’s own judgment for one’s own seat, and so that is what floats to the top: not our best and brightest but our most durable campaigners, sustained by a system in which twenty percent approval coexists comfortably with ninety percent re-election. The one-hit wonders of markets are washed out by time; the incumbents of politics are fortified by it. Same variable, opposite sign, because the underlying selection points the opposite way.

The practical conclusion follows directly, and it is more hopeful than the diagnosis sounds. If the problem were the people, we would be stuck, because moral exhortation has never improved a candidate pool. But the problem is the filter, and filters can be redesigned. Anything that restores the three conditions moves the system toward honesty: make more elections genuinely competitive so the

trials mean something; build scoreboards — retrospective review, transparent scoring, published results — so performance is measured; and raise the rewards for independent judgment relative to tribal loyalty so the talented find the trade worth making. None of this is easy, and incumbents will resist all of it, for reasons this paper has made clear. But the direction is not mysterious. Every arena that gets selection right teaches the same lesson, and it is the lesson I would leave with the reader: stop asking why the cream is not rising, and start asking what the filter is actually skimming. The cream always rises to the top — in systems built to let it. Build the system, and time, the most honest judge we have, will do the rest.