NBER WORKING PAPER SERIES

PAYING FOR QUALITY VS. PAYING FOR RANK:
WHEN PURIFYING A METRIC BACKFIRES

Joshua S. Gans
ⓡ
Scott Duke Kominers

Working Paper 35363
http://www.nber.org/papers/w35363

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2026

We are grateful to David Laibson, Nat Malkus, Jeff Miron, Alexandru Nichifor, and Jesse Shapiro for helpful conversations. We also thank Refine.ink, ChatGPT 5.5/Pro, Claude Opus 4.7-8 and Fable 5 for research assistance. Responsibility for all errors remains our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w35363

Paying for Quality vs. Paying for Rank: When Purifying a Metric Backfires
Joshua S. Gans ⓡ Scott Duke Kominers
NBER Working Paper No. 35363
June 2026
JEL No. C78, D47, D83, I23

ABSTRACT

Matching markets allocate scarce opportunities using performance metrics that agents can game. When does making a metric less gameable actually reduce gaming? We show that the answer hinges on how the market converts evaluations into stakes. Where evaluations assign prices — wages, credit terms, score-indexed payments — partially purifying the metric raises gaming precisely when gameable variation dominates it: the market's trust in the metric then rises faster than its exposure to gaming falls. By contrast, where evaluations assign positions — fixed seats allocated by rank — purification never backfires: ranking cancels trust from the incentive, leaving a rank price that deflates with the metric's dispersion, so mean gaming falls monotonically. We solve both markets in closed form under a first-order equilibrium concept; transfers microfound the first regime, rank tournaments the second. A prize-curvature index then determines when gaming improves rather than harms sorting, and when perfecting the metric is itself sorting-suboptimal.

Joshua S. Gans
University of Toronto
Rotman School of Management
and NBER
joshua.gans@rotman.utoronto.ca

ⓡ

Scott Duke Kominers
Harvard University
Harvard Business School
kominers@fas.harvard.edu

1 Introduction

Many matching markets allocate scarce opportunities as a function of an established performance metric. A postgraduate programme looks at students' grade-point averages; teaching hospitals are evaluated according to risk-adjusted outcome scores; a lender assesses borrowers' credit scores; and teachers and schools are assessed based on their students' standardised test performance. In each case, the metric summarises dispersed information, and the demand side of the market uses it to rank potential matches.

Because the metric is consequential, those being measured respond to it. Students enrol in easier-graded classes; hospitals select less-risky patients or shift coding intensity; borrowers churn accounts to flatten their credit scores; teachers teach to the test.[1] None of these responses necessarily changes the latent quality that the other side of the market is trying to infer: they move the residual non-quality component of the released statistic. We call this behaviour *metric gaming*. It is distinct from manipulation of the matching mechanism itself, because the downstream matching rule can be entirely non-manipulable with respect to the preferences determined by the reported metric, even while the reported metric itself is still shaped upstream, through a gameable process.

Performance metrics differ in their gameability: a raw grade loads one-for-one on course choice and grading leniency, while a difficulty-adjusted grade subtracts a noisy proxy for those margins. The natural ordering is by how much of the gameable action the metric still loads on, from raw to *completely purified* — where the released statistic, conditional on the agent's latent type, no longer depends on the gaming action. Most practical reforms sit in the interior: the gameable component is usually observed only with noise, so institutions settle for partial purification.

This paper asks the design question every such reform faces: *does making the metric less gameable actually reduce gaming*? At first read, one might wonder why there is even a question here — but the issue is that two different effects collide: mechanically, purification lowers the metric's *exposure* to the gameable action, so a unit of gaming buys less metric movement. But then in equilibrium, a purer metric carries a larger share of latent quality per unit of movement, so the market rationally *trusts* it more — and a more trusted metric is more worth moving. Whether the reform reduces gaming depends on whether the reduction of exposure or the increased trust in the metric dominates.

Our main result is that the impact of purification on gaming is determined not by the

[1] Education provides prominent examples of such responses and of ordinal approaches designed to address them. Neal (2010) discusses test-based accountability and advocates contests among schools serving similar students, while Barlevy and Neal (2012) propose rewarding teachers by within-comparison-set percentile rank. These designs anticipate the protocol contrast at the centre of this paper.

form of the metric, but rather by the structure of the downstream market that makes use of it. In markets that *pay for quality* — the value of a better evaluation is a price, as when wages or interest rates are attached to the posterior — partial purification raises equilibrium gaming precisely when gameable variation dominates the metric in the benchmark case (Theorem 1): trust can rise faster than exposure falls. In markets that *pay for rank* — a fixed distribution of positions allocated by relative standing, as in promotion ladders or centralised assignment — the backfire is impossible: for every parameter configuration, informative gaming ability included, and every median-thinning prize distribution, every step of purification strictly reduces the equilibrium distortion (Theorem 2).[2]

The mechanism behind the contrast is an elasticity race, and neither result involves anyone being fooled: the market correctly accounts for the equilibrium distortion, and the residual incentive is private because moving one's own metric still moves one's own market-assessed posterior. Both regimes are microfounded in the same closed matching market, and the question that separates them is concrete: *what does being believed to be better actually buy a candidate, and what happens to the value of that purchase as the market becomes better informed?* With transfers, stability forces each position's wage to adjust until the position is content with its hire, so the value of being believed better is a price — for the median agent, the median prize — which does not move as the market becomes better informed. Without transfers, being believed better buys only a climb past rivals, so its value is the crowding of rival beliefs around the agent's own — and a better-informed market spreads beliefs out, leaving any given improvement fewer rivals to overtake. Ranking thus cancels the market's trust in the metric from the incentive altogether; what remains is a *rank price* — the density of competing reports — which deflates with the metric's standard deviation, whereas trust responds to its variance. A price holds its value as the field spreads; a crowd thins. We call the stabilising force the *rank-price feedback*, and a reduced-form proposition makes the comparison exact: gaming is non-monotone along the reform path exactly when the elasticity ψ at which the price of report movement deflates with the report's variance exceeds a threshold lying strictly above $\frac{1}{2}$ for all primitives; quality pricing has $\psi = 1$, rank pricing $\psi = \frac{1}{2}$ — exactly the boundary elasticity, on the safe side.[3]

[2]The unconditional claim concerns the mean distortion; the total cost of gaming falls monotonically as well whenever gaming ability is not too negatively correlated with latent quality, a bound admitting every non-negative correlation.

[3]The two microfoundations are asymmetric by construction: the rank-market theorem is global, while the transfer-market backfire is exact in the concentrated-prize limit and confirmed computationally beyond it (Figure 2, Remark 1). Closing the market also pays a technical dividend: each fixed point collapses to a single scalar equation — globally unique in the rank market for all primitives, and unique under transfers except at extreme measurement noise.

What gaming does to the market's *ranking* is a separate question, and the second half of the paper answers it. At a locally linear schedule, equilibrium gaming takes the form of a common shift that the market subtracts, and thus is rank-neutral. That knife-edge is unreachable in both closed markets: with prizes that thin out above the median, equilibrium schedules are locally convex, convexity makes high-quality agents game hardest, and the amplification stretches latent quality against the noise floor, *raising* sorting above the linear benchmark. The strength of the effect is governed by a single primitive of the prize distribution — a *prize-curvature index* — and when prize dispersion crosses an explicit threshold, complete purification ceases to be sorting-optimal even though the purification technology is noiseless: the sorting optimum moves to strictly positive exposure (Theorem 3). Because the closed market prices match output and action costs in the same units, these statements aggregate to a total-surplus criterion rather than a heuristic (Proposition 7).

Methodologically, the closed markets are solved under a *first-order equilibrium* concept — agents optimise against the second-order expansion of the matching schedule their own behaviour induces, with the expansion point consistent with the equilibrium posteriors — which a supplemental appendix shows is the correct local model: exactly Bayes-consistent, with optimisation errors an order smaller than the curvature effects being measured.[4]

The characterisation has empirical content because the dividing line is observable. Whether the reward for a better evaluation is a price or a position is a fact about institutions, not about preferences or information. Credit scores inform interest rates, which are quality-priced and hump-prone. Centralised assignment that allocates a fixed set of seats, by contrast, is rank-priced, and thus immune.[5] The same purified transcript can be both — hump-prone in the hands of wage-setting employers and monotone in the hands of seat-constrained graduate programmes. The model thus converts a design controversy into a measurement question: before predicting whether a purer metric will tame or inflame gaming, ask what the metric buys.

Related literature. The closest antecedents of the effect we highlight are price-mediated throughout, and our quality-priced regime nests their central comparative statics while the rank regime reverses its scope. In career-concerns and signal-jamming models (Holmström,

[4]We also report the approximation quality against the true schedules wherever we plot: in the configurations shown, re-optimising every type against the exact schedule raises expected payoffs by well under 1% of the relevant prize scale.

[5]With scarce seats, the admission margin attaches a fixed stake to the cutoff. Immunity survives to first order in stakes — in aggregate, the cutoff density is itself a rank price — while large admission stakes can revive a modest interior peak in mean gaming (Supplemental Appendix C.2).

1999; Stein, 1989; Gibbons and Murphy, 1992), effort or manipulation responds to a posterior-sensitivity coefficient times a fixed downstream price; our knife-edge first-order condition is theirs with an exposure dial, and the non-monotonicity of effort in information structures has a precedent in the work of Dewatripont et al. (1999). The manipulable-information literature studies scores used by price-setting receivers: manipulation muddying inference (Frankel and Kartik, 2019), commitment to underweight manipulable data (Frankel and Kartik, 2022), linear scoring of strategic agents (Ball, 2025), falsification-proof tests (Pérez-Richet and Skreta, 2022), and consumer scores that feed pricing (Bonatti and Cisternas, 2020). In accounting, Ewert and Wagenhofer (2005) show that tightening standards can increase earnings management because the market's earnings-response coefficient rises — the partial-purification backfire in a price-mediated setting, anticipated in the reporting-bias framework of Fischer and Verrecchia (2000). Relative to all of these, our contribution is to make the stakes process the object of analysis: we microfound the fixed price in a transferable-utility matching market, solve the non-transferable rank market in which the same reform is uniformly safe, and locate both on an elasticity frontier.

On the matching side, Ostrovsky and Schwarz (2010) study what evaluations schools choose to disclose; we study the upstream margin on which the evaluation itself is contaminated. Pathak and Sönmez (2013) and Azevedo and Budish (2019) compare mechanisms by their manipulability with respect to agents' reports. Our agents face mechanisms that may be entirely non-manipulable downstream while the residual action margin operates upstream, and our rank-market theorem gives such ordinal matching mechanisms a new defence: they cannot be made more gameable by partial measurement reform. The transferable-utility analysis builds on assortative-matching payoffs in the tradition of Becker (1973) and Shapley and Shubik (1971), as well as the supply and demand analysis of matching markets in the style of Azevedo and Leshno (2016). The broader performance-measurement literature (Holmström and Milgrom, 1991; Baker, 1992; Campbell, 1979; Goodhart, 1975) supplies the multitasking backdrop. Unlike Spence-style signalling (Spence, 1973), the gaming action in our setting carries no intrinsic information about quality, so complete purification can eliminate the incentive. The equilibrium gaming order is not a Blackwell order (Blackwell, 1953): behaviour changes with the experiment. The grading application connects to the grade-inflation literature (Sabot and Wakeman-Linn, 1991; Bar et al., 2009; Chan et al., 2007; Popov and Bernhardt, 2013) and to eigengrade-style adjustment procedures (Gans ⓡ Kominers, 2026), which are related methodologically to spectral-ranking approaches (e.g., Page et al. (1999)). Finally, the rank market is a tournament, and contest theory supplies both its first-order condition and the design proposals closest to our message.[6]

The rest of the paper is organised as follows. Section 2 sets up the environment and derives the marginal-incentive decomposition whose middle term — the price of report movement — is the paper's organising object. Section 3 establishes how purification can backfire when the market prices quality, Section 4 shows the impossibility of backfiring in rank markets, and Section 5 the backfire frontier and its transferable-utility microfoundation. Section 6 turns to sorting and welfare, and Section 7 concludes with the design implications. Appendix A contains omitted proofs; the Supplemental Appendix collects the notation table, pairwise comparisons, rationed markets, the formal validity of the quadratic specification, applications, multiplicity details, and auxiliary proofs.

2 Environment

2.1 Agents, positions, actions, and reports

We work in a matching setting with a continuum of agents of mass 1 and, in the baseline formulation, a continuum of positions of mass 1.[7] An agent has latent *(match-relevant) quality* $\theta \in \mathbb{R}$ and private *action type* $z \in \mathbb{R}$, drawn from a commonly known joint distribution with finite first moments. A position has desirability q, distributed according to a continuous distribution H with finite mean and strictly positive density h on its support. Positions prefer agents with higher posterior expected quality, and agents prefer positions with higher desirability. Table A.1 in the Supplemental Appendix collects notation.

In this section and the next, we examine agent behaviour under price-taking, summarising the downstream environment by an increasing placement-value schedule Q — so that an agent assigned posterior expected quality m receives expected value $Q(m)$, and each individual agent takes this schedule as fixed. Then, in Sections 4 and 5, we close the matching market by deriving Q from the market itself, under the two allocation protocols

[6] The marginal incentive in Section 4 is the prize-spread-times-rival-density condition of Lazear and Rosen (1981), its cancellation of common components echoes relative-performance evaluation (Holmström, 1982; Green and Stokey, 1983; Nalebuff and Stiglitz, 1983), and the continuum is a large contest in the sense of Olszewski and Siegel (2016). Prize-distribution curvature is the design instrument of Moldovanu and Sela (2001); Olszewski and Siegel (2020), chosen there to maximise effort; our χ_H points it at sorting under gaming. The cheating-in-contests literature (Gilpatric, 2011; Kräkel, 2007) fixes the detection regime; our contribution is the comparative static along the purification path. Closest in prescription, Barlevy and Neal (2012) pay teachers by within-comparison-set percentile rank because ordinal schemes resist scale corruption, Neal (2010) advocates contests among comparable schools, and Li and Qiu (2023) implement allocation through coarse-ranking contests — design counterparts of Theorem 2, which obtains the immunity as an equilibrium property of any market that pays in positions.

[7] Supplemental Appendix C relaxes the equal-mass normalisation and allows rationing.

that the paper contrasts.

Before the report is realised, each agent chooses an *action* $x \in \mathbb{R}$ with private value $v(x,z)$. The maintained interpretation is that x is a residual margin that affects the report but is not itself part of the latent quality the market is trying to infer.[8]

Assumption 1 (Single-peaked private action payoffs)**.** For each z, the function $v(\cdot,z)$ is twice continuously differentiable and strictly concave, with unique maximiser $x = z$: $v_x(z,z) = 0$ and $v_{xx}(x,z) < 0$ for all x.

Assumption 1 gives z a natural interpretation: it is *the action the agent would choose absent any market reward for the public report*. The action-space gaming amount is the deviation $a = x - z$ from that privately optimal action.

A *reporting rule* is indexed by $\lambda \in [0,1]$. The *report* is

$$S_\lambda = S_\lambda(x) = \theta + \rho(\lambda)x + \varepsilon_\lambda, \tag{1}$$

where ε_λ is a mean-zero measurement shock whose conditional distribution may depend on λ and on (θ,z) but is independent of the action choice conditional on (θ,z), and $\rho : [0,1] \to [0,1]$ is continuous and weakly decreasing with $\rho(0) = 1$ and $\rho(1) = 0$. The parameter λ indexes the degree of purification, and $\rho(\lambda)$ is the report's residual causal *exposure* to the targeted gaming margin. A designer changes ρ through the formula, verification regime, or residualisation rule used to construct the statistic: if a raw metric is $S^{raw} = \theta + x + \varepsilon$ and the designer observes a noisy proxy $G = x + u$ for the gameable component, then $S_\lambda = S^{raw} - \lambda G = \theta + (1-\lambda)x + (\varepsilon - \lambda u)$, so $\rho(\lambda) = 1 - \lambda$ while purification also changes residual measurement noise. Allowing the distribution of ε_λ to vary with λ captures the fact that better purification need not be free.

The timing is standard: the reporting rule λ is fixed and publicly known; each agent observes (θ,z) and chooses x; the report is realised; the market forms posterior beliefs; agents and positions are matched.

A *posterior rule* is a Borel measurable function $m_\lambda : \mathbb{R} \to \mathbb{R}$. For type (θ,z), the expected market payoff from choosing action x is

$$R_\lambda(x;\theta,z) = \mathbb{E}[Q(m_\lambda(S_\lambda)) \mid \theta, z, x]\,. \tag{2}$$

Definition 1 (Equilibrium)**.** Fix λ and a matching schedule Q. An equilibrium consists of a measurable strategy $\sigma_\lambda(\theta,z)$ and a posterior rule m_λ such that (i) for almost every type,

[8] If an application has an action with both productive and distortive components, x is the non-quality component after the productive part has been netted out.

$\sigma_\lambda(\theta, z)$ maximises $R_\lambda(x; \theta, z) + v(x, z)$ over $x \in \mathbb{R}$, and (ii) the posterior rule is Bayesian given the population strategy: $m_\lambda(s) = \mathbb{E}[\theta \mid S_\lambda = s, \lambda, \sigma_\lambda]$ for equilibrium-almost every report s.

Each agent is infinitesimal and takes the posterior rule and the schedule as given. Values of m_λ at off-path reports are part of the assessment, exactly as in signalling models.[9]

2.2 Metric gaming and complete purification

Three pieces of terminology organise the analysis. A reporting rule λ is *(locally) gameable* at action x for type (θ, z) if $\partial R_\lambda(x; \theta, z)/\partial x \neq 0$, and *(locally) upward gameable* if that derivative is strictly positive. A reporting rule is *gaming-neutral* if, for almost every payoff-relevant type, $R_\lambda(x; \theta, z)$ is constant in x on the relevant action set, and *strongly gaming-neutral* if, for almost every type and all relevant x, x', the conditional law of the report is unchanged: $\mathcal{L}(S_\lambda \mid \theta, z, x) = \mathcal{L}(S_\lambda \mid \theta, z, x')$. Strong neutrality is a property of the report-generating process; gaming neutrality is a property of incentives in the market where the report is used. The former implies the latter for every posterior rule and every schedule (Lemma A.1, Appendix A), which is what lets the complete-purification endpoint be evaluated without reference to beliefs or to the downstream market. Finally, a family of reporting rules $(S_\lambda)_{\lambda \in [0,1]}$ is a *complete purification family* if there exists a regular equilibrium assessment at $\lambda = 0$ under which S_0 is locally gameable for a positive-measure set of payoff-relevant types, and the endpoint report S_1 is strongly gaming-neutral, so that the endpoint can be evaluated before any equilibrium analysis.

Corollary 1 (Complete purification removes the targeted gaming margin)**.** *At $\lambda = 1$ the report $S_1 = \theta + \varepsilon_1$ is strongly gaming-neutral, and therefore gaming-neutral for every posterior rule and every matching schedule. Under Assumption 1, for almost every type, the unique optimal action at $\lambda = 1$ is $x = z$: equilibrium gaming is* 0.

Corollary 1 is the endpoint against which partial reforms are measured.[10] Everything that follows concerns the interior, where exposure is positive, beliefs respond to the reform, and the equilibrium can move against the designer.

[9] The general results below are conditional on an equilibrium and a selected assessment regular on the reports reached by the deviations evaluated; no general existence theorem is asserted. Section 3 characterises equilibria explicitly, and the closed markets of Sections 4–5 deliver uniqueness.

[10] For the proof: since $\rho(1) = 0$, the conditional law of $S_1 = \theta + \varepsilon_1$ given (θ, z) does not depend on the action; Lemma A.1 upgrades strong neutrality to gaming neutrality, the market term in the objective is constant in x, and Assumption 1 gives $x = z$.

2.3 The marginal incentive and the price of report movement

The paper's organising identity factors an individual agent's incentive to game into three terms.

Proposition 1 (Exposure–sensitivity decomposition)**.** *Under the regularity conditions of Assumption A.1 (Appendix A), for every x in the interior of the feasible interval,*

$$\frac{\partial R_\lambda(x;\theta,z)}{\partial x} = \underbrace{\rho(\lambda)}_{\textit{exposure}}\ \mathbb{E}\left[\underbrace{m'_\lambda(S_\lambda)}_{\textit{posterior sensitivity}}\ \underbrace{Q'(m_\lambda(S_\lambda))}_{\textit{matching slope}}\ \Big|\ \theta,z,x\right]. \tag{3}$$

The first factor is technological: it is what the reform controls. The product inside the expectation — posterior sensitivity times matching slope — is a market outcome: it is the value, to the agent, of moving the report by 1 unit. We call it the *price of report movement*. The two market structures studied in the sequel share everything in (3) except this object: paying for quality makes it trust — a Bayesian regression coefficient — times a fixed price of posterior quality; paying for rank, Section 4 shows, collapses it to the density of competing reports, from which trust cancels out entirely. The contest between exposure and the price of report movement, along the reform path, is the subject of the next three sections.

3 The Gaussian-Quadratic Market and the Quality-Priced Benchmark

We now impose a parametric structure that yields closed-form equilibria while retaining two policy-relevant features: the privately preferred action may be correlated with latent quality, and purification may change residual measurement noise. The matching payoff in this section is a local price schedule with free slope and curvature; Sections 4 and 5 derive both parameters from market structure. The present section's results should thus be read as the analysis of any market in which the slope is pinned by an outside price.

3.1 Specification

Assumption 2 (Gaussian types and noise)**.** The pair (θ, z) is jointly normal with mean 0, variances $\sigma_\theta^2 > 0$ and $\sigma_z^2 \geq 0$, and covariance $\sigma_{\theta z}$ with $(\sigma_{\theta z})^2 \leq \sigma_\theta^2 \sigma_z^2$. For each λ, the shock ε_λ in (1) is normal with mean 0 and variance $\sigma_\varepsilon^2(\lambda) \geq 0$, independent of (θ, z). Write

$r_\lambda = \rho(\lambda)$ and

$$C_\lambda = \sigma_\theta^2 + r_\lambda \sigma_{\theta z}, \qquad V_\lambda^{\mathrm{o}} = \sigma_\theta^2 + 2 r_\lambda \sigma_{\theta z} + r_\lambda^2 \sigma_z^2, \qquad V_\lambda = V_\lambda^{\mathrm{o}} + \sigma_\varepsilon^2(\lambda), \tag{4}$$

so that $C_\lambda = \mathrm{Cov}(\theta, \theta + r_\lambda z)$ and $V_\lambda^{\mathrm{o}} = \mathrm{Var}(\theta + r_\lambda z)$. For every λ the report has positive variance and is positively informative: $V_\lambda > 0$ and $C_\lambda > 0$.

Joint normality makes beliefs tractable: posteriors are linear in the report, so the market's response is summarised by a single regression coefficient. Two restrictions recur and deserve names. Gaming ability is *uninformative* about agents' quality when $\sigma_{\theta z} = 0$: the privately preferred action carries no information about latent quality, so the gameable component enters the report as pure contamination. The reporting family is *noise-stationary* when $\sigma_\varepsilon^2(\lambda) \equiv \sigma_\varepsilon^2$, and *linear* when $\rho(\lambda) = 1 - \lambda$.

Assumption 3 (Quadratic private payoffs and quadratic matching payoff)**.** The agent's non-market payoff is $v(x, z) = -\frac{\kappa}{2}(x - z)^2$ with $\kappa > 0$, and the matching payoff is

$$Q(m) = \beta m + \frac{\gamma}{2} m^2, \qquad \beta > 0, \quad \gamma \in \mathbb{R}.$$

Here $\beta = Q'(0)$ is the local price of posterior quality and $\gamma = Q''(0)$ says whether that price is rising or falling in the posterior; the affine case $\gamma = 0$ is the knife-edge. The quadratic is the payoff primitive of this section: agents optimise against it exactly, in the linear-quadratic tradition.[11]

Assumption 4 (Selected Gaussian posterior assessment)**.** For every conjectured affine strategy, the selected posterior assessment is, on the support of every conditional signal law generated by a feasible deviation, the affine Bayesian version

$$m_\lambda(s) = W\big(s - \mathbb{E}[S_\lambda]\big), \qquad W = \frac{\mathrm{Cov}(\theta, S_\lambda)}{\mathrm{Var}(S_\lambda)}.$$

If $\sigma_\varepsilon^2(\lambda) > 0$ this convention is payoff-equivalent to any Bayesian version that agrees almost everywhere; if $\sigma_\varepsilon^2(\lambda) = 0$ it fixes off-path beliefs, exactly as in signalling models.

The selection commits the market to one belief function: every deviation is evaluated against the same linear regression posterior, on and off the equilibrium path, so beliefs can neither manufacture nor destroy equilibria.

[11]For $\gamma \neq 0$ it is not globally increasing, so it is to be read as the second-order expansion of an increasing schedule about the prior mean of posteriors; Supplemental Appendix D makes that reading a theorem, with approximation errors one order smaller than the curvature effects of interest.

3.2 Affine equilibria

An *affine equilibrium* at reporting rule λ is an equilibrium with action strategy $x_\lambda(\theta, z) = \alpha_0 + \alpha_\theta \theta + \alpha_z z$; it is a *translation equilibrium* if $\alpha_\theta = 0$ and $\alpha_z = 1$, so that $x_\lambda(z) = z + \Delta_\lambda$ for a constant shift Δ_λ. The translation case is the benchmark to keep in view: when the equilibrium response is a common shift, the market simply differences it out when forming beliefs, no agent's relative standing moves, and the gaming is pure waste — a rat race rather than a re-ranking. Affine conjectures reproduce themselves: against a linear posterior, the quadratic objective has an affine best response, and Bayesian consistency then reduces the model to a single scalar unknown.

Proposition 2 (Characterisation of selected-assessment affine equilibria)**.** *Under Assumptions 2, 3, and 4, fix λ with $r_\lambda > 0$, let W be the slope of the linear equilibrium posterior, and write $D_\lambda = \kappa - \gamma W^2 r_\lambda^2$. If $D_\lambda > 0$, the unique best response to that posterior is affine, with*

$$\alpha_z = \frac{\kappa}{D_\lambda}, \qquad \alpha_\theta = \frac{\gamma W^2 r_\lambda}{D_\lambda}, \qquad \alpha_0 = \frac{\beta W r_\lambda}{\kappa}. \tag{5}$$

The induced report, net of its mean, is

$$S_\lambda - \mathbb{E}[S_\lambda] = g_\theta\,(\theta + r_\lambda z) + \varepsilon_\lambda, \qquad g_\theta = 1 + r_\lambda \alpha_\theta = \frac{\kappa}{D_\lambda}, \tag{6}$$

the equilibrium distortion loads on the same composite index as the signal,

$$a_\lambda(\theta, z) = x_\lambda(\theta, z) - z = \alpha_0 + \alpha_\theta(\theta + r_\lambda z), \tag{7}$$

and Bayesian consistency delivers the scalar fixed point

$$W = \frac{g_\theta\, C_\lambda}{g_\theta^2\, V_\lambda^{\circ} + \sigma_\varepsilon^2(\lambda)}, \qquad g_\theta = \frac{\kappa}{\kappa - \gamma W^2 r_\lambda^2}. \tag{8}$$

The selected-assessment affine equilibria at λ are exactly those induced by the solutions W of (8) *satisfying $D_\lambda > 0$, each with coefficients* (5)*.*

The interpretation is a feedback loop between two scalar objects. Given trust W, individual optimality determines how agents lean into the curvature of the schedule: a convex schedule rewards being believed to be very good more than proportionally, so higher-quality agents game harder, and the report's loading on quality is amplified, $g_\theta > 1$; a concave schedule compresses rewards at the top and yields $g_\theta < 1$. Given g_θ, Bayesian consistency determines how much trust a report so loaded earns. We refer to g_θ as the *amplification*.

Existence is general; uniqueness is not. For every γ an affine equilibrium exists, with $g_\theta \in (0,1)$ under a concave schedule and $g_\theta > 1$ under a convex one, and near the knife edge it is unique and depends smoothly on γ, with $g_\theta(0) = 1$ and $W(0) = W_\lambda := C_\lambda / V_\lambda$; away from sufficient conditions recorded in Supplemental Appendix F, a curved schedule can support up to three affine equilibria. When we refer to *the* affine equilibrium away from the knife edge, we mean the *principal branch*: the unique smooth continuation of the knife-edge equilibrium through $\gamma = 0$. It is the only branch that extends to complete purification — every other root appears through an interior fold of the reform path — hence the unique selection under which comparative statics are continuous in λ from the purified endpoint, and it is the unique equilibrium at every configuration plotted in the paper.[12]

3.3 The knife-edge: quality-priced gaming and the backfire

When the schedule is locally linear, the quality-dependent tilt disappears and the equilibrium is a common shift. Setting $\gamma = 0$ gives $D_\lambda = \kappa$, $g_\theta = 1$, and $\alpha_\theta = 0$; the fixed point (8) has a unique solution, and the unique selected-assessment affine equilibrium is the translation equilibrium

$$x_\lambda(z) = z + \Delta_\lambda, \qquad \Delta_\lambda = \frac{\beta}{\kappa} r_\lambda W_\lambda = \frac{\beta}{\kappa} \frac{r_\lambda(\sigma_\theta^2 + r_\lambda \sigma_{\theta z})}{V_\lambda}, \qquad W_\lambda = \frac{C_\lambda}{V_\lambda}. \tag{9}$$

The knife-edge is exact: for any interior exposure and informative signal, $g_\theta = 1$ if and only if $\gamma = 0$.[13]

In (9), gaming equals exposure times trust times the price β, divided by the marginal cost κ. The equilibrium is the joint solution of two conditions. The first is an *incentive locus* from the individual best response, relating trust W to a candidate distortion, $\Delta = x_\lambda(z) - z$,

$$W = \frac{\kappa}{\beta r_\lambda} \Delta, \tag{10}$$

a ray through the origin in (Δ, W) space, rotating steeper as exposure falls. The second is a

[12] Supplemental Appendix F states and proves the existence and uniqueness result. Closing the market removes the issue altogether: the rank market of Section 4 has a globally unique equilibrium for all primitives, and under transfers uniqueness fails only when measurement noise exceeds roughly forty-five times the report's type variance (Proposition 5).

[13] Immediate from Proposition 2: $D_\lambda = \kappa$ gives $\alpha_z = 1$, $\alpha_\theta = 0$, $g_\theta = 1$, and $\alpha_0 = \beta r_\lambda W / \kappa$, while (8) reduces to $W = C_\lambda / V_\lambda$. Conversely $g_\theta = 1$ forces $\gamma W^2 r_\lambda^2 = 0$, hence $\gamma = 0$ when $r_\lambda > 0$ and $W \neq 0$.

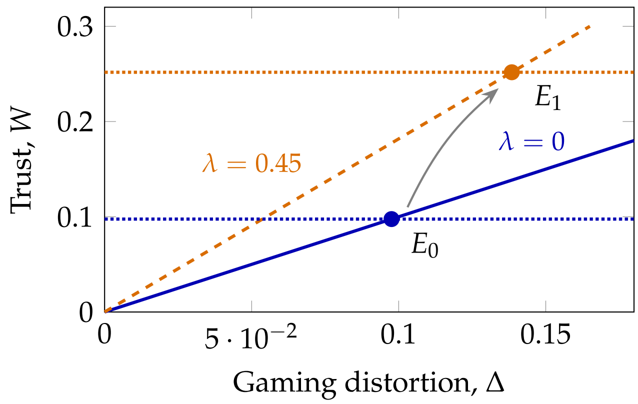

Figure 1: The knife-edge backfire: partial purification rotates the incentive locus and lifts the trust schedule; here the trust shift dominates, and equilibrium gaming rises from E_0 to E_1.

trust schedule from Bayesian consistency, written here for uninformative ability,

$$W_\lambda = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\varepsilon^2 + r_\lambda^2 \sigma_z^2}, \tag{11}$$

a horizontal line whose height rises as exposure falls: the market trusts the purer report more. Figure 1 shows both conditions at $\lambda = 0$ and $\lambda = 0.45$: purification rotates the locus toward smaller Δ but lifts the trust schedule, and the lift can move the intersection to the *right* — the formal source of the backfire, a market rationally trusting the purer report more and feeding that trust back into the private incentive to move it.[14]

Whether the trust shift dominates is a property of primitives, and the leading case isolates it cleanly.[15]

Theorem 1 (Backfire in quality-priced markets)**.** *Suppose, in addition to Assumptions 2, 3, and 4, that $\gamma = 0$, gaming ability is uninformative, and the reporting family is noise-stationary and linear. Then*

$$\Delta_\lambda = \frac{\beta}{\kappa} \frac{\sigma_\theta^2 (1-\lambda)}{\sigma_\theta^2 + \sigma_\varepsilon^2 + (1-\lambda)^2 \sigma_z^2}. \tag{12}$$

If $\sigma_z^2 > \sigma_\theta^2 + \sigma_\varepsilon^2$, then Δ_λ is strictly increasing on $[0, \lambda^*)$, strictly decreasing on $(\lambda^*, 1]$, and attains its unique maximum at $\lambda^* = 1 - \sqrt{(\sigma_\theta^2 + \sigma_\varepsilon^2)/\sigma_z^2}$. If $\sigma_z^2 \leq \sigma_\theta^2 + \sigma_\varepsilon^2$, then Δ_λ is weakly decreasing on $[0, 1]$.

Theorem 1 is the quality-priced backfire result. Its scope is any market in which the price of posterior quality, β, is held fixed along the reform: a population small relative to

[14]Parameters in Figure 1: $\sigma_\theta^2 = 1$, $\sigma_z^2 = 9$, $\sigma_\varepsilon^2 = 0.25$, $\beta/\kappa = 1$.

[15]Supplemental Appendix G records the general derivative of Δ_λ along the reform path, allowing informative ability and noise injection, together with the converse result that globally affine posterior rules force affine strategies.

the destination market pricing it, a payment schedule indexed to the posterior, or — as Section 5 shows — transferable-utility matching, where stability itself pins β to an invariant prize level. Within that scope, the condition is exact: purification backfires precisely when gameable variation dominates the metric, $\sigma_z^2 > \sigma_\theta^2 + \sigma_\varepsilon^2$, which is when the raw metric was so heavily discounted that partial purification buys back trust faster than it removes exposure to gaming. The mechanics sit in the product $r_\lambda W_\lambda$: purification lowers exposure but raises trust, and when σ_z^2 is large a modest reduction in r_λ raises trust by more than it lowers exposure, so the product rises; once r_λ is small, further purification mainly removes exposure and gaming falls, the marginal return reaching 0 at $r_\lambda = 0$ regardless of trust. In elasticity terms, exposure falls with elasticity 1 while trust — a covariance over a variance — rises with elasticity up to 2: it is because trust responds to the report's *variance* that the race can be lost, a point that becomes the hinge of the analysis once stakes are formed inside the market.

Whatever the market nets out on average, each agent bears the private cost of distorting their action.

Corollary 2 (Action-cost loss)**.** *Under Assumptions 2, 3, and 4, the private action-cost loss in the affine equilibrium is*

$$L_\lambda = \mathbb{E}\big[v(z,z) - v(x_\lambda(\theta,z),z)\big] = \tfrac{\kappa}{2}\,\mathbb{E}\big[a_\lambda(\theta,z)^2\big] = \underbrace{\tfrac{\kappa}{2}\,\alpha_0^2}_{\substack{\text{common-shift}\\ \text{cost}}} + \underbrace{\tfrac{\kappa}{2}\,\alpha_\theta^2\,V_\lambda^{\mathrm{o}}}_{\substack{\text{curvature-induced}\\ \text{dispersion}}}. \tag{13}$$

At the knife-edge, the second term vanishes and $L_\lambda = \frac{\kappa}{2}\Delta_\lambda^2$ inherits the monotonicity of Theorem 1.

The loss in (13) is an action-cost loss, not total welfare.[16] The common-shift component is anticipated by the market and buys no average placement; the curvature-induced dispersion is the cost of the quality-dependent tilt that moves sorting.

4 Matching Markets that Pay for Rank

The analysis so far treats the local price β and curvature γ as primitives. In a closed matching market, neither is free. When positions cannot pay transfers, stable matching in the balanced market is assortative in posterior rank, and an agent with posterior expected quality m receives the position at the corresponding desirability quantile, $Q_\lambda(m) = H^{-1}(F_\lambda(m))$,

[16] For the proof: by (7), $\mathbb{E}[a_\lambda^2] = \alpha_0^2 + \alpha_\theta^2 V_\lambda^{\mathrm{o}}$, and substituting the quadratic payoff gives (13). The loss omits sorting effects, which Section 6 prices; in the closed market, the two will be denominated in the same units.

where F_λ is the equilibrium distribution of posteriors (Lemma C.1, Supplemental Appendix C; balanced case used throughout this section). The schedule the agents game against is, therefore, manufactured from the distribution of posteriors — which the reporting rule shapes — and the distribution of prizes. This section completes that loop, and the solution answers two questions the reduced form leaves open: what slope and curvature a rank market actually generates, and whether the backfire of Theorem 1 survives when stakes are determined inside the market.

Throughout this section, the market is assumed to be balanced, Assumptions 2–4 are maintained, the family is noise-stationary with $\sigma_\varepsilon^2(\lambda) \equiv \eta > 0$, and we parameterise by residual exposure $r_\lambda \in [0,1]$. In a selected-assessment affine equilibrium, the posterior is the linear rule $m_\lambda(S_\lambda) = W(S_\lambda - \mathbb{E}[S_\lambda])$ with $W > 0$ — the maintained informativeness $C_\lambda > 0$ and $g_\theta > 0$ make the posterior a strictly increasing transform of the report — so the posterior itself is Gaussian,

$$m_\lambda(S_\lambda) \sim N(0, I_\lambda), \qquad I_\lambda = \operatorname{Var}\big(\mathbb{E}[\theta \mid S_\lambda]\big) = W g_\theta C_\lambda,$$

where the second equality follows from $W = g_\theta C_\lambda / \operatorname{Var}(S_\lambda)$. The induced schedule inherits its local shape from this Gaussian posterior distribution and from the prize density h, which we require to be continuously differentiable and strictly positive at the median prize $q^{\text{med}} = H^{-1}(\frac{1}{2})$.

4.1 The induced schedule and the rank price

The first step computes the schedule the market manufactures: its slope, which is what a unit of report movement buys, and its curvature, which will determine how equilibrium gaming loads on quality.

Lemma 1 (Induced slope, curvature, and the rank price)**.** *In the balanced closed market without transfers, in any selected-assessment affine equilibrium, the schedule $Q_\lambda = H^{-1} \circ F_\lambda$ is twice continuously differentiable at the prior mean $m = 0$, with*

$$\beta_\lambda = Q_\lambda'(0) = \frac{1}{h(q^{\text{med}})\sqrt{2\pi I_\lambda}}, \qquad \gamma_\lambda = Q_\lambda''(0) = -\beta_\lambda^2 \, \frac{h'(q^{\text{med}})}{h(q^{\text{med}})}. \tag{14}$$

Consequently the combination

$$\gamma_\lambda \, I_\lambda = \chi_H := -\frac{h'(q^{\text{med}})}{2\pi \, h(q^{\text{med}})^3} \tag{15}$$

is a primitive of the prize distribution alone, invariant along the reform path. Moreover,

$$\beta_\lambda W = \frac{1}{h(q^{\mathrm{med}})\sqrt{2\pi\,\mathrm{Var}(S_\lambda)}} : \tag{16}$$

the price of report movement depends on beliefs only through the variance of the report.

Each part of Lemma 1 carries an economically meaningful force. The slope β_λ is the *rank price*: the number of position quantiles a unit of posterior quality buys, times the prize spacing per quantile. It falls with I_λ — the *rank-price feedback*: a better-informed market spreads posteriors, so any given posterior improvement climbs over fewer competitors and buys a cheaper upgrade. No counterpart exists in Section 3, where β is fixed.

The curvature result (15) says that what the matching schedule does to gaming incentives is, locally, a property of prize inequality. We call χ_H the *prize-curvature index*: it is positive exactly when the density of desirability is decreasing at its median, so that prizes thin out as an agent climbs — the winner-take-most case — and negative when positions bunch at the top.[17] The identity $\gamma_\lambda I_\lambda = \chi_H$ is the bookkeeping that makes the fixed point below scalar: the market manufactures exactly as much curvature as its informativeness permits. The identity (16) is the deepest of the three and drives the gaming results. The posterior is a monotone transform of the report, so ranking agents by posterior is the same as ranking them by report: the cardinal scale of beliefs — the object the trust channel of Theorem 1 operates on — cancels from the marginal incentive entirely. What replaces it is the density of competing reports at the agent's position. In a rank tournament, the market's trust in the metric is irrelevant to the private return from gaming it; what matters is how crowded the report distribution is.[18]

4.2 Equilibrium

Because the matching schedule is now an equilibrium object, the fixed point of Proposition 2 acquires one more loop: strategies determine posteriors, posteriors determine the schedule's local shape through Lemma 1, and the shape feeds back into strategies. The

[17] For exponential prizes with $1 - H(q) = e^{-\xi q}$, the index is $\chi_H = 2/(\pi\xi)$, proportional to the standard deviation of desirability; for Pareto prizes with scale q_m and tail index a it equals $2(a+1)2^{1/a}q_m/(\pi a^2)$, again growing with dispersion.

[18] This is the prize-spread-times-rival-density first-order condition of Lazear and Rosen (1981), with the Bayesian layer cancelling because ranking by posterior is ranking by report; see Olszewski and Siegel (2016) for the continuum as a large-contest limit and Brown (2011) for field evidence that competitor strength affects performance incentives. Prize-distribution shape is contest theory's canonical design instrument for effort (Moldovanu and Sela, 2001; Olszewski and Siegel, 2020); related work studies the risk-taking and selection effects generated by contest incentives (Hvide and Kristiansen, 2003). Our index χ_H instead uses the local shape of the prize distribution to characterise the gaming response (Section 6).

natural solution concept evaluates the quadratic model at the slope and curvature the market itself generates.[19]

Definition 2 (First-order closed-market equilibrium)**.** Fix λ with $r_\lambda > 0$. A *first-order closed-market equilibrium* is a selected-assessment affine equilibrium of the Gaussian-quadratic model with payoff parameters $(\beta, \gamma) = (\beta_\lambda, \gamma_\lambda)$ such that β_λ and γ_λ are the slope and curvature of the schedule induced, through the market's allocation protocol, by that same equilibrium's posterior distribution.

The definition asks agents to optimise against the second-order expansion of the schedule their own behaviour creates, with the expansion point — the median — consistent with the posteriors they generate.[20] Under the rank protocol the induced coefficients are (14); Section 5 applies the same definition under transfers. The next result shows that the fixed point in $(\beta, \gamma, W, g_\theta)$ collapses to one scalar equation, and that closing the market is a force for uniqueness rather than against it.

Proposition 3 (The rank market solved)**.** *Let the prize density satisfy $\chi_H > 0$ and fix λ with $r_\lambda > 0$. Write $b_\lambda = \chi_H r_\lambda^2/(\kappa C_\lambda)$. In any first-order closed-market equilibrium under the rank protocol, the amplification is linear in trust,*

$$g_\theta = 1 + b_\lambda W, \tag{17}$$

and W solves the cubic

$$b_\lambda^2 V_\lambda^{\mathrm{o}} W^3 + 2b_\lambda V_\lambda^{\mathrm{o}} W^2 + \big(V_\lambda^{\mathrm{o}} + \eta - b_\lambda C_\lambda\big)W - C_\lambda = 0. \tag{18}$$

1. (Existence and global uniqueness.) *The cubic* (18) *has exactly one positive root, for every parameter configuration; it identifies the unique first-order closed-market equilibrium, which satisfies $D_\lambda = \kappa/g_\theta > 0$ and lies on the principal branch.*

2. (Endogenous convexity.) *The equilibrium has $g_\theta > 1$ strictly: with a median-thinning prize density, the rank market's schedule is locally convex at every interior exposure, and the linear-price case $g_\theta = 1$ is unattainable.*

[19] Supplemental Appendix D licenses this; the affine strategies are within roughly 1% of optimal at every plotted configuration.

[20] The closed-market results of this and the next section are exact characterisations of this first-order object, not of equilibrium under the exact nonlinear schedule, whose existence we do not claim (Supplemental Appendix D licenses the approximation). The definition fixes $r_\lambda > 0$; every closed-market object below extends to $r_\lambda = 0$ by continuity of the principal branch.

3. (Explicit formulas.) $I_\lambda = W g_\theta C_\lambda$, *the schedule coefficients are* (14), *and the equilibrium distortion has*

$$\alpha_0 = \frac{r_\lambda}{\kappa\, h(q^{\mathrm{med}})\sqrt{2\pi\big(g_\theta^2 V_\lambda^{\mathrm{o}} + \eta\big)}}, \qquad \alpha_\theta = \frac{g_\theta - 1}{r_\lambda} = \frac{\chi_H\, r_\lambda}{\kappa\, C_\lambda}\, W. \tag{19}$$

Two features deserve comment. First, uniqueness here is global and unconditional, in contrast to the fixed-(β,γ) model, where uniqueness beyond the knife edge required side conditions and could genuinely fail (Supplemental Appendix F). The discipline comes from the market's budget constraint on prizes: in the reduced form, conjectured trust and conjectured curvature can amplify one another because the stakes are exogenous, whereas in the rank market, a more trusted report spreads posteriors, deflates the rank price through Lemma 1, and caps how much amplification any level of trust can finance. Multiplicity in the reduced form is thus partly an artefact of holding the schedule fixed. Second, part (ii) settles which case of the curvature analysis is relevant: with prizes that thin at the median, the closed market is always strictly inside the convex regime. The knife-edge $\gamma = 0$ is not merely non-generic; it is unreachable. Along the reform path the equilibrium amplification is strictly increasing in exposure and falls to its knife-edge value $g_\theta = 1$ as the metric is purified, and the common-shift distortion falls monotonically with it (Theorem 2). The force is the rank price (16), which depends on beliefs only through the report's standard deviation and so deflates monotonically as purification concentrates the report distribution; trust W itself need not be monotone once gaming ability is informative.[21]

4.3 No backfire under rank pricing

Theorem 1 showed that, with a fixed price of posterior quality, partial purification can raise equilibrium gaming. The rank market changes the third factor in the decomposition (3), and the change is fatal to the backfire.

Theorem 2 (Rank-price feedback: no backfire in the rank market)**.** *Let $\chi_H > 0$. In the first-order closed-market equilibrium under the rank protocol:*

1. (Common shift.) *The common-shift distortion $\alpha_0(r_\lambda)$ in* (19) *is strictly increasing in exposure on $(0, 1]$, for every parameter configuration $(\sigma_\theta^2, \sigma_z^2, \sigma_{\theta z}, \eta, \kappa)$ admitted by Assumption 2 — informative gaming ability included — and every prize density with $\chi_H > 0$.*

[21] The leading trust response at complete purification is $W'(r_\lambda)\big|_{r_\lambda=0} = \sigma_{\theta z}(\eta - \sigma_\theta^2)/(\sigma_\theta^2 + \eta)^2$, so a purer report earns strictly *less* trust whenever $\sigma_{\theta z}$ and $\eta - \sigma_\theta^2$ share a sign—for instance at $\sigma_\theta^2 = \sigma_z^2 = 1$, $\sigma_{\theta z} = \frac{1}{2}$, $\eta = 2$, where W rises from $\frac{1}{3}$ at $r_\lambda = 0$. Because ranking cancels the cardinal scale of beliefs, this leaves the common-shift monotonicity of Theorem 2 intact.

2. (Dispersion and total loss.) *If in addition* $\sigma_{\theta z} \geq -\frac{2\sqrt{2}}{3}\sigma_\theta\sigma_z$ *— gaming-ability correlation is no more negative than* $-2\sqrt{2}/3 \approx -0.94$*, covering every* $\sigma_{\theta z} \geq 0$ *— the quality-loading cost* $\alpha_\theta^2 V_\lambda^{\mathrm{o}}$ *is strictly increasing in exposure on* $(0, 1]$*, and with it the total action-cost loss* L_λ *of Corollary 2 and the mean-squared distortion* $\mathbb{E}[a_\lambda^2]$.

Every step of partial purification strictly reduces the common-shift distortion — and, under the condition of part 2, the total cost of gaming: the non-monotonicity of Theorem 1 cannot arise in the balanced rank market.

The intuition is the elasticity comparison promised in Section 3, and it locates exactly where Theorem 1 lives. In the quality-priced market, the price of report movement is trust times a fixed price, and trust is a regression coefficient: a covariance divided by the report's *variance*. As purification removes gameable variation, that denominator can fall fast — with elasticity approaching 2 when gameable variation dominates — so trust can outrun the one-for-one loss of exposure. In the rank market, Lemma 1 replaces trust times price by the rank price in report space, an inverse *standard deviation*: by (16), the private return is governed by how crowded the report distribution is, and purification concentrates reports with an elasticity bounded by $\frac{1}{2}$ of the trust channel's. Ranking allocations halve the feedback from purifying the metric, and the halved feedback never outruns exposure. The curvature channel — the response of g_θ to exposure — pushes in the backfire's direction, exactly as under transfers, but the rank-price feedback caps it: the proof bounds the amplification response by $r_\lambda g_\theta g_\theta' V_\lambda^{\mathrm{o}} < g_\theta^2 C_\lambda + \eta$, so the channel narrows the monotonicity margin without ever overturning it. The dispersion margin of part 2 is the same force one derivative up: purification unwinds the amplification itself, $g_\theta \downarrow 1$, so the quality-dependent tilt — and with it the curvature-induced cost of Corollary 2 — shrinks alongside the common shift.[22]

The result sharpens, rather than retracts, the cautionary message of Theorem 1. The backfire is a *fixed-stakes* phenomenon: it belongs to environments where the value of being believed to be better is pinned outside the ordinal ranking of agents. In a balanced rank tournament — a promotion ladder, a fixed set of positions filled purely by relative standing — the rank-price feedback is automatic, and partial purification monotonically reduces gaming.[23] Between the two poles sit the rationed markets of Supplemental Appendix C:

[22] Part 1 places no restriction on $\sigma_{\theta z}$: ranking cancels the cardinal scale of beliefs, so correlated gaming ability enters the common-shift margin only through $C_\lambda > 0$. The dispersion margin is sharper — the quality-loading cost rises with exposure exactly when $C_\lambda V_\lambda^{\mathrm{o}} g_\theta^2 + \eta\,(2V_\lambda^{\mathrm{o}} g_\theta - C_\lambda) > 0$, which the correlation bound of part 2 secures by keeping $2V_\lambda^{\mathrm{o}} - C_\lambda \geq 0$ along the path. Only strong negative correlation with large noise inverts it (e.g. $(\sigma_\theta^2, \sigma_z^2, \sigma_{\theta z}, \eta) = (1, 1, -0.96, 50)$, $\chi_H = \kappa$, for $r_\lambda \in [0.6, 0.85]$); part 1 is unaffected even then.

[23] Part 1 is relative-performance-evaluation logic (Holmström, 1982; Green and Stokey, 1983) taken one

with scarce positions, the extensive-margin cutoff term attaches a locally fixed stake — the value of being matched at all — to a density.[24] The next section makes the two poles precise and locates them on a single frontier.

5 Markets that Pay for Quality, and the Backfire Frontier

We have now produced two answers to its opening question: in the quality-priced benchmark, purification backfires whenever gameable variation dominates the metric (Theorem 1); in the rank market, it never does (Theorem 2). This section isolates the object that separates them, characterises the frontier between backfiring and safe stakes processes, and then microfounds the quality-priced regime inside the same closed matching market as Section 4 — so that the two theorems differ, in the end, only in whether positions can pay transfers.

5.1 Stakes processes and the backfire threshold

A *stakes process* assigns to each reporting rule a market payoff that is affine in the agent's report, with slope $P\big(\mathrm{Var}(S_\lambda)\big)$, where $P : (0,\infty) \to (0,\infty)$ is continuously differentiable and weakly decreasing; its *stakes-deflation elasticity* is

$$\psi(V) = -\frac{d\ln P(V)}{d\ln V} \ \geq 0.$$

The *translation environment* is the knife-edge specification with uninformative ability and noise-stationary measurement, so that in the translation equilibrium, the report variance is $V(r_\lambda) = \sigma_\theta^2 + \eta + r_\lambda^2\sigma_z^2$ and each price-taking agent's first-order condition gives

$$\Delta_\lambda = \frac{r_\lambda}{\kappa}\, P\big(V(r_\lambda)\big). \tag{20}$$

This is a reduced form, and deliberately so: it strips the market down to the one question that matters for the backfire, namely how fast the price of report movement deflates

reform-derivative up, and its contractual kinship is Malcomson (1984): rank contracts never consult the cardinal scale, so the recalibrated trust that drives Theorem 1 is incentive-inert. In the designed contest of Barlevy and Neal (2012) a common-scale manipulation earns nothing even privately; here it retains a positive private return while remaining collectively futile, which is why the rank market still games in equilibrium yet never games more as the metric is purified.

[24]Supplemental Appendix C.2 resolves what this implies: to first order in stakes, the aggregate distortion is still governed by the population density at the cutoff — itself a rank price — and falls monotonically, while at large admission stakes the equilibrium feedback revives a modest interior peak in mean gaming, so scarcity is the one margin on which a rank market partially inherits the backfire.

as the report becomes more informative. Because a common shift moves the mean of the report distribution but not its variance, $V(r_\lambda)$ is invariant to Δ_λ and (20) is an exact translation equilibrium for every stakes process; no fixed-point machinery is needed. A single elasticity then decides the comparative static.

Proposition 4 (The backfire frontier)**.** *In the translation environment with stakes process P and $\sigma_z^2 > 0$:*

1. *Δ_λ is strictly increasing in exposure at r_λ if $2\psi\big(V(r_\lambda)\big)\,\sigma_z^2 r_\lambda^2 / V(r_\lambda) < 1$, and gaming is non-monotone along the reform path if and only if the reverse strict inequality holds for some $r_\lambda \in (0, 1]$.*

2. *If ψ is constant, gaming is non-monotone if and only if*
$$\psi > \psi^* := \frac{1}{2} + \frac{\sigma_\theta^2 + \eta}{2\sigma_z^2},$$
in which case Δ_λ has a unique interior peak at $r^* = \sqrt{(\sigma_\theta^2 + \eta)/\big((2\psi - 1)\sigma_z^2\big)}$. As ψ falls to ψ^*, the peak exits through the raw end of the path ($r^* \to 1$).

3. *$\psi^* > \frac{1}{2}$ for all primitives. Hence, any stakes process with $\psi \le \frac{1}{2}$ everywhere induces gaming that is monotone in exposure for* every *parameter configuration, whereas for any $\psi > \frac{1}{2}$ there are primitives under which purification backfires.* Quality pricing*, $P(V) = \beta\sigma_\theta^2/V$ with $\psi = 1$, backfires exactly when $\sigma_z^2 > \sigma_\theta^2 + \eta$, recovering Theorem 1;* rank pricing*, $P(V) \propto V^{-1/2}$ with $\psi = \frac{1}{2}$, never backfires.*

Proof. Write $s = r_\lambda^2$. From (20), $\frac{d \ln \Delta}{d \ln r} = 1 - 2\psi(V)\sigma_z^2 s / V(s)$, which gives part (i); the supremum of $\sigma_z^2 s/V(s)$ over $s \in (0, 1]$ is attained at $s = 1$ because $s/V(s)$ is increasing. For constant ψ, the sign of the derivative is the sign of $\sigma_\theta^2 + \eta + (1 - 2\psi)\sigma_z^2 s$, which is positive everywhere when $\psi \le \frac{1}{2}$ and otherwise crosses 0 exactly once, at $s^* = (\sigma_\theta^2 + \eta)/\big((2\psi - 1)\sigma_z^2\big)$; the crossing lies inside $(0, 1)$ if and only if $2\psi\sigma_z^2/(\sigma_\theta^2 + \eta + \sigma_z^2) > 1$, which rearranges to $\psi > \psi^*$; at $\psi = \psi^*$ the crossing sits at the endpoint $s^* = 1$ and Δ_λ is still monotone on the path. Part (iii): $\psi^* - \frac{1}{2} = (\sigma_\theta^2 + \eta)/(2\sigma_z^2) > 0$; for $\psi = 1$ the condition $\psi > \psi^*$ rearranges to $\sigma_z^2 > \sigma_\theta^2 + \eta$; for $\psi = \frac{1}{2} \le \psi^*$ monotonicity follows from part (i). □

The interpretation is an elasticity race with a handicap. The share of the report's variance that is gameable, $\sigma_z^2 r_\lambda^2 / V$, is bounded below 1, and the price of report movement rises with elasticity 2ψ times that share while exposure falls with elasticity 1: a backfire requires the stakes to respond to the report's variance more than half-for-half. The constant $\frac{1}{2}$ is the exponent gap between a variance and a standard deviation, and the two protocols sit exactly on either side of it; by Lemma 1 the rank market prices at $P(V) = 1/\big(h(q^{\text{med}})\sqrt{2\pi V}\big)$ — the boundary elasticity, on the safe side, for every prize density — and Theorem 2 is the exact-equilibrium counterpart with the curvature channel active. It remains to show that quality pricing is not merely a reduced form, and this is what transfers do.

5.2 Transfers price quality

We now hold the matching environment of Section 4 fixed — the same balanced market, the same prize distribution, the same action environment — and change only the allocation protocol: positions and agents trade with transfers, the surplus from matching a position of desirability q to an agent with posterior expected quality m being mq, and the allocation is stable (assortative) with its supporting wage schedule w_λ.[25]

Proposition 5 (Transferable utility prices quality)**.** *In the balanced closed market with transfers:*

1. (Wage integrates the prize schedule.) *For every prize distribution and every λ,*

$$w'_\lambda(m) = H^{-1}\big(F_\lambda(m)\big); \tag{21}$$

i.e., the marginal wage of an agent is the desirability of the position their posterior buys. Consequently, the wage schedule is globally convex for every prize distribution, with

$$\beta_\lambda^{\text{TU}} = w'_\lambda(0) = q^{\text{med}}, \qquad \gamma_\lambda^{\text{TU}} = w''_\lambda(0) = \frac{1}{h(q^{\text{med}})\sqrt{2\pi I_\lambda}},$$

so that $\gamma_\lambda^{\text{TU}}\sqrt{I_\lambda} = \chi_{\text{TU}} := 1/\big(\sqrt{2\pi}\,h(q^{\text{med}})\big)$ is a primitive of the prize distribution, invariant along the reform path. The slope is the median prize, invariant to the informativeness of the report; equivalently, the population-average marginal wage is the mean prize, $\mathbb{E}[w'_\lambda(m)] = \bar{q}$, for every λ. The median agent's stakes process is therefore $P(V) = q^{\text{med}}\,W$, which at the knife-edge is quality pricing with $\psi = 1$.

2. (Solved fixed point.) *The first-order closed-market equilibria under transfers are characterised by the roots of*

$$\mathcal{T}_\lambda(g_\theta) = (g_\theta - 1)\big(g_\theta^2 V_\lambda^{\text{o}} + \eta\big)^{3/2} - \frac{\chi_{\text{TU}}\,C_\lambda\,r_\lambda^2}{\kappa}\,g_\theta^2 = 0, \qquad g_\theta > 1. \tag{22}$$

A root exists for all primitives and every root satisfies $D_\lambda = \kappa/g_\theta > 0$; the root is unique whenever $\eta \le \varphi^* V_\lambda^{\text{o}}$, where $\varphi^* = \min_{g>2} g^2(2g-1)/(g-2) \approx 44.9$, and otherwise

[25] We normalise desirability so that the median prize is strictly positive, $q^{\text{med}} > 0$: the slope of the induced schedule below is q^{med} itself, and Assumption 3 requires a positive slope.

multiple roots can arise, in which case we select the smooth continuation of the knife-edge root (Supplemental Appendix F).

3. (Backfire under transfers.) *Let gaming ability be uninformative, $\sigma_{\theta z} = 0$. As $\chi_{\mathrm{TU}} \to 0$ with $q^{\mathrm{med}} > 0$ fixed, the equilibrium converges to the translation equilibrium with $\beta = q^{\mathrm{med}}$, so in the concentrated-prize limit gaming is non-monotone exactly when $\sigma_z^2 > \sigma_\theta^2 + \eta$, recovering Theorem 1. For sufficiently small positive χ_{TU}, uniform convergence preserves each strict case—non-monotone when $\sigma_z^2 > \sigma_\theta^2 + \eta$, monotone when $\sigma_z^2 < \sigma_\theta^2 + \eta$—while the exact backfire threshold sits strictly below the knife-edge value $\sigma_\theta^2 + \eta$, as quantified in Remark 1.*

Part 1 is the economic heart of the section. Under transferable utility, the wage schedule is the *integral* of the prize schedule: match stability forces each position to be content with its hire, so the marginal wage at posterior m equals the desirability of the position assigned there — precisely the rank market's *level* schedule. Integration shifts the information-dependence of the schedule up one derivative, out of the slope and into the curvature: the marginal stake of the median agent is the median *prize level*, fixed by what partners are worth, while under the rank protocol it is a prize *density* — the chance of leapfrogging a competitor — which deflates as beliefs spread. A level attached to the Bayesian multiplier W gives $\psi = 1$; a density gives $\psi = \frac{1}{2}$, because the posterior multiplier cancels against the posterior's own scale, as in (16); the half-unit gap is the entire difference between Theorem 1 and Theorem 2.

Part 2 records what integration does to the fixed point: the deflation now acts on the curvature rather than the slope, and this weaker feedback leaves room for multiple roots at extreme measurement noise, among which the smooth continuation of the knife-edge root is selected.

Part 3 and Remark 1 (below) add a refinement that the reduced form cannot see: under transfers, the curvature channel is an accelerant. Endogenous convexity inflates the report's variance most where exposure is high, depressing trust there and steepening the trust schedule, so non-monotonicity appears below the knife-edge threshold. In the rank market the same convexity is present — Proposition 3 shows it is unavoidable — but there it is dominated by the rank-price feedback. *Convexity widens the backfire region when stakes are priced, and cannot create one when stakes are ranked.*[26]

[26]The quadratic licence is itself prize-dependent: with pure exponential prizes the integrated wage schedule has cubic tails, the dimensionless curvature does not shrink, and the first-order model is an organising device rather than a quantitative approximation — one reason part (iii) of Proposition 5 works through the concentrated-prize limit (Supplemental Appendix D.1).

Remark 1 (Curvature expands the backfire region). Near the concentrated-prize limit, the critical contamination level $\sigma^2_{z,\mathrm{crit}}(\chi_{\mathrm{TU}})$ at which non-monotonicity appears is smooth in χ_{TU}, with

$$\sigma^2_{z,\mathrm{crit}}(\chi_{\mathrm{TU}}) = \sigma^2_\theta + \eta - \frac{\sigma^2_\theta(\sigma^2_\theta + 2\eta)}{\kappa\big(2(\sigma^2_\theta + \eta)\big)^{3/2}}\,\chi_{\mathrm{TU}} + O(\chi^2_{\mathrm{TU}}),$$

strictly below the knife-edge threshold for small $\chi_{\mathrm{TU}} > 0$ (proof in Supplemental Appendix F). Away from the limit the threshold keeps falling: at $\sigma^2_\theta = \kappa = 1$, $\eta = 0.15$ (knife-edge threshold $\sigma^2_z = 1.15$), it decreases monotonically to about 0.24 as χ_{TU} rises to 20, the local slope matching the expansion.

Remark 2 (Where actual markets sit). The characterisation turns on an observable feature of institutions: whether the reward to a better evaluation is a price or a position. Quality-priced environments — wages bargained against a score, interest rates set on credit ratings, procurement awards with score-contingent payments — have $\psi \approx 1$ and backfire when contamination is large. Rank-priced environments — admission to a fixed set of seats, promotion ladders with a fixed distribution of posts, tournaments with a fixed prize ladder — have $\psi = \frac{1}{2}$ and are immune on this margin, whatever the contamination. The same metric can face both: a purified transcript used by wage-setting employers and by seat-constrained graduate programmes is hump-prone in the first use and monotone in the second, so the aggregate gaming response to a grading reform is a mixture weighted by which downstream market the marginal student is gaming for. Two boundary cases sit outside the scalar class — rationed markets, where a fixed admission stake attaches to a density, and prize structures redesigned as informativeness improves, where ψ can exceed 1; both are taken up in the design discussion (Section 7).

Figure 2 summarises the positive analysis: panel (a) shows one market, one prize distribution, and opposite comparative statics under the two protocols; panel (b) shows the frontier falling toward $\frac{1}{2}$ as contamination grows without ever reaching it, which is why rank pricing is safe uniformly over primitives rather than merely in examples.[27] Whether making a metric less gameable reduces gaming is therefore not a property of the metric alone; it is a property of the market on the other side of it — and the relevant feature of that market is observable.

[27]Panel (a): $\sigma^2_\theta = 1$, $\sigma^2_z = 5$, $\eta = 0.15$, $\kappa = 1$, the same prize distribution in both markets ($q = 5 + \mathrm{Exp}(1)$, $\chi_{\mathrm{TU}} \approx 0.80$), each curve normalised to its own maximum; the transfer-market peak at $\lambda \approx 0.53$ sits close to the knife-edge prediction $\lambda^* \approx 0.52$, and the affine strategies are within 0.04 per cent of optimal (Supplemental Appendix D.1). Panel (b) plots the frontier ψ^* of Proposition 4.

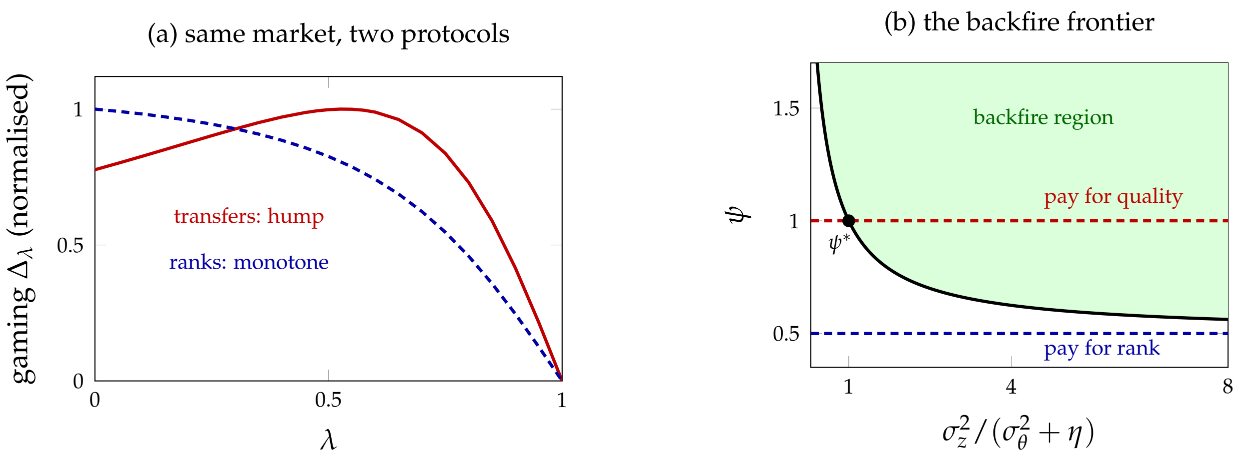

Figure 2: When purifying a metric backfires. *Panel (a)*: the common-shift distortion along the reform path in two markets that share every primitive and differ only in the allocation protocol — hump-shaped under transfers (solid), strictly decreasing under ranks (dashed). *Panel (b)*: the backfire frontier ψ^* of Proposition 4; purification backfires in the shaded region.

6 Sorting Efficiency and Welfare

The backfire results are incentive results, not welfare theorems. Whether gaming distorts the market's *ranking* is a separate question, and this section answers it for both protocols at once. The analysis applies to any selected-assessment affine equilibrium — whether the payoff coefficients (β, γ) are primitives, as in Section 3, or induced by the market, as in Sections 4 and 5 — because the sorting objects depend only on the equilibrium population strategy and the Gaussian structure.

6.1 The posterior-information index

In an affine equilibrium the rank-relevant signal is $Y = g_\theta(\theta + r_\lambda z) + \varepsilon_\lambda$, where $r_\lambda \in [0, 1]$ is residual exposure, $\eta(r_\lambda) = \mathrm{Var}(\varepsilon_\lambda)$ the noise variance, and $g_\theta \geq 0$ the equilibrium amplification.[28] It is convenient to regard the second-moment objects of Assumption 2 as functions of residual exposure: writing $C(r_\lambda) = C_\lambda$, $V^o(r_\lambda) = V^o_\lambda$, and $\eta(r_\lambda) = \sigma^2_\varepsilon(\lambda)$, we have $\mathrm{Cov}(\theta, Y) = g_\theta C(r_\lambda)$ and $\mathrm{Var}(Y) = g^2_\theta V^o(r_\lambda) + \eta(r_\lambda)$, and we assume $C(r_\lambda) > 0$. Throughout this section "reducing exposure" means moving r_λ toward 0, the direction of increasing purification; unlike Sections 4 and 5, we allow η to vary with r_λ, since reforms built from noisy proxies inject noise as they remove exposure.

For the balanced closed market, the sorting criterion is the expected product of latent

[28] Along a reform path we evaluate (g_θ, W) on the principal branch (Section 3), which in the closed markets is the unique equilibrium.

quality and the desirability of the position assigned by posterior rank:

$$\mathcal{E}(r_\lambda) = \mathbb{E}\Big[\theta\, H^{-1}(F_Y(Y))\Big]\,. \tag{23}$$

Joint normality reduces the criterion to a single scalar index.

Proposition 6 (Posterior-information index governs sorting)**.** *Consider the Gaussian-quadratic specification, a selected-assessment affine equilibrium with amplification g_θ, and the criterion* (23)*; assume the prize distribution has finite second moment. Then $K_H = \mathbb{E}[T\, H^{-1}(\Phi(T))]$, with T standard normal, is finite and strictly positive, and*

$$\mathcal{E}(r_\lambda) = K_H\, \frac{\mathrm{Cov}(\theta, Y)}{\sqrt{\mathrm{Var}(Y)}} = K_H\, \frac{g_\theta\, C(r_\lambda)}{\sqrt{g_\theta^2 V^{\mathrm{o}}(r_\lambda) + \eta(r_\lambda)}}. \tag{24}$$

Sorting efficiency is therefore ordered by the posterior-information index

$$I_\gamma(r_\lambda) = \mathrm{Var}(\mathbb{E}[\theta \mid Y]) = \frac{C(r_\lambda)^2}{V^{\mathrm{o}}(r_\lambda) + \eta(r_\lambda)/g_\theta^2}, \tag{25}$$

which is the linear-price index $I(r_\lambda) = C(r_\lambda)^2/\big(V^{\mathrm{o}}(r_\lambda) + \eta(r_\lambda)\big)$ with the noise $\eta(r_\lambda)$ replaced by $\eta(r_\lambda)/g_\theta^2$. When $\eta(r_\lambda) > 0$, $I_\gamma(r_\lambda)$ is strictly increasing in g_θ; hence sorting strictly exceeds the linear-price benchmark under a locally convex schedule ($g_\theta > 1$) and falls strictly short of it under a locally concave one ($g_\theta < 1$). At the knife-edge $\gamma = 0$, $g_\theta = 1$: the gaming response is a common shift and is rank-neutral.

The index isolates two channels. The first runs through exposure and noise and, at the knife edge, alone determines the *shape* of sorting along the reform: with uninformative gaming ability, monotone under fixed noise and interior-peaked under noise injection, while informative ability can create an interior peak of its own through collateral information, as signed below. The second runs through the amplification g_θ and determines the *level* of sorting relative to the linear case at the same exposure: a convex schedule makes high-quality agents game hardest, stretching the quality dimension against the noise floor and improving the ranking, while a concave schedule compresses it. The amplification channel is silent without measurement noise: if $\eta(r_\lambda) = 0$ then $I_\gamma = C(r_\lambda)^2/V^{\mathrm{o}}(r_\lambda)$ for every γ, because a uniform stretching is visible to ranks only against a fixed noise floor.

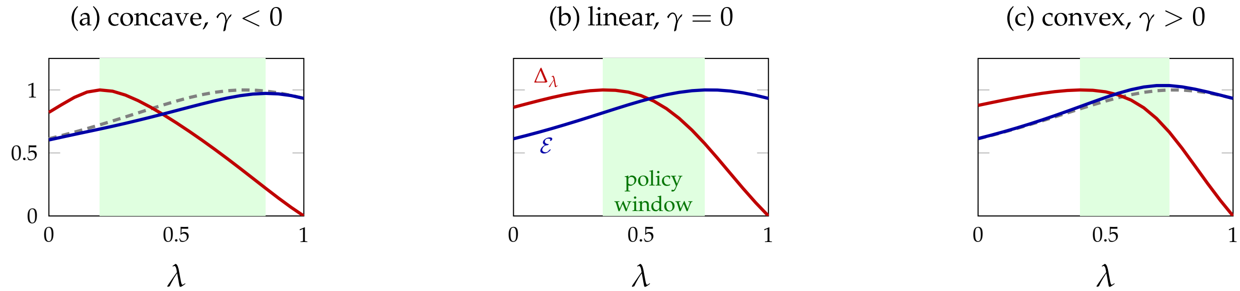

Figure 3: Sorting (blue) and gaming (red) across schedule curvature in the quality-priced market when purification injects noise; the dashed grey curve repeats the linear-case ($\gamma = 0$) sorting curve as a benchmark. sorting peaks in the interior, and the policy window (green) is bounded.

6.2 The knife edge: three named forces

At the knife-edge, the amplification channel is shut, $g_\theta = 1$, and the exposure–noise channel can be signed completely.

Suppose $\gamma = 0$ and the Gaussian distribution is nonsingular, $\sigma_\theta^2\sigma_z^2 - \sigma_{\theta z}^2 > 0$. With η differentiable, sorting falls locally as exposure is reduced if and only if $2\sigma_{\theta z}\eta(r_\lambda) - (\sigma_\theta^2 + \sigma_{\theta z}r_\lambda)\eta'(r_\lambda) > 2r_\lambda(\sigma_\theta^2\sigma_z^2 - \sigma_{\theta z}^2)$ (a one-line Gaussian computation, Supplemental Appendix G), and the condition isolates three named forces. *Contamination removal* is the baseline: with uninformative ability and fixed noise, reducing exposure cannot lower sorting, and strictly raises it when $\sigma_z^2 > 0$ and $r_\lambda > 0$ — even while, by Theorem 1, gaming effort may be rising; the part of gaming that rises is the common shift, which never enters the ranking. *Collateral information* qualifies the baseline on the type side: with fixed noise and $\sigma_{\theta z} > 0$ the gameable component carries genuine information, and sorting falls as exposure is reduced once $r_\lambda < \sigma_{\theta z}\eta/(\sigma_\theta^2\sigma_z^2 - \sigma_{\theta z}^2)$. *Noise injection* qualifies it on the technology side: with uninformative ability, sorting falls whenever $-\eta'(r_\lambda) > 2\sigma_z^2 r_\lambda$ — for $\eta(r_\lambda) = \eta_0 + \tau(1 - r_\lambda)^2$ with $\tau > 0$, a peak at residual exposure $r^* = \tau/(\sigma_z^2 + \tau)$, interior along the linear family — the interior peak of Figure 3.

Figure 3 displays the noise-injection case across the three curvature regimes, each panel plotting both the common-shift gaming distortion and sorting efficiency against the degree of purification.[29]

[29] The two curves use different normalisations: sorting $\mathcal{E}$ is placed on a scale shared across the panels (the linear-case maximum), so its height can be read against the dashed grey benchmark, which repeats the linear sorting curve in the concave and convex panels, while the gaming curve Δ_λ is scaled to its own maximum within each panel, displaying the timing and shape of the hump rather than its level; the green band is the policy window. Injected noise $\sigma_\varepsilon^2(\lambda) = \eta_0 + \tau\lambda^2$ with $\sigma_\theta^2 = 1$, $\sigma_z^2 = 5$, $\sigma_{\theta z} = 0$, $\kappa = 1$, $\beta = 1.5$, $\eta_0 = 0.15$, $\tau = 1.5$, $\gamma = -9, 0, 7$, on the principal branch. With fixed noise ($\tau = 0$) sorting would instead be monotone in every panel, the plotted curvature lying below the threshold $\gamma^* \approx 22$ of Theorem 3(ii).

6.3 Curvature and the prize-dispersion threshold

Under a convex schedule, the amplification rises with exposure, and against a fixed noise floor, that stretching of the quality dimension is itself a sorting force. The next result makes the trade-off exact, first for fixed curvature and then in the closed-market setting, where the strength of the convexity force is a primitive of the prize distribution.

Theorem 3 (Sorting thresholds)**.** *Let gaming ability be uninformative, let measurement noise be fixed, $\eta(r_\lambda) \equiv \eta > 0$, and let the prize density satisfy $\chi_H > 0$ for the rank-market claims.*

1. (Gaming raises closed-market sorting.) *In the first-order closed-market equilibrium under either protocol, sorting strictly exceeds the linear-price benchmark at the same exposure, $I_\gamma(r_\lambda) > C(r_\lambda)^2/(V^o(r_\lambda)+\eta)$, for every $r_\lambda \in (0,1]$.*

2. (Fixed curvature.) *In the quality-priced market with fixed (β,γ), define*

$$\gamma^* = \frac{\sigma_z^2\,\kappa\,(\sigma_\theta^2+\eta)^2}{2\,\eta\,\sigma_\theta^4}.$$

   If $\gamma < \gamma^*$, complete purification is locally sorting-optimal: $I_\gamma(r_\lambda) < I_\gamma(0)$ for all small $r_\lambda > 0$. If $\gamma > \gamma^*$, sorting is strictly higher at small positive exposure than at complete purification.

3. (Closed markets: prize-dispersion thresholds.) *Complete purification is locally sorting-optimal in the rank market if and only if*

$$\chi_H \le \frac{\kappa\,\sigma_z^2\,(\sigma_\theta^2+\eta)}{2\,\eta},$$

   and in the transfer market if and only if $\chi_{\mathrm{TU}} \le \kappa\sigma_z^2(\sigma_\theta^2+\eta)^{3/2}/(2\eta\sigma_\theta^2)$. With exponential prizes, the rank-market condition for complete purification to cease being sorting-optimal reads $\xi < 4\eta/\big(\pi\kappa\sigma_z^2(\sigma_\theta^2+\eta)\big)$: the standard deviation of prizes must exceed $\pi\kappa\sigma_z^2(\sigma_\theta^2+\eta)/(4\eta)$.

The interpretation is a two-force race at the end of the reform. Restoring a little exposure lets contamination back in at rate σ_z^2, but also reactivates quality-dependent gaming, which stretches latent quality against the noise floor — a stretch funded, in the closed markets, by prize dispersion through χ_H or χ_{TU}. Below the threshold contamination dominates and complete purification stays locally sorting-optimal; above it the separation from quality-dependent gaming outweighs the contamination it rides in on, so complete purification ceases to be sorting-optimal even though purification is noiseless: every sorting maximiser retains strictly positive exposure, interior or at the raw end according to Corollary 3.[30] The separation is not free — it is generated by the costly dispersion term of Corollary 2 — so part (i) concerns the ranking, not welfare.[31] Figure 4 shows the rank market above its threshold: sorting dominates the linear benchmark everywhere and peaks strictly inside the path, at prize dispersion roughly three times that of quality.[32]

Corollary 3 (An interior sorting optimum)**.** *In the rank market, maintain the assumptions of Theorem 3 and let $\chi_H > \kappa\sigma_z^2(\sigma_\theta^2+\eta)/(2\eta)$. If in addition $\sigma_z^2 \geq \eta$, then sorting strictly improves at both ends of the reform path, and every global sorting maximiser is interior: $\lambda^*_{\mathcal{E}} \in (0,1)$.*

Proof. Above the threshold, the proof of Theorem 3(iii) gives $I_\gamma(s) > I_\gamma(0)$ for all small $s > 0$, so no maximiser sits at complete purification. At the raw end, $I_\gamma = \sigma_\theta^4/(V^{\mathrm{o}} + \eta/g_\theta^2)$ falls in exposure at $r_\lambda = 1$ if $\sigma_z^2 > \eta g'_\theta/g_\theta^3$ there. The bound (A.3) gives $g'_\theta < (g_\theta^2\sigma_\theta^2 + \eta)/(r_\lambda g_\theta V^{\mathrm{o}})$, and $(g_\theta^2\sigma_\theta^2+\eta)/g_\theta^4 < \sigma_\theta^2+\eta$ for $g_\theta > 1$, so at $r_\lambda = 1$,

$$\frac{\eta\, g'_\theta}{g_\theta^3} < \frac{\eta(\sigma_\theta^2+\eta)}{V^{\mathrm{o}}(1)} = \frac{\eta(\sigma_\theta^2+\eta)}{\sigma_\theta^2+\sigma_z^2} \leq \sigma_z^2,$$

where the last step rearranges to $\sigma_z^2(\sigma_\theta^2+\sigma_z^2) \geq \eta(\sigma_\theta^2+\eta)$, immediate from $\sigma_z^2 \geq \eta$ because $x \mapsto x(\sigma_\theta^2 + x)$ is increasing. Hence sorting also strictly improves with the first step of purification away from the raw metric, and every maximiser of the continuous I_γ on $[0,1]$ is interior. □

The condition is a race at the raw end between the two things the first increment of purification does to the ranking signal. It removes contamination, at rate σ_z^2; it also unwinds amplification, which costs sorting at rate $\eta g'_\theta/g_\theta^3$ — and the rank-price feedback that disciplines gaming in Theorem 2 caps that response at $\eta(\sigma_\theta^2+\eta)/(\sigma_\theta^2+\sigma_z^2)$. Contamination at least as strong as the noise floor settles the race in favour of purification at the raw end, while the prize-dispersion threshold settles it against purification at the fully purified end: every optimum is pinned strictly inside.

[30] At the boundary $\sigma_z^2 = 0$, rank-market sorting is strictly increasing in exposure and the raw metric itself is sorting-optimal.

[31] Two boundary checks anchor the thresholds: as $\eta \to 0$, $\gamma^* \to \infty$ and both prize-dispersion thresholds diverge — without a noise floor, the amplification is invisible to ranks — and at the primitives of Figure 3, $\gamma = 7$ while $\gamma^* \approx 22$, so with fixed noise sorting would be monotone there: the interior peaks of the figure are entirely the work of noise injection.

[32] Figure 4: exponential prizes with $\xi = 0.32$, so $\chi_H \approx 1.99$ against a threshold of 1; $\sigma_\theta^2 = \sigma_z^2 = \eta = \kappa = 1$; the peak exceeds the complete-purification level by about 4 per cent.

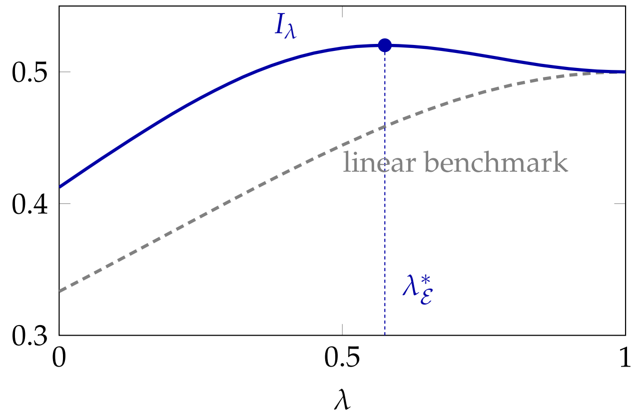

Figure 4: An interior sorting optimum from prize dispersion alone: rank-market sorting I_λ (solid) strictly dominates the linear-price benchmark (dashed) at every exposure and peaks at $\lambda^*_{\mathcal{E}} \approx 0.58$, although purification is noiseless.

6.4 Total surplus and the policy window

For the surplus calculation, we interpret $v(x, z)$ as a real resource payoff in the units of match output θq, and transfers, where present, as pure reallocations: the heuristic trade-off between sorting and action costs then becomes a theorem.

Proposition 7 (Total surplus in the closed market)**.** *In the balanced closed market under either protocol, with match output θq accruing to each matched pair, total surplus equals*

$$\mathcal{S}(r_\lambda) = \mathcal{E}(r_\lambda) - L(r_\lambda) + \mathbb{E}[v(z, z)] :$$

sorting revenue minus the action-cost loss, transfers (where present), and cancelling. Consequently, under the rank protocol with $\chi_H > 0$, noise-stationary purification, and $\sigma_{\theta z} \geq -\frac{2\sqrt{2}}{3}\sigma_\theta\sigma_z$, let $\lambda^*_{\mathcal{E}}$ be a global maximiser of $\mathcal{E}$ on $[0, 1]$: every purification level below $\lambda^*_{\mathcal{E}}$ is weakly dominated by $\lambda^*_{\mathcal{E}}$ itself — strictly so when it leaves strictly more exposure — and the surplus maximum is attained on $[\lambda^*_{\mathcal{E}}, 1]$.

Proof. Aggregate match output under assortative matching on posteriors is $\mathbb{E}[m\, q(m)] = \mathbb{E}[\mathbb{E}[\theta \mid S_\lambda]\, H^{-1}(F_\lambda(m))] = \mathbb{E}[\theta\, H^{-1}(F_Y(Y))] = \mathcal{E}(r_\lambda)$ by iterated expectations, since the assignment is a function of the report. Agents' action payoffs contribute $\mathbb{E}[v(x_\lambda, z)] = \mathbb{E}[v(z, z)] - L(r_\lambda)$, and transfers are a within-pair reallocation. For the second claim, a global maximiser $\lambda^*_{\mathcal{E}}$ exists because $\mathcal{E}$ is continuous in λ along the unique equilibrium branch. Fix $\lambda < \lambda^*_{\mathcal{E}}$. Global optimality gives $\mathcal{E}(\lambda) \leq \mathcal{E}(\lambda^*_{\mathcal{E}})$, and part 2 of Theorem 2 gives $L(\lambda) \geq L(\lambda^*_{\mathcal{E}})$, since $r_\lambda \geq r_{\lambda^*_{\mathcal{E}}}$ and the total action-cost loss is strictly increasing in exposure. Hence $\mathcal{S}(\lambda) \leq \mathcal{S}(\lambda^*_{\mathcal{E}})$, strictly when $r_\lambda > r_{\lambda^*_{\mathcal{E}}}$, and the maximum of $\mathcal{S}$ is attained on $[\lambda^*_{\mathcal{E}}, 1]$. $\square$

Combining the gaming and sorting results yields the policy geometry of Figure 3(b): with noise-injecting purification, gaming peaks at λ^*_Δ and falls while sorting peaks later at $\lambda^*_{\mathcal{E}}$, opening a *policy window* in which gaming is past its peak while sorting still improves (with noise-stationary purification below the curvature threshold the window runs to complete purification). In the rank market the window is degenerate in the best way: gaming falls monotonically from the start, so the designer's only question is where sorting peaks, which Proposition 7 bounds from below.[33]

7 Design Implications and Conclusion

Four lessons for the design of performance metrics follow from the analysis, in increasing order of novelty.

First, complete purification retains a privileged status in every regime: by Corollary 1 it eliminates the targeted gaming incentive regardless of beliefs, prices, or protocol, and action-independent coarsening of a completely purified statistic preserves the neutrality (Lemma C.3, Supplemental Appendix C), so removing the gameable margin separates from choosing how finely to disclose what remains. The backfire results we find reflect the impact of purification *along the path of purification*, not at the fully-purified destination.

Second, pairwise comparisons of feasible metrics should use equilibrium objects, not exposures: under quality pricing a lower-exposure metric can be more gameable in equilibrium, while under rank allocation the exposure and equilibrium orders agree at common residual noise, though a sufficiently less noisy metric can be more gameable under either protocol (Supplemental Appendix B). The reliability of the hypothesis that "a less-gameable metric will be less-gamed" is itself dependent on the downstream matching protocol.

Third, the grades application inherits a sharp prediction. Difficulty-adjusted grades and eigengrade-style procedures (Gans ⓡ Kominers, 2026) are partial exposure reductions, so the equilibrium response depends on which downstream market the marginal student is gaming for: where transcripts feed wage-setting employers, partial purification raises course-shopping exactly when grade variation is dominated by non-ability components; where they feed seat-constrained admissions, every increment of purification reduces aggregate gaming at moderate stakes — a grading reform praised by admissions partners and blamed by the placement office (Supplemental Appendix E).

[33] The quadratic specification is the correct first-order model of any smooth, strictly increasing schedule whose second and third derivatives are uniformly controlled in the curvature scale (Proposition D.5, Supplemental Appendix D); schedules with heavier-than-quadratic growth, such as the wage induced by exponential prizes, fall outside this class. Within it, $\gamma = Q''(0)$ is the only feature of a curved schedule that matters locally.

Fourth, the stakes process is itself a design margin. Converting score-indexed payments into rank-based promotion moves ψ from 1 toward $\frac{1}{2}$ and shrinks the backfire region of Proposition 4 — at the cost, by Theorem 3, of changing how gaming feeds back into sorting — while prize structures redesigned to sharpen as information improves push ψ above 1 and widen it.[34] The boundary case is rationing: at first order in stakes the aggregate backfire cannot arise, the population density at the admission cutoff being itself a rank price, but large admission stakes with severe scarcity restore a modest interior peak in mean gaming, concentrated on applicants near the cutoff (Supplemental Appendix C.2).

This paper asked when making a metric less gameable reduces gaming, and found the answer to be a property of the downstream market, not of the metric alone: where evaluations buy prices, partial reform backfires exactly when gameable variation dominates the metric, whereas where they buy positions, ranking cancels trust from the incentive and every step of purification reduces mean gaming. Transfers and rank tournaments microfound the two regimes within one matching market, separated by an elasticity threshold lying strictly above the rank market's elasticity everywhere.[35] Both protocols nonetheless generate convex schedules whose amplification raises sorting above the linear benchmark, so that — match output and action costs being priced in the same units — whether complete purification is sorting-optimal turns on a single prize-dispersion index, as a statement about total surplus rather than a heuristic (Proposition 7).

Three margins remain partly open. The first is the rationed market beyond the contest's concavity bound, where admission contests develop jump responses.[36] Multiple simultaneous gaming margins, with purification procedures that trade exposure across them, would test whether the exposure–price decomposition aggregates.[37] And the stakes process itself is chosen by institutions: treating ψ as a design instrument, jointly with the disclosure and purification choices studied here, is the natural continuation.

[34] Each direction has been walked in practice: percentile-rank teacher pay (Barlevy and Neal, 2012), proposed against the proficiency-rate inflation documented by Neal (2010); and broiler contracts switching between tournaments and fixed performance standards (Knoeber and Thurman, 1994). Forced-ranking promotion versus rating-indexed bonuses is the same choice inside a firm.

[35] In the classical ledger comparing tournaments with piece rates — common-shock filtering (Holmström, 1982), ordinal enforceability (Malcomson, 1984), sabotage and collusion (Lazear, 1989) — this is a new row: robustness of the gaming margin to reform of the measurement itself.

[36] There the contest is an all-pay contest with heterogeneous contestants; the payoff characterisation of Siegel (2009) and the large-contest limit of Olszewski and Siegel (2016) are the natural tools (Supplemental Appendix C.2).

[37] Two are first in line. Productive effort: under quality pricing the trust channel would amplify real investment in the same proportion as gaming, adding a countervailing welfare term where the two actions are cost-separable. Dispersion: rank rewards invite risk-taking (Hvide, 2002), and the return to manipulating one's own report variance is governed by the convexity that χ_H summarises at the median — signing that margin is open.

Appendix A Proofs Omitted from the Main Text

Assumption A.1 (Local regularity)**.** Fix λ, a type (θ, z), and a compact interval $I \subset \mathbb{R}$ of feasible actions. (i) *(Specified beliefs.)* The selected posterior assessment m_λ is specified at every signal value that can be reached by some action in I. (ii) *(Smoothness.)* The assessment m_λ and the matching schedule Q are continuously differentiable on an open set containing the support of $\theta + \rho(\lambda)x + \varepsilon_\lambda$ for every $x \in I$. (iii) *(Dominated derivative.)* There exists an integrable random variable D_I such that, almost surely, $\sup_{x \in I}\left|Q'(m_\lambda(\theta + \rho(\lambda)x + \varepsilon_\lambda))\, m'_\lambda(\theta + \rho(\lambda)x + \varepsilon_\lambda)\right| \le D_I$.

The only substantive content is part (i): a deviation can reach off-path report values where Bayes' rule is silent, so the derivative must refer to a definite belief function; with full support — as in the Gaussian specification with positive measurement noise — part (i) has no bite.

Lemma A.1 (Strong neutrality implies gaming neutrality)**.** *If a reporting rule is strongly gaming-neutral, then it is gaming-neutral for every fixed posterior rule m_λ and every matching schedule Q for which the expectations in* (2) *are well-defined.*

The proofs — dominated convergence under Assumption A.1 for Proposition 1, action-invariance of the conditional law for the lemma — are in Supplemental Appendix G.

Proof of Proposition 2. Conjecture a linear posterior $m_\lambda(s) = W(s - \mathbb{E}[S_\lambda])$, consistent with joint normality under any affine strategy. An agent of type (θ, z) choosing x generates S_λ with conditional mean satisfying $\mathbb{E}[S_\lambda \mid \theta, z, x] - \mathbb{E}[S_\lambda] = \theta + r_\lambda x - \mathbb{E}[S_\lambda] =: \xi$, so $m_\lambda(S_\lambda) = W(\xi + \varepsilon_\lambda)$ and the market payoff is $\beta W \xi + \frac{\gamma}{2} W^2(\xi^2 + \sigma_\varepsilon^2(\lambda))$. Adding $-\frac{\kappa}{2}(x - z)^2$ gives an objective with second derivative $\gamma W^2 r_\lambda^2 - \kappa = -D_\lambda$. When $D_\lambda > 0$ the objective is strictly concave and the first-order condition $\kappa(x - z) = \beta W r_\lambda + \gamma W^2 r_\lambda(\theta + r_\lambda x - \mathbb{E}[S_\lambda])$ has the unique affine solution with $\alpha_z = \kappa / D_\lambda$ and $\alpha_\theta = \gamma W^2 r_\lambda / D_\lambda$. Imposing $\mathbb{E}[S_\lambda] = r_\lambda \alpha_0$ and using $D_\lambda + \gamma W^2 r_\lambda^2 = \kappa$ gives $\alpha_0 = \beta W r_\lambda / \kappa$, proving (5). The report loading on θ is $g_\theta = 1 + r_\lambda \alpha_\theta = \kappa / D_\lambda$ and on z is $r_\lambda \alpha_z = r_\lambda g_\theta$, giving (6); moreover $\alpha_z - 1 = r_\lambda \alpha_\theta$, so $a_\lambda = \alpha_0 + \alpha_\theta(\theta + r_\lambda z)$, which is (7). The rank-relevant signal is $Y = g_\theta(\theta + r_\lambda z) + \varepsilon_\lambda$ with $\mathrm{Cov}(\theta, Y) = g_\theta C_\lambda$ and $\mathrm{Var}(Y) = g_\theta^2 V_\lambda^{\mathrm{o}} + \sigma_\varepsilon^2(\lambda)$, so Bayesian consistency $W = \mathrm{Cov}(\theta, Y) / \mathrm{Var}(Y)$ is the first equation of (8). Conversely, any selected-assessment affine equilibrium faces the quadratic objective above, and $D_\lambda < 0$ leaves no maximiser while $D_\lambda = 0$ leaves a nonzero slope for almost every type; hence $D_\lambda > 0$ and W solves (8). $\square$

Proof of Theorem 1. Setting $\gamma = 0$, $\sigma_{\theta z} = 0$, $\rho(\lambda) = 1 - \lambda$, and $\sigma_\varepsilon^2(\lambda) \equiv \sigma_\varepsilon^2$ in (9) yields (12). With $r = 1 - \lambda$ and $\Delta(r) = (\beta/\kappa)\, \sigma_\theta^2 r / (\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_z^2 r^2)$,

$$\frac{d\Delta}{dr} = \frac{\beta}{\kappa} \frac{\sigma_\theta^2(\sigma_\theta^2 + \sigma_\varepsilon^2 - \sigma_z^2 r^2)}{(\sigma_\theta^2 + \sigma_\varepsilon^2 + \sigma_z^2 r^2)^2}.$$

If $\sigma_z^2 > \sigma_\theta^2 + \sigma_\varepsilon^2$ the sign changes once at $r^* = \sqrt{(\sigma_\theta^2 + \sigma_\varepsilon^2)/\sigma_z^2} \in (0, 1)$, giving $\lambda^* = 1 - r^*$; otherwise the derivative is nonnegative on $[0, 1]$ and Δ_λ is weakly decreasing in λ. □

Proof of Lemma 1. On the balanced matched interior, $Q'_\lambda(m) = f_\lambda(m)/h(Q_\lambda(m))$, the $\mu = 1$ case of the rank-price term in Proposition C.3 (Supplemental Appendix C), where f_λ is the density of the $N(0, I_\lambda)$ posterior. At $m = 0$ the agent holds the median posterior and is assigned the median prize, $Q_\lambda(0) = q^{\text{med}}$, and $f_\lambda(0) = (2\pi I_\lambda)^{-1/2}$, giving the first expression in (14). Differentiating once more, $Q''_\lambda(m) = Q'_\lambda(m)\big[f'_\lambda(m)/f_\lambda(m) - \big(h'(Q_\lambda(m))/h(Q_\lambda(m))\big)Q'_\lambda(m)\big]$, and Gaussian symmetry gives $f'_\lambda(0) = 0$, yielding the second expression; multiplying by I_λ and substituting $\beta_\lambda^2 I_\lambda = 1/(2\pi h(q^{\text{med}})^2)$ gives (15). Finally, the posterior is the increasing affine transform $m_\lambda = W(S_\lambda - \mathbb{E}[S_\lambda])$ of the Gaussian report, so $f_\lambda(0) = f_{S_\lambda}(\mathbb{E}[S_\lambda])/W = (2\pi \operatorname{Var}(S_\lambda))^{-1/2}/W$, and (16) follows. □

Proof of Proposition 3. By Proposition 2, an affine equilibrium at $(\beta_\lambda, \gamma_\lambda)$ has $g_\theta = \kappa/(\kappa - \gamma_\lambda W^2 r_\lambda^2)$. Substituting $\gamma_\lambda = \chi_H/I_\lambda$ from (15) and $I_\lambda = W g_\theta C_\lambda$ gives $\gamma_\lambda W^2 r_\lambda^2 = \chi_H W r_\lambda^2/(g_\theta C_\lambda)$, so the optimality condition reads $g_\theta \kappa - \chi_H W r_\lambda^2/C_\lambda = \kappa$, which is (17). Bayesian consistency is $W(g_\theta^2 V_\lambda^{\text{o}} + \eta) = g_\theta C_\lambda$; substituting (17) and collecting powers of W gives (18). The coefficient sequence of (18) has signs $(+, +, \pm, -)$ and therefore exactly one sign change whatever the sign of the third coefficient, so by Descartes' rule of signs there is exactly one positive root. A root with $W < 0$ is not an equilibrium ($g_\theta > 0$ would contradict $\operatorname{sign}(W) = \operatorname{sign}(C_\lambda) > 0$ from consistency, and $g_\theta < 0$ violates the second-order condition $D_\lambda = \kappa/g_\theta > 0$); the positive root has $D_\lambda = \kappa/g_\theta > 0$. As $b_\lambda \to 0$ the cubic degenerates to $(V_\lambda^{\text{o}} + \eta)W = C_\lambda$, the knife-edge benchmark, so the unique root is the continuous continuation of the knife-edge equilibrium and lies on the principal branch. Part (ii) is immediate from (17) with $b_\lambda > 0$, $W > 0$. For part (iii), $\alpha_0 = \beta_\lambda W r_\lambda/\kappa$ combines with (16) and $\operatorname{Var}(S_\lambda) = g_\theta^2 V_\lambda^{\text{o}} + \eta$ to give (19), and $\alpha_\theta = (g_\theta - 1)/r_\lambda$ restates the loading identity $g_\theta = 1 + r_\lambda \alpha_\theta$. □

Proof of Theorem 2. Write $r = r_\lambda$, $g = g_\theta(r)$, $V^{\text{o}}(r) = \sigma_\theta^2 + 2r\sigma_{\theta z} + r^2\sigma_z^2$, and $C(r) = \sigma_\theta^2 + r\sigma_{\theta z}$, with $C > 0$ by Assumption 2. Combining (17) with Bayesian consistency $W =$

$gC/(g^2V^{\mathrm{o}}+\eta)$, the covariance cancels and the equilibrium satisfies the scalar identity

$$(g-1)\big(g^2V^{\mathrm{o}}+\eta\big)=\omega\, r^2 g, \qquad \omega := \frac{\chi_H}{\kappa},$$

with $g>1$ on $(0,1]$ by Proposition 3(ii). Both parts use two consequences. First, since $V^{\mathrm{o}\prime}(r)=2(\sigma_{\theta z}+r\sigma_z^2)$, the variance kernel net of its own exposure response is the covariance:

$$V^{\mathrm{o}}(r)-r\big(\sigma_{\theta z}+r\sigma_z^2\big)=C(r). \tag{A.1}$$

Second, differentiating the scalar identity in r, eliminating ω through $\omega r^2 g=(g-1)(g^2V^{\mathrm{o}}+\eta)$ and $2\omega rg=2(g-1)(g^2V^{\mathrm{o}}+\eta)/r$, and simplifying with (A.1),

$$g'\,M=\frac{2(g-1)\big(g^2C+\eta\big)}{r}>0, \qquad M:=\frac{g^2V^{\mathrm{o}}+\eta}{g}+2(g-1)gV^{\mathrm{o}}, \tag{A.2}$$

so $g'>0$; and because M exceeds $2(g-1)gV^{\mathrm{o}}$ by $(g^2V^{\mathrm{o}}+\eta)/g>0$,

$$rgg'V^{\mathrm{o}}=\big(g^2C+\eta\big)\,\frac{2(g-1)gV^{\mathrm{o}}}{M}<g^2C+\eta. \tag{A.3}$$

Part 1. By (19), α_0 is strictly increasing in r if and only if $\varphi(r)=r^2/(g^2V^{\mathrm{o}}+\eta)$ is. The sign of $\varphi'(r)$ is the sign of

$$2r\big(g^2V^{\mathrm{o}}+\eta\big)-r^2\big(2gg'V^{\mathrm{o}}+g^2V^{\mathrm{o}\prime}\big)=2r\Big[\big(g^2C+\eta\big)-rgg'V^{\mathrm{o}}\Big]>0,$$

where the equality uses (A.1) and the inequality is (A.3).

Part 2. Let $D(r)=\alpha_\theta^2V^{\mathrm{o}}=(g-1)^2V^{\mathrm{o}}/r^2$. Since $rV^{\mathrm{o}\prime}-2V^{\mathrm{o}}=-2C$ by (A.1), substituting (A.2) for g' gives

$$D'(r)=\frac{g-1}{r^3}\Big[2rg'V^{\mathrm{o}}-2(g-1)C\Big]=\frac{2(g-1)^2}{r^3}\left[\frac{2V^{\mathrm{o}}\big(g^2C+\eta\big)}{M}-C\right],$$

and a direct computation yields

$$2V^{\mathrm{o}}\big(g^2C+\eta\big)-CM=\frac{CV^{\mathrm{o}}g^2+\eta\big(2V^{\mathrm{o}}g-C\big)}{g}.$$

The quadratic $q(r):=2V^{\mathrm{o}}-C=2\sigma_z^2r^2+3\sigma_{\theta z}r+\sigma_\theta^2$ is nonnegative on $[0,1]$ under the stated condition: for $\sigma_{\theta z}\geq 0$ every term is nonnegative with $q(0)=\sigma_\theta^2>0$, and for $0>\sigma_{\theta z}\geq -\frac{2\sqrt{2}}{3}\sigma_\theta\sigma_z$ the discriminant $9\sigma_{\theta z}^2-8\sigma_\theta^2\sigma_z^2$ is nonpositive, so $q\geq 0$ everywhere.

Hence $2V^{\mathrm{o}}g - C = 2V^{\mathrm{o}}(g-1) + q(r) > 0$ for $g > 1$, the display is strictly positive, and $D'(r) > 0$ on $(0,1]$. Corollary 2 then makes $L_\lambda = \frac{\kappa}{2}\big(\alpha_0^2 + D\big)$, and hence $\mathbb{E}[a_\lambda^2] = \alpha_0^2 + D$, strictly increasing in exposure. For general $\sigma_{\theta z}$ the displayed numerator gives the exact local direction of the dispersion margin, as recorded in footnote 22. □

Proof of Proposition 5. (i) Stability with transfers requires each position to solve $\max_m\{mq - w_\lambda(m)\}$, with first-order condition $q = w'_\lambda(m)$ at the chosen m; assortativity assigns position $H^{-1}(F_\lambda(m))$ to posterior m, giving (21) and pinning w_λ up to a constant that does not affect incentives. Since $H^{-1} \circ F_\lambda$ is increasing, w_λ is convex; differentiating (21) at $m = 0$ and using $F_\lambda = N(0, I_\lambda)$ gives $w''_\lambda(0) = f_\lambda(0)/h(q^{\mathrm{med}})$, while $w'_\lambda(0) = H^{-1}(\frac{1}{2}) = q^{\mathrm{med}}$; the average-slope identity is $\mathbb{E}[w'_\lambda(m)] = \mathbb{E}[H^{-1}(F_\lambda(m))] = \int_0^1 H^{-1}(u)\,du = \bar{q}$. The median agent's marginal market return to the report is $w'_\lambda(0)\,W = q^{\mathrm{med}}W$, and at the knife edge $W = C_\lambda/V$, which is $P(V) \propto V^{-1}$.

(ii) By Proposition 2, $g_\theta = \kappa/(\kappa - \gamma W^2 r_\lambda^2)$, so $D_\lambda = \kappa/g_\theta$; substituting $\gamma^{\mathrm{TU}} = \chi_{\mathrm{TU}}/\sqrt{I_\lambda}$, $I_\lambda = W g_\theta C_\lambda$, and Bayesian consistency $W = g_\theta C_\lambda/(g_\theta^2 V_\lambda^{\mathrm{o}} + \eta)$ yields (22) after collecting terms. $\mathcal{T}_\lambda(1) < 0$ and $\mathcal{T}_\lambda(g) \to \infty$, so a root exists; at any root, $\mathcal{T}'_\lambda = \big(g^2 V^{\mathrm{o}} + \eta\big)^{1/2} B(g)$ with $B(g) = (g^2 V^{\mathrm{o}} + \eta)(2-g)/g + 3(g-1)gV^{\mathrm{o}}$, which is positive for $g \le 2$ and, for $g > 2$, positive whenever $\eta(g-2) < g^2 V^{\mathrm{o}}(2g-1)$; minimising $g^2(2g-1)/(g-2)$ over $g > 2$ (the minimiser is $g^* = (13 + \sqrt{105})/8$) gives the constant φ^*. If $\eta < \varphi^* V_\lambda^{\mathrm{o}}$ the inequality is strict, so $B > 0$ on $(1,\infty)$ and $\mathcal{T}'_\lambda > 0$ throughout, whence the root is unique. If $\eta = \varphi^* V_\lambda^{\mathrm{o}}$, then $B > 0$ on $(1,2]$ while on $(2,\infty)$ we have $B \ge 0$ with equality only at g^*; thus $\mathcal{T}'_\lambda \ge 0$ on $(1,\infty)$ with a single isolated zero, so $\mathcal{T}_\lambda$ is still strictly increasing and has at most one root. Uniqueness therefore holds whenever $\eta \le \varphi^* V_\lambda^{\mathrm{o}}$.

(iii) From (22), $g_\theta - 1 = O(\chi_{\mathrm{TU}})$ uniformly in $r_\lambda \in [0,1]$, hence W converges uniformly to the knife-edge $C_\lambda/(V_\lambda^{\mathrm{o}} + \eta)$ and $\alpha_0 = q^{\mathrm{med}} W r_\lambda/\kappa$ converges uniformly, together with its derivative in r_λ (both are polynomial expressions in a branch that depends smoothly on χ_{TU}), to the translation distortion of Theorem 1 with $\beta = q^{\mathrm{med}}$. When $\sigma_z^2 > \sigma_\theta^2 + \eta$ the limit has a strict interior maximum, and when $\sigma_z^2 < \sigma_\theta^2 + \eta$ it is strictly increasing on $(0,1]$ with strictly positive derivative at $r_\lambda = 1$; uniform convergence preserves each strict property for χ_{TU} small.

The frontier expansion stated in Remark 1 is proved in Supplemental Appendix F. □

The proof of Proposition 6 and the knife-edge sorting computation of Section 6 are mechanical Gaussian calculations, given in Supplemental Appendix G.

Proof of Theorem 3. Part (i). By (25), $I_\gamma = C^2/(V^{\mathrm{o}} + \eta/g_\theta^2)$, strictly increasing in g_θ for $\eta > 0$; both protocols have $g_\theta > 1$ at every interior exposure (Proposition 3(ii) and

Proposition 5(ii), the latter because $\mathcal{T}_\lambda(1) < 0$ forces the root above 1), which gives strict dominance over the $g_\theta = 1$ value.

Part (ii). With $\sigma_{\theta z} = 0$ and fixed η, write the fixed point (8) as $\Phi(g, W; r_\lambda) = 0$ with $\Phi_1 = g(\kappa - \gamma r_\lambda^2 W^2) - \kappa$ and $\Phi_2 = W((\sigma_\theta^2 + \sigma_z^2 r_\lambda^2)g^2 + \eta) - \sigma_\theta^2 g$. At $r_\lambda = 0$ the unique solution is $(1, W_0)$ with $W_0 = \sigma_\theta^2/(\sigma_\theta^2 + \eta)$, and the Jacobian is triangular with nonzero diagonal, so the implicit function theorem gives a unique smooth branch near $r_\lambda = 0$; the consistency curve $W = \sigma_\theta^2 g/((\sigma_\theta^2 + \sigma_z^2 r_\lambda^2)g^2 + \eta)$ is bounded while, off $g = 1$, the optimality curve $W = \sqrt{\kappa(g-1)/(\gamma r_\lambda^2 g)}$ diverges as $r_\lambda \to 0$, confining all solutions to that branch for small r_λ. Since Φ depends on r_λ only through $s = r_\lambda^2$, $\Phi_1 = 0$ gives $g(s) = 1 + (\gamma W_0^2/\kappa)s + o(s)$, so $\eta/g^2 = \eta - (2\eta\gamma W_0^2/\kappa)s + o(s)$ and

$$I_\gamma(s) = \frac{\sigma_\theta^4}{\sigma_\theta^2 + \eta + \left(\sigma_z^2 - 2\eta\gamma W_0^2/\kappa\right)s + o(s)},$$

whose s-coefficient is positive (complete purification locally optimal) if and only if $\gamma < \gamma^* = \sigma_z^2\kappa/(2\eta W_0^2) = \sigma_z^2\kappa(\sigma_\theta^2 + \eta)^2/(2\eta\sigma_\theta^4)$.

Part (iii). The same expansion applies with the endogenous curvature evaluated along each branch. In the rank market, (17) gives $g(s) = 1 + (\chi_H/(\kappa\sigma_\theta^2))W_0\, s + o(s)$, so the s-coefficient of $V^{\mathrm{o}} + \eta/g^2$ is $\sigma_z^2 - 2\eta\chi_H W_0/(\kappa\sigma_\theta^2) = \sigma_z^2 - 2\eta\chi_H/(\kappa(\sigma_\theta^2 + \eta))$, nonnegative if and only if $\chi_H \le \kappa\sigma_z^2(\sigma_\theta^2 + \eta)/(2\eta)$; substituting $\chi_H = 2/(\pi\xi)$ gives the exponential form. Under transfers, (22) gives $(g-1)(\sigma_\theta^2 + \eta)^{3/2} = \chi_{\mathrm{TU}}\sigma_\theta^2 s/\kappa + o(s)$, so $g(s) = 1 + \chi_{\mathrm{TU}}\sigma_\theta^2 s/(\kappa(\sigma_\theta^2 + \eta)^{3/2}) + o(s)$ and the s-coefficient is $\sigma_z^2 - 2\eta\chi_{\mathrm{TU}}\sigma_\theta^2/(\kappa(\sigma_\theta^2 + \eta)^{3/2})$, nonnegative if and only if $\chi_{\mathrm{TU}} \le \kappa\sigma_z^2(\sigma_\theta^2 + \eta)^{3/2}/(2\eta\sigma_\theta^2)$. In both protocols, the endogenous curvature varies with s only at order s, which leaves the first-order expansion unchanged; when the s-coefficient is nonzero, its sign settles the claim.

At either threshold the s-coefficient vanishes and the comparison moves to second order: the branch expansion $g(s) = 1 + cs + ds^2 + o(s^2)$ has $c = \sigma_z^2/(2\eta)$ and $d = -c^2$ at both thresholds, giving $V^{\mathrm{o}}(s) + \eta/g(s)^2 = \sigma_\theta^2 + \eta + 5\eta c^2 s^2 + o(s^2) > \sigma_\theta^2 + \eta$ for all small $s > 0$ (computation in Supplemental Appendix G): at equality, complete purification remains locally sorting-optimal, which completes the "if and only if". □

References

AZEVEDO, E. M. (2014): "Imperfect Competition in Two-Sided Matching Markets," *Games and Economic Behavior*, 83, 207–223.

AZEVEDO, E. M. AND E. BUDISH (2019): "Strategy-proofness in the Large," *Review of Economic Studies*, 86, 81–116.

AZEVEDO, E. M. AND J. D. LESHNO (2016): "A Supply and Demand Framework for Two-Sided Matching Markets," *Journal of Political Economy*, 124, 1235–1268.

BAKER, G. P. (1992): "Incentive Contracts and Performance Measurement," *Journal of Political Economy*, 100, 598–614.

BALL, I. (2025): "Scoring Strategic Agents," *American Economic Journal: Microeconomics*, 17, 97–129.

BAR, T., V. KADIYALI, AND A. ZUSSMAN (2009): "Grade Information and Grade Inflation: The Cornell Experiment," *Journal of Economic Perspectives*, 23, 93–108.

BARLEVY, G. AND D. NEAL (2012): "Pay for Percentile," *American Economic Review*, 102, 1805–1831.

BECKER, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

BLACKWELL, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

BODOH-CREED, A. L. AND B. R. HICKMAN (2018): "College Assignment as a Large Contest," *Journal of Economic Theory*, 175, 88–126.

BONATTI, A. AND G. CISTERNAS (2020): "Consumer Scores and Price Discrimination," *Review of Economic Studies*, 87, 750–791.

BROWN, J. (2011): "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars," *Journal of Political Economy*, 119, 982–1013.

CAMPBELL, D. T. (1979): "Assessing the Impact of Planned Social Change," *Evaluation and Program Planning*, 2, 67–90.

CHAN, W., L. HAO, AND W. SUEN (2007): "Grades, Course Evaluations, and Academic Choice," *Journal of Economic Behavior & Organization*, 63, 272–286.

CHE, Y.-K. AND I. L. GALE (1998): "Caps on Political Lobbying," *American Economic Review*, 88, 643–651.

DEWATRIPONT, M., I. JEWITT, AND J. TIROLE (1999): "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, 66, 183–198.

DUBEY, P. AND J. GEANAKOPLOS (2010): "Grading Exams: 100, 99, 98, ... or A, B, C?" *Games and Economic Behavior*, 69, 72–94.

EWERT, R. AND A. WAGENHOFER (2005): "Economic Effects of Tightening Accounting Standards to Restrict Earnings Management," *The Accounting Review*, 80, 1101–1124.

FISCHER, P. E. AND R. E. VERRECCHIA (2000): "Reporting Bias," *The Accounting Review*, 75, 229–245.

FRANKEL, A. AND N. KARTIK (2019): "Muddled Information," *Journal of Political Economy*, 127, 1739–1776.

——— (2022): "Improving Information from Manipulable Data," *Journal of the European Economic Association*, 20, 79–115.

GALE, D. AND L. S. SHAPLEY (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

GANS, J. S. ⓡ S. D. KOMINERS (2026): "What Does A Grade Mean? Informativeness and Strategic Manipulation of Grading Systems," Working Paper 35183, National Bureau of Economic Research.

GIBBONS, R. AND K. J. MURPHY (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence," *Journal of Political Economy*, 100, 468–505.

GILPATRIC, S. M. (2011): "Cheating in Contests," *Economic Inquiry*, 49, 1042–1053.

GOODHART, C. A. E. (1975): "Problems of Monetary Management: The U.K. Experience," *Papers in Monetary Economics*, 1, 1–20.

GREEN, J. R. AND N. L. STOKEY (1983): "A Comparison of Tournaments and Contracts," *Journal of Political Economy*, 91, 349–364.

HOLMSTRÖM, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–340.

——— (1999): "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 66, 169–182.

HOLMSTRÖM, B. AND P. MILGROM (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, & Organization*, 7, 24–52.

HVIDE, H. K. (2002): "Tournament Rewards and Risk Taking," *Journal of Labor Economics*, 20, 877–898.

HVIDE, H. K. AND E. G. KRISTIANSEN (2003): "Risk Taking in Selection Contests," *Games and Economic Behavior*, 42, 172–179.

KNOEBER, C. R. AND W. N. THURMAN (1994): "Testing the Theory of Tournaments: An Empirical Analysis of Broiler Production," *Journal of Labor Economics*, 12, 155–179.

KRÄKEL, M. (2007): "Doping and Cheating in Contest-Like Situations," *European Journal of Political Economy*, 23, 988–1006.

LAZEAR, E. P. (1989): "Pay Equality and Industrial Politics," *Journal of Political Economy*, 97, 561–580.

LAZEAR, E. P. AND S. ROSEN (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 841–864.

LI, Y. AND X. QIU (2023): "Mechanism Design under Costly Signaling: The Value of Non-Coordination," arXiv:2302.09168v3, revised February 2026.

MALCOMSON, J. M. (1984): "Work Incentives, Hierarchy, and Internal Labor Markets," *Journal of Political Economy*, 92, 486–507.

MEYER, M. A. (1991): "Learning from Coarse Information: Biased Contests and Career Profiles," *Review of Economic Studies*, 58, 15–41.

MOLDOVANU, B. AND A. SELA (2001): "The Optimal Allocation of Prizes in Contests," *American Economic Review*, 91, 542–558.

MOLDOVANU, B., A. SELA, AND X. SHI (2007): "Contests for Status," *Journal of Political Economy*, 115, 338–363.

NALEBUFF, B. J. AND J. E. STIGLITZ (1983): "Prizes and Incentives: Towards a General Theory of Compensation and Competition," *Bell Journal of Economics*, 14, 21–43.

NEAL, D. (2010): "Aiming for Efficiency Rather than Proficiency," *Journal of Economic Perspectives*, 24, 119–132.

Neal, D. and J. Root (2024): "The Provision of Information and Incentives in School Assignment Mechanisms," Working Paper 32378, NBER.

Neal, D. and D. W. Schanzenbach (2010): "Left Behind by Design: Proficiency Counts and Test-Based Accountability," *Review of Economics and Statistics*, 92, 263–283.

Olszewski, W. and R. Siegel (2016): "Large Contests," *Econometrica*, 84, 835–854.

——— (2020): "Performance-Maximizing Large Contests," *Theoretical Economics*, 15, 57–88.

Ostrovsky, M. and M. Schwarz (2010): "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, 2, 34–63.

Page, L., S. Brin, R. Motwani, and T. Winograd (1999): "The PageRank Citation Ranking: Bringing Order to the Web," Technical Report 1999-66, Stanford InfoLab.

Pathak, P. A. and T. Sönmez (2013): "School Admissions Reform in Chicago and England: Comparing Mechanisms by Their Vulnerability to Manipulation," *American Economic Review*, 103, 80–106.

Pérez-Richet, E. and V. Skreta (2022): "Test Design under Falsification," *Econometrica*, 90, 1109–1142.

Popov, S. V. and D. Bernhardt (2013): "University Competition, Grading Standards, and Grade Inflation," *Economic Inquiry*, 51, 1764–1778.

Rosen, S. (1986): "Prizes and Incentives in Elimination Tournaments," *American Economic Review*, 76, 701–715.

Roth, A. E. (1985): "The College Admissions Problem Is Not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 36, 277–288.

Sabot, R. and J. Wakeman-Linn (1991): "Grade Inflation and Course Choice," *Journal of Economic Perspectives*, 5, 159–170.

Shapley, L. S. and M. Shubik (1971): "The Assignment Game I: The Core," *International Journal of Game Theory*, 1, 111–130.

Siegel, R. (2009): "All-Pay Contests," *Econometrica*, 77, 71–92.

Spence, M. (1973): "Job Market Signaling," *Quarterly Journal of Economics*, 87, 355–374.

Stein, J. C. (1989): "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," *Quarterly Journal of Economics*, 104, 655–669.

Supplemental Appendix

Table A.1: Notation used in the paper.

Primitives	
θ	Latent match-relevant quality of an agent.
z	Private action type; the action the agent would choose absent any market reward.
q	Desirability of a position, with distribution H, density h, median q^{med}, mean $\bar{q}$.
x	Action chosen by the agent before the report is realised.
$v(x,z)$	Private (non-market) payoff to action x for an agent of action type z.
κ	Private-payoff curvature, $v(x,z) = -\frac{\kappa}{2}(x-z)^2$.
$\sigma_\theta^2, \sigma_z^2, \sigma_{\theta z}$	Variances and covariance of (θ, z).
$\sigma_\varepsilon^2(\lambda),\ \eta$	Variance of the measurement shock; η when noise-stationary.
Reporting rule and report	
$\lambda \in [0,1]$	Purification index; $\lambda = 0$ raw, $\lambda = 1$ completely purified.
$\rho(\lambda),\ r_\lambda$	Residual exposure of the report to the gameable action.
S_λ	Public report, $S_\lambda = \theta + \rho(\lambda)x + \varepsilon_\lambda$.
$C_\lambda,\ V_\lambda^{\text{o}},\ V_\lambda$	$\text{Cov}(\theta, \theta + r_\lambda z)$; $\text{Var}(\theta + r_\lambda z)$; total report variance.
Beliefs, matching, and stakes	
$m_\lambda(\cdot),\ W$	Posterior rule; slope of the linear posterior ("trust").
$Q(\cdot)$	Placement-value schedule; $\beta = Q'(0)$, $\gamma = Q''(0)$.
$R_\lambda(x;\theta,z)$	Expected market payoff at action x.
$P(V),\ \psi,\ \psi^*$	Stakes process; stakes-deflation elasticity; backfire threshold.
Equilibrium objects	
g_θ	Amplification: equilibrium loading of the report on quality, $g_\theta = 1 + r_\lambda \alpha_\theta$.
$\Delta_\lambda,\ \alpha_0,\ \alpha_\theta$	Translation shift; common-shift and quality-loading coefficients.
$I_\lambda,\ I_\gamma(r_\lambda)$	Posterior-information index, $\text{Var}(\mathbb{E}[\theta \mid Y])$.
L_λ	Private action-cost loss.
$\mathcal{E}(r_\lambda),\ K_H$	Latent sorting efficiency and its distribution constant.
$\mathcal{S}(r_\lambda)$	Total surplus in the closed market, $\mathcal{E} - L$ up to a constant.
$\mathcal{M}_j,\ \mathcal{M}_j^{\text{rank}}$	Pairwise gaming indices, quality- and rank-priced (Supplemental Appendix B).
Closed markets	
$F_\lambda,\ f_\lambda$	Equilibrium distribution and density of posterior expected quality.
$\beta_\lambda,\ \gamma_\lambda$	Induced slope and curvature of the closed-market schedule.
χ_H	Prize-curvature index (rank protocol), $-h'(q^{\text{med}})/(2\pi h(q^{\text{med}})^3)$.
χ_{TU}	Prize-curvature index (transfers), $1/(\sqrt{2\pi}\, h(q^{\text{med}}))$.
$w_\lambda,\ \mathcal{T}_\lambda$	TU wage schedule; TU fixed-point function (22).
φ^*	Uniqueness constant for (22), $\min_{g>2} g^2(2g-1)/(g-2) \approx 44.9$.
$\mu,\ q_0$	Mass of positions; unmatched outside option (Supplemental Appendix C).
$Q_\lambda^\mu(m),\ \Gamma_\lambda^\mu(m)$	Unbalanced schedule and matched-interior rank-price term.
First-order validity (Supplemental Appendix D)	
$Q_t,\ t,\ \gamma_0,\ k_2,\ k_3$	Family of increasing schedules with curvature scale t and derivative bounds.
$(x_t, m_t),\ x_t^{\text{BR}},\ t_0,\ K_0$	Quadratic-benchmark equilibrium; exact best response; approximation constants.

Appendix B Comparing Reporting Rules

Many policy problems are pairwise rather than continuous: an institution has two feasible reporting rules and wants to know which is less gameable in equilibrium. Consider two reports, indexed by $j \in \{A, B\}$:

$$S^j = \theta + \rho_j x + \varepsilon_j, \qquad \rho_j \geq 0, \tag{B.1}$$

where $\varepsilon_j \sim N(0, \sigma^2_{\varepsilon,j})$ is independent of (θ, z). We use the Gaussian-quadratic affine-price benchmark of Section 3 with $\gamma = 0$, including the affine Gaussian assessment selection, and compare the translation equilibria; the comparison is therefore the quality-priced ($\psi = 1$) comparison, and Proposition B.2 below develops the rank-priced counterpart through the rank price of Lemma 1. Assume each report is positively informative: $\sigma^2_\theta + \rho_j \sigma_{\theta z} > 0$.

Fix the joint distribution of (θ, z), the private payoff function, and the local matching slope β; say report B is *weakly less gameable in equilibrium* than report A if the corresponding translation-equilibrium action-cost losses satisfy $L_B \leq L_A$, and *strictly less gameable* if the inequality is strict.

Proposition B.1 (Pairwise gaming index)**.** *Under the Gaussian-quadratic affine-price assumptions, with the affine Gaussian assessment selection adapted to report j, define*

$$\mathcal{M}_j = \frac{\rho_j(\sigma^2_\theta + \rho_j \sigma_{\theta z})}{\sigma^2_\theta + 2\rho_j \sigma_{\theta z} + \rho_j^2 \sigma_z^2 + \sigma^2_{\varepsilon,j}}. \tag{B.2}$$

Then the unique translation-equilibrium distortion under report j is $\Delta_j = (\beta/\kappa)\mathcal{M}_j$, the corresponding action-cost loss is $L_j = (\beta^2/2\kappa)\mathcal{M}_j^2$, and report B is weakly less gameable than report A if and only if $\mathcal{M}_B \leq \mathcal{M}_A$.

Proof. Apply the translation equilibrium (9) to report j: the posterior coefficient is $W_j = (\sigma^2_\theta + \rho_j \sigma_{\theta z})/(\sigma^2_\theta + 2\rho_j \sigma_{\theta z} + \rho_j^2 \sigma_z^2 + \sigma^2_{\varepsilon,j})$, so $\Delta_j = (\beta/\kappa)\rho_j W_j = (\beta/\kappa)\mathcal{M}_j$ and Corollary 2 gives $L_j = (\kappa/2)\Delta_j^2$. Positive informativeness makes $\mathcal{M}_j \geq 0$, so $L_B \leq L_A$ if and only if $\mathcal{M}_B \leq \mathcal{M}_A$. $\square$

The index $\mathcal{M}_j$ combines direct exposure, the covariance structure, and the report's residual variance; a lower-exposure report need not have a lower index. The equilibrium gaming order is also distinct from an informativeness or garbling order: a report can be more informative about θ at fixed actions and still be more gameable in equilibrium, because the remaining exposed component becomes more valuable when the market trusts the report more.

Corollary B.1 (Lower exposure can increase equilibrium gaming)**.** *Suppose $\sigma_{\theta z} = 0$, $\sigma^2_{\varepsilon,A} = \sigma^2_{\varepsilon,B} = \sigma^2_\varepsilon$, and $0 < \rho_B < \rho_A$. Then report B has lower direct exposure than report A, but induces more equilibrium gaming if and only if*

$$\sigma_z^2 \rho_A \rho_B > \sigma_\theta^2 + \sigma_\varepsilon^2. \tag{B.3}$$

Proof. Under the stated assumptions $\mathcal{M}_j = \rho_j \sigma_\theta^2 / (\sigma_\theta^2 + \sigma_\varepsilon^2 + \rho_j^2 \sigma_z^2)$; cross-multiplying $\mathcal{M}_B > \mathcal{M}_A$ and rearranging yields $(\rho_A - \rho_B)(\sigma_z^2 \rho_A \rho_B - (\sigma_\theta^2 + \sigma_\varepsilon^2)) > 0$, which with $\rho_A > \rho_B$ is (B.3). □

Corollary B.1 is the pairwise analogue of Theorem 1. The endpoint case is different: if $\rho_B = 0$, then $\mathcal{M}_B = 0$, so the completely purified report is unambiguously less gameable.

The rank protocol has its own pairwise comparison, and it is sharper in one direction and weaker in another.

Proposition B.2 (Pairwise comparisons under rank pricing)**.** *Let gaming ability be uninformative and let $0 < \rho_B < \rho_A$.*

1. (Common noise.) *Suppose $\sigma^2_{\varepsilon,A} = \sigma^2_{\varepsilon,B} = \eta > 0$ and $\chi_H > 0$. Then the two reports are two exposures of the balanced rank market of Section 4, and by Theorem 2 report B carries strictly less equilibrium gaming than report A — in the common-shift distortion and in the total action-cost loss — at every level of contamination σ_z^2.*

2. (Noise asymmetry.) *Under the rank-priced stakes process $P(V) = 1/\big(h(q^{\mathrm{med}})\sqrt{2\pi V}\big)$ of Lemma 1, each report admits an exact translation equilibrium with distortion $\Delta_j = (\rho_j/\kappa)\, P(V_j) \propto \mathcal{M}_j^{\mathrm{rank}}$, where*

$$\mathcal{M}_j^{\mathrm{rank}} = \frac{\rho_j}{\sqrt{\sigma_\theta^2 + \rho_j^2 \sigma_z^2 + \sigma_{\varepsilon,j}^2}}, \tag{B.4}$$

and report B induces a strictly larger equilibrium distortion and action-cost loss than report A if and only if

$$\rho_B^2 \big(\sigma_\theta^2 + \sigma_{\varepsilon,A}^2\big) > \rho_A^2 \big(\sigma_\theta^2 + \sigma_{\varepsilon,B}^2\big). \tag{B.5}$$

Proof. Part 1 reads the two reports as the noise-stationary rank market at exposures $\rho_B < \rho_A$: Theorem 2 makes the common-shift distortion, and with uninformative ability the total action-cost loss, strictly increasing in exposure. For part 2, a common shift moves the mean of S^j but not its variance, so the first-order condition $\kappa \Delta_j = \rho_j P(V_j)$ with $V_j = \sigma_\theta^2 + \rho_j^2 \sigma_z^2 + \sigma_{\varepsilon,j}^2$ is an exact translation equilibrium, which is (B.4) up to the constant

$1/\left(h(q^{\text{med}})\sqrt{2\pi}\right)$. Cross-multiplying $\mathcal{M}_B^{\text{rank}} > \mathcal{M}_A^{\text{rank}}$ gives $\rho_B^2 V_A > \rho_A^2 V_B$, in which the contamination terms $\rho_A^2\rho_B^2\sigma_z^2$ cancel, leaving (B.5). With common noise, (B.5) would require $\rho_B^2 > \rho_A^2$, which fails for $\rho_B < \rho_A$, consistent with part 1. □

The two indices formalise how far the naive comparison can be trusted under each protocol. Under quality pricing, contamination itself generates common-noise reversals: trust does the work, and (B.3) binds exactly when gameable variation dominates the report. Under rank pricing, contamination enters the report only through exposure, so it rescales both sides of the pairwise comparison symmetrically and drops out: with common residual noise the equilibrium order coincides with the exposure order — the pairwise counterpart of Theorem 2 — and by (B.5) a reversal requires the less-exposed metric to be less noisy by more than the square of the exposure ratio. The caveat has practical weight, because purer metrics are often also less noisy: a rank market can then game the purer metric harder — not because anyone trusts it more, but because its reports crowd more densely.

Appendix C Closed Matching Markets: Rationing, Atoms, and Coarsening

Sections 4 and 5 use the balanced closed market, in which stable matching is assortative and the induced schedule is $H^{-1} \circ F_\lambda$. This appendix establishes that foundation in general: it allows the mass of positions to differ from the mass of agents, handles atoms in the posterior distribution, derives the marginal incentive with a rationing cutoff, and records the coarsening result used in Section 7. The balanced case used in the main text is the special case $\mu = 1$ of Lemma C.1.

Appendix C.1 Stable matching by posterior rank

Normalise the mass of agents to 1 and let the mass of positions be $\mu > 0$. The distribution H is the distribution of desirability conditional on being a position, so the measure of positions with desirability at most q is $\mu H(q)$. Agents who are not matched receive outside option q_0, with $q_0 \leq \underline{q}$, so every position is weakly preferable to being unmatched. Positions rank agents by posterior expected quality and are indifferent among agents with the same posterior; an unfilled position receives an outside option below the value of being filled by any agent. Stability uses the usual no-blocking-pair condition under these preferences; indifference within equal-posterior tie blocks does not create blocking pairs.

Let F_λ denote the equilibrium distribution of posterior expected quality $m_\lambda(S_\lambda)$ under reporting rule λ. Quantile functions are interpreted as generalised inverses, $G^{-1}(u) = \inf\{y : G(y) \geq u\}$. Assignments are measure-preserving matchings between agents ranked by posterior expected quality and positions ranked by desirability; uniqueness is up to null sets, with uniform tie-breaking within positive-mass posterior classes. For $\mu > 0$, define the lower agent-rank and position-quantile cutoffs

$$c_A(\mu) = (1-\mu)^+, \qquad c_P(\mu) = \left(1 - \frac{1}{\mu}\right)^+. \tag{C.6}$$

Lemma C.1 (Unbalanced closed-market schedule and tie blocks)**.** *Fix λ, a position mass $\mu > 0$, and an equilibrium distribution F_λ of posterior expected quality. Let $\phi_\mu(p) = c_P(\mu) + (p - c_A(\mu))/\mu$ for agent ranks $p \in [c_A(\mu), 1]$.*

1. *A stable closed-market matching exists. Every stable matching is assortative on the matched set. If $\mu < 1$, only agents with posterior ranks at least $c_A(\mu) = 1 - \mu$ are matched. If $\mu > 1$, only positions with desirability quantiles at least $c_P(\mu) = 1 - 1/\mu$ are filled.*

2. *If F_λ is atomless, the stable assignment is unique up to null sets, and an agent with posterior expected quality m receives expected placement value*

$$Q^\mu_\lambda(m) = \begin{cases} q_0, & F_\lambda(m) < c_A(\mu), \\ H^{-1}\left(c_P(\mu) + \dfrac{F_\lambda(m) - c_A(\mu)}{\mu}\right), & F_\lambda(m) \geq c_A(\mu). \end{cases} \tag{C.7}$$

   In particular, when $\mu = 1$ this becomes $Q^1_\lambda(m) = H^{-1}(F_\lambda(m))$.

3. *If F_λ has an atom at m, let $F_\lambda(m^-) = \lim_{t \uparrow m} F_\lambda(t)$. If $F_\lambda(m) \leq c_A(\mu)$, agents in the atom are unmatched and receive q_0. If $F_\lambda(m) > c_A(\mu)$, let $\underline{p}_\lambda(m) = \max\{F_\lambda(m^-), c_A(\mu)\}$ and $\overline{p}_\lambda(m) = F_\lambda(m)$. Under uniform tie-breaking within the posterior class, their expected placement value is*

$$\bar{Q}^\mu_\lambda(m) = \frac{1}{F_\lambda(m) - F_\lambda(m^-)} \left[\left(\underline{p}_\lambda(m) - F_\lambda(m^-)\right) q_0 + \mu \int_{\phi_\mu(\underline{p}_\lambda(m))}^{\phi_\mu(\overline{p}_\lambda(m))} H^{-1}(u)\, du \right]. \tag{C.8}$$

Proof. First, construct the candidate assortative matching: match the upper $\min\{1, \mu\}$ mass of agents, ranked by posterior quality, to the upper $\min\{1, \mu\}$ mass of positions, ranked by desirability, using the rank map ϕ_μ; if an atom of agents has the same posterior, assign the relevant block of position quantiles uniformly within that atom. The construction is measure preserving because an agent-rank increment dp is matched to a position-quantile increment du with $dp = \mu\, du$; the matched agent-rank interval is $[c_A(\mu), 1]$ and the matched position-quantile interval is $[c_P(\mu), 1]$.

Stability of the candidate and the converse are the standard assortative checks, in both directions: a crossing pair among matched pairs blocks, an unmatched agent ranks below every matched agent's posterior when $\mu < 1$, an unfilled position ranks below every filled one when $\mu > 1$, and ties do not create strict blocks. If F_λ is atomless, an agent with posterior m has rank $F_\lambda(m)$, which yields (C.7). If F_λ has an atom at m, agents in the class occupy the rank block $[F_\lambda(m^-), F_\lambda(m)]$; the part below $c_A(\mu)$ is unmatched, the rest is mapped through ϕ_μ, and uniform tie-breaking gives (C.8). □

A simple sufficient condition for atomlessness is that S_λ have a density and the posterior rule be strictly monotone on the relevant support; in the Gaussian translation benchmark, $m_\lambda(s) = W_\lambda(s - r_\lambda \Delta_\lambda)$ with $W_\lambda > 0$, so the condition holds.

Assumption C.1 (Closed-market local regularity and cutoff regularity)**. (i)** Fix λ, μ, a type (θ, z), and a compact interval $I \subset \mathbb{R}$ of feasible deviations. The distribution F_λ is atomless and continuously differentiable with density f_λ on an open interval containing every matched-interior posterior value reached by actions in I; the density h is strictly positive and continuous on the corresponding position support; the selected posterior assessment m_λ is continuously differentiable on an open set containing the signal values reached by actions in I; and there exists an integrable random variable D_I^μ dominating $\sup_{x \in I} |\Gamma_\lambda^\mu(m_\lambda(\theta + \rho(\lambda)x + \varepsilon_\lambda))\, m_\lambda'(\theta + \rho(\lambda)x + \varepsilon_\lambda)|$, where Γ_λ^μ is defined in (C.10).

(ii) When $\mu < 1$ and a rationing cutoff can be reached by deviations in I, let $G_\lambda(s) = F_\lambda(m_\lambda(s)) - c_A(\mu)$ on the relevant signal region and $\mathcal{Z}_\lambda = \{s : G_\lambda(s) = 0\}$. The set $\mathcal{Z}_\lambda$ is finite; G_λ is continuously differentiable near its elements with $G_\lambda'(s_i) \neq 0$; and the conditional law of S_λ given (θ, z, x) has a density $\ell_\lambda(\cdot \mid \theta, z, x)$ continuous at every $s_i \in \mathcal{Z}_\lambda$, generated locally by the location shift $S_\lambda = \theta + \rho(\lambda)x + \varepsilon_\lambda$: $\ell_\lambda(s \mid \theta, z, x) = \bar{\ell}_\lambda(s - \rho(\lambda)x \mid \theta, z)$ near the cutoff preimages.

Lemma C.2 (Derivative of a shifted cutoff probability)**.** *Fix a finite set of regular boundaries $\mathcal{C} = \{s_1, \ldots, s_K\}$ for a set $\mathcal{A} = \{s : G(s) > 0\}$, with G continuously differentiable near each s_i and $G'(s_i) \neq 0$, and suppose $\mathcal{A}$ is, up to endpoints at infinity, a finite union of open intervals whose finite endpoints are the s_i. Let $S(x) = U + \rho x$, where $\rho \geq 0$ and U has a density $\bar{\ell}$ continuous at $s_i - \rho x$ for each i. Then*

$$\frac{d}{dx} \Pr(S(x) \in \mathcal{A}) = \rho \sum_{i=1}^{K} \operatorname{sgn}(G'(s_i))\, \ell_x(s_i), \qquad \ell_x(s) = \bar{\ell}(s - \rho x).$$

Proof. Write $\mathcal{A}$ as a finite union of disjoint open intervals (a_k, b_k), so $\Pr(S(x) \in \mathcal{A}) = \sum_k \int_{a_k-\rho x}^{b_k-\rho x} \bar{\ell}(u)\,du$. Differentiating each interval integral where the finite endpoint densities are continuous gives $\rho \sum_k [\bar{\ell}(a_k - \rho x) - \bar{\ell}(b_k - \rho x)]$; a finite lower endpoint of a matched interval has $G'(s_i) > 0$ and contributes $+\rho \ell_x(s_i)$, an upper endpoint $-\rho \ell_x(s_i)$. □

Proposition C.3 (Local marginal incentive in an unbalanced closed market)**.** *Suppose Assumption C.1 holds. On matched interiors,*

$$(Q^{\mu}_{\lambda})'(m) = \frac{f_{\lambda}(m)}{\mu\, h(Q^{\mu}_{\lambda}(m))}, \tag{C.9}$$

and define the matched-interior rank-price term

$$\Gamma^{\mu}_{\lambda}(m) = \begin{cases} \dfrac{f_{\lambda}(m)}{\mu\, h(Q^{\mu}_{\lambda}(m))}, & F_{\lambda}(m) > c_A(\mu), \\ 0, & F_{\lambda}(m) < c_A(\mu). \end{cases} \tag{C.10}$$

If no rationing cutoff is crossed by deviations in I, then for every x in the interior of I,

$$\frac{\partial R^{\mu}_{\lambda}(x;\theta,z)}{\partial x} = \rho(\lambda)\mathbb{E}\left[\Gamma^{\mu}_{\lambda}(m_{\lambda}(S_{\lambda}))m'_{\lambda}(S_{\lambda}) \mid \theta, z, x\right]. \tag{C.11}$$

When $\mu < 1$ and the cutoff preimages satisfy Assumption C.1(ii)*, the derivative exists and equals* (C.11) *plus the cutoff-crossing term*

$$\rho(\lambda)\big(H^{-1}(0) - q_0\big) \sum_{s_i \in \mathcal{Z}_{\lambda}} \operatorname{sgn}\big(G'_{\lambda}(s_i)\big)\, \ell_{\lambda}(s_i \mid \theta, z, x). \tag{C.12}$$

If $q_0 = H^{-1}(0)$, the cutoff-crossing term vanishes; with a unique increasing cutoff it reduces to $\rho(\lambda)(H^{-1}(0) - q_0)\, \ell_{\lambda}(s^c_{\lambda} \mid \theta, z, x)$.

Proof. On matched atomless regions, Lemma C.1 gives $Q^{\mu}_{\lambda}(m) = H^{-1}(c_P(\mu) + (F_{\lambda}(m) - c_A(\mu))/\mu)$; differentiating with the inverse-function rule gives (C.9), and on unmatched interiors the value is the constant q_0. Away from cutoffs the composition $Q^{\mu}_{\lambda}(m_{\lambda}(s))$ is continuously differentiable with derivative $\Gamma^{\mu}_{\lambda}(m_{\lambda}(s))m'_{\lambda}(s)$, and the dominated-convergence argument of Proposition 1 yields (C.11). For the cutoff term, decompose the schedule locally into a matched-interior continuously differentiable part plus the jump $J\mathbf{1}\{S_{\lambda} \in \mathcal{A}_{\lambda}\}$ with $J = H^{-1}(0) - q_0 \geq 0$ and $\mathcal{A}_{\lambda} = \{s : G_{\lambda}(s) > 0\}$; Assumption C.1(ii) allows Lemma C.2 to be applied to the jump probability under the local location shift, giving (C.12). □

Closed markets thus add two terms to the exposure–sensitivity decomposition: a *rank-price* term on matched interiors — whose balanced case, combined with the linear posterior, is the trust-cancellation identity (16) of the main text — and a *cutoff-crossing* term when the market rations. The cutoff term is the reason rationed markets fall outside the scalar characterisation of Section 5: it attaches a fixed jump J, the value of being matched at all, to an agent-specific density $\ell_\lambda(s^c_\lambda \mid \theta, z, x)$, so the effective stakes process is type-dependent and combines a rank-priced intensive margin with a fixed-stakes extensive margin. The next subsection resolves the conjecture this invites — that the fixed stake should restore the backfire of Theorem 1 — in two halves: not at first order in stakes, where the aggregate recombines the agent-specific densities into a population density that is itself a rank price, and yes at large admission stakes, where equilibrium feedback restores a modest interior peak in mean gaming.

Appendix C.2 The admission margin: small stakes versus large stakes

We linearise the rationed market in the scale of stakes: replace every desirability q and the outside option q_0 by τq and τq_0, and expand the equilibrium around $\tau = 0$. At $\tau = 0$ no agent games; the report is $S^0_\lambda = \theta + r_\lambda z + \varepsilon_\lambda \sim N(0, V^{\mathrm{o}}_\lambda + \eta)$ with density f_S, the posterior is exactly linear, the matched set is the top μ of reports, and the cutoff is $s^c_\lambda = z_\mu \sqrt{V^{\mathrm{o}}_\lambda + \eta}$ with $z_\mu = \Phi^{-1}(1-\mu)$. At order τ, each type's distortion is κ^{-1} times the marginal return of Proposition C.3, evaluated at this no-gaming benchmark.

Proposition C.4 (No aggregate revival at first order in stakes)**.** *Let gaming ability be uninformative and the family noise-stationary with $\eta > 0$; fix $\mu \in (0,1)$, a prize distribution H with support bounded below, and an admission stake $J = H^{-1}(0) - q_0 \geq 0$. The order-$\tau$ distortion of type (θ, z) is*

$$a_1(\theta, z) = \frac{r_\lambda}{\kappa}\,\mathbb{E}\left[\frac{f_S(S^0_\lambda)}{\mu\, h\big(Q^\mu_\lambda(m_\lambda(S^0_\lambda))\big)}\,\mathbf{1}\{S^0_\lambda > s^c_\lambda\}\;\Big|\;\theta, z\right] + \frac{r_\lambda}{\kappa}\,J\,\phi_\eta\big(s^c_\lambda - \theta - r_\lambda z\big), \tag{C.13}$$

with ϕ_η the $N(0,\eta)$ density, and its population mean is

$$\mathbb{E}[a_1] = \frac{r_\lambda}{\kappa\sqrt{V^{\mathrm{o}}_\lambda + \eta}}\,\Lambda, \qquad \Lambda = \int_{1-\mu}^{1} \frac{\phi\big(\Phi^{-1}(p)\big)}{\mu\, h\big(H^{-1}\big(\frac{p-1+\mu}{\mu}\big)\big)}\,dp + J\,\phi(z_\mu), \tag{C.14}$$

where Λ is invariant along the reform path. Hence $\mathbb{E}[a_1]$ is strictly increasing in exposure on $(0,1]$, for every rationing level μ, prize distribution H, and stake J: to first order in stakes, partial purification monotonically reduces aggregate gaming in the rationed rank market. With identical *positions and $J > 0$ — the pure admission contest, a degenerate-prize limit outside the maintained positive-density assumption, in which the integral term is absent — the mean-squared distortion $\mathbb{E}[a_1^2]$ is strictly increasing in exposure as well.*

Proof. Scaling prizes by τ scales the matched-interior term Γ_λ^μ and the jump J linearly, so the order-τ first-order condition sets κa_1 equal to r_λ times the bracket of (C.11)–(C.12) at the no-gaming benchmark. The density identity $f_m(m_\lambda(s))\, m_\lambda'(s) = f_S(s)$ — the unbalanced analogue of the trust cancellation (16) — turns $\Gamma_\lambda^\mu(m_\lambda(s))\, m_\lambda'(s)$ into $f_S(s)/\big(\mu h(Q_\lambda^\mu(m_\lambda(s)))\big)$, giving (C.13). For the mean, iterated expectations and the change of variables $p = F_S(s)$ give

$$\mathbb{E}\left[\frac{f_S(S_\lambda^0)}{\mu h(\cdot)}\mathbf{1}\{S_\lambda^0 > s_\lambda^c\}\right] = \int_{s_\lambda^c}^{\infty} \frac{f_S(s)^2}{\mu\, h\big(H^{-1}\big(\frac{F_S(s)-1+\mu}{\mu}\big)\big)}\, ds = \frac{1}{\sqrt{V_\lambda^{\mathrm{o}}+\eta}} \int_{1-\mu}^{1} \frac{\phi\big(\Phi^{-1}(p)\big)}{\mu\, h\big(H^{-1}\big(\frac{p-1+\mu}{\mu}\big)\big)}\, dp,$$

using $f_S(F_S^{-1}(p)) = \phi(\Phi^{-1}(p))/\sqrt{V_\lambda^{\mathrm{o}}+\eta}$, while the Gaussian convolution $\mathbb{E}_{\theta,z}\big[\phi_\eta(s_\lambda^c - \theta - r_\lambda z)\big] = f_S(s_\lambda^c) = \phi(z_\mu)/\sqrt{V_\lambda^{\mathrm{o}}+\eta}$ handles the cutoff term, which is (C.14); and $r/\sqrt{V^{\mathrm{o}}(r)+\eta}$ has derivative $(\sigma_\theta^2+\eta)/(V^{\mathrm{o}}+\eta)^{3/2} > 0$. For the contest second moment, $\phi_\eta^2 = (2\sqrt{\pi\eta})^{-1}\phi_{\eta/2}$ and the same convolution give

$$\mathbb{E}[a_1^2] = \Big(\frac{r_\lambda J}{\kappa}\Big)^2 \frac{1}{2\sqrt{\pi\eta}}\,\frac{1}{\sqrt{u}}\,\phi\Big(z_\mu\sqrt{\tfrac{V_\lambda^{\mathrm{o}}+\eta}{u}}\Big), \qquad u := V_\lambda^{\mathrm{o}} + \tfrac{\eta}{2},$$

whose logarithmic derivative in r_λ is $2/r_\lambda - \sigma_z^2 r_\lambda/u + z_\mu^2\eta\sigma_z^2 r_\lambda/(2u^2) > 0$, the first two terms summing positive because $2u > \sigma_z^2 r_\lambda^2$. $\square$

The economics of (C.14) is the resolution of the conjecture. Individually, an applicant prices the admission stake with their *own* noise density, which does not deflate as the metric is purified — a $\psi = 0$ stakes process, even safer than rank pricing. What compresses along the reform is the field: purification packs the population more densely around the cutoff, and in aggregate the agent-specific densities recombine into the population report density there, $f_S(s_\lambda^c) \propto (V_\lambda^{\mathrm{o}}+\eta)^{-1/2}$. In aggregate, the rationed market is again a $\psi = \frac{1}{2}$ stakes process: at first order in stakes, rationing leaves the market on the safe side of the frontier of Proposition 4.

Remark C.1 (Incidence: individual humps under a monotone aggregate). What rationing changes at small stakes is incidence, not direction. In the pure contest, type (θ, z) has $a_1 \propto r_\lambda\, \phi_\eta(d)$ with cutoff distance $d = s_\lambda^c - \theta - r_\lambda z$, and gaming falls with exposure at r_λ whenever $(-d/\eta)\big(z - z_\mu\sigma_z^2 r_\lambda/\sqrt{V_\lambda^{\mathrm{o}}+\eta}\big) > 1/r_\lambda$. A high-$z$ applicant who sits safely above the cutoff under the raw metric drifts into the contested window as purification strips the gameable advantage $r_\lambda z$: their own-noise density rises faster than exposure falls, and their gaming is hump-shaped along the reform. At $\sigma_\theta^2 = \kappa = 1$, $\sigma_z^2 = 5$, $\eta = 0.15$, $\mu = 0.3$, the type $(\theta, z) = (0, 2)$ has a_1 peaking near $\lambda \approx 0.45$ at roughly 2.3 times its raw-metric level, even though the population mean falls monotonically throughout.[38]

Beyond first order the conjecture has content. The contest objective is strictly concave provided $\kappa > r_\lambda^2 J e^{-1/2}/(\eta\sqrt{2\pi})$, and within that region the exact pure-contest equilibrium is a one-dimensional fixed point: the distortion depends on the type only through $w = \theta + r_\lambda z$, and the cutoff clears capacity. Solving it numerically, mean gaming acquires an interior peak when the stake is large and seats are scarce: at $\sigma_\theta^2 = \kappa = 1$, $\sigma_z^2 = 5$, $\eta = 1$, $\mu = 0.1$, the peak appears once J exceeds roughly 0.7 of the concavity bound, and at J equal to 0.85 (0.97) of the bound mean gaming peaks at $\lambda \approx 0.10$ (0.15), about half a per cent (1%) above its raw-metric level. The effect strengthens with scarcity and contamination, disappears by $\mu = 0.3$ at these stakes, and the action-cost loss remained monotone in every computed configuration. The backfire at the admission margin is therefore a genuinely large-stakes phenomenon — absent from the linearised market for every (μ, H, J), present but modest near the concavity bound — and the regime beyond that bound, where admission contests develop jump responses, is the margin we leave open.[39]

Appendix C.3 Coarsening after complete purification

Lemma C.3 (Coarsening preserves strong gaming neutrality)**.** *Let T be a strongly gaming-neutral statistic and let M be any public message generated from T through a Markov kernel $K(dm \mid T)$ that does not depend on the action except through T. Then M is strongly gaming-neutral.*

Proof. Fix a type in the full-measure set on which $\mathcal{L}(T \mid \theta, z, x)$ is action-invariant and a measurable set B in the message space. By iterated expectations, $\Pr(M \in B \mid \theta, z, x) =$

[38] This is the incentive geometry of 'educational triage' under threshold-based accountability: instructional effort concentrated on students near the proficiency cutoff at the expense of those far from it (Neal and Schanzenbach, 2010; Neal, 2010). The NCLB cutoff was exogenous — a fixed stake with no market-clearing feedback, so nothing protected the aggregate — whereas here the cutoff clears a fixed capacity, and that single difference is what restores aggregate monotonicity at moderate stakes.

[39] The configuration — severe scarcity, concentrated terminal prizes — is that of elimination ladders (Rosen, 1986). Read as enforcement, the computations say that improved testing reduces mean cheating monotonically in balanced ladders yet can locally raise it in winner-take-all settings: a path statement absent from the cheating-in-contests literature (Gilpatric, 2011; Kräkel, 2007), which fixes the detection regime. Thin, high-stakes fields are also where sabotage and collusion revive (Lazear, 1989), so scarcity weakens rank immunity in every direction at once.

$\mathbb{E}[K(B \mid T) \mid \theta, z, x]$; the inner object is a measurable function of T alone, and the conditional distribution of T is action-invariant, so the expectation is too. □

Complete purification and subsequent disclosure are therefore separable: one may first remove the gaming margin and then choose how much of the purified statistic to reveal. Coarsening a raw gameable signal is different: it dampens incentives only by reducing responsiveness to both non-quality actions and genuine information.[40]

Appendix D First-Order Validity of the Quadratic Specification

The quadratic matching payoff of Assumption 3 is the payoff primitive of Sections 3–5: agents optimise against $Q(m) = \beta m + \frac{\gamma}{2}m^2$ exactly, and for $\gamma \neq 0$ that function is not globally increasing, whereas every schedule the model is meant to describe is. This appendix fills the gap. We show that the quadratic model is the correct first-order model of *any* smooth, strictly increasing schedule with the same local slope and curvature: its principal-branch equilibrium remains exactly Bayes-consistent in the model with the increasing schedule, and is an approximate equilibrium of it, with a strategy error of order t^2 and a payoff error of order t^4 in the curvature scale t, whereas the curvature effects the quadratic model is used to measure are of order t. Subsection Appendix D.1 reports the approximation quality at the configurations plotted in the paper.

The structural observation that makes this work is an asymmetry between the two sides of the equilibrium fixed point. The market side is exact: with an affine population strategy and Gaussian primitives the report is Gaussian, so the Bayes posterior is exactly linear regardless of the matching schedule. The agent side is where the approximation lives: against a non-quadratic schedule, the best response to linear beliefs is not exactly affine. The proposition below bounds that error.

To formalise "a general increasing schedule with the same local slope and curvature", fix $\beta > 0$ and $\gamma_0 \in \mathbb{R}$ and consider a family of placement schedules $\{Q_t\}_{t \in (0,\bar{t}]}$, each strictly increasing and three times continuously differentiable, satisfying, after the normalisation

[40]Coarse ranks carry their own design literature: Meyer (1991) on coarse and biased contests, Moldovanu et al. (2007) on status categories, and Dubey and Geanakoplos (2010), who show that coarse grading of *raw* scores can raise effort when students care about rank — an effort objective, complementary to the neutrality result here, which concerns coarsening *after* complete purification. A cap is a coarsening that creates an atom at the boundary, and caps can raise aggregate outlays in all-pay contests (Che and Gale, 1998): the contest counterpart of a grade ceiling.

$Q_t(0) = 0,$

$$Q_t'(0) = \beta, \qquad Q_t''(0) = \gamma_0\, t, \qquad \sup_m |Q_t''(m)| \le k_2\, t, \qquad \sup_m |Q_t'''(m)| \le k_3\, t^2, \quad \text{(D.15)}$$

for constants k_2, k_3. The canonical example is the scaled family $Q_t(m) = t^{-1}Q(tm)$ for a single strictly increasing $Q \in C^3$ with $Q(0) = 0$, $Q'(0) = \beta$, $Q''(0) = \gamma_0$, and bounded second and third derivatives.

Two objects are compared. The *quadratic benchmark* is the model of Section 3 with curvature $\gamma = \gamma_0 t$; for small t its selected-assessment affine equilibrium is unique by Proposition F.6(ii) and we write it (x_t, m_t), with posterior slope W_t and amplification $g_\theta(t)$. The *curved-schedule model* replaces the quadratic by Q_t and changes nothing else; $x_t^{\mathrm{BR}}(\theta, z)$ denotes the exact best response to (m_t, Q_t).

Proposition D.5 (First-order validity of the quadratic specification)**.** *Fix λ with $r_\lambda > 0$ and $\sigma_\varepsilon^2(\lambda) > 0$, maintain Assumptions 2 and 4 together with the quadratic private cost, and let $\{Q_t\}$ satisfy* (D.15)*. There exist $t_0 \in (0, \bar{t}\,]$ and $K_0 < \infty$, depending only on the primitives and on (γ_0, k_2, k_3), such that for every $t \le t_0$ the quadratic benchmark has a unique selected-assessment affine equilibrium (x_t, m_t), and:*

1. (Exact market-side consistency.) *Under the population strategy x_t, the report is Gaussian and m_t is the exact Bayes posterior in the curved-schedule model, for the selected continuous Gaussian posterior version at every report value.*

2. (Approximate optimality, with orders.) *For every type, the objective in the curved-schedule model is strictly concave, with second derivative in $[-2\kappa, -\kappa/2]$, and its unique maximiser satisfies*

$$\left|x_t^{\mathrm{BR}}(\theta, z) - x_t(\theta, z)\right| \le K_0\, t^2\big(1 + \theta^2 + z^2\big),$$

$$0 \le U_t\big(x_t^{\mathrm{BR}}\big) - U_t\big(x_t\big) \le K_0\, t^4\big(1 + \theta^2 + z^2\big)^2,$$

and the expected payoff shortfall over the type distribution is at most $K_0\, t^4$.

3. (First-order formulas.) *Along the branch, $g_\theta(t) - 1 = (W_\lambda^2 r_\lambda^2/\kappa)\gamma_0 t + O(t^2)$ and $\alpha_\theta(t) = (W_\lambda^2 r_\lambda/\kappa)\gamma_0 t + O(t^2)$, and the same expansion describes sorting in the curved-schedule model under population behaviour x_t:*

$$I_{\gamma_0 t}(r) = I(r) + \frac{2C(r)^2 \eta(r)}{\big(V^{\mathrm{o}}(r) + \eta(r)\big)^2}\, \frac{W_\lambda^2 r^2}{\kappa}\, \gamma_0 t + O(t^2).$$

If, in addition, $\sigma_\varepsilon^2(\cdot)$ is continuous with $\inf_\lambda \sigma_\varepsilon^2(\lambda) = \underline{\eta} > 0$, then t_0 and K_0 can be chosen uniformly over $\lambda \in [0,1]$.

Proof. Fix λ and abbreviate $r = r_\lambda > 0$ and $\eta = \sigma_\varepsilon^2(\lambda) > 0$. By Proposition F.6(ii) there is a neighbourhood of $\gamma = 0$ on which the fixed point (8) has a unique solution $(g_\theta(\gamma), W(\gamma))$, depending smoothly on γ; choose t_2 so that $\gamma_0 t$ lies in a compact subinterval for all $t \le t_2$, and set $\bar{g} = \sup_{t\le t_2} g_\theta(t)$, $\bar{W} = \sup_{t\le t_2} W_t$.

Part (i). Under the affine population strategy the report is an affine function of the Gaussian vector $(\theta, z, \varepsilon_\lambda)$, so $\mathbb{E}[\theta \mid S_\lambda = s]$ is the linear regression $W_t(s - \mathbb{E}[S_\lambda]) = m_t(s)$; no step involves the matching schedule, and $\eta > 0$ makes m_t the continuous version at every report value.

A Taylor remainder. Let $\zeta_t(m) = Q_t(m) - \beta m - \frac{\gamma_0 t}{2} m^2$. By (D.15), $\zeta_t(0) = \zeta_t'(0) = \zeta_t''(0) = 0$ and $|\zeta_t'''| \le k_3 t^2$, so integrating,

$$|\zeta_t''(m)| \le k_3 t^2 |m|, \qquad |\zeta_t'(m)| \le \tfrac{k_3}{2}\, t^2 m^2, \qquad |\zeta_t(m)| \le \tfrac{k_3}{6}\, t^2 |m|^3; \tag{D.16}$$

likewise $|Q_t'(m)| \le \beta + k_2 t|m|$, so all expectations below are finite and derivatives pass through them.

Part (ii): uniform strong concavity. Writing $\xi = \theta + rx - \mathbb{E}[S_\lambda]$, the curved-schedule objective U_t has $\partial^2 U_t/\partial x^2 = W_t^2 r^2\, \mathbb{E}[Q_t''(m_t(S_\lambda)) \mid \theta, z, x] - \kappa \in [-\kappa - k_2 t\bar{W}^2,\ k_2 t\bar{W}^2 - \kappa]$, so for $t \le t_0 := \min\{t_2, \kappa/(2k_2\bar{W}^2)\}$ the objective is strictly concave with second derivative in $[-2\kappa, -\kappa/2]$ and has a unique maximiser.

Part (ii): from a first-order-condition gap to a strategy gap. Let $\tilde{U}_t$ denote the quadratic-benchmark objective against the same posterior. Then $U_t - \tilde{U}_t = \mathbb{E}[\zeta_t(m_t(S_\lambda)) \mid \theta, z, x]$, and (D.16) with $\mathbb{E}[m_t(S_\lambda)^2 \mid \theta, z, x] = W_t^2(\xi^2 + \eta)$ gives

$$\left| \frac{\partial U_t}{\partial x} - \frac{\partial \tilde{U}_t}{\partial x} \right| = W_t r \left| \mathbb{E}[\zeta_t'(m_t(S_\lambda)) \mid \theta, z, x] \right| \le \frac{k_3}{2}\, t^2\, W_t^3 r\, (\xi^2 + \eta).$$

Since $x_t(\theta, z)$ maximises $\tilde{U}_t$, at $x = x_t(\theta, z)$ the conditional mean report is $\xi = g_\theta(t)(\theta + rz)$, so $|\partial U_t/\partial x(x_t)| \le \frac{k_3}{2} t^2 \bar{W}^3 (2\bar{g}^2(\theta^2 + z^2) + \eta)$. Strong concavity converts the first-order-condition gap into a maximiser gap, $|x_t - x_t^{\mathrm{BR}}| \le (2/\kappa)|\partial U_t/\partial x(x_t)|$, giving the strategy bound with $K_0 = (k_3\bar{W}^3/\kappa)\max\{2\bar{g}^2, \eta\}$. A second-order Taylor expansion of U_t about x_t^{BR}, whose second derivative is bounded by 2κ, gives $0 \le U_t(x_t^{\mathrm{BR}}) - U_t(x_t) \le \kappa(x_t^{\mathrm{BR}} - x_t)^2$, and the Gaussian type distribution has finite fourth moments; enlarging K_0 absorbs the constants.

Part (iii). Along the branch, $g_\theta(\gamma) = \kappa/(\kappa - \gamma r^2 W(\gamma)^2)$ gives $g_\theta'(0) = r^2 W_\lambda^2/\kappa$; smoothness bounds the second derivative on the compact subinterval, and evaluating at $\gamma = \gamma_0 t$

gives the expansions of $g_\theta(t) - 1$ and $\alpha_\theta(t) = (g_\theta(t) - 1)/r$. For sorting, write (25) as $I = C(r)^2/(V^{\mathrm{o}}(r) + \eta(r)/g_\theta^2)$; its derivative at $g_\theta = 1$ is $2C(r)^2\eta(r)/(V^{\mathrm{o}}(r) + \eta(r))^2$, and the chain rule gives the display. The rank-relevant signal, and hence the index, is determined by the population strategy and the Gaussian primitives, which by part (i) are the same objects in the two models.

Uniformity. At any solution of (8) with $g_\theta > 0$, the consistency equation and the arithmetic–geometric mean inequality give $W \le C_\lambda/(2\sqrt{V_\lambda^{\mathrm{o}}\underline{\eta}})$, bounded over $\lambda \in [0,1]$; the elimination identity recorded in Supplemental Appendix F then confines all solutions, for $|\gamma|$ small uniformly in λ, to a neighbourhood of $(1, W_\lambda)$ in which a quantitative implicit-function argument yields a branch radius, local uniqueness, and derivative bounds uniform in λ. Every constant above depends on λ only through objects ranging over a compact set. At any λ with $r_\lambda = 0$ the claims hold trivially. □

This is the precise sense in which $\gamma = Q''(0)$ is the only feature of an increasing curved schedule with the uniformly controlled derivatives of part (i) that matters locally, as asserted in Section 6: the market side transfers exactly, while the agent side is approximate with errors one order smaller than the curvature effects of interest.

Remark D.2 (Exact equilibria of the curved-schedule model). Proposition D.5 deliberately stops short of an existence theorem for the curved-schedule model itself: an exact equilibrium there is a fixed point in an infinite-dimensional space, since a non-affine population strategy generates a non-Gaussian report, hence a non-linear posterior, hence a new best response. We conjecture that a contraction argument delivers, for small t, an exact equilibrium within $O(t^2)$ of x_t; the approximate-equilibrium statement already carries the first-order content the paper uses.

Appendix D.1 Approximation quality at the plotted configurations

The first-order closed-market equilibria of Sections 4 and 5 apply the quadratic benchmark at the slope and curvature induced by the market. Table D.2 reports, for each configuration plotted in the paper, the expected payoff gain from re-optimising every type against the *exact* induced schedule ($H^{-1} \circ F_\lambda$ under ranks; its integral under transfers), holding the equilibrium posterior fixed — the natural ϵ for which the first-order equilibrium is an ϵ-equilibrium of the true model.

Two lessons follow. First, in the configurations the paper plots, the affine strategies are within roughly 1% of optimal or better, with the gap declining monotonically along the reform path and concentrated in tail types, consistent with the $(1 + \theta^2 + z^2)^2$ payoff

Table D.2: Expected payoff gap between the affine strategy and the exact best response.

Protocol / prizes	Configuration	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 0.8$	$\lambda \approx 0.9$
Rank, exponential ($\xi = 0.32$)	Fig. 4	0.96%	0.45%[a]	0.05%	$< 0.05\%$
Rank, exponential ($\xi = 0.12$)	$\sigma_z^2 = 5$, $\eta = 0.15$	1.23%	0.89%	0.31%	0.06%
Transfers, concentrated	Fig. 2(a)	0.04%	0.04%	0.01%	$< 0.01\%$
Transfers, exponential ($\xi = 0.12$)	not plotted	32%	26%	5%	—

Entries are the expected payoff shortfall of the affine strategy, over the type distribution, as a percentage of the mean prize (rank rows) or of the wage spread over ± 2 posterior standard deviations (transfer rows); the concentrated configuration has $q_0 = 5$, $\delta = 1$. The transfer/exponential row is outside the licensed regime of Proposition D.5: the induced wage has cubic tails, so the unrestricted quadratic-cost objective is unbounded above and that entry is a re-optimisation over the bounded posterior-standard-deviation action window—an indication that the affine approximation degrades, not a global ϵ-equilibrium gap.

envelope of Proposition D.5. Second, the transferable-utility market with *pure* exponential prizes is not in the licensed regime: the wage schedule there integrates an exponential prize schedule and so has cubic tails, the dimensionless curvature does not shrink with ξ, and the quadratic model is then an organising device rather than a quantitative approximation. This is why Proposition 5(iii) handles general transferable-utility prizes through the concentrated-prize limit — where the licence is restored, as the third row shows — rather than through the exponential example.[41]

Appendix E Applications

The framework is best read as a theory of evaluation-then-allocation environments: an institution first produces a public evaluation, and a matching or allocation rule then converts that evaluation into scarce opportunities. The characterisation of Section 5 adds a classification step to each application: identify what the evaluation buys, and the comparative statics of partial reform follow. The subsections that follow are interpretive mappings of the formal results onto institutional settings, not additional solved models, and two scope notes apply throughout. First, the framework applies to the residual non-quality component of the action; when the same action also improves latent quality, only

[41] With an unbounded action set, the agent's exact objective against this cubic-tailed wage is itself unbounded above, so no global exact best response exists, and the transfer/exponential entry in Table D.2 is a windowed re-optimisation reported only to show that the affine approximation degrades there. By contrast the *rank* schedule under exponential prizes grows only quadratically—the quantile $H^{-1} \circ F_\lambda$ inherits the Gaussian square in the exponent—so its exact best response is well defined and the rank/exponential rows are bona fide ϵ-gaps.

the residual is identified. Second, the non-monotonicity is a property of partial reform: complete purification removes the targeted margin in every regime, and Lemma C.3 shows that action-independent coarsening of a completely purified statistic preserves neutrality.

Appendix E.1 Grades, course choice, and eigengrades

Grades are the leading direct application. Let θ denote latent student quality as perceived by employers or graduate programmes. The raw transcript also reflects course difficulty, grading standards, course choice, and instructor leniency; the residual action x is the part of those choices that moves measured grades without moving latent quality. Difficulty-adjusted grades and eigengrade-style procedures (Gans ⓡ Kominers, 2026) are exposure reductions: they separate student effects from course or grader effects using overlapping transcript data, usually partially, and their role in the framework is not to reveal more information but to reduce the loading of the report on a non-quality component.

The two backfire theorems then deliver the prediction developed in Section 7: where transcripts feed wage-setting employers, a partial correction can earn the transcript enough additional trust to outpace the fall in exposure, and the equilibrium return to course-shopping on the still-exposed component rises exactly when grade variation is dominated by non-ability components; where transcripts feed seat-constrained admissions, every increment of purification reduces aggregate gaming at moderate admission stakes, with the large-stakes caveat of Supplemental Appendix C.2.[42] This complements the grade-inflation literature (Sabot and Wakeman-Linn, 1991; Bar et al., 2009; Chan et al., 2007; Popov and Bernhardt, 2013), which studies how grading standards and course evaluations shape grades; we abstract from those institutional details to isolate the equilibrium return to the remaining gaming margin after partial correction. Transcript production is also multi-agent — students, instructors, departments — and the scalar action should be read either as one targeted margin at a time or as a reduced-form residual.

Appendix E.2 School choice through the rank lens

School choice is a structural analogue in which the strategic report is an ordered list rather than a scalar, and the mechanism maps lists, priorities, capacities, and tie-breakers into assignments (Gale and Shapley, 1962; Roth, 1985; Pathak and Sönmez, 2013). Fix a school s and a student i, let θ_{is} denote the part of the student's claim that the mechanism is intended

[42]Bodoh-Creed and Hickman (2018) model the admissions side quantitatively as a large contest with pre-college investment; their investment margin is productive, which isolates by contrast the wasteful margin studied here. Dubey and Geanakoplos (2010) treat the grading rule itself as a contest-design choice.

to respect, and x_{is} a rank-position or list-construction margin that moves assignment probability without changing θ_{is}. Where assignment probabilities are differentiable in a reduced-form index $S_{is}^{\lambda} = \theta_{is} + \rho(\lambda)x_{is} + \varepsilon_{is}^{\lambda}$, the local return to the strategic margin is

$$\frac{\partial R_{is}^{\lambda}}{\partial x_{is}} = \underbrace{\rho(\lambda)}_{\text{exposure}} \times \underbrace{\frac{\partial \Pr_i^{\lambda}(s)}{\partial S_{is}^{\lambda}}}_{\text{assignment sensitivity}} \times \underbrace{(u_{is} - u_{i,\text{next}})}_{\text{value of the marginal school}},$$

the school-choice analogue of (3). The Boston and first-preference-first mechanisms correspond to high exposure; equal-preference and deferred-acceptance reforms reduce it; unrestricted student-proposing deferred acceptance removes the targeted student-side margin altogether and is the analogue of complete purification. The new observation supplied by Section 4 is that assignment systems are *rank-priced*: seats are fixed and allocated by relative standing, the assignment sensitivity is a density of competing claims, and the value of the marginal school does not inflate as the system becomes better informed. The rank-market logic therefore predicts that partial reforms of the evaluation inputs to centralised assignment — purifying priority scores or test components — reduce input-gaming monotonically, in contrast to the decentralised, price-mediated settings where Theorem 1 bites. This complements Pathak and Sönmez (2013): our margin operates upstream, on the evaluation that enters the mechanism, with the rationing caveat of Supplemental Appendix C.2 where seats are scarce.[43]

Appendix E.3 Capacity choice and imperfect competition

Capacity choice in two-sided matching is a second structural analogue (Azevedo, 2014), in which the strategic margin is a real allocation variable that moves market-clearing cutoffs rather than an evaluative report. Building on the supply-and-demand representation of Azevedo and Leshno (2016), let $P_i(q)$ denote firm i's cutoff at capacity vector q. In differentiable regions, marginal revenue from capacity satisfies

$$MR_i(q) = P_i(q) - \underbrace{\sum_{j \neq i} \big(\bar{P}_{ij}(q) - P_i(q)\big) M_{ij}(q) \left(-\frac{\partial P_j(q)}{\partial q_i}\right)}_{\text{capacity-gaming wedge}},$$

[43]Neal and Root (2024) study the same mechanisms as joint providers of information and incentives, emphasising parents' limited information about schools and the use of mechanism outputs — submitted rankings, enrolment — to reward or de-authorise schools. Those proposals inherit the present dichotomy one level up: per-pupil funding formulas price school quality and are hump-prone on whatever margin schools game, while authorisation contests among comparable schools price rank.

where $M_{ij}(q)$ is the mass of workers on the margin between firms i and j — those who prefer j and are marginal at its cutoff, so that own-cutoff entrants are valued at $P_i(q)$ in the direct term and only rivals' cutoff movements carry a composition value — and $\bar{P}_{ij}(q)$ the corresponding average productivity: cutoff sensitivity, mass at the margin, and composition value play the roles of posterior sensitivity, density, and matching value. Writing effective capacity as $\tilde{q}_i^\lambda = \bar{q}_i - \rho(\lambda)x_i$, the endpoint $\lambda = 1$ represents verified capacity or commitment, and the local return to withholding, net of the direct value of the marginal position, is $\rho(\lambda)$ times the wedge. Personalised wages sharpen the analogy: efficient for fixed capacities, they can raise the return to capacity withholding through rivals' bidding, so coarser uniform wages can dominate *because* they blunt the residual margin. Large-market strategy-proofness (Azevedo and Budish, 2019) removes an individual's ability to move aggregate cutoffs, not the incentive to move one's own evaluative signal — the margin this paper locates.

Appendix F Equilibrium Multiplicity and the Principal Branch

This appendix states and proves the existence and uniqueness result for the fixed-(β, γ) model summarised in Section 3, records the multiplicity that motivates the principal-branch selection, and proves the frontier expansion of Remark 1.

Proposition F.6 (Existence, second-order condition, and uniqueness)**.** *Maintain the assumptions of Proposition 2 with $\sigma_\varepsilon^2(\lambda) > 0$ and $r_\lambda > 0$. (i) For every γ there is an affine equilibrium whose posterior slope solves* (8) *and at which $D_\lambda > 0$; concave schedules have $g_\theta \in (0,1)$ and convex schedules $g_\theta > 1$. (ii) There is a neighbourhood of $\gamma = 0$ on which the equilibrium is unique and depends smoothly on γ, with $g_\theta(0) = 1$ and $W(0) = W_\lambda := C_\lambda / V_\lambda$. (iii) The convex-schedule equilibrium is unique if $\sigma_\varepsilon^2(\lambda) \leq V_\lambda^\circ$, and the concave-schedule equilibrium is unique if $\sigma_\varepsilon^2(\lambda) \geq V_\lambda^\circ$.*

Proof. Write the two scalar relations between W and $g \equiv g_\theta$: the *consistency curve* $W = B(g) := C_\lambda g / (V_\lambda^\circ g^2 + \sigma_\varepsilon^2(\lambda))$ and, inverting the best-response relation on the admissible domain, the *optimality curve* $W = M(g) := \sqrt{\kappa(g-1)/(\gamma r_\lambda^2 g)}$, real for $g \in (0,1)$ when $\gamma < 0$ and for $g \in (1, \infty)$ when $\gamma > 0$. At any intersection $\gamma r_\lambda^2 W^2 = \kappa(g-1)/g < \kappa$, so $D_\lambda = \kappa / g > 0$.

Existence. For $\gamma = 0$ the fixed point forces $g = 1$ and $W = C_\lambda / V_\lambda$. For $\gamma < 0$, M is continuous and strictly decreasing on $(0,1)$ with $M(0^+) = +\infty$, $M(1^-) = 0$, while

$B(0^+) = 0 < B(1)$; for $\gamma > 0$, M is continuous and strictly increasing on $(1, \infty)$ with $M(1^+) = 0$, $M(\infty) = \sqrt{\kappa/(\gamma r_\lambda^2)}$, while $B(1) > 0 = B(\infty)$. In each case $M - B$ changes sign and a crossing exists.

Local uniqueness. Treat (8) as $\Phi(g, W; \gamma) = 0$ with $\Phi_1 = g(\kappa - \gamma r_\lambda^2 W^2) - \kappa$ and $\Phi_2 = W(V_\lambda^{\mathrm{o}} g^2 + \sigma_\varepsilon^2(\lambda)) - C_\lambda g$. At $\gamma = 0$ the unique solution is $(1, W_\lambda)$ and the Jacobian in (g, W) is lower-triangular with nonzero diagonal (κ, V_λ); the implicit function theorem gives a unique smooth branch. The branch exhausts the equilibria near the knife edge: B is bounded, $B(g) \leq C_\lambda / (2\sqrt{V_\lambda^{\mathrm{o}} \sigma_\varepsilon^2(\lambda)})$, while for any $\delta \in (0, 1)$, $M(g)^2 \geq \kappa\delta/((1+\delta)|\gamma| r_\lambda^2)$ on $|g - 1| \geq \delta$, so for $|\gamma|$ small no crossing has $|g - 1| \geq \delta$, and continuity of B places every solution inside the implicit-function neighbourhood.

Global sufficient condition. B is single-peaked with $B'(g) \propto \sigma_\varepsilon^2(\lambda) - V_\lambda^{\mathrm{o}} g^2$. If $\sigma_\varepsilon^2(\lambda) \leq V_\lambda^{\mathrm{o}}$ then B is strictly decreasing on $(1, \infty)$ while M is strictly increasing there, so the convex crossing is unique; symmetrically for the concave case. □

The sufficient conditions in part (iii) are not superfluous. Eliminating W between the two relations in (8) shows that interior affine equilibria of the fixed-(β, γ) model solve

$$\gamma = \frac{\kappa\,(g-1)\big(V_\lambda^{\mathrm{o}} g^2 + \sigma_\varepsilon^2(\lambda)\big)^2}{r_\lambda^2\, C_\lambda^2\, g^3},$$

and the right-hand side need not be monotone on the relevant domain: when $\sigma_\varepsilon^2(\lambda)$ is large relative to V_λ^{o} a convex schedule can admit up to three selected-assessment affine equilibria, and when it is small relative to V_λ^{o} the same is true of a concave schedule. For instance, with $\kappa = r_\lambda = C_\lambda = V_\lambda^{\mathrm{o}} = 1$ and $\sigma_\varepsilon^2(\lambda) = 20$, every $\gamma \in [62, 70]$ supports three affine equilibria, all satisfying $D_\lambda > 0$. The principal branch — the unique smooth continuation of the knife-edge equilibrium delivered by part (ii) — selects among them (Section 3).

Closing the market shrinks the scope for multiplicity. Under the rank protocol, Proposition 3 shows the equilibrium is globally unique for *all* primitives. Under transfers, the fixed-point function $\mathcal{T}_\lambda$ of (22) has a unique root whenever $\eta \leq \varphi^* V_\lambda^{\mathrm{o}}$ with $\varphi^* \approx 44.9$; the bound is not vacuous — at $\sigma_\theta^2 = \kappa = 1$, $r_\lambda = 0.75$, $\sigma_z^2 = 0.18$, $\eta \approx 70$ (so $\eta / V_\lambda^{\mathrm{o}} \approx 64$) and $\chi_{\mathrm{TU}} \approx 279$, the function $\mathcal{T}_\lambda$ has three roots — but configurations violating it require measurement noise an order of magnitude beyond any plotted in the paper.

The frontier expansion of Remark 1. Abbreviate $\chi = \chi_{\mathrm{TU}}$ and write $V^{\mathrm{o}} = \sigma_\theta^2 + \sigma_z^2 r^2$, suppressing the λ-subscript on r. The implicit function theorem applied to (22) — whose g_θ-derivative at the knife-edge root is $(V^{\mathrm{o}} + \eta)^{3/2} > 0$ — makes the branch jointly smooth in (r, χ) with $g_\theta - 1 = \chi\, \sigma_\theta^2 r^2 / \big(\kappa(V^{\mathrm{o}} + \eta)^{3/2}\big) + O(\chi^2)$, so that, in C^1 on $[0, 1]$,

$$rW = \varphi(r)\big[1 + \chi\, \psi(r)\big] + O(\chi^2), \qquad \varphi(r) = \frac{\sigma_\theta^2\, r}{V^{\mathrm{o}} + \eta}, \qquad \psi(r) = \frac{\sigma_\theta^2\, r^2\, (\eta - V^{\mathrm{o}})}{\kappa\, (V^{\mathrm{o}} + \eta)^{5/2}}.$$

The knife-edge factor has $\varphi'(r) = \sigma_\theta^2(\sigma_\theta^2 + \eta - \sigma_z^2 r^2)/(V^{\mathrm{o}} + \eta)^2$: it is positive on $[0, 1 - \delta]$ and its r-derivative at $r = 1$ equals $-\sigma_\theta^2/(2(\sigma_\theta^2 + \eta)) < 0$ when $\sigma_z^2 = \sigma_\theta^2 + \eta$, both uniformly for (σ_z^2, χ) near $(\sigma_\theta^2 + \eta, 0)$. Hence $\partial_r(rW)$ changes sign at most once, near $r = 1$, and non-monotonicity on $(0, 1]$ is equivalent to $F(\sigma_z^2, \chi) := \partial_r(rW)\big|_{r=1} < 0$. At $(\sigma_\theta^2 + \eta, 0)$, $F = 0$ and $\partial_{\sigma_z^2} F = \partial_{\sigma_z^2} \varphi'(1) = -\sigma_\theta^2/(4(\sigma_\theta^2 + \eta)^2) \neq 0$, so the critical frontier is smooth in χ with

$$\frac{d\sigma_{z,\mathrm{crit}}^2}{d\chi}\bigg|_{\chi=0} = -\frac{\varphi(1)\, \psi'(1)}{\partial_{\sigma_z^2} \varphi'(1)} = 2(\sigma_\theta^2 + \eta)\, \psi'(1) = -\frac{\sigma_\theta^2(\sigma_\theta^2 + 2\eta)}{\kappa\big(2(\sigma_\theta^2 + \eta)\big)^{3/2}},$$

using $\partial_\chi F = (\varphi\psi)'(1) = \varphi(1)\psi'(1)$ at the critical point (where $\varphi'(1) = 0$), $\varphi(1) = \sigma_\theta^2/(2(\sigma_\theta^2 + \eta))$, and $\psi'(1) = -2\sigma_\theta^2(\sigma_\theta^2 + 2\eta)(\sigma_\theta^2 + \eta)\big(2(\sigma_\theta^2 + \eta)\big)^{-7/2}/\kappa$, all evaluated at $\sigma_z^2 = \sigma_\theta^2 + \eta$. This is the expansion displayed in Remark 1.

Appendix G Auxiliary Proofs

This appendix collects the proofs of auxiliary results whose statements appear in the main paper or in Appendix A: the exposure–sensitivity decomposition and the neutrality lemma, the reform-path derivative and the converse affine-strategy lemma for the quality-priced market, and the sorting computations.

Proof of Proposition 1. Let $g_\lambda(s) = Q(m_\lambda(s))$, so $R_\lambda(x; \theta, z) = \mathbb{E}[g_\lambda(\theta + \rho(\lambda)x + \varepsilon_\lambda) \mid \theta, z, x]$, where the conditional law of ε_λ given (θ, z) is independent of the chosen action. Fix x in the interior of I and $h \neq 0$ small enough that $x + h \in I$. By the mean value theorem, for almost every realisation of ε_λ there is t_h between 0 and h with

$$\frac{g_\lambda(\theta + \rho(\lambda)(x + h) + \varepsilon_\lambda) - g_\lambda(\theta + \rho(\lambda)x + \varepsilon_\lambda)}{h} = \rho(\lambda)\, g_\lambda'(\theta + \rho(\lambda)(x + t_h) + \varepsilon_\lambda).$$

Because $g_\lambda'(s) = Q'(m_\lambda(s)) m_\lambda'(s)$, Assumption A.1 provides an integrable dominating variable, and dominated convergence permits differentiation under the expectation, yielding (3). $\square$

Proof of Lemma A.1. Fix a type at which strong neutrality holds and two actions x, x'. For any fixed m_λ and Q, the random variable $Q(m_\lambda(S_\lambda))$ is a measurable transformation of S_λ, so equality of the conditional laws of S_λ under x and x' implies equality of the conditional laws of $Q(m_\lambda(S_\lambda))$, hence of the expectations whenever well-defined. $\square$

The reform path in the quality-priced market. The translation-equilibrium shift (9) is differentiable along any differentiable reform path, with

$$\frac{d\Delta_\lambda}{d\lambda} = \frac{\beta}{\kappa V_\lambda^2}\Big\{r'_\lambda\big[(\sigma_\theta^2+\eta_\lambda)(\sigma_\theta^2+2r_\lambda\sigma_{\theta z})+r_\lambda^2(2(\sigma_{\theta z})^2-\sigma_\theta^2\sigma_z^2)\big]-\eta'_\lambda r_\lambda(\sigma_\theta^2+r_\lambda\sigma_{\theta z})\Big\}, \tag{G.17}$$

where $\eta_\lambda = \sigma_\varepsilon^2(\lambda)$; this follows from the quotient rule applied to (9) and shows how informative ability ($\sigma_{\theta z} > 0$) and noise injection ($\eta'_\lambda > 0$) weaken the trust effect.

Lemma G.4 (Globally affine posterior rules imply an affine strategy)**.** *Under Assumptions 2 and 3, suppose an equilibrium has a posterior rule $m_\lambda(s) = n_\lambda + Ws$ on a Borel set $\mathcal{B}_\lambda$ such that, for almost every type and every feasible action,* $\Pr(S_\lambda \in \mathcal{B}_\lambda \mid \theta, z, x) = 1$. *Then the equilibrium strategy is affine, and if in addition $\gamma = 0$ it is a translation; any such equilibrium is among those characterised in Proposition 2.*

Proof. Given the affine posterior on a set of probability 1 under every feasible deviation, the agent's objective is quadratic in x with second derivative $\gamma W^2 r_\lambda^2 - \kappa$. As in the proof of Proposition 2, equilibrium forces $\kappa - \gamma W^2 r_\lambda^2 > 0$, the objective is strictly concave, its first-order condition is linear in (x, θ), and the best response is affine (θ-coefficient 0 when $\gamma = 0$). Bayesian consistency identifies the coefficients with (5). $\square$

Proof of Proposition 6. Because (θ, Y) is jointly normal, $\mathbb{E}[\theta \mid Y] = (\mathrm{Cov}(\theta, Y)/\mathrm{Var}(Y))Y$. Let $T = Y/\sqrt{\mathrm{Var}(Y)}$, standard normal, so $F_Y(Y) = \Phi(T)$. Iterated expectations in (23) give $\mathcal{E} = (\mathrm{Cov}(\theta, Y)/\sqrt{\mathrm{Var}(Y)})\,\mathbb{E}[T\,H^{-1}(\Phi(T))]$, which is (24); K_H is finite because $\mathbb{E}\big|T\,H^{-1}(\Phi(T))\big| \le (\mathbb{E}[T^2])^{1/2}\,(\mathbb{E}[q^2])^{1/2}$ by Cauchy–Schwarz, $\Phi(T)$ being uniform, and $K_H > 0$ because $H^{-1}\circ\Phi$ is strictly increasing. Substituting $\mathrm{Cov}(\theta, Y) = g_\theta C(r_\lambda)$ and $\mathrm{Var}(Y) = g_\theta^2 V^{\mathrm{o}}(r_\lambda) + \eta(r_\lambda)$ and cancelling g_θ^2 yields (25). Finally $\partial I_\gamma/\partial g_\theta = 2C(r_\lambda)^2\eta(r_\lambda)g_\theta/(g_\theta^2 V^{\mathrm{o}}(r_\lambda) + \eta(r_\lambda))^2 > 0$ when $\eta(r_\lambda) > 0$. $\square$

The knife-edge sorting condition of Section 6. At $g_\theta = 1$, $I(r_\lambda) = C(r_\lambda)^2/V(r_\lambda)$ with $V = V^{\mathrm{o}} + \eta$, and

$$I'(r_\lambda) = \frac{\sigma_\theta^2 + \sigma_{\theta z} r_\lambda}{V(r_\lambda)^2}\big[2r_\lambda(\sigma_{\theta z}^2 - \sigma_\theta^2\sigma_z^2) + 2\sigma_{\theta z}\eta(r_\lambda) - (\sigma_\theta^2 + \sigma_{\theta z} r_\lambda)\eta'(r_\lambda)\big];$$

since $\sigma_\theta^2 + \sigma_{\theta z} r_\lambda > 0$, sorting falls as exposure is reduced exactly when $I'(r_\lambda) > 0$, which is the condition stated in Section 6; the three named cases substitute the stated primitives.

The equality case in Theorem 3(iii). Each fixed point is a scalar relation in (g, s) whose g-derivative at $(g, s) = (1, 0)$ equals $\sigma_\theta^2 + \eta > 0$ in the rank market and $(\sigma_\theta^2 + \eta)^{3/2} > 0$ under transfers, so each branch is smooth near $s = 0$ and we may write $g(s) = 1 + cs + ds^2 + o(s^2)$. Matching powers of s in the scalar rank identity $(g - 1)(g^2 V^{\mathrm{o}} + \eta) = (\chi_H/\kappa)\, s\, g$ (proof of Theorem 2) at $\chi_H = \kappa\sigma_z^2(\sigma_\theta^2 + \eta)/(2\eta)$ gives $c = \sigma_z^2/(2\eta)$ and $d = -c^2$; matching powers in (22) at the transfer threshold gives the same pair (c, d). Hence $\eta/g^2 = \eta\big(1 - 2cs + (3c^2 - 2d)s^2\big) + o(s^2)$ and

$$V^{\mathrm{o}}(s) + \frac{\eta}{g(s)^2} = \sigma_\theta^2 + \eta + 5\eta c^2 s^2 + o(s^2) > \sigma_\theta^2 + \eta$$

for all small $s > 0$, as used in the proof of Theorem 3.